Rosdiazli Ibrahim ·
K. Porkumaran · Ramani Kannan ·
Nursyarizal Mohd Nor ·
S. Prabakar   *Editors*

International Conference on Artificial Intelligence for Smart Community

AISC 2020, 17–18 December, Universiti Teknologi Petronas, Malaysia

Springer

Lecture Notes in Electrical Engineering

Volume 758

The book series *Lecture Notes in Electrical Engineering* (LNEE) publishes the latest developments in Electrical Engineering—quickly, informally and in high quality. While original research reported in proceedings and monographs has traditionally formed the core of LNEE, we also encourage authors to submit books devoted to supporting student education and professional training in the various fields and applications areas of electrical engineering. The series cover classical and emerging topics concerning:

- Communication Engineering, Information Theory and Networks
- Electronics Engineering and Microelectronics
- Signal, Image and Speech Processing
- Wireless and Mobile Communication
- Circuits and Systems
- Energy Systems, Power Electronics and Electrical Machines
- Electro-optical Engineering
- Instrumentation Engineering
- Avionics Engineering
- Control Systems
- Internet-of-Things and Cybersecurity
- Biomedical Devices, MEMS and NEMS

For general information about this book series, comments or suggestions, please contact leontina.dicecco@springer.com.

To submit a proposal or request further information, please contact the Publishing Editor in your country:

China

Jasmine Dou, Editor (jasmine.dou@springer.com)

India, Japan, Rest of Asia

Swati Meherishi, Editorial Director (Swati.Meherishi@springer.com)

Southeast Asia, Australia, New Zealand

Ramesh Nath Premnath, Editor (ramesh.premnath@springernature.com)

USA, Canada

Michael Luby, Senior Editor (michael.luby@springer.com)

All other Countries

Leontina Di Cecco, Senior Editor (leontina.dicecco@springer.com)

** This series is indexed by EI Compendex and Scopus databases. **

Rosdiazli Ibrahim · K. Porkumaran ·
Ramani Kannan · Nursyarizal Mohd Nor ·
S. Prabakar
Editors

International Conference on Artificial Intelligence for Smart Community

AISC 2020, 17–18 December, Universiti Teknologi Petronas, Malaysia

Springer

Editors
Rosdiazli Ibrahim
Department of Electrical and Electronics
Engineering
Universiti Teknologi PETRONAS
Seri Iskandar, Perak, Malaysia

Ramani Kannan
Department of Electrical and Electronics
Engineering
Universiti Teknologi PETRONAS
Seri Iskandar, Perak, Malaysia

S. Prabakar
Department of Biomedical Engineering
Sona College of Technology
Salem, India

K. Porkumaran
Department of Electrical and Electronics
Engineering
Sri Sairam Engineering College
Chennai, India

Nursyarizal Mohd Nor
Department of Electrical and Electronics
Engineering
Universiti Teknologi PETRONAS
Seri Iskandar, Perak, Malaysia

Contents

Design of PID Controller for Integrating Processes with Inverse Response

A. Parvathi, M. Rathaiah, R. Kiranmayi, and K. Nagabhushanam

Abstract In chemical process control designing, forms showing an underlying inverse reaction alters the final state can as often as possible in industries. Procedures displaying such attributes are called inverse reaction procedures. Moreover, inverse reaction procedures can incorporate an integrator in their corresponding model. There are just a couple of studies tending to control of integrating/capacitive procedures with inverse reaction are in literature. Existing techniques, generally, propose the utilization of surely PID-controllers for the control of converse reaction forms with an integrator. However, it has been demonstrated by concentrates in the literature that PID controllers bring about unsuitable closed loop exhibitions for these forms. Expansion of inverse reaction makes the control of an integrating procedure progressively hard to control. Then, a filter is used as shown in proposed structure, performs in exceptionally acceptable closed loop reactions for a class of these procedures. This paper describes the utilization of PID controllers with filters to improve the response of integrating/capacitive forms with reverse reaction. Simulation models are given to represent the prevalence of the proposed technique compared to existing methods.

Keywords Integrating process · Inverse reaction · Servo operation · PID filter · Set point filter · Stability

1 Introduction

The standard PID (Proportional-Integral-Derivative) controller is having three terms in its implementation. This control is well known in process automation and chemical industries due to its easier implementation and improved performances. Its tuning parameters can be discovered by utilizing one of the hypothetical methodologies accepting that the procedure model and by utilizing an ordinary procedure model and one of a few tuning rules, which can be found in the writing. In general, an error

A. Parvathi (✉) · M. Rathaiah · R. Kiranmayi · K. Nagabhushanam
Department of EEE, JNTUA College of Engineering (Autonomous), Ananthapuramu, Andhra Pradesh, India

signal which is the difference between the desired and actual output of the processes is an input for the PID controller.

It offers numerous options to the researchers for tuning. The earlier tuning principles presented by Ziegler and Nichols [1]. They described many tuning procedures for different types of systems. This tuning approach becomes complicated when the process having the integrating terms and zeros on the right half of the s-plane. The intricacy becomes more significant when the time delay (dead time) is related to the procedure. If the process has large time delay in their model then this type of design procedure may not give the accurate desired performance of the processes considered. Therefore a new developments or modification in the basic structure of PID controller/tuning plan is required.

Procedures with inverse reaction attributes are regularly experienced in industries, for example, the degree of the drum kettle in a refining segment, the left temperature of a cylindrical exothermic reactor. Integrating procedures with inverse reaction can likewise be knowledgeable about mechanical applications, for example, level control of a kettle steam drum. The model of this type of process includes the zero of transfer function on the RHS.

It is noted that PID-type controllers bring about unacceptable performance exhibitions in controlling the processes like integrating type. Numerous researchers have projected different approaches for tuning this type of process models as: Jin and Liu [2] proposed an upgraded inside the model controller and determined scientific PID tuning rules for incorporating forms. Anil and Sree [3] planned a PID controller for different types of industrial processes with time deferring based on directsynthesis approach. Ajmeri and Ali [4] utilized a direct way to combine and deal with tuning PID controllers for complete integrating, and twofold request integrating procedures with huge time delays.

Then again, there are very few works tending to the control of integrating procedures with reverse/inverse reaction and dead time. Luyben [5] proposed a PI and additionally a PID controller in the traditional single-input single-yield (SISO) framework to control incorporating forms with converse reaction. Pai et al. [6] have used PID for inverse operating integrating process models.

In recent years, Kaya [7] has projected a PI–PD form of controller to improve the performance of such (integrating with inverse operating) processes using Smith indicator plan.

The target of this work is to structure an effective control methodology for the stable-integrating processes of first order with delay in operating time. The goal is cultivated by utilizing a PID controller along filter is designed. The controller parameters are inferred as far as procedure parameters with the assistance of polynomial strategy.

The remaining portion of paper is structured as follows: The projected control scheme is explained in part II, control system design in part III, and simulation studies in part IV. Finally conclusion is presented.

2 Structure of Proposed Control System

The projected control loop structure for (capacity type) integrating processes is given in Fig. 1. This control loop consist of controller—$G_c(s)$, disturbance signal—d, process with inverse response—$G_p(s)$ and set point filter—$F(s)$. In this structure r—denotes the reference and Y—denotes the output response.

In this study the process considered is (capacity type) integrating type which has one pole (type-1) at origin of s-plane. The form of integrating (capacity type) processes with inverse response is given as:

$$G_p(s) = \frac{K_p(-\alpha s + 1)}{s(\tau s + 1)} e^{-\theta s} \quad \alpha > 0; \quad \tau, \theta > 0 \tag{1}$$

where, K_p—is the steady-state gain, τ—is time constant, and θ—is the time delay.

The servo and regulatory responses are derived as presented in Eqs. (2) and (3) respectively.

$$\frac{Y(s)}{r(s)} = \frac{F(s)G_c(s)G_p(s)}{1 + G_c(s)G_p(s)} \tag{2}$$

$$\frac{Y(s)}{d(s)} = \frac{G_p(s)}{1 + G_c(s)G_p(s)} \tag{3}$$

For deriving the controller parameters, the expression of process is rearranged as shown in Eq. (4).

$$G_p(s) = \frac{k(-\alpha s + 1)}{s(\tau s + 1)} e^{-\theta s} \tag{4}$$

$$G_p(s) = \frac{k(\alpha s + 1)(-\alpha s + 1)}{s(\alpha s + 1)(\tau s + 1)} e^{-\theta s}$$

The controller used in projected approach is of the form of Eq. (5).
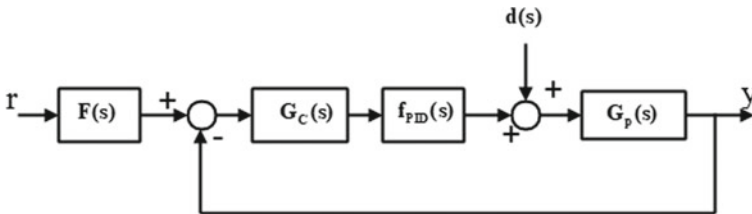
Fig. 1 Structure of proposed closed loop PI control system

$$G_c(s) = \left(k_p + \frac{k_i}{s} + k_d s \right) \tag{5}$$

These sections explains the steps to intend the overall controller which includes set point filter F(s), tuning rules, selecting λ value and performance measurement.

2.1 Set Point Filter

It is clear from the controller presents zeroes in the servo reaction. The zeroes are the main reason for undesired over shoot in the servo reaction prompting motions and the settling time is very high. The presented strategy utilized a set point channel as appeared in Eq. (6) so as to defeat the impact of zeroes presented by the controller in servo reaction [8].

$$F(s) = \frac{1}{\left(k_d \tau_i s^2 + k_p \tau_i s + 1 \right)} \tag{6}$$

2.2 Tuning of PID Controller

The form of controller is,

$$G_c(s) = \left(k_p + \frac{k_i}{s} + k_d s \right) \left(\frac{\alpha_2 s^2 + \alpha_1 s + 1}{\beta_2 s^2 + \beta_1 s + 1} \right) \tag{7}$$

To rearrange the inference, the controller parameters α1 α2 and β_1, β_2 are accepted as appeared below.

$$\alpha_1 = \left(\frac{\theta + 2s}{2} \right)$$

$$\alpha_2 = \left(\frac{\theta + 2s}{4} \right)^2$$

$$\beta_1 = \frac{8\lambda\tau + 4\lambda\tau - 6\theta\tau + 20\tau s + 2\theta s - \theta^2 + 8s^2}{32\tau - 8\lambda + 6\theta + 12s}$$

$$\beta_2 = \frac{s(-6\theta\tau + 4\lambda\tau + 8\lambda s + 8\lambda\tau - 12\tau s - 4\theta s - \theta^2 - 4s^2)}{32\tau - 8\lambda + 6\theta + 12s}$$

2.3 Selection of Tuning Parameter (Λ)

The suggested scope of M_S is 1.2–2, which is a trade off between the speed of reaction and powerful stability. Be that as it may, numerous multiple times, researchers think about qualities past the most extreme point on account of precarious and incorporating forms in order to accomplish desired speed of reaction. For the projected technique, the loop transfer function (L) is,

$$L = G_c(s)G_p(s)$$

$$M_s = \max\left(\left|1 + G_p(s)G_c(s)\right|\right)^{-1} \tag{8}$$

2.4 Performance Measurement

The performance indexes considered in the current study are integral of absolute error (IAE), integral of the square error (ISE). Lessening the IAE and ISE is the goal of present paper.

These are expressed as:

$$IAE = \int_0^\infty (|E(t)|) dt \tag{9}$$

$$ISE = \int_0^\infty \left(E^2(t)\right) dt \tag{10}$$

The total variation (TV) of the input $u(t)$,

$$TV = \sum_{i=1}^\infty |u_{i+1} - u_i| \tag{11}$$

3 Simulation Results and Discussions

Example 1 The model is given to describe the controller technique.

$$G_p(s) = \frac{0.547(-0.418s + 1)}{s(1.06s + 1)} \tag{12}$$

The PID filter is computed as given in Eq. (7)

$$f_{PID}(s) = \frac{0.0548s^2 + 0.468s + 1}{0.0433s^2 + 0.5215s + 1} \tag{13}$$

The set point filter of the process is calculated as given in Eq. (6).

$$F(s) = \frac{1}{2.9216s^2 + 3.815s + 1} \tag{14}$$

This procedure TF (transfer function) has a place with a mechanical contextual investigation of heater steam drum and the level is constrained by controlling evaporator. The PI-PD controller parameter utilized by Kaya [7] is used for comparison purpose. Controller parameters for the current strategy recommended are $K_c = 2.3174$, ki = 0.6075 and kd = 1.7748. In this it is used that $\lambda = 1.500$.

Figure 2 shows reactions for all plan techniques to both a change in reference and a stage upsets with size of -0.5, at time instant $t = 25$ s. The corresponding control signal has been shown in Fig. 3.

Execution particulars for all structure techniques are condensed in Table 1. In all design methods, for both nominal and distributed system the method that present in this work has low values of ISE, IAE and TV.

Example 2 The model is given to describe the controller technique.
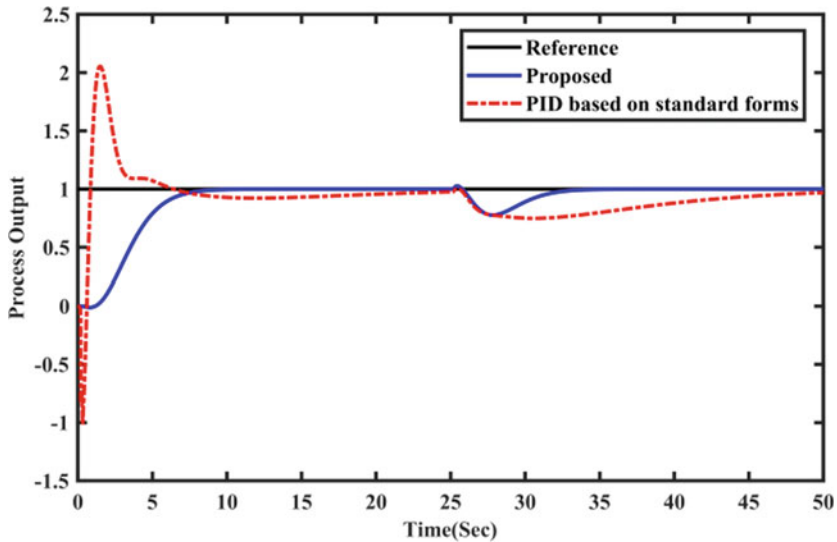
Fig. 2 Closed loop response of example 1
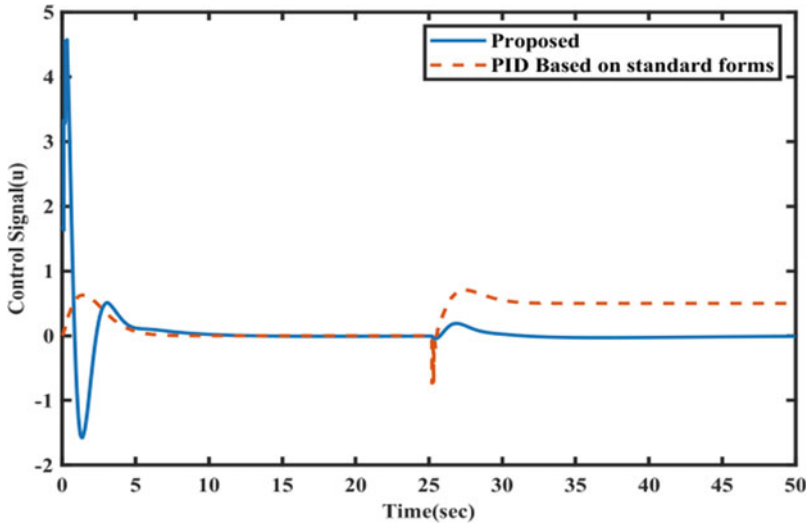
Fig. 3 Control signal of example 1

Table 1 Performance indices of example 1

Control scheme	IAE	ISE	TV
Proposed method	$5.52e^{-05}$	$5.45e^{-08}$	0.003
PID based on standard forms	1.7319	1.1135	0.0055

$$G_p(s) = \frac{0.5(-0.5s + 1)}{s(0.5s + 1)(0.4s + 1)(0.1s + 1)}e^{-0.7s} \tag{15}$$

In the design, the model is used as:

$$G_p(s) = \frac{0.5183(-0.4699s + 1)}{s(1.1609s + 1)}e^{-0.81s} \tag{16}$$

The PID filter is computed as given in Eq. (7)

$$f_{PID}(s) = \frac{0.1914s^2 + 0.875s + 1}{0.0974s^2 + 0.677s + 1} \tag{17}$$

The set point filter of the process is calculated as given in Eq. (6).

$$F(s) = \frac{1}{6.1703s^2 + 6.475s + 1} \tag{18}$$

This procedure move capacity has a place with a mechanical contextual investigation of heater steam drum and the level is constrained by controlling evaporator.
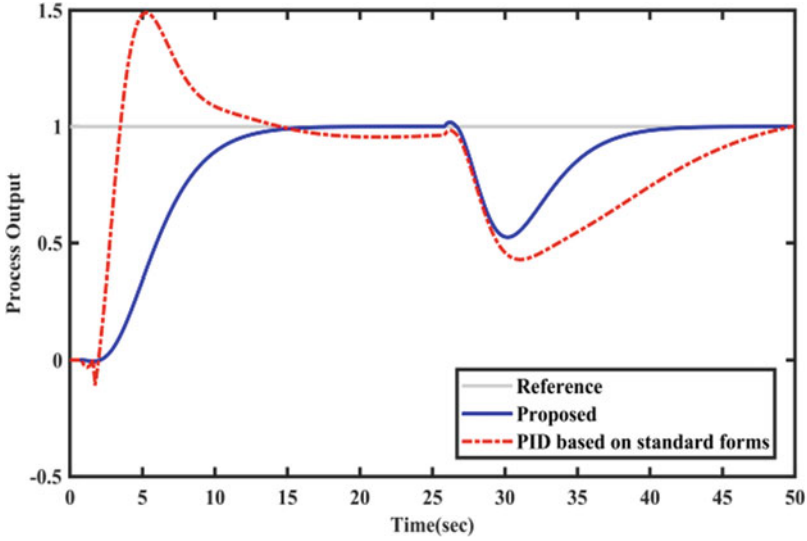
Fig. 4 Closed loop response of example 2

The PI-PD controller parameter utilized by Kaya [7] is used for comparison purpose. Controller parameters for the current strategy recommended are $Kc = 1.1292$, $ki = 0.1744$ and $kd = 1.0761$. In this it is used that $\lambda = 1.500$.

Figure 4 shows reactions for all plan techniques to both a change in reference and a stage upsets with size of -0.5 at time instant(t) $= 25$ s. The corresponding control signal has been shown in Fig. 5.

Execution particulars for all structure techniques are condensed in Table 2. In all methods, for both nominal and distributed system the method that present in this work has low values of ISE, IAE and TV.

Example 3 The model is given to describe the controller technique.

$$G_p(s) = \frac{(-s + 1)}{s(10s + 1)} e^{-7s} \tag{19}$$

The PID filter is computed as given in Eq. (7)

$$f_{PID}(s) = \frac{0.218s^2 + 0.112s + 1}{0.084s^2 + 0.571s + 1} \tag{20}$$

The set point filter of the process is calculated as given in Eq. (6).

$$F(s) = \frac{1}{1.4285s^2 + 3.2075s + 1} \tag{21}$$
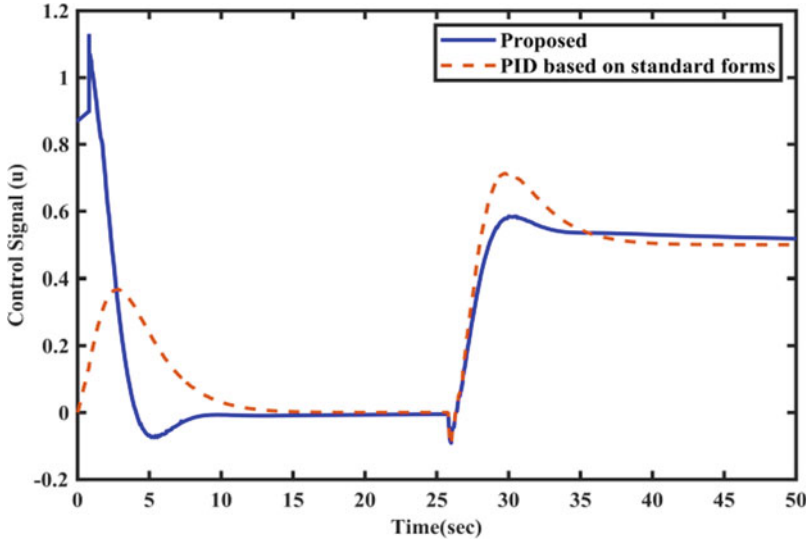
Fig. 5 Control signal of example 2

Table 2 Performance indices of example 2

Control scheme	IAE	ISE	TV
Proposed method	1.918	0.2358	0.0028
PID based on standard forms	2.5006	1.6787	0.0032

The PI-PD controller parameter utilized by Kaya [7] is used for comparison purpose. Controller parameters for the current strategy recommended are $K_c = 0.119$, $k_i = 0.0371$ and $k_d = 0.053$. In this it is used that $\lambda = 1.500$.

Figure 6 shows reactions for all plan techniques to both one unit reference change and a stage upsets with size of -0.5, at time instant $t = 250$ s. Execution particulars for all structure techniques are condensed in Table 3. In all design methods, for both nominal and distributed system the method that present in this work has low values of ISE, IAE and TV (Fig. 7).

4   Conclusions

This paper portrayed the controller plan for integrating procedure with backwards reaction and pole at RHS. The PID controller has been planned in an adjusted structure with arrangement of forward path. The set point channel/filter additionally used to improve the reaction of process because of step changes in reference signal. The proposed technique is connected to the two distinctive relative investigations and their reactions have been appeared in Figs. 2 and 3. The reaction obviously clears that the
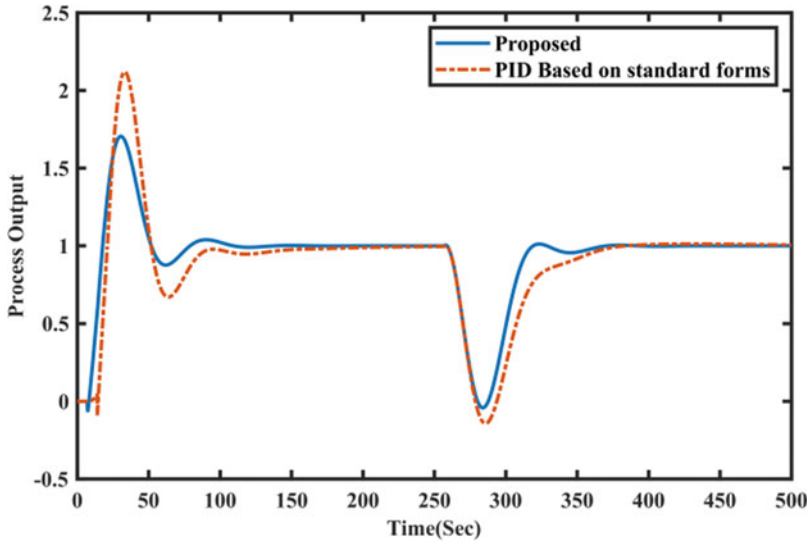
Fig. 6 Closed loop response of example 3

Table 3 Performance indices of example 3

Control scheme	IAE	ISE	TV
Proposed method	1.545	0.358	0.127
PID based on standard forms	2.899	1.223	0.872
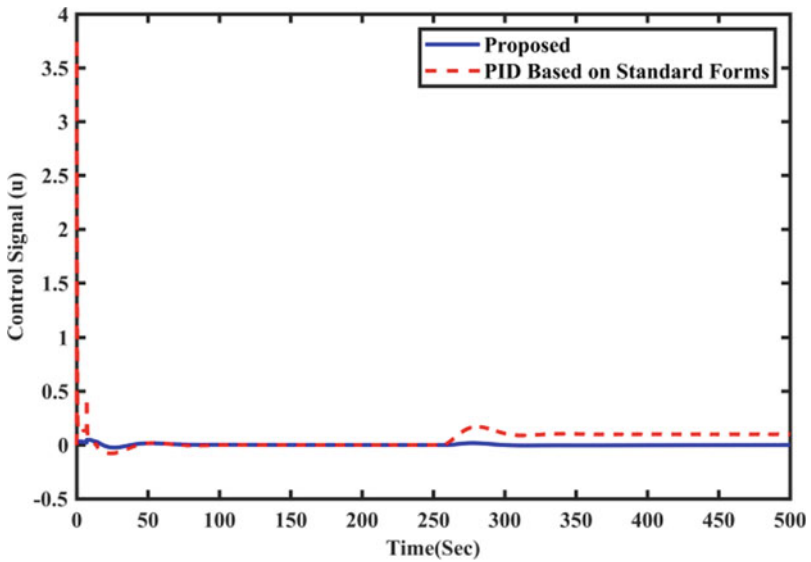
Fig. 7 Control signal of example 3

proposed strategy has improved reaction than the strategies utilized for comparison. The proposed design plan is additionally simpler than other related methodologies for integrating/capacitive type processes.

References

1. Ziegler JG, Nichols NB (1942) Optimum settings for automatic controllers. Am Soc Mech Eng 64:759–768
2. Jin QB, Liu Q (2014) Analytical IMC-PID design in term of performance/robustness tradeoff for integrating processes: from 2-Dof to1-Dof. J Process Control 24:22–32
3. Ch, Anil, Sree RP (2015) Tuning of PID controllers for integrating processes using direct synthesis method. ISA Trans 57:211–219
4. Ajmeri M, Ali A (2015) Simple tuning rules for integrating processes with large time delay. Asian J Control 17:2033–2040
5. Uma S, Chidambaram M, Rao AS (2010) Set point weighted modified Smith predictor with filter controllers for non-minimum-phase (NMP) integrating processes. Chem Eng Res Des 88:592–601
6. Pai NS, Chang SC, Huang CT (2010) Tuning PI/PID controllers for integrating processes with dead-time and inverse response by simple calculations. J Process Control 20:726–733
7. Kaya I (2018) Controller design for integrating processes with inverse response and dead time based on standard forms. Electr Eng 100(3):2011–2022
8. Kumar, Praveen M, Manimozhi M (2018) A new control scheme for integrating processes with inverse response and time delay. Chem Product Process Modeling 13(4)

Environmental Feasibility Survey of Solar Photovoltaic Cells

Sai Goutm Golive, B. Vijaya Krishna, B. Parama Sivam, and K. Ramash Kumar

Abstract Sustainable power source advancements presents an outflow free of methods for the energy collecting in the current power energy situation requesting practical necessities. Be that as it may, the creation cycle of RETs particularly sunlight based photovoltaic (SPV) is material what's more, energy costly cycle. In this paper such data and the energy concentrated creation cycles of Solar photovoltaics frameworks are examined. The frameworks is broke down up to a creation stage with center given to the energy and material separation between various phases of creation. The assessment of an outflows through creation step is completed for a 1 kWp Solar photovoltaics framework. A correlation is made through various kinds of advancements i.e., mono Si, poly Si, slim film CdTe and strip Silicon. Likewise, correlation of maintainability parts of housetop and ground layed establishments was completed. In strip Silicon, energy process represented higher discharges as opposed to material in instance of different innovations. In the layed construct, in rooftop construct brought about least discharges among material investment funds while, layed mounting constructs are ecologically costly because of high material utilization. Additionally, the emanation portion of materials utilized underway stage is decreased directly from mono crystalline Si to strip Si when in rooftop attaching is utilized. Consequently, slight film CdTe, lace Si consists a naturally less expensive alternatives of outfitting sun based energy and creation cycles of advances like mono crystalline Si, poly crystalline Si should be enhanced to make them ecologically reasonable.

S. G. Golive (✉) · B. Vijaya Krishna
Department of Electrical and Electronics Engineering, Bapatla Engineering College, Bapatla, Andhra Pradesh, India
e-mail: saigoutham.golive@becbapatla.ac.in

B. Parama Sivam
Department of Electrical and Electronics Engineering, Annamalai University, Chennai, India

K. Ramash Kumar
Department of Electrical and Electronics Engineering, Dr. N.G.P. Institute of Technology, Coimbatore 641048, India

13

1 Introduction

An Earth-wide temperature boost and environmental evolve have shifted worldwide energy index to move from outflow escalated fossil powered assets to sustainable power sources [1]. Flourishing with the maintainability and green advancement to moderate atmosphere change issues non-customary and sustainable power source innovations are given a help via financial approaches around the globe [2]. All through the gross RESs are encouraged to chop downwards discharges through liberated and changed approaches, for example, power buy agreements, premium feed in duties because of their emanation easy nature of energy reaping [3]. In any case, to choose maintainability parts of any RES life cycle emanations need to be review as opposed to counting emanations that are stayed away from in operated stage.

Life cycle investigation presents a valuable device in breaking down emanations over the whole life cycle by means of utilizing products and power sources i.e., life cycle stock in various periods of creation, utilize and removal of any item [4]. The assembly phase of photovoltaics boards incorporate three arranges to be specific wafer creation, cell manufacture and board fabricating. The SPV board is the most costly piece of absolute PV framework monetarily as naturally examine high measure of energy and input sources of info. Numerous examinations have assessed the ecological effects of the PhotoVoltaics boards in the entire existence with outflows per KWh source created as a mark [5, 6]. Notwithstanding, the separation of emanations among various creation stages is the way to distinguish the territory in which betterment is needed to extend PV more monetarily and naturally manageable innovation. The existence cycle investigation apparatus is related with four phases (Fig. 1) [4]. 1. Objective and degree definition consists illustrating the target, framework limits concerning LCI and measures, characterizing the practical section like natural or financial grade 2. Life cycle stock examination counting information assortment, isolation and standardization and so on. 3. The existence cycle examination stage incorporate assessment of either financial or natural parts of the item and 4. Translation of results to upgrade manageability lists of the item in various periods of life assessment. Over last decades numerous investigations helped out life assessment of SPV along records similar to energy restitution time frame and outflows per KWh energy bridled. The main mono Silicon boards brought about a normal 2–4 years of EPBT for various degrees of illumination and outflow level of 160 to 300 g—CO_2 eq./KWh were accounted for [5–8] The variety upon the mark determined is because of various regardless of whether conditions and varieties underway cycles of SPV. Too exemplified energy differs as per sort of mounting utilized. Generally the housetop mount and open ground surrounding frameworks had higher cycle and exemplified energies contrasted with veneer surrounded frameworks [9]. Be that as it may, the veneer surrounded constructs may bring about higher
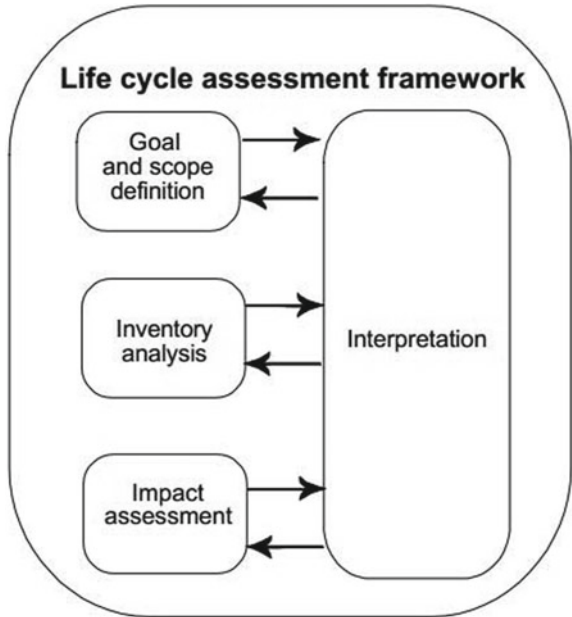
compensation periods because of restricted sun based radiation [10]. Counting the current examinations on life assessment investigation of PV, this examination thinks about the cycle energy and typified energy of ascend constructs in assessing SPV framework maintainability angles. Dissimilar to the vast majority of the investigations an incorporated way to deal with derive stage savvy discharges at all steps and parts of SPV framework is concocted with this examination. The extent of its investigation was limited to creation phase of PV frameworks barring the capacity components similar to batteries and power modules. To measure outflows underway period of SPV ECO-indicator instrument was utilized and useful unit is huge amounts of CO2 per 1 kWp SPV framework creation. Rest of the paper is sorted out as includes: Methodology received to assess life cycle emanations is portrayed in segment II. Segment III diagrams the portrayal of 1 kWp framework for various kinds of PV boards, body of construct and ascend constructs. The segment IV the outcomes are introduced and examined as for generally speaking framework just as every segment.

2 Methodology Adopted

The supportability parts of Photo framework are assessed utilizing ECO-indicator generally known as Eco Indicator 99 eco it configuration instrument [11]. The ECO-it is a straight forward programming to adaptable information base with utilize. There are two pointers together effection climate specifically ReCiPe (RCP) strategy and

Fig. 2 Analysis stages of ECO-it eco design tool [11]

CO_2 counterparts. The ReCiPe (RCP) esteems are unit results which mean the seriousness to outflows with choosen it emor movement. Large data, high the rapid articulated on effect on climate. The ECO-it device comprises off our areas/phases of assessment. The initial stage gives data about the framework and any definitions with respect to the frame work, practical element can be characterized to assist an ce the professional in evaluation. The subsequent stage was the creation period of the item/administration whose step and cycles for the frame work can be comprehended individually to assess coming about emanations. The ECO-it was encouraged to incorporate various components like mass, zone and energy indifferent decimal standards. Any product, energy and cycle in flux into the frame work or item can be determined via five classifications of LCI gave i.e., product, energy, relocate, handling and administration/framework (Fig. 2). When the information from LCI is began coming about emanations can be seen as far as CO_2 counter parts or ECO-it scores. The score of each cycle or part can be utilized to assess the effect of specific segment/measure in the entire life (Fig. 3).

3   Model Representation

In this complete scenario, life rotation discharges for a 1 kWp Solar framework is dissected as four kinds of boards, three sorts of ascend cohesive and regularly utilized BOS. The virtual and specialized attributes of a wide range of Solar boards utilized
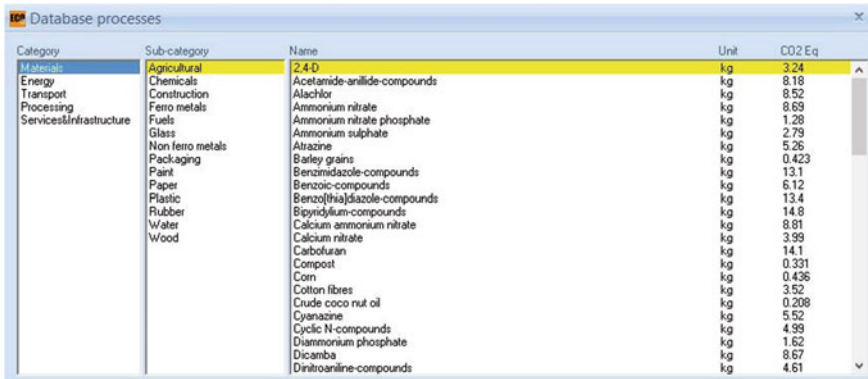
Fig. 3 LCI inventory inclusion in ECO-it eco design tool [11]

Table 1 SPV panel characteristics

	Cell type	Manufacturer Module Rating (Wp)	Cells per module	Module dimensions (mm × mm)	Weight (kg)	Modules per 1 kWp	Module area per 1 kWp (m^2)
1	M ono *Si*	SunWorld [12] 250	60	1675 × 951	46.7	7	6.7
2	poly *Si*	SunWorld [13] 250	60	1675 × 961	39.7	7	6.7
3	CdTe	FirstSolar [14] 92.5	216	1200 × 600	26.45	11	7.78
4	Ribbon	Evergreensolar [15] 205	114	1650.5 × 951.3	41	5	7.65

in this examination are specified in Table 1. The average LCI of SPV boards per 1m^2 zone of board is embraced [9] in discussion with zone needed for 1 kWp for different sorts of SPV boards dissected in this investigation. For the ascend framework, three kinds of ascend courses of action are examined in particular, in-rooftop, on-rooftop and ground surrounded constructs. As the products and zone necessity differs amid the diverse ascend surrounded considered, the in general emanations may likewise shift. The boundaries of the ascend constructs are specified in Table 2. Since the existence cycle stock of three ascend construction it tends to be seen that, ground ascend constructs are material serious and in rooftop framework can spare material contribution surrounded rooftop tile sparing. Be that as it may, the energy repay of in rooftop frameworks should be lower contrasted with that of on rooftop frameworks because of skewed environmental and air refrigeration on lower edge of PV boards [12]. To examine outflows owed to BOS of SPV framework, the framework is larger than usual by half [11]. Thusly, outflow investigation for 1 kWp will in the long run outcome in discharges for 1.5 KVA of inverter limit. In the association part, LCI of Direct wiring is additionally remembered for BOS itself instead of take up as

Table 2 Mounting constructs LCI [9]

Mounting type	On roof	In roof	Ground
Product maker	PhonixSonnenstrom AG¨	Schletter	Springerville
Steel (kg)	0	0	4.01
Stainless steel (kg)	0.49	0.28	0
Aluminium (kg)	0.54	1.21	0
Concrete (kg)	0	0	8.03

a different substance take up as a immaterial effect contrasted with different BOS components, for example, inverter [16].

4   Results and Discussion

The 1 kWp SPV framework is planned from support to entryway regarding emanations during assembling/creation phase of cell, ascending and body system of the framework. The framework limit for module is found from wafer manufacturing leaving Si raw material product and energy inputs.

4.1   Emanation Investigation for SPV Modules

The vitality and material concentrated assembling measure of SPV module applies impressive test in prepare SPV a practical energy reaping innovation. The mono translucent innovation is product just as energy serious there by bringing about most noteworthy discharges of 1257 kg of CO_2 for every 1 kWp of creation. While, strip Si with low product and energy utilization fueled by present day creation innovation brought about best natural alternative with 311 kg of CO_2 per 1 kWp of creation (Fig. 4).
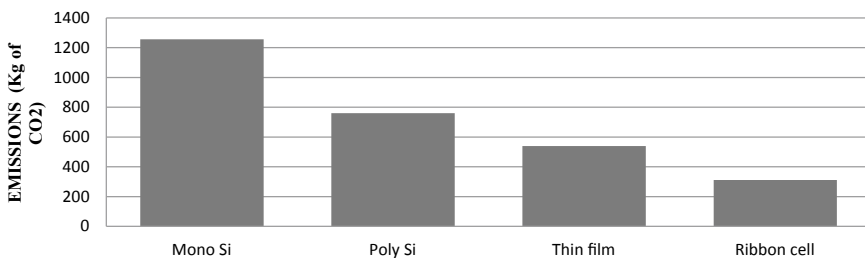
Fig. 4 Emission equivalents of different SPV technologies

(1) LCA of mono-Si:

By a wide margin, the mono Si innovation is effectively the most earth costly because of huge product and energy inputs. The wafer creation stage comprises over 61% of complete discharges and the cell phase represents simple 2.9% of the discharges and left emanations are from cell bundling into module. All things considered the discharge because of the PV module creation in mono Si innovation can be classified into five sorts in particular, gases, solids, fluids, power and direct emanations to air. While, a portion of these are restricted to specific phases of creation. For instance, the fluid and gases utilization is practically immaterial in the event of module readiness. It tends to be seen that, aside from direct discharges what's more, discharges from solids utilized in PV board creation, wafer creation stage is answerable for larger part of the discharges. The wafer phase of PV board has represented 35%, 71%, 94%, 85% of emanations from solids, fluid, gases and power respectively. Producing of silicon material in wafer production stage is related with over 65% of the emissions proceeded in wafer phase. Therefore, powerful wafer fabricating/creation innovation is the need of great importance to reduce outflows through material contributions to wafer stage. It is likewise saw that, over 60% of the outflows from solids were in module phage. This can be ascribed to the fact that, module planning related with board bundling is material escalated with front and back metelled, BSF coating on back end, aluminum utilized for bundling and so forth. In the cell manufacture stage, back metelled was the key cycle which caused lion's share of the outflows in this stage. In this manner, the emanation investigation of mono Si PV board surmisings the urge of abridgement in energy just as material contributions to wafer creation stage (Fig. 5).

(2) LCA of Poly-Si:

Lately, poly Si innovation developed as impressively proficient option for exorbitant and product concentrated mono Si innovation. When contrasted with the

Fig. 5 Emissions of mono Si panel at different production stages
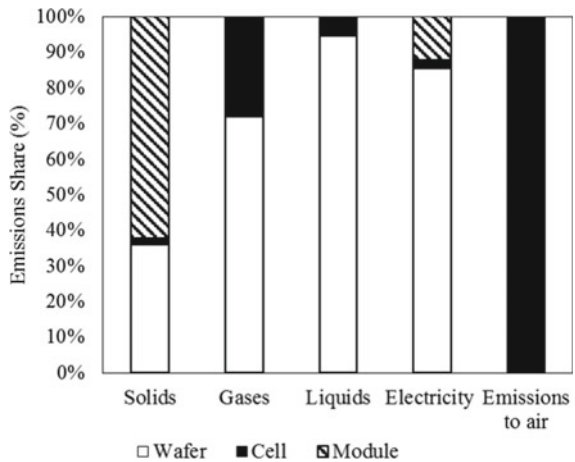
creation outflows of mono Si, emanations in wafer creation phase are simple 46.7% and module bundling had brought about most noteworthy discharges of over 48% of complete poly Si creation discharges. The principle distinction in the creation measure is in to wafer phase of creation. As an aggregate decrease of 530 kg of CO_2 for each 1 kWp wafer creation is seen which was proportionate to 46.7% contrasted with 69.2% of mono Si. In this manner, poly Si wafer creation measure is naturally 22.4% suggested over mono Si wafer creation. Additionally, when contrasted with mono Si the emanations from wafer creation measure are decreased by 7.3%, 58.7%, 7.27% and 22.88% in the event of solids, fluids, gases and power individually (Fig. 6).

(3) LCA of Thin Film CdTe:

Additionally, the slight film modules had came up long way from their presentation in 1975. The record effectiveness of slight film CdTe modules is accounted for as 21.5% [18]. The primary bit of leeway of the flimsy film modules is the product adequacy and wide scope of activity in both diffused and direct retention range [19]. Notwithstanding, the effectiveness is lower contrasted with mono and poly Si partners, there with heightening geological necessity to tackle same sum of energy contrasted with previous PV advancements. Creation of 1 kWp of flimsy film PV module brought about 475 kg-eq of CO_2 which is 62%, 34.6% lower contrasted with mono and poly Si advances separately (Fig. 1). The discharges from creation phase of Cd-Te innovation are partitioned into three classes to be specific module measure, plant compound utilization and plant equipment use. The module creation measure involve 98% of absolute discharges while plant compound utilization and equipment use represented 1.4%, 0.5% individually (Fig. 7). In the module cycle, material and energy inputs involve emanation portions of 60, 40% individually.

Fig. 6 Emissions of mono Si panel at different production stages
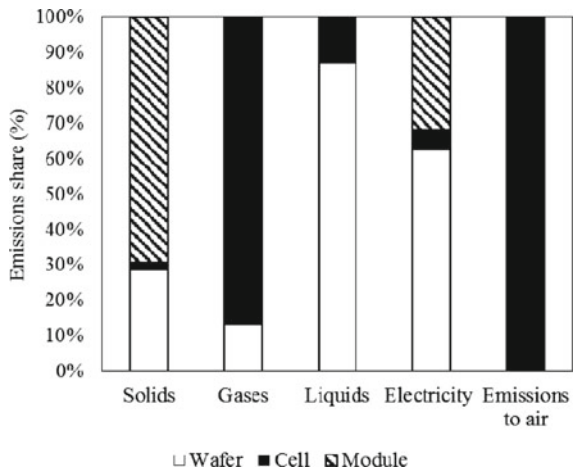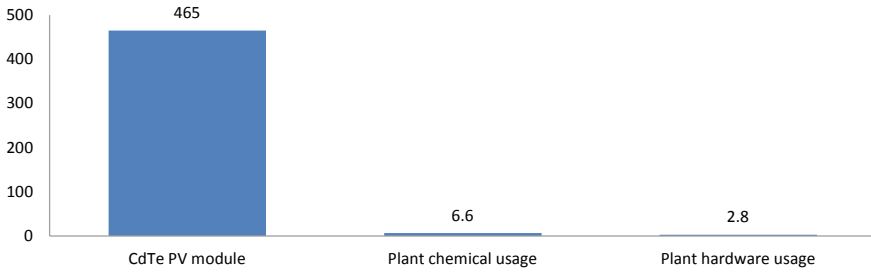
Fig. 7 LCA of thin film technology
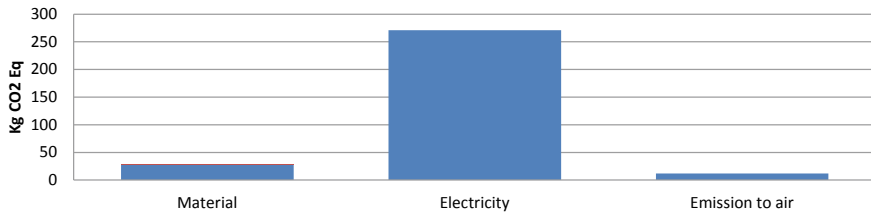
Fig. 8 LCA of ribbon Si wafer technology

(4) LCA of Ribbon Silicon Wafer:

The strip silicon wafer is main of the most recent PV advances empowering tremendous product decrease alongside quick fabricating measure. Nonetheless, the effectiveness of lace silicon wafer is exceptionally low contrasted with slight film too. This is because of the way that, strip Si wafer have extremely dainty silicon construct with lower likelihood of photon assimilation there by settling on productivity [15]. For 1 kWp creation of lace Si wafer the outflows are assessed from LCI accessible. The lace Si wafer creation stage brought about 311 kg of CO_2-eq which was 75%, 57% and 34.5% lower contrasted with mono, poly and flimsy film innovations individually. In contrast to mono, poly and tin film innovations, energy source is liable for greater part of emanations instead of material contributions as if there should arise an occurrence of another PV advances. Energy input oneself is answerable for over 87% of absolute discharges during wafer creation stage. While product and direct emanations to air settled at 9% and 3.8% separately (Fig. 8).

4.2   LCA of Mounting Constructs

When the PV producing was finished, establishment measure is the key territory which devours impressive measure of materials through ascend constructs [9]. The ordinary LCI of three ascend constructs used in this examination is represented in
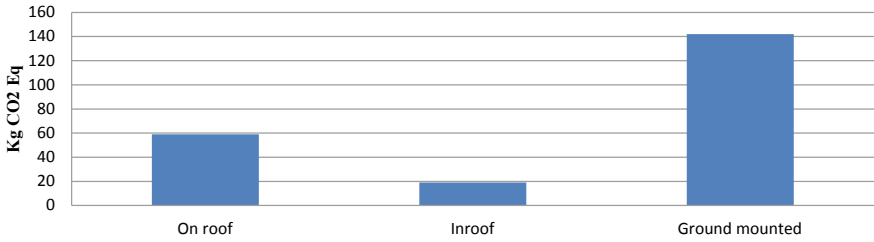
Fig. 9 LCA of mounting constructs

Table 2. In light of the undertaking type mounting constructs are separated into three sorts specifically on-rooftop constructs which are sent in average rooftop top PV frameworks, in rooftop constructs in which boards are coordinated to rooftop top supplanting the tiles there by offering material reserve funds in building development and the third sort is ground ascend constructs which require extra material as concrete utilized for readiness of establishments. The outflows during establishment measure are thought to be exclusively from exemplified energy of product utilized for ascending constructs. The discharges three kinds of discharges are represented in Fig. 9. Ground layed constructs brought about most elevated emanations of 135 kg of CO_2–eq. thought about to 61 and 13 kg for on rooftop and in rooftop partners. The contrast in emanations if there should be an occurrence of in-rooftop construct can be credited to the way that, 1 m^2 of PV boards conceivably can supplant 40 rooftop tile there by decreasing epitomized energy of in general framework (Fig. 10).

LCA of generally framework To comprehend the portion of various framework parts in emanations of entire framework consisting casing and BOS, LCA of entire framework is to be performed. Here, LCA of entire framework is assessed for a 1 kWp framework including ascending and BOS prerequisites. The emanations of 1 kWp framework are investigated with in rooftop framework ascending for all PV advances as appeared in Fig. 11. The emanation portion of PV module with the entire framework is diminished from 96.3% for mono Si to 86.8% for lace Si wafer innovation. While emanations with poly and slight film advances remained at 94%, 90% separately.
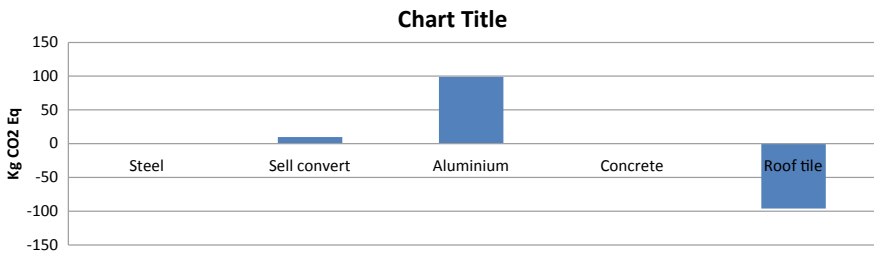
Fig. 10 LCA of in-roof mounting constructs

5 Conclusion

Energy collecting through sustainable power source innovations is moderately discharge free contrasted with non-renewable energy source advancements like coal, flammable gas, atomic frameworks and so on. Notwithstanding, material furthermore, energy costly cycles utilized underway periods of RET advancements like SPV presents maintainability challenges on their arrangement. In this manner, manageability examination of SPV assistance in distinguishing energy and product costly stages upon by congratulating fabricating/creation advances to improve the related cycles so as to decrease natural weight. This work represents LCA of SPV creation and establishment stages for various SPV innovations and ascend constructs. Key discoveries of the problem close the strip Si wafer innovation as best natural alternative for saddling sun based energy. What's more, for strip Si innovation energy utilization is mindful for greater part of discharges though, its material's exemplified energy that mutual greater part of outflows if there should be an occurrence of mono, poly what's more, dainty film advancements. At long last, it is reasoned that, the blend of lace Si innovation with in rooftop ascending might lessen by and large SPV framework emanations by 73% contrasted with mono Silicon innovation.

References

1. The Emissions Gap Report 2012. United Nations Environment Programme (UNEP), Nairobi. Advancing State Clean Energy Funds Options for Administration and Funding, U.S. Environmental Protection Agency's Climate Protection Partnerships Division, May 2008
2. Available [online] http://www.epa.gov/greenpower/documents/purchasingguideforweb.pdf
3. Guide to Purchasing Green Power Renewable Electricity, Renewable Energy Certificates, and On-Site Renewable Generation
4. Environmental management—Life Cycle Assessment—Principles and framework, ANSI/ISO 14040, American National Standard (1997)
5. Kreith F, Norton P, Brown D (1990) A comparison of CO_2 emissions from fossil and solar power plants in the United States. Energy 15:1181–1198
6. Kannan R, Leong KC, Osman R, Ho HK, Tso (2006) CPO Life Cycle assessment study of solar PV systems: an example of a 27 kWp distributed solar PV system in Singapore. Solar Energy 80:55–63
7. Nawaz I, Tiwari GN (2006) Embodied energy analysis of photovoltaic (PV) system based on macro- and micro-level. Energy Policy 34:3144–3152
8. Raghu N, Kshatri SS, Boyina VK (2020) Impact of temperature and irradiance on life time of Pv inverters. Int J Res 2:283–291
9. Knapp K, Jester T (2001) Empirical investigation of the energy payback time for photovoltaic modules. Sol Energy 71:165–172
10. Fthenakis V, Kim HC, Frischknecht R, Raugei M, Sinha P, Stucki M (2011) Life cycle inventories and life cycle assessments of photovoltaic systems. In: International energy agency photovoltaic power systems programme, October 2011
11. Eco-indicator 99 manual for designers, A damage oriented method for life cycle impact assessment, Ministry of Housing, Spatial planning and environment, The Netherlands, October 2000 (Online): www.pre-sustainability.com/download/EI99Manual.pdf

12. Vijaya Krishna B, Venkata Prasanth B, Sujatha P (2018) Design and implementation of DPFC for multi-bus Power system. Int J Eng Technol 7(2.8):18–29
13. Sunmodule SW 250 POLY (33mm frame). (Online): http://www.solarworld-usa.com//media/www/files/datasheets/sunmodule-pro-series/33mm-frame/sunmodule-solar-panel-pro-series-poly-33mm-frame-ds.pdf
14. Vijaya Krishna B, Venkata Prasanth B, Sujatha P (2018) Enhancement of power quality in distribution system using hybrid seven level H-bridge inverter based Dpfc. J Electrical Eng 2:61–71
15. Vijaya Krishna B, Venkata Prasanth B, Sujatha P (2018) MA TLAB/Simulink study of of multi-level inverter topologies using minimized quantity of switches. Int J Eng Technol 7(1.5):209–216
16. Evergreensolar, ES-A Series Photovoltaic panels. [Online] http://www.evergreensolar.com/upload2/InstallManuals/2011-JANES-AIMWETLSM-0029-USRevE

Intelligent Control Techniques for Parameter Tuning of PID Controller for LFC with Emphasis on Genetic Algorithm

B. Venkata Prasanth, Raja Sathish Kumar, Vijaya Krishna Boyina, and K. Ramash Kumar

Abstract The theme of this work is carried out to reduce the deviations involved in LFC of a developed system by via GA technique (intelligent controller) deployed PID controller. In MATLAB environment, the system is modelled and as well as simulated for conventional controller, Fuzzy, PSO and GA based PID tuning parameters. Simulation results shows that there is an improved dynamic response of the system when subjected to different step perturbations with proposed GA-PID controller.

1 Introduction

The goal of LFC during transient conditions is to maintain constant frequency, power exchanges and tackling system complex model as well as variations involved [1–3]. In general case, the habitual apparent power produced by generators must be met in accordance with the load power variations. The principal task before a power system engineers is to provide good quality of power supply generated by a mixture of renewable and non-renewable sources of energy to the utility customers without any distortions. Load frequency control to be referred as by maintaining frequency within permissible limits by regulating both wattage input power generation and demand

B. Venkata Prasanth
QIS College of Engineering & Technology, Ongole, India

R. Sathish Kumar
Keshav Memorial Institute of Technology, Hyderabad, India

V. K. Boyina (✉)
Bapatla Engineering College, Bapatla, India
e-mail: vijayakrishna.boyina@becbapatla.ac.in

K. Ramash Kumar
Department of Electrical and Electronics Engineering, Dr.N.G.P. Institute of Technology, Coimbatore 641048, India

[4]. The variables (Δf and ΔP_{tie}) effects during power transient load power conditions. Automatic control is more efficient method of load frequency control because manual controls are sluggish and involve inherent human time lags. The problem of ALFC resolves into not only measuring Δf but also analysing the measured change from a reference measurement value. Keep view on this correction is initiated by using control strategy to keep the system original measured value. The risk of the interconnected system increases because of system design deals with more time varying parameters. Valid assumptions are to be incorporated in design of a system controller to enhance overall performance. Tyreus-Luyben, Cohen-Coon, Fertik, Z-N and Integral Control Methods, are considered for controlling the LFC of isolated single area system [3, 5, 6]. Integral gain of conventional controller restricts the dynamic performance of the system. An optimal control scheme based PSO which enhances the gain of controller for addressing LFC in power systems of single area or multi area type power systems. The authors [7, 8] proposed the AGC by using a optimal controller for two area power system as well as compared the results. Intelligent control technique needs to be deployed to achieve further improvement in power system dynamics [9–12]. Among them GA has been used to deal complex optimization problems which uses H2/H∞ Controller [6–8].

2   Model Investigation and Controllers

The power plant model shown in Fig. 1 consists of governor, turbine, a generator and load with speed regulation as feedback etc., are characterized by transfer functions. The values of gain and time constants chosen for the model are listed out in Table 1.

The state variables are listed below for the model shown above in Fig. 1.
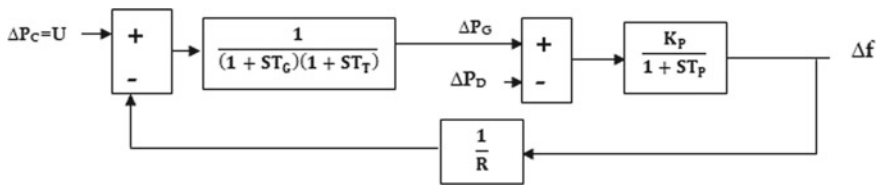
Fig. 1   Single area thermal power generation model

Table 1 Power system parameters	Name of the component	Gain	Time constant (s)
	Governor	$K_G = 1$	$T_G = 0.08$
	Turbine	$K_T = 1$	$T_T = 0.3$
	Power System/Generator	$K_P = 120$	$T_P = 20$
	Speed Regulation	$R = 2.4$	

$$\dot{x}_1 = \left(\frac{-1}{T_g}\right)x_1 + \left(\frac{K_P}{T_P}\right)x_3 - K_P \Delta P_D$$

$$\dot{x}_2 = x_6$$

$$\dot{x}_3 = \left(\frac{1}{T_t}\right)x_2 + \left(\frac{-1}{T_t}\right)x_3$$

$$\dot{x}_4 = x_5$$

$$\dot{x}_5 = \left(\frac{-K_P}{T_P T_t}\right)x_2 + \left(\frac{K_P}{T_P T_t^2}\right)x_3 + \left(\frac{-1}{T_g}\right)x_5 + \left(\frac{K_P}{T_P T_t}\right)x_6$$

$$\dot{x}_6 = \left(\frac{k_i}{T_g}\right)x_1 + \left(\frac{k_p R - 1}{T_g R}\right)x_4 + \left(\frac{k_d}{T_g}\right)x_5 + \left(\frac{-1}{T_g}\right)x_6$$

In Power system, invariably the nature of the elements is non-linear. At a particular working point linearization to be done to get a linear model from a nonlinear model. The actual conditions and practical conditions are differing in general scenario. Classical algorithms are used to design a load frequency with PID controller in terms of state space variables for the modelled system [13, 14]. The system dynamics are tabulated for uncontrolled, conventional, Fuzzy rule based, PSO based and GA based PID Controller in Sect. 3. The intelligent controllers implemented for the model to get hold of better tuning of gains are listed below.

PID Controller:

Almost in all control applications, PID controllers are extensively used because of its robustness and system performance. The proto type of PID controller has parameters namely K_P, K_i and K_d which are shown in the Fig. 2. To achieve better system performance within desirable limits; it is necessary to tune the K_P, K_i and K_d terms
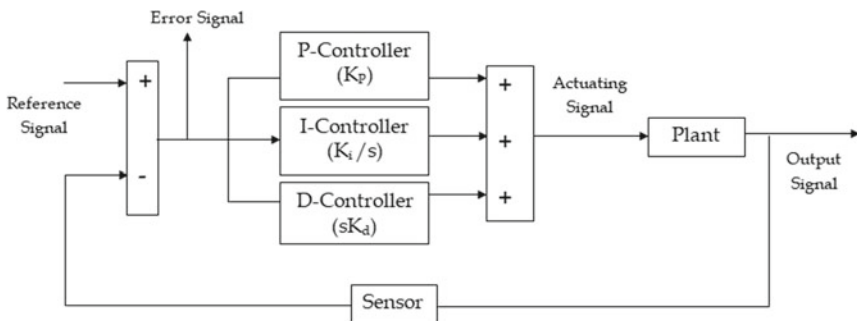
Fig. 2 Proto type PID controller

in the controller by appropriate optimization technique. Improper tuning leads to adverse effects on system performance.

Fuzzy Logic Controller:

The general Fuzzy based logic controller consists of fuzzification, rule based control and defuzzification shown in the Fig. 3. The logic is much closer to the human logics and is better as compared with classical systems. We cannot achieve satisfactory solutions from regular PID controllers due to the involvement of time varying variables while designing the system. The linguistic variables from [3] used for input e(t) and $[\dot{e}(t)]$ for the chosen model as well as for output also under fuzzification and rule based fuzzy table framed and shown in Table 2 for the model shown in Fig. 1 based on If and Then rules. Centroid method used for defuzzification process to transform the output linguistic variable to a real value signal.

PSO Algorithm:

PSO is also one of the soft computing techniques to acquire better tuning values for parameters involved in PID controller to solve more complex engineering problems. The procedure for minimization of performance index which is integral time absolute error (ITAE) i.e., ITAE = $\int$ t e(t)dt and frequency deviation is an error function is obtained by uniformly distributed particles, particle's position and particles velocities & weights. Figure 4 shows the flow chart of PSO algorithm.
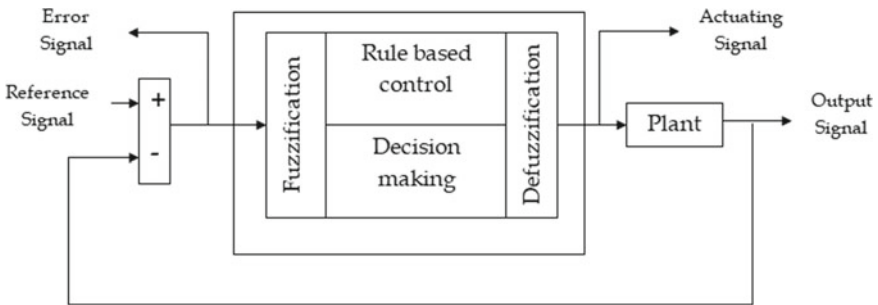
Fig. 3 Fuzzy logic controller

Table 2 Fuzzy rule based table

$\dot{e}$	e				
	NL	NS	ZZ	PS	PL
NL	S	S	M	M	L
NS	S	M	M	L	VL
ZZ	M	M	L	VL	VL
PS	M	L	VL	VL	VVL
PL	L	VL	VL	VVL	VVL

Fig. 4 PSO algorithm flow
chart
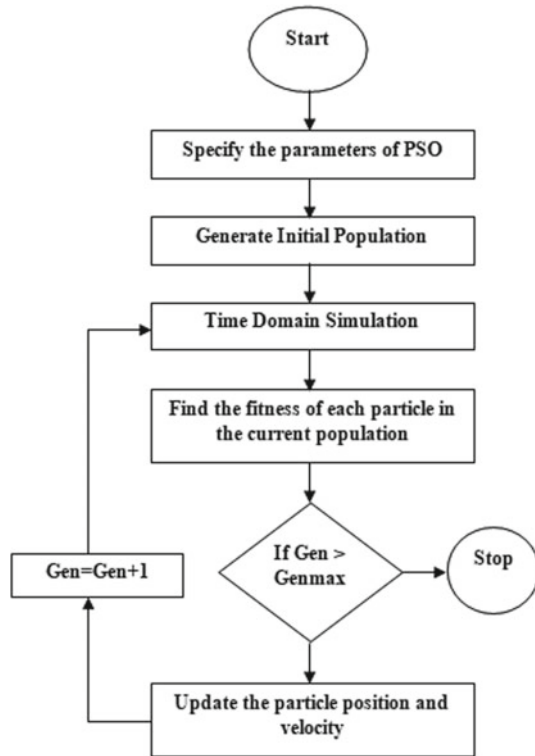
Genetic Algorithm:

To solve more complex problems GA isused and the process involved represented
by flow chart which is shown in Fig. 5. The procedure for minimization of perfor-
mance index which is integral time absolute error (ITAE) = ∫ t e(t)dt and frequency
deviation is an error function is obtained by reproduction, crossover and mutation.

3 Simulation Results

The dynamic response of the system is studied by applying step change in load
perturbation ($\Delta P_D = 0.01, 0.02, 0.03, 0.04$ and 0.05) to the generator. The dynamics
are studied in terms of time domain specifications with GA, PSO and Fuzzy Logic
PID Controllers as well as concluded from the responses shown in Figs. 6, 7, 8,
9 and 10 that GA-PID controller perks up better transient behavior among all the
mentioned controllers (Table 3).
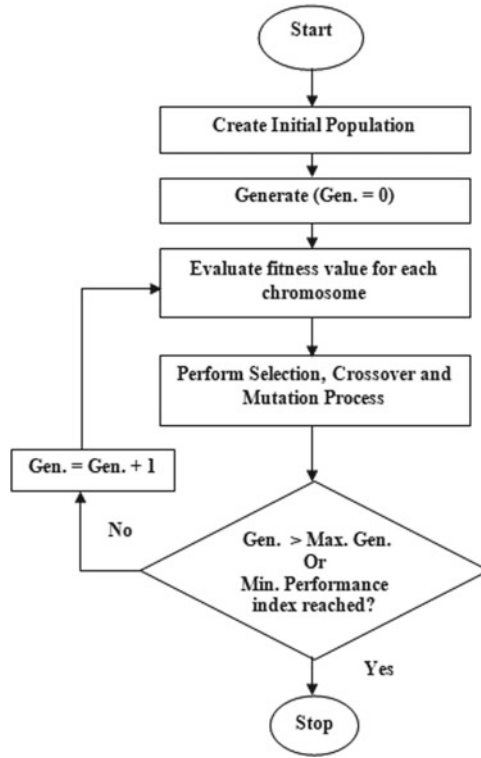
Fig. 5 Genetic algorithm flow chart
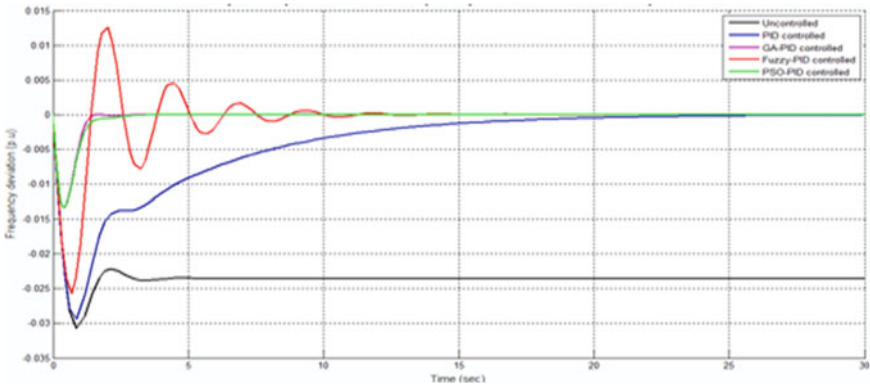
Fig. 6 Dynamic response of isolated thermal system with a disturbance of $\Delta P_D = 0.01$
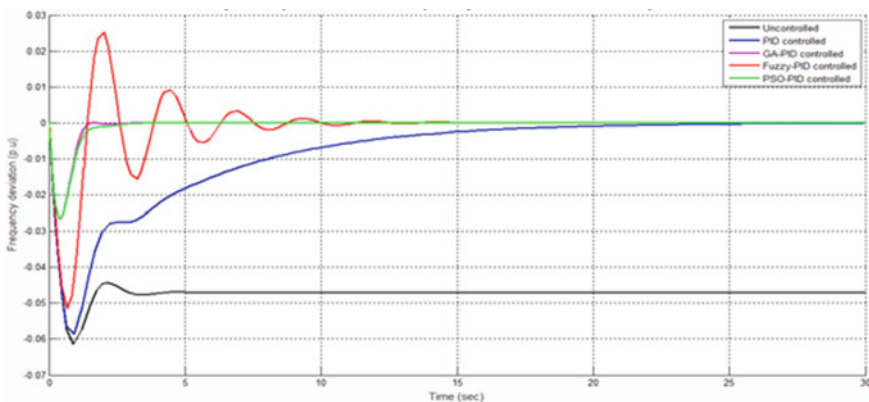
Fig. 7 Dynamic response of isolated thermal system with a disturbance of $\Delta P_D = 0.02$
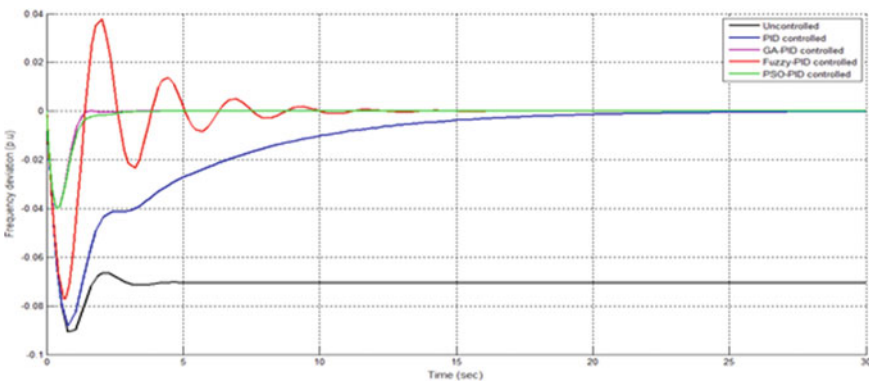
Fig. 8 Dynamic response of isolated thermal system with a disturbance of $\Delta P_D = 0.03$
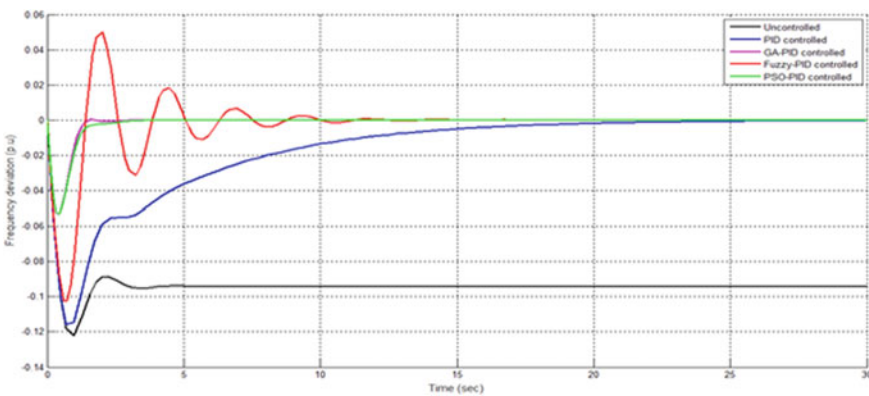
Fig. 9 Dynamic response of isolated thermal system with a disturbance of $\Delta P_D = 0.04$
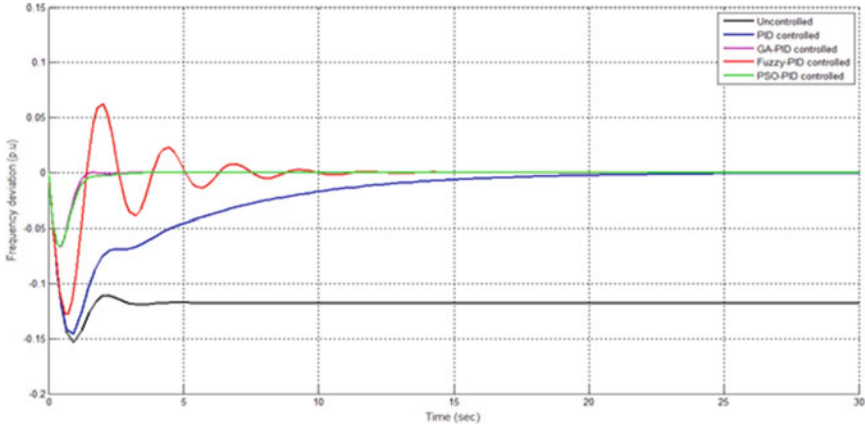
Fig. 10 Dynamic response of isolated thermal system with a disturbance of $\Delta P_D = 0.05$

Table 3 Comparison of different controllers

Disturbance with $\Delta P_D = 0.01$			
Type of Controller	Undershoot	Static Error	Settling Time
Uncontrolled	−0.03069	−0.02352	4.255
PID controlled	−0.02942	0	23.780
Fuzzy-PID controlled	−0.02716	0	6.755
PSO-PID controlled	−0.01329	0	3.195
GA-PID controlled	−0.01325	0	2.744
Disturbance with $\Delta P_D = 0.02$			
Type of Controller	Undershoot	Static Error	Settling Time
Uncontrolled	−0.06139	−0.04697	4.439
PID controlled	−0.05855	0	24.030
Fuzzy-PID controlled	−0.05355	0	7.641
PSO-PID controlled	−0.02658	0	3.459
GA-PID controlled	−0.02638	0	3.226
Disturbance with $\Delta P_D = 0.03$			
Type of Controller	Undershoot	Static Error	Settling Time
Uncontrolled	−0.09076	−0.07044	4.817
PID controlled	−0.08821	0	24.860
Fuzzy-PID controlled	−0.08035	0	8.032
PSO-PID controlled	−0.04003	0	3.765
GA-PID controlled	−0.03999	0	3.507

(continued)

Table 3 (continued)

Disturbance with $\Delta P_D = 0.01$			
Type of Controller	**Undershoot**	**Static Error**	**Settling Time**
Disturbance with $\Delta P_D = 0.04$			
Type of Controller	**Undershoot**	**Static Error**	**Settling Time**
Uncontrolled	−0.12201	−0.09405	5.045
PID controlled	−0.11560	0	25.310
Fuzzy-PID controlled	−0.10250	0	11.49
PSO-PID controlled	−0.05332	0	4.566
GA-PID controlled	−0.05228	0	4.255
Disturbance with $\Delta P_D = 0.05$			
Type of Controller	**Undershoot**	**Static Error**	**Settling Time**
Uncontrolled	−0.15340	−0.11770	5.685
PID controlled	−0.14590	0	26.451
Fuzzy-PID controlled	−0.12780	0	12.760
PSO-PID controlled	−0.06599	0	5.349
GA-PID controlled	−0.06579	0	5.021

4   Conclusion

A single area Load frequency control problem with PID controller is modelled and state variables are derived. Conventional controller, Fuzzy based, PSO based and GA based PID controllers are applied to the generator to get dead beat response from the system. These controllers are used to minimize the system transient behaviour pertaining to step load disturbances i.e., ΔP_D (p.u.) $= 0.01$ to 0.05 in steps of 0.01. GA based PID tuned controller improves the system transient behaviour in contrast with other controllers implemented. The proposed GA controller is more robust in nature to minimize the system transient behaviour. There is a scope for further research to where the robustness of power system can be improved by new soft computing techniques which deals optimization of complex problems.

References

1. Sathish Kumar R (2018) LFC problem by using improved genetic algorithm tuning PID controller. Int J Pure Appl Math 120:7899–7908
2. Prasanth BV (2018) Tuning of PID controller by improved GA technique for an Isolated Nuclear Power System. Int J Pure Appl Math 120:8011–8020
3. Fagna R (2017) Load frequency control of single area thermal power plant using type 1 fuzzy logic controller. Sci J Circuits Syst Signal Process 6: 50–56
4. Vijaya Krishna B (2018) Enhancement of power quality in distribution system using hybrid seven level H-bridge inverter based DPFC. J Electr Eng 2:61–71

5. Prasanth BV (2015) PSO-PID tuning technique for load frequency control problem. Int J Global Innov 3:204–209
6. Rajani C (2015) Multi-objective design of load frequency problem using genetic algorithm based PID controller. Int J Adv Eng Global Technol 3:307–318
7. Kumar RS (2014) Load frequency control of two area power system with fuzzy logic and ANN control. Int J Adv Eng Global Technol 2:143–147
8. Prasanth BV (2013) New control strategy for improvement of transient behavior in multi area interconnected power system with emphasis on robust genetic algorithm. Int J Eng ResTechnol (IJERT) 2:3989–3993
9. Subramanyam B (2009) Comparative study of different control strategies for a load frequency problem with emphasis on new fuzzy logic controller. IJEEE Spring Edn 1:25–33
10. Vijaya Krishna B (2018) MATLAB/simulink study of multi-level inverter topologies using minimized quantity of switches. Int J Eng Technol 7:201–216
11. Vijaya Krishna B (2018) Design and implementation of DPFC for multi-bus power system. Int J Eng Technol 7:18–29
12. Vijaya Krishna B (2016) Designing of multilevel DPFC to improve power quality. In: International Conference on Electrical, Electronics and Optimization Techniques (ICEEOT) @ IEEE
13. Wood AJ, Wollenberg BF (1996) Power generation operation and control. Wiley, New Jersey
14. Kennedy J (1995) Particle swarm optimization. In: Proc IEEE Int Conf Neural Netw 4:1942–1948

Simplified Decoupler Based Fractional Order PID Controller for Two Variable Fractional Order Process

A. Divya Teja, R. Kiranmayi, K. Nagabhushanam, and N. Swathi

Abstract In recent years, more research work has been proved the advantages of using fractional order modeling and control techniques. This paper describes the design of fractional order proportional integral derivative (FO-PID) controller along with simplified decoupling method of two variable fractional order processes. The structure of two variable processes is different from single variable process. The interaction effects occur for two variable fractional order processes. So to reduce the interaction effect, the process is decoupled by simplified decoupling method. The maximum sensitivity based frequency domain strategy is suitable for parameters tuning of FO-PID controller. The described controller yield better execution for set point-tracking. Disturbances like white noise are included in the system so as to exhibit the FO-PID method shows better result in rejection of disturbance. The parameter uncertainties are added to the process. The course of action is more robust to such dissimilarities. Simulation results will exhibit the better achievement of the proposed method.

Keywords Fractional order control · Simplified decoupling · Fractional order proportional integral derivative time delay · Two-input-two-output process · Inverted decoupling · Fractional calculus

1 Introduction

One of the oldest control method proposed in 1920s is PID control method. Mostly it is utilized in several industrial fields because of good performance, simple design and small settling time [1, 2]. Many of the PID control methods are modelled as integer order system. In some cases the systems should be modelled as fractional order, for the systems the fractional PID controller is employed. Many scientists employ fractional order controllers for many applications to obtain desired performance of the systems [3–5]. In 1994 I. Podlubny [6] introduced a PID controller for a fractional

A. Divya Teja (✉) · R. Kiranmayi · K. Nagabhushanam · N. Swathi
Department of EEE, JNTUA College of Engineering (Autonomous), Ananthapuramu, Andhra Pradesh, India

order processes. He described the PID controller in generalized form introducing fractional calculus and compare with the classical PID controller to exhibit good response of the controller. The PID controllers contains three parameters (K_p, K_i, and K_d) to tune but in FOPID controllers contains other two parameters λ and μ. λ is fractional integration order and that of μ is fractional derivative order. Therefore, the FOPID contains five parameters (K_p, K_i, K_d, λ and μ) for tuning purpose. This makes the design of controller more flexible. Many literatures will mainly focus on tuning method of FO-PID. For the tuning of FOPID literatures [7–9] demonstrates different tuning methods.

In this paper, a simplified decoupler based FOPID controller is introduced for two variable processes. For two variable process the interactions between the variables must exist. So to eliminate the interactions for two control loops, simplified decoupling is used. The frequency domain tuning method is used to design the FOPID controller and it was proposed by Monje et al. [8].

The left over portion of this paper is arranged as follows: In unit 2, introducing the fractional order control, the fractional order PID controller and the structure of the fractional order of two variable processes. In unit 3, the structure of simplified decoupling, the frequency tuning strategy and the design of proposed controller are given. For simulation study the numerical examples are considered with and without time delay processes and comparisons of the presented method and FOIMC method are structured in unit 4. Conclusion and future scope are present in unit 5.

2 Fractional Order Control

2.1 Fractional Calculus

The concept of fractional calculus is obtained from mathematical branch. It describes the integrals and derivative theory of non-integer order. Fractional calculus has developed for a long time. Researchers have done their work in several areas of engineering and technology by using fractional calculus during the last few decades. There are many other definitions of fractional derivatives and integrals. Riemann and Liouville had described fractional order integral definition using Taylor series

$$_aD_t^{-v}\mathrm{f}(t) = \frac{1}{\Gamma(\gamma)} \int_a^t (t-\tau)^{v-1}\mathrm{f}(\tau)d\tau \tag{1}$$

The fractional calculus operator $_aD_t^{-v}$ represents fractional derivative and f(t) is an function of integer. $\Gamma(\gamma)$ defines gamma function.
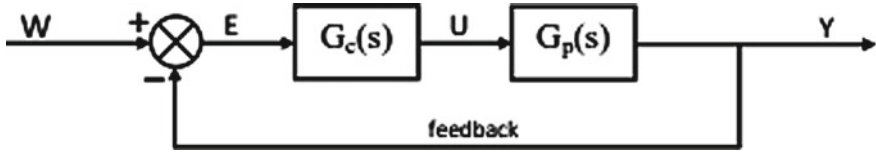
Fig. 1 The FO-PID standard structure

2.2 Fractional Order PID Controller

Podulbuny has first described the FOPID controller. It was an expansion of conventional PID controller using fractional calculus. The fractional order PID based control method is presented in Fig. 1. $G_p(s)$ is the process, $G_c(s)$ is the FOPID controller. By adding three control actions, the expression for the control method is obtained and it is described in time domain below:

$$u(t) = K_p e(t) + K_i \int e(t)dt + \frac{de(t)}{dt} \tag{2}$$

The expression of PID controller in s-domain is

$$G_c(s) = \frac{u(S)}{e(s)} K_p + K_i \frac{1}{s} + K_p s \tag{3}$$

2.3 Fractional Order TITO Process

A two variable fractional order linear system is represented by the following fractional order differential equations:

$$_0D_t^\gamma x(t) = Ax(t) + Bu(t), \tag{4}$$

$$Y(t) = Cx(t) \tag{5}$$

Assuming initial conditions are zero and by applying Laplace transform, the representation of state space equations can be transformed as

$$s^\gamma Y(s) = (s) U(s) \tag{6}$$

By substituting (4) and (5) in (6), the following equation can be obtained as:

$$G_p(s) = C(I-A)^{-1}B = \begin{bmatrix} G_{p11}(s) & G_{p12}(s) \\ G_{p21}(s) & G_{p22}(s) \end{bmatrix} \tag{7}$$

where each $G_{ij}(s)$ is a fractional order transfer function and defined as follows:

$$G_{ij}(s) = \frac{b_{ij1}s^{\beta_{ij1}} + b_{ij0}}{a_{ij2}s^{\alpha_{ij2}} + a_{ij1}s^{\alpha_{ij1}} + a_{ij0}}, \quad i,j = 1,2 \tag{8}$$

where β_{ij} and α_{ij} are the fractional order, a_{ij} and b_{ij} are the coefficients. The structure of fractional order two variable process is introduced in Fig. 1.

3 Simplified Decoupling Based FO-TITO-PID Controller Design

To design the FO-TITO-PID controller, two steps are required.

Step1: FO-TITO process is decoupled with simplified one.

Step2: FO-PID controller is designed for FO-TITO process.

The interactions are present between two control loops of the TITO process can be controlled by decoupling method. According to [10] the decoupling methods are ideal decoupling, simplified decoupling and inverted decoupling. Ideal decoupling is most complex in design and the inverted decoupling has some limitations to the process i.e., the time delay for the diagonal element matrix must be minimum value, the outputs of the control method exhibit large variations. To overcome the limitations, simplified decoupling method is chosen. The structure of simplified decoupling represent in Fig. 2.
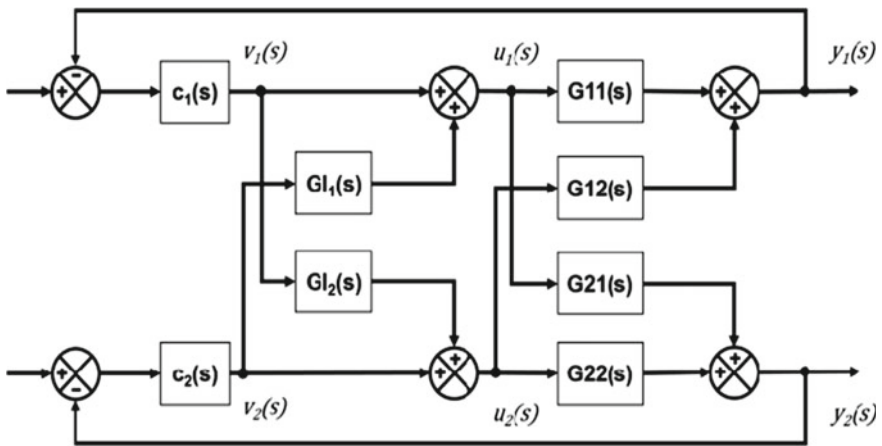
Fig. 2 The structure of the simplified decoupling FO-PID method

The new decoupler can be obtained as:

$$Gl_1(s) = -\frac{G_{12}(s)}{G_{11}(s)} \tag{9}$$

$$Gl_2(s) = -\frac{G_{21}(s)}{G_{22}(s)} \tag{10}$$

The output of c_1, c_2 and the input u_1, u_2 of $G(s)$ have the relationship as:

$$u_1 = c_1 + c_2 \frac{G_{12}(s)}{G_{11}(s)} \tag{11}$$

$$u_2 = c_2 + c_1 \frac{G_{21}(s)}{G_{22}(s)} \tag{12}$$

The relationship between the inputs u_1, u_2 of $G(s)$ and the output is:

$$u_1 G_{11}(s) + u_2 G_{12}(s) = y_1 \tag{13}$$

$$u_2 G_{22}(s) + u_1 G_{21}(s) = y_2 \tag{14}$$

Substitute (10) into (11), the result will be:

$$c_1\left(G_{11}(s) - \frac{G_{12}(s)\,G_{21}(s)}{G_{22}(s)}\right) = y_1 \tag{15}$$

$$c_2\left(G_{22}(s) - \frac{G_{12}(s)\,G_{21}(s)}{G_{11}(s)}\right) = y_2 \tag{16}$$

The generalized loop transfer function is

$$G(s) = \begin{bmatrix} \left(G_{11}(s) - \frac{G_{12}(s)\,G_{21}(s)}{G_{22}(s)}\right) & 0 \\ 0 & \left(G_{22}(s) - \frac{G_{12}(s)\,G_{21}(s)}{G_{11}(s)}\right) \end{bmatrix} \tag{17}$$

Substituting (9) into (10), the decoupler can be rewrite as

$$Gl_1(s) = -\frac{\left(b_{121}s^{\beta_{121}} + b_{120}\right)\left(a_{112}s^{\alpha_{112}} + a_{111}s^{\alpha_{111}} + a_{110}\right)}{\left(b_{111}s^{\beta_{111}} + b_{110}\right)\left(a_{122}s^{\alpha_{112}} + a_{121}s^{\alpha_{121}} + a_{120}\right)} \tag{18}$$

$$Gl_1(s) = -\frac{\left(b_{211}s^{\beta_{211}} + b_{210}\right)\left(a_{222}s^{\alpha_{222}} + a_{221}s^{\alpha_{221}} + a_{220}\right)}{\left(b_{221}s^{\beta_{221}} + b_{220}\right)\left(a_{212}s^{\alpha_{112}} + a_{211}s^{\alpha_{211}} + a_{210}\right)} \tag{19}$$

3.1   Controller Design

A fractional order two variable processes is considered. After applying the simplified decoupling to the two variable process, the interactions are present between two control loops has been eliminated. The FOPID control method discussed in Sect. 2 obtained the following expressions:

$$C_{11}(s) = K_{p1} + K_{i1}\frac{1}{s^{\lambda 1}} + K_{d1}s^{\mu 1} \tag{20}$$

$$C_{22}(s) = K_{p2} + K_{i2}\frac{1}{s^{\lambda 2}}K_{d2}s^{\mu 2} \tag{21}$$

In controller design, parameter tuning is of great importance because of changes in operating conditions. To solve this problem tuning the parameters based on maximum sensitivity approach, the controller values are obtained. It is defined as

$$M_s = \max_{\omega \in [0, +\infty]} \left| \frac{1}{1 + C(j\omega)G(j\omega)} \right| \tag{22}$$

4   Simulation Study

In this section, the TITO-FOIMC method will compare with the proposed method.
The model of simulink diagram is shown in Fig. 2.

4.1   Example 1

A FO-TITO process without time delay is considered as

$$G(s) = \begin{bmatrix} \frac{1.2}{2s^{0.5}+1} & \frac{0.6}{3s^{0.7}+1} \\ \frac{0.5}{s^{0.8}+1} & \frac{1.5}{3s^{0.6}+1} \end{bmatrix} \tag{23}$$

The new decoupler can be obtained as

$$Gl_{12}(s) = -\frac{2s^{0.5}+1}{2(1+3s^{0.7})} \tag{24}$$

$$Gl_{21}(s) = -\frac{3s^{0.6}+1}{3(1+s^{0.8})} \tag{25}$$

The TITO-FOIMC method is compare with the proposed method in this part. A maximum sensitivity based parameter tuning method is used to design the FOIMC controller. The maximum sensitivity value of loop1 is equal to 1, the value of M_{s2} for loop 2 is equal to 1. The transfer functions of TITO-IMC controller are given below:

$$C_{IMC-1}(s) = \frac{2s^{0.5}}{1.2(1 + 0.667s)} \qquad (26)$$

$$C_{IMC-2}(s) = \frac{3s^{0.6} + 1}{1.5(1 + 0.5843s)} \qquad (27)$$

The FO-PID controller is designed based on frequency domain tuning approach. By tuning the five parameters represented above.

The following FOPID controllers can be obtained as

$$C_{11}(s) = 10.7613 + \frac{33.307}{s^{0.8143}} \qquad (28)$$

$$C_{22}(s) = 10.1953 + \frac{30.723}{s^{0.7573}} \qquad (29)$$

The output behaviour of the FO-PID method is obtained in Fig. 3. At two instants at t = 0 s and at t = 4 s, a unit step input is applied and also for the two inputs, a disturbance of the magnitude 0.1 are added at t = 10 s. From Fig. 3 the simplified
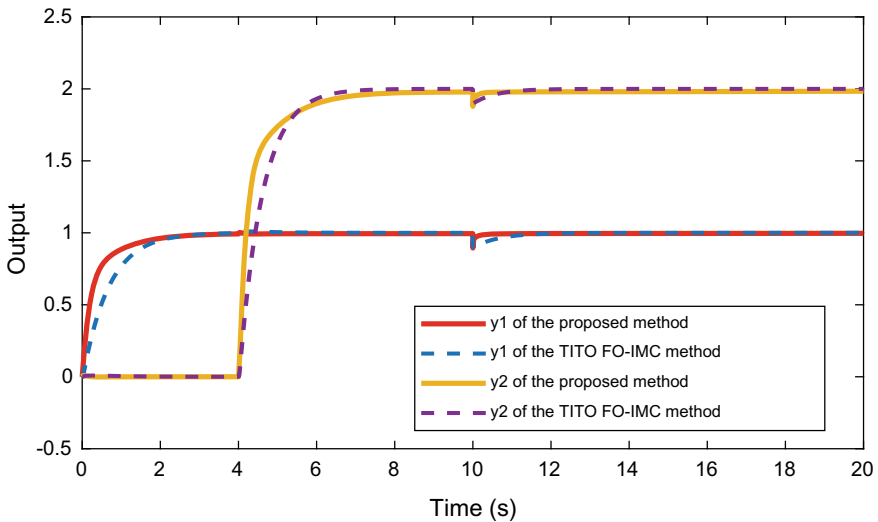
Fig. 3 Step response of FOIMC and the proposed method

Table 1 Time domain specifications of proposed and FO-IMC method

S. No	Control method		Rise time (t_r)	Peak time (t_p)	Settling time (t_s)	Overshoot (M_p)
1	FO-IMC method	Y1	1.4716	0.9985	4.1943	11.0936
		Y2	1.2911	0.002	20	10.5483
2	Proposed method	Y1	1.0648	0.5544	4.0333	10.1183
		Y2	1.1695	0	9.9982	10.0673

decoupling of the FO-PID controller exhibit less overshoot when compared to FO-IMC method. The output response of the proposed method reaches the steady state from the disturbance faster than FO-IMC method.

The time domain specifications are listed in Table 1. The values of rise time, settling time and peak overshoot have decreased when compare to the FO-IMC method. The presented method having quicker settling time and shows better results.

The parameter uncertainties (50%) are added to the process to test the robustness of the system. The following results are obtained in Figs. 4 and 5. In the case of model mismatch with parameter uncertainties, the presented method is more robust than FO-IMC method. The output response of the proposed method will not undergo any change for parameter uncertainties.

Sometimes external noises are included in the system while operation of process. In order to reject noise in the system and to exhibit the better results of the presented design, a Gaussian white noise is included to the step input. The result is obtained in Fig. 6. From the figure, the proposed method performs well in noise rejection.
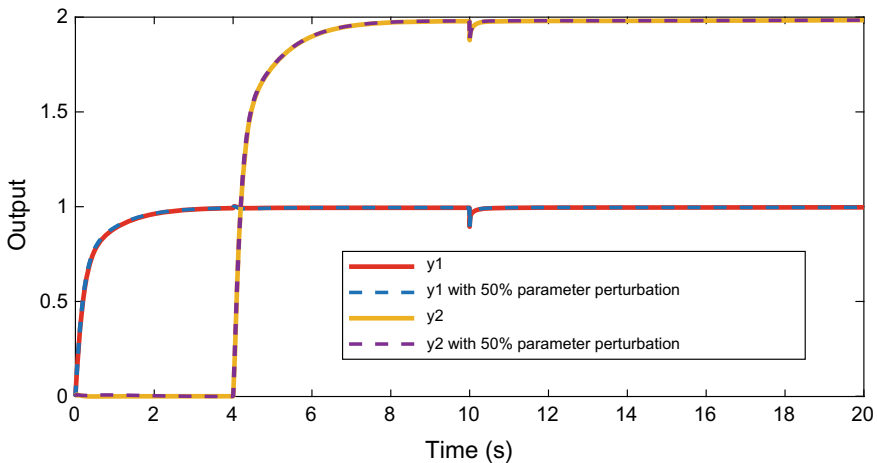
Fig. 4 Step response of the proposed method with 50% parameter perturbation
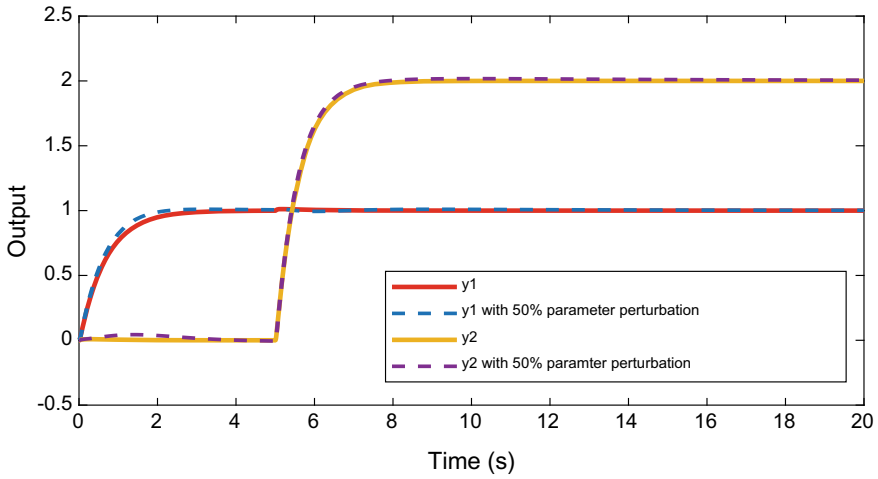
Fig. 5 Step response of FOIMC method with 50% parameter perturbation
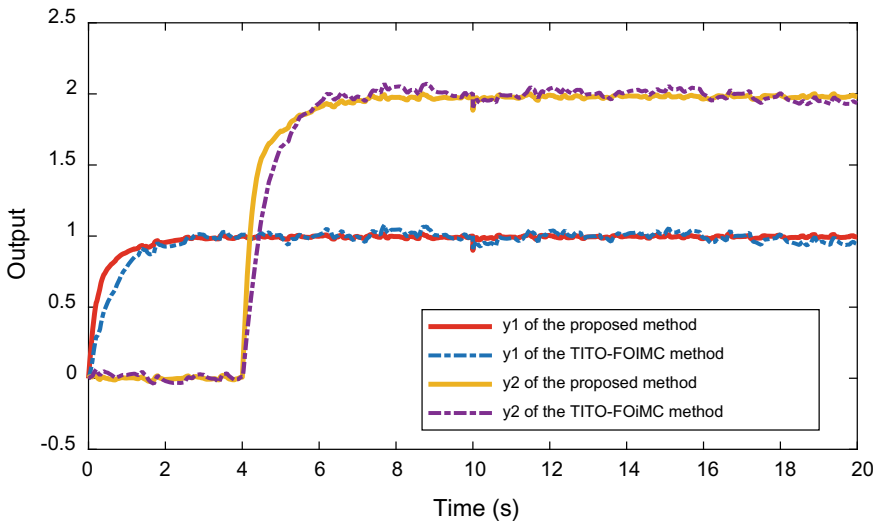
Fig. 6 Tracking and disturbance rejection performance of the proposed method with white noisy

4.2 Example 2

A FO-TITO process with time delay is considered:

$$G(s) = \begin{bmatrix} \frac{1.2e^{-0.2s}}{2s^{0.5}+1} & \frac{0.6e^{-0.3s}}{3s^{0.7}+1} \\ \frac{0.5e^{-0.4s}}{s^{0.8}+1} & \frac{1.5e^{-0.3s}}{3s^{0.6}+1} \end{bmatrix} \tag{30}$$

The new decoupler can be obtained as

$$Gl_{12}(s) = \frac{2s^{0.5} + 1}{2(1 + 3s^{0.7})} e^{-0.1s} \tag{31}$$

$$Gl_{21}(s) = -\frac{3s^{0.6} + 1}{3(1 + s^{0.8})} e^{-0.1s} \tag{32}$$

The TITO-FOIMC method is compare with the proposed method in this part. A maximum sensitivity based parameter tuning method is used to design the FOIMC controller. The Maximum sensitivity value of loop1 is equal to 1.2785. The value of M_{s2} for loop 2 is equal to 1.2557.

The transfer functions of TITO-IMC controller are given below:

$$C_{IMC-1}(s) = \frac{2s^{0.5} + 1}{1.2(1 + 0.6885s)} \tag{33}$$

$$C_{IMC-2}(s) = \frac{3s^{0.6} + 1}{1.5(1 + 0.6521s)} \tag{34}$$

The following FOPID controllers can be obtained as

$$C_{11}(s) = 1.2 + \frac{7.35}{s^{0.93}} \tag{35}$$

$$C_{22}(s) = 1 + \frac{5.12}{s^{0.81}} \tag{36}$$

The output behaviour of the FO-PID method is obtained in Fig. 7. At two instants at t = 0 s and at t = 4 s a unit step input is applied and also for the two inputs, a disturbance of the maitude 0.1 are added at t = 10 s. From Fig. 7 the simplified decoupling of the FO-PID controller exhibit less overshoot when compared to FO-IMC method. The output response of the presented method has attained steady state from the disturbance faster than FO-IMC method. The time domain specifications are listed in Table 2. The values of rise time, settling time and peak overshoot have decreased when compare to the FO-IMC method. The presented method having quicker settling time and shows better results.

The parameter uncertainties (50%) are added to the process to test the robustness of the system. The following results are obtained in Figs. 8 and 9. In the case of model mismatch with parameter uncertainties, the presented method is more robust than FO-IMC method. The output response of the proposed method will not undergo any change for parameter uncertainties.

Sometimes external noises are included in the system while operation of process. In order to reject noise in the system and to exhibit the better results of the presented
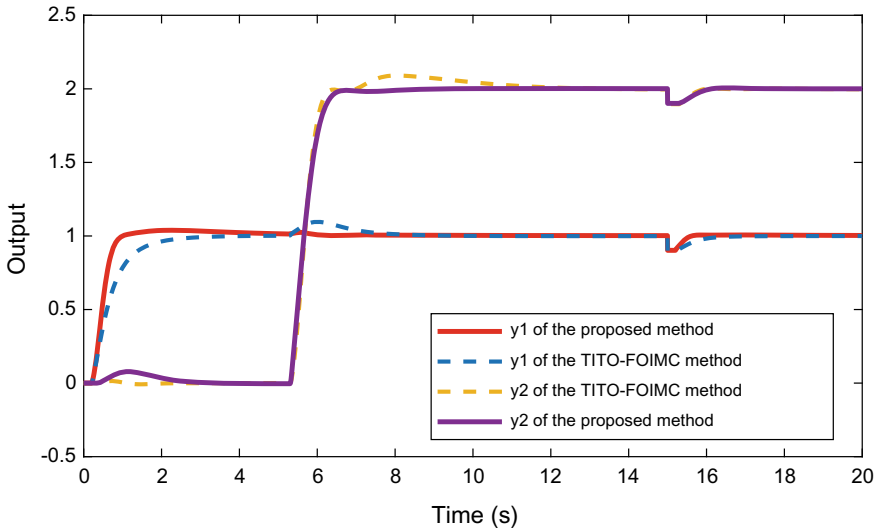
Fig. 7 Step response of TITO-FOIMC and the proposed method

Table 2 Time domain specifications of proposed and FO-IMC method for time delay process

S. No	Control method		Rise time (t_r)	Peak time (t_p)	Settling time (t_s)	Overshoot (M_p)
1	FO-IMC method	Y1	1.1056	6	15.9399	9.6176
		Y2	0.7162	16.5009	15.6956	4.6633
2	Proposed method	Y1	0.4491	2.2389	15.6025	3.5137
		Y2	0.5921	8.0387	9.9982	0.3095

design, a Gaussian white noise is included to the step input. The result is obtained in Fig. 10. From the figure, the proposed method performs well in noise rejection.

5 Conclusion

A simplified decoupling FO-PID controller is projected for FO-TITO process. Decoupling interconnection must introduce between controller and process so as to avoid the interactions within the two primary loops. The Fractional order PID is decoupled by simplified decoupler and determine the two controllers separately. Simulations results were proved that the proposed method does not undergo any change to parameter uncertainties. Whenever a disturbance is added to two inputs, the output response of the proposed method reaches steady state faster than FO-IMC method. The time domain specification of presented design method is much
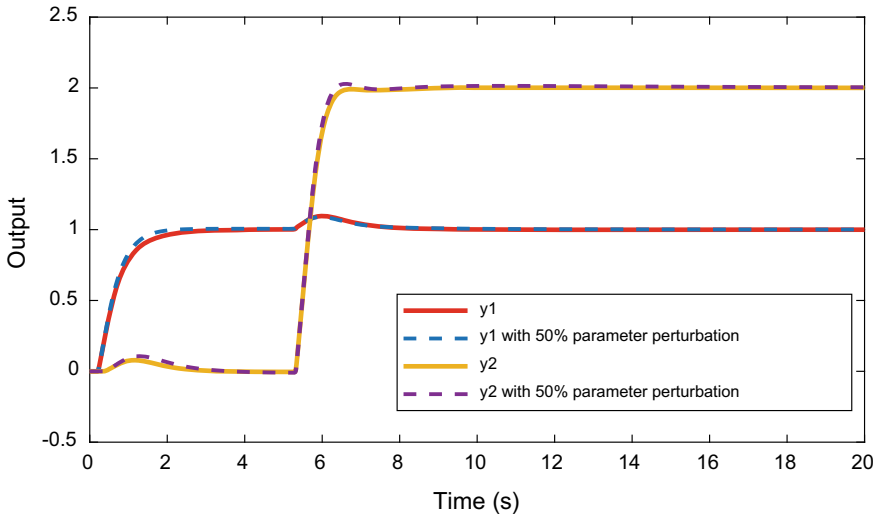
Fig. 8 Step response of the proposed method with 50% parameter perturbation with time delay
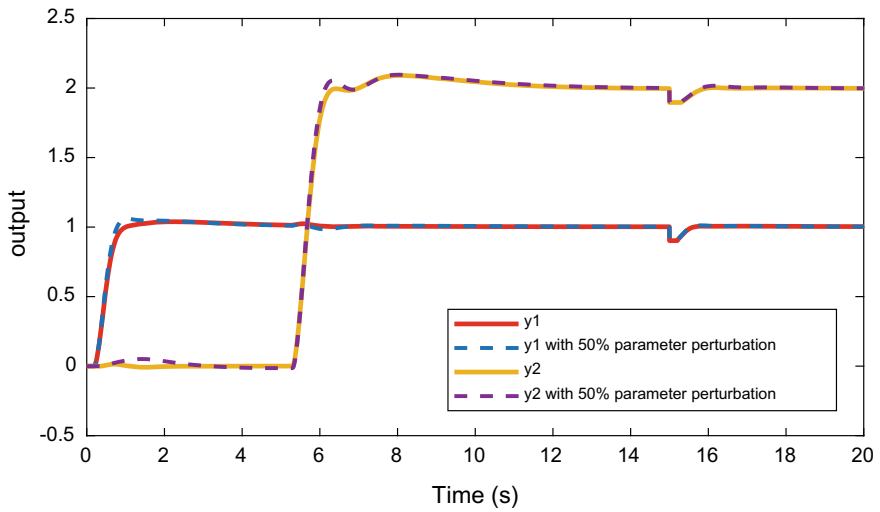
Fig. 9 Step response of the TITO-FOIMC method with 50% parameter perturbation with time delay

better than FO-IMC method. In few words, the FO-PID controller presented in this paper has produced robust performance for both set point-tracking and rejection in disturbance. The proposed method will extend to three variable FO-TITO process for future work.
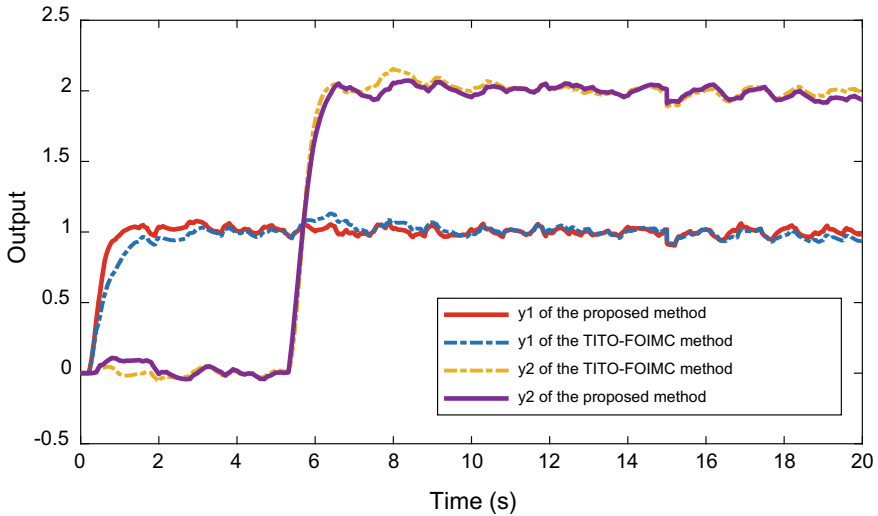
Fig. 10 Tracking and disturbance rejection performance of the proposed method with white noisy

References

1. Xue Y, Zhao H, Yang Q (2008) Self-tuning of PID parameters based on the modified particle swarm optimization, IEEE
2. Maiti D, Acharya A, Chakraborty M, Konar A (2005) Tuning PID and PID$^\delta$ Controllers using the integral time absolute error criterion, IEEE, Aug 2005
3. Monje CA, Vinagre BM, Feliu V, Chen Y (2008) Tuning and auto-tuning of fractional order controllers for industry applications. Control Eng Prac 16(7):798–812
4. Chen Y, Bhaskaran T, Xue D (2008) Practical tuning rule development for fractional order proportional and integral controllers. J Computat Nonlinear Dyn 3(2):021403
5. Zhao C, Xue D, Chen Y (2005) A fractional order PID tuning algorithm for a class of fractional order plants. In: 2005 IEEE international conference mechatronics and automation, vol 1. IEEE, pp 216–221
6. Podlubny I (1994) Fractional-order systems and fractional-order controllers. Inst Exp Phys Slovak Acad Sci, Kosice
7. Polubny (1999) Fractional-order systems and PI$^\lambda$D$^\mu$ controller. IEEE Trans Automatic Control 44:208–214
8. Monje CA, Vinagre BM, Feliu V (2008) Tuning andauto-tuning of fractional order controllers for industry applications. Control Eng Pract 16(7):798–812
9. Samko SG, Kilbas AA, Marichev OI (1993) Fractional integrals and derivatives. Theory and applications, Gordon and Breach, Yverdon
10. Li Z, Chen YQ (2014) Ideal, simplified and inverted decoupling of fractional order TITO processes. In: IFAC Proceedings Volumes 2014, vol 47, no 3, pp 2897–2902

Improved Centralized PID Controller with Disturbance Rejection for LTI Multivariable Processes

S. Anitha, R. Kiranmayi, K. Nagabhushanam, and K. Firdose Kowser Ahamadia

Abstract Decoupling control is a notable strategy for MIMO processes to dispose of the impacts of undesirable process interactions. It is a compelling instrument for MIMO process; outer disturbances actually have erase severe impacts on control execution of the closed loop structures. The extended 2-DOF control calculation for the MIMO measures in which the greater part of the controllers to decouple the frameworks first and afterward the controller is intended for the decoupled loops. This eventually builds the intricacy of the system design. This method requires the decoupled PID controller i.e., 'n' diagonal controllers for 'n' variable processes in addition to two off diagonal lead/lag filters for MIMO processes to achieve the desired performance. But the decoupler requires the exact mathematical model of the processes. If there is any errors occur in the modeling of physical processes, it may not give the satisfactory performance. To overcome these limitations of decoupled control system, a centralized PID controller is proposed. A centralized control system requires, 'n^2' PID controllers for 'n' variable processes will give better response irrespective of model errors. Among these, main diagonal controllers improve the performance and off diagonal reduce the combinability effects of variables in the MIMO processes. The PID controller will be tuned by frequency response approaches to reduce the steady state errors like IAE. The proposed method with disturbance approach in the loop will give improved performance than existing approaches.

Keywords Interaction · MIMO process variables · Decoupler · Centralized control · Diagonal loop · Off diagonal loop

1 Introduction

In present day design a controller in industry, vast number modules of mechanical plants were displayed LTI processes. As indicated by the quantity of variables present in the system, the LTI processes can be separated into SISO processes and Multi input

S. Anitha (✉) · R. Kiranmayi · K. Nagabhushanam · K. Firdose Kowser Ahamadia
Department of EEE, JNTUA College of Engineering (Autonomous), Ananthapuramu, Andhra Pradesh, India

and Multi Output (MIMO) processes. Because of the presence of associations and coupling in MIMO processes, there are a few problems to increase the difficulty of designing use of SISO structures into design of MIMO structures directly [1].

In multiple loop control approach, the MIMO techniques are treated as combinations of multiple-single controlling loops, and a main controller is organized and designed on each controlling loop by in view of loop variable associations [2]. Due to their reasonable sensitivity, and structure straight and direct forwardness, multiple loop controlling has been a lot of acknowledge by controlling loop control industry and broad attempt has been determined to get better the introduction of multiple-loop PI/PID controllers. Different control structure, such as, detuning factor procedures, chronological loop closing systems, self-governing control methods and decentralized control methodologies, etc., have been proposed consistently. Exactly when the associations along with different variables are nominal a multiple-loop PID/PI controller is conventionally enough, in such case, the off diagonal- decentralized control approach procedure may not succeed to give acceptable reaction if there be present genuine circle interchanges [3].

Decoupling (Off-diagonal) control is a popular method for MIMO processes to take out the impacts of process variable collaborations. This at last builds the difficulty in designing a controller. In decoupling (Off-diagonal) PID-like controller have intended for the MIMO systems directly with straightforwardness procedure as in case of SISO systems. The methodology describes how to decouple (off-diagonal elements) and control approach of the systems in a single step. Despite the fact that decoupling control is a viable approach in favour of MIMO type processes, outer disturbances actually have consequences for control execution of the structure. To acquire a superior in the interruption limination, disturbance observers (DOB) be presented [1]. DOBs have been recognized for their capacities in minimizing the effect of disturbance.

Accordingly, various logical decentralized control schemes are used to tune and to make sure system steadiness, which basis design procedure more simple and reduces costs.

For the MIMO structures with extraordinary loop variable associations, the decoupling (off-diagonal) control format shave been regularly utilized [4]. The decoupling (off diagonal) control as a rule necessitates two phases: (1) arrangement of decoupler off diagonal elements to restrict the interrelation among control loops, (2) preparation of the standard loop regulators for as a rule for carrying out control algorithm.

The important purposes of the decoupling (Off-diagonal) control approach are, (1) it allows the usage of SISO control strategies, (2) if there ought to be an event of sensor issues, it is decently effortless to adjust loop physically, adjust individual loop is reasonably exaggerated by the breakdown. Regardless, such an arrangement may outcome especially stable control composition, especially, when the structure estimation is high. Thus, the large numbers of investigators are generally focused on TITO structures.

For more number of control variables methods, Wang et al. [5]projected certain procedures for fully coupled PID or other than PID regulators with deliberate arrangement strategy and for the most part system structures [6]. Regardless, owing to

the mathematic assessment of the decouplers (off diagonal) for large number variable system, the vitality of like a control structure can't be definite. Subsequently, stationary decoupling come close to which can in a general sense improve control shows with their robustness like so as to of multiple-loop control structures are obtained beside the production to control loop relationship for large variable MIMO structures.

Decoupling systems only may not give the guaranteed stability in a control strategy. In spite of the way that static off diagonal element is simple but difficult to design and complete, they may possibly not commonly give sensible robustness of loop but there subsist significant interactions. However, a combination of decoupler with diagonal PID will give a centralized control scheme and will improve the response of the system.

In this projection, another off diagonal control procedure [7] is along with frequency response based PID controller is anticipated. The primary objective of this direct methodology is to get better the trouble dismissal when arrangement has serious outside disturbance influences, and to simplify the control algorithm.

The organization of the proposal is structured as: Section 2 discusses the illustration of IMC based centralized control system. Section 3 depicts the centralized control arrangement design using frequency response approach, Section 4 presents the simulation of case studies and projected algorithm justification, and summary are depicted in Sect. 5.

2 Representation of IMC Based Centralized Control System

A graphic description of broad and projected closed loop IMC control arrangement of multiple control variable strategies are showed up in Figs. 1 and 2 independently. In
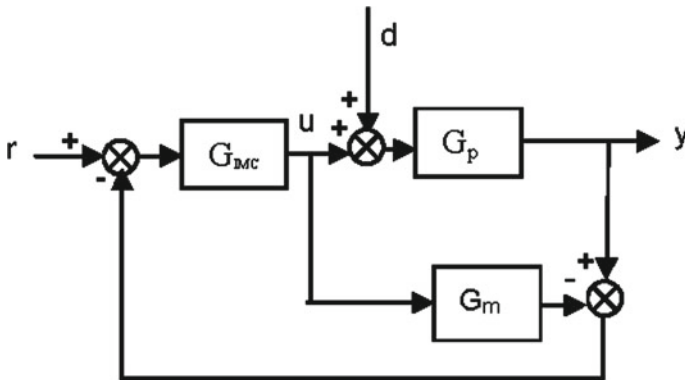
Fig. 1 General IMC structure with disturbance 'd'
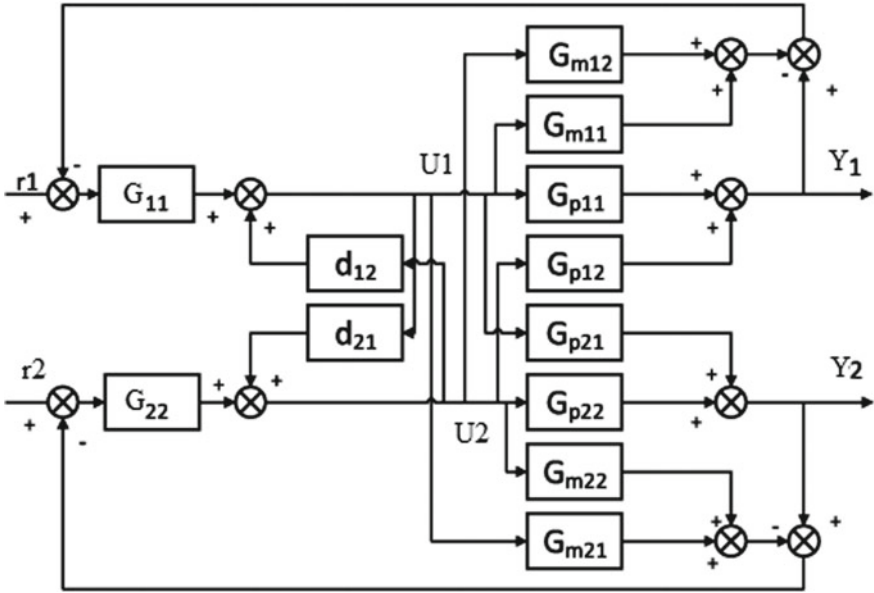
Fig. 2 Centralized PID Control system for TITO system

projected approach, the diagonal regulator is used for excusal of loop internal variable combinability and desirable set value/reference feed is for next the set value. The parts exist in projected control system are decentralized- main loop controller Gc(s), Gm(s) is inside model and process Gp(s).

For ease, it is acknowledged that the number of restricted variables in multiple variable techniques is comparable to numbered controlled components. This allows mixing of single (main) variable control with a single variable in the course of an investigation of controller.

In common the basic exhibition of the transfer function of the 2×2 (TITO) process is:

$$G_p(s) = \begin{bmatrix} G_{p_{11}}(s) & G_{p12}(s) \\ G_{p_{21}}(s) & G_{p22}(s) \end{bmatrix} \tag{1}$$

Where, the process $G_{pij}(s)$ is embedded as first order process with delay time (FOPDT), i.e.,

$$G_{pij}(s) = \frac{K}{\tau s + 1} e^{-\theta s} \tag{2}$$

The RGA of the Eq. (1) can be written as:

$$RGA = G(0) \otimes G(0)^{-T} \tag{3}$$

The arrangement of centralized (fully coupled) controller is of the outline as:

$$G_c(s) = \begin{bmatrix} G_{c_{11}}(s) & G_{c_{12}}(s) \\ G_{c_{21}}(s) & G_{c_{22}}(s) \end{bmatrix} \tag{4}$$

The relation amid input and output can be written as [9]:

$$Y(s) = G_p(s)U(s) \tag{5}$$

$$Y(s) = \begin{bmatrix} y_1(s) \\ y_2(s) \end{bmatrix}, \quad U(s) = \begin{bmatrix} u_1(s) \\ u_2(s) \end{bmatrix} \tag{6}$$

Here, Y(s), U(s) are specifies an output, input vectors of the process correspond-ingly. The large variable processing out variable association for the process with two variables is able to be given as:

$$Y_1(s) = G_{p,11}(s)u_1(s) + G_{p,12}(s)u_2(s) \tag{7}$$

$$Y_2(s) = G_{p,21}(s)u_1(s) + G_{p,22}(s)u_2(s) \tag{8}$$

In the TITO (two variable) structure, while the resulting loop is off (closed), the commitment from u_i to y_i has two different ways of sign transmission is feasible. The grouping of two conduction ways is measured as productive open control loop components. In case the later analysis controller would be in modified configuration, with $y_{s2} = 0$, by then the as a rule closed loop model everywhere in the scope of y_1 and u_1 is known by

$$\frac{y_1(s)}{u_1(s)} = G_{p11}(s) - \frac{G_{p12}(s)G_{p21}(s)G_{c2}(s)}{1 + G_{c2}(s)G_{p22}(s)} \tag{9}$$

These analytical expressions can be produced as:

$$\frac{y_1(s)}{u_1(s)} = G_{p11}(s) - \frac{G_{p12}(s)G_{p21}(s)(G_{c2}(s)G_{p22}(s))}{G_{p22}(1 + G_{c2}(s)G_{p22}(s))} \tag{10}$$

Likewise, for additional loops be as:

$$\frac{y_2(s)}{u_2(s)} = G_{p22}(s) - \frac{G_{p21}(s)G_{p12}(s)(G_{c1}(s)G_{p11}(s))}{G_{p11}(1 + G_{c1}(s)G_{p11}(s))} \tag{11}$$

In this projected two variable structures, open-loop components amid output controlled variable (y_i) and controlled variable (ui) not simply depend upon the looking at move limit model (g_{ii}) yet what's more depend upon various methods and regulators in each and every former controlling loop. It deduces the regulation of

single controller is incomprehensible independently and this depends upon various controllers. An abnormal relations of Eqs. (10) and (11) be able to modified by tolerating two notions: First, the perfect controller assume for the former loop (the output achieves steady and fast state with no transients) was used to smooth out the Eqs. (8 and 9), that is,

$$\hat{G}_{11}(s) = \frac{y_1(s)}{u_1(s)} = G_{p11}(s) - \frac{G_{p12}(s)G_{p21}(s)}{G_{p22}(s)} \tag{12}$$

Similarly, the other ETFs are,

$$\hat{G}_{22}(s) = \frac{y_2(s)}{u_2(s)} = G_{p22}(s) - \frac{G_{p12}(s)G_{p21}(s)}{G_{p11}(s)} \tag{13}$$

Here, G_{11} and G_{22} are the incredible open-loop effective transfer function (EOTF). These EOTFs are confounded models, and it is difficult to clearly use them for the controller structure. The decouplers are resolved for the off diagonal positions as:

$$D_{12}(s) = -\frac{G_{p,12}(s)}{G_{p,11}(s)} \tag{14}$$

$$D_{21}(s) = -\frac{G_{p,21}(s)}{G_{p,22}(s)} \tag{15}$$

The diagonal controller and off diagonal decouplers together forms a centralized control system.

3 Centralized Control System Design

The most of the industrial processes are represented by second order plus delay model as,

$$G_{ij}(s) = \frac{b_{0,ij}}{a_{2ij}s^2 + a_{1ij}s + 1} e^{-\theta_{ij}s} \tag{16}$$

Similarly, ETF is represented as,

$$\hat{G}_{ij}(s) = \frac{G_{ij}(0)}{a_{2ij}s^2 + a_{1ij}s + 1} e^{-\hat{\theta}_{ij}s} \tag{17}$$

The Multiloop decentralized controller would then be able to be easily planned by solitary loop advances dependent on relating ETFs. At this moment, we utilize the increase and stage edges approach. It is essential in light of the fact that the

recurrence reaction technique gives great execution despite indistinctness in together plant model and aggravations. The PID-type controller of every circle is assumed of accompanying standard structure:

$$G_{c,i}(s) = k_{p,i} + \frac{k_{i,i}}{s} + k_{d,i}s \tag{18}$$

The controller preserved as,

$$G_{c,i}(s) = \frac{k_{d,i}s^2 + k_{p,i}s + k_{i,i}}{s} \tag{19}$$

Then the open loop system model becomes,

$$G_{c,i}(s)\hat{G}_{ij}(s) = k\frac{\hat{G}_{ij}(0)}{s}e^{-\hat{\theta}_{ij}s} \tag{20}$$

Abbreviating the gain (GM) and phase margin (PM) provisions as $A_{m,i}$ and $\psi_{m,i}$, and their intersect frequencies as $\omega_{g,i}$ and $\omega_{p,i}$, correspondingly.

$$\arg[G_{c,i}(j\omega_{g,i})\hat{G}_{ij}(j\omega_{g,i})] = -\pi \tag{21}$$

$$A_{m,i}\left|G_{c,i}(j\omega_{g,i})\hat{G}_{ij}(j\omega_{g,i})\right| = 1 \tag{22}$$

$$\left|g_{c,i}(j\omega_{p,i})\hat{g}_{ii}(j\omega_{p,i})\right| = 1 \tag{23}$$

$$\psi_{m,i} = \pi + \arg[G_{c,i}(j\omega_{p,i})\hat{G}_{ij}(j\omega_{p,i})] \tag{24}$$

By simplification we obtain,

$$\omega_{g,i}\hat{\theta}_{ij} = \frac{\pi}{2} \qquad A_{m,i} = \frac{\omega_{g,i}}{k\hat{G}_{ij}(0)}$$

$$k\hat{G}_{ii}(0) = \omega_{p,i} \quad \psi_{m,i} = \frac{\pi}{2} - \omega_{p,i}\hat{\theta}_{ij}$$

Which results,

$$\psi_{m,i} = \frac{\pi}{2}\left(1 - \frac{1}{A_{m,i}}\right), \quad k = \frac{\pi}{2A_{m,i}\hat{\theta}_{ij}\hat{G}_{ij}(0)}$$

Table 1 Typical gain (GM) and Phase margins (PM)

$\psi_{m,i}$	$\pi/4$	$\pi/3$	$3\pi/8$	$2\pi/5$
$A_{m,i}$	2	3	4	5

As a result of this plan, the increase and phase (PM) and gain margins (GM) are consistent to one another, some conceivable phase and gain margin choices are provided in Table 1.

The controller considerations are obtained by,

$$\begin{bmatrix} k_{p,i} \\ k_{i,i} \\ k_{d,i} \end{bmatrix} = \frac{\pi}{2A_{m,i}\hat{\theta}_{ij}\hat{G}_{ij}(0)} \begin{bmatrix} a_{1,ij} \\ 1 \\ a_{2,ij} \end{bmatrix} \tag{25}$$

The complete design steps of proposed centralized control system are given below.

Step 1: Choose the minimum phase multivariable process and get the mathematical model of the process $G_p(s)$.

Step 2: Determine the control loops by using relative gains of the variables in the process as in Eq. (3).

Step 3: Approximate the mathematical model of equivalent processes and determine the decoupler elements in off diagonal positions as in Eq. (16).

Step 4: Design the PID controllers of main diagonal using the frequency response tuning method as given in Eq. (25).

Step 5: An exhibition of the projected unified control is assessed regarding indispensable of integral over time of working time, i.e., necessary of IAE (Integral Absolute Error), ISE (Integral Square Error) as given in Eqs. (25) and (26). The lesser the estimations of these IAE and ISE shows better exhibition of control algorithm.

The expressions for two variable-TITO process are:

$$IAE = \int_0^\infty (|E_1(t)| + |E_2(t)|)dt \tag{26}$$

$$ISE = \int_0^\infty (E_1^2(t) + E_2^2(t))dt \tag{27}$$

4 Simulation Results

Example 1: A Wood–Berry twofold refining area methodology is utilized as the two variable (TITO) systems that have(been broadly thought about by most of the experts. This technique can be made out of the trade work structure as given by Eq. (28).

$$G_p(s) = \begin{bmatrix} \frac{12.8e^{-3s}}{16.7s+1} & \frac{-18.9e^{-3s}}{21s+1} \\ \frac{6.6e^{-7s}}{10.9s+1} & \frac{-19.4e^{-3s}}{14.4s+1} \end{bmatrix} \tag{28}$$

Since it is 2×2 process, there two loops should be controlled. The control loops are determined by RGA method. Based on the outcome of RGA the pairing of control loops have selected. The RGA of the given process is,

$$RGA = \begin{bmatrix} 2.0094 & -1.0094 \\ -1.0094 & 2.0094 \end{bmatrix} \tag{29}$$

From the RGA [8, 9], it is concluded that the diagonal loops are need to have controller for improving the performance of the process. The effective models of the corresponding processes are:

$$G_{11}^{eff}(s) = \frac{y_1(s)}{u_1(s)} = \frac{12.8e^{-3s}}{16.7s+1} - \frac{18.9 \times 6.6(14.4s+1)e^{-7s}}{19.4(21s+1)(10.9s+1)} \tag{30}$$

$$G_{22}^{eff}(s) = \frac{y_2(s)}{u_2(s)} = \frac{-19.4e^{-3s}}{14.4s+1} - \frac{18.9 \times 6.6(16.7s+1)e^{-9s}}{12.8(21s+1)(10.9s+1)} \tag{31}$$

The parameters for controller are obtained as:

$$G_{c,proposed}(s) = \begin{bmatrix} \frac{1.3046s+0.0781}{10s+1} & 0 \\ 0 & \frac{-0.7422s-0.0515}{6s+1} \end{bmatrix} \tag{32}$$

An inverted decoupler of the process is designed and arranged in off diagonal to reduce the inter combinability among the variable of the process. This is given in equation below.

$$D(s)_{inverted} = \begin{bmatrix} 1 & \frac{(189s+18.9)e^{-2s}}{(21s+1)} \\ \frac{(-39.66s-6.6)e^{-6s}}{(10.9s+1)} & 1 \end{bmatrix} \tag{33}$$

For comparison, a decoupled control system with disturbance observer is used in this paper. Also the method proposed by Shen et al. [10] has been used. The simulation results of all these control method have shown. The first output of the proposed method including the comparison methods is shown in Fig. 3 and second output is shown in Fig. 4. Also the control signals are depicted in Figs. 5 and 6 correspondingly which compares the smoothness of the controller output.

For an intention of MATLAB simulation, a unit value change to the set or reference input assistance for the control pair loop 1 at time $t = 0$ s and for later (second) control loop 2 at time $t = 100$ s is incorporated. A step value change in disturbance signal is also further incorporated to the simulation at $t = 250$ s. The index of presentation of two control loops i.e., loop 1 and loop 2 with projected method and comparison approaches is provided in Table 2. This proves the projected control algorithm is
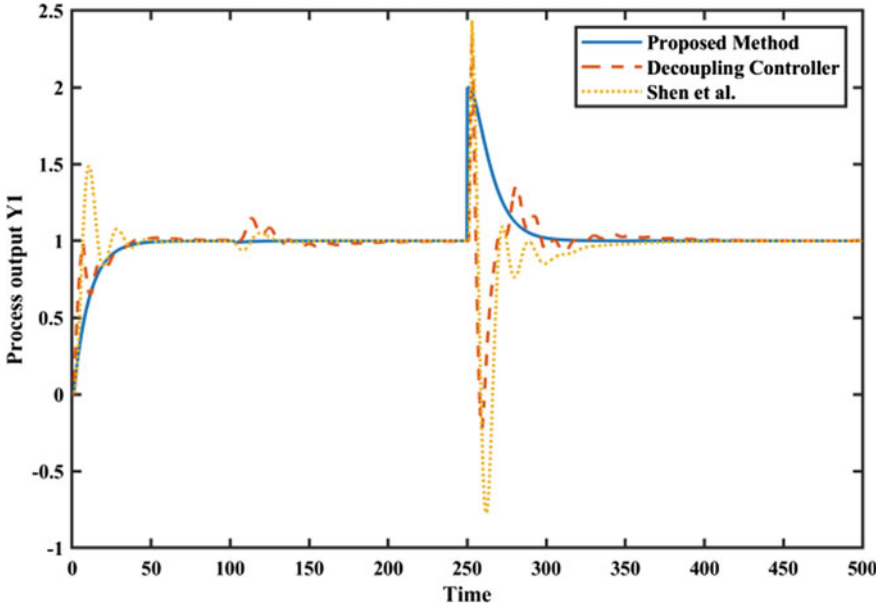
Fig. 3 Closed loop response of first output (Y_1) of WB column
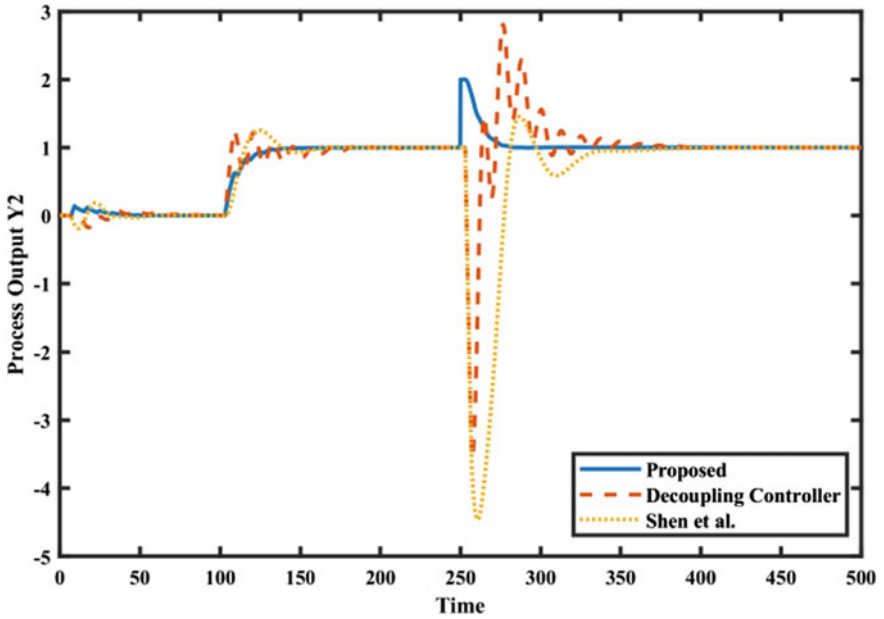
Fig. 4 Closed loop response of second output (Y_2) of WB column
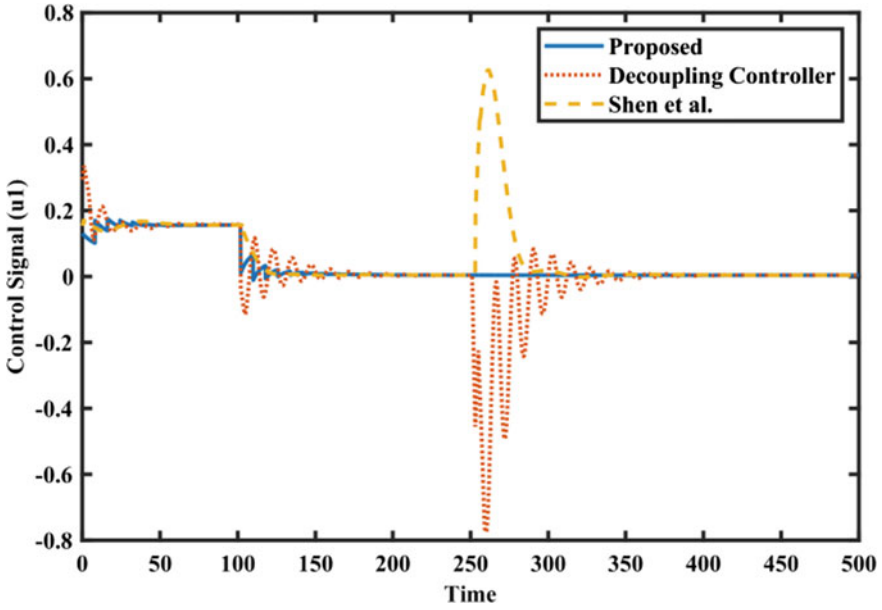
Fig. 5 Control signal of first loop ($Y_{s1} - U_1$) of WB column
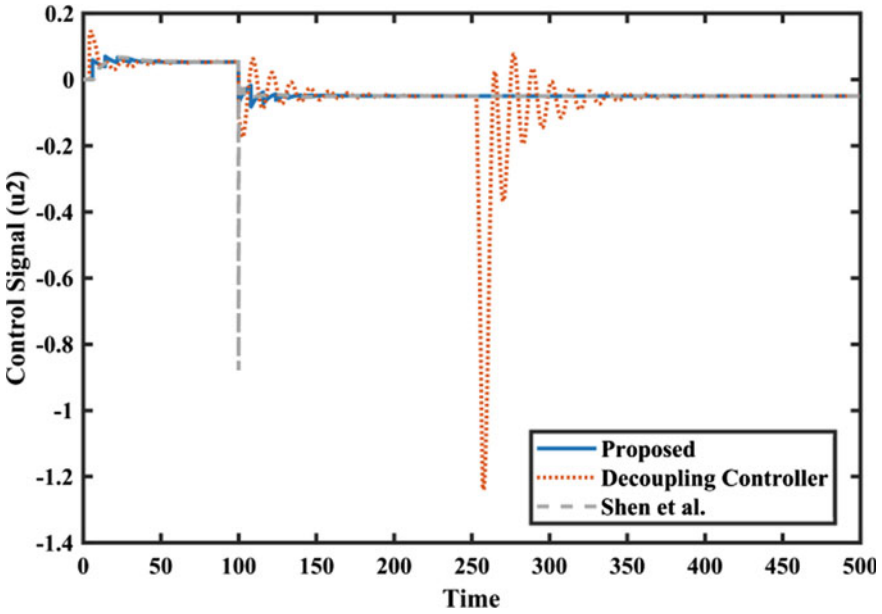
Fig. 6 Control response of secondary loop ($Y_{s2} - U_2$) of WB process

Table 2 Performance indices of WB

Controller method	Control loop	Set Point IAE	Disturbance IAE
Proposed Method	Y_1-U_1 Y_2-U_2	4.267 5.892	10.528 46.251
Decoupling Controller	Y_1-U_1 Y_2-U_2	6.576 8.355	14.768 54.737
Shen et al Method	Y_1-U_1 Y_2-U_2	12.789 13.962	27.137 97.860

enhanced the whole performance of the two variable process by attenuating influence of disturbance. *The proposed method* shows IAE for set point is 4.267 and 5.892 respectively for two channels. Similarly IAE for disturbance is 10.528 and 46.251 for the two channels respectively. These values show improvement as compared to existing method used for comparisons.

Example 2: The transfer function of the Shell Heavy Oil process (HOP) by [1] is known as:

$$G_p(s) = \begin{bmatrix} \frac{4.05e^{-27s}}{27s+1} & \frac{1.77e^{-28s}}{21s+1} \\ \frac{5.39e^{-18s}}{50s+1} & \frac{5.72e^{-14s}}{60s+1} \end{bmatrix} \tag{34}$$

Since it is 2×2 process, there two loops should be controlled. The control loops are determined by RGA method. Based on the outcome of RGA the pairing of control loops have selected. The RGA [9] of the given process is,

$$RGA = \begin{bmatrix} 1.7002 & -0.7002 \\ -0.7002 & 1.7002 \end{bmatrix} \tag{35}$$

From the RGA, it is concluded that the diagonal loops are need to have controller for improving the performance of the process. The effective models of the corresponding processes are:

$$G_{11}^{eff}(s) = \frac{y_1(s)}{u_1(s)} = \frac{4.05e^{-27s}}{27s+1} - \frac{1.77 \times 5.39(60s+1)e^{-32s}}{5.72(21s+1)(50s+1)} \tag{36}$$

$$G_{22}^{eff}(s) = \frac{y_2(s)}{u_2(s)} = \frac{5.72e^{-14s}}{60s+1} - \frac{1.77 \times 5.39(27s+1)e^{-32s}}{4.05(21s+1)(50s+1)} \tag{37}$$

The parameters for controller are obtained as:

$$G_{c,proposed}(s) = \begin{bmatrix} \frac{0.6663s+0.2469}{19s+1} & 0 \\ 0 & \frac{10.488s+0.1748}{26s+1} \end{bmatrix} \tag{38}$$

An inverted decoupler of the process is designed and arranged in off diagonal to reduce the inter combinability among the variable of the process. This is given in equation below.

$$D(s)_{inverted} = \begin{bmatrix} 1 & \frac{(-33.63s-1.77)e^{-2s}}{(60s+1)} \\ \frac{(-140.14s-5.39)e^{-4s}}{(50s+1)} & 1 \end{bmatrix} \tag{39}$$

For comparison, a decoupled control system with disturbance observer is used in this paper. Also the method ID IMC has been used. The simulation results of all these control method have shown. The first output of the proposed method including the comparison methods is shown in Fig. 7 and second output is shown in Fig. 8. Also the control signals are shown in Figs. 9 and 10 respectively which compares the smoothness of the controller output.

For an intention of MATLAB simulation, a unit value change to the set or reference input assistance for the control pair loop 1 at time $t = 0$ s and for later (second) control loop 2 at time $t = 200$ s is incorporated. A step value change in disturbance signal is also further incorporated to the simulation at $t = 800$ s.

The index of presentation of two control loops i.e., loop 1 and loop 2 with projected method and comparison approaches is provided in Table 3.

This proves the projected control algorithm is enhanced the whole performance of the two variable process by attenuating influence of disturbance.
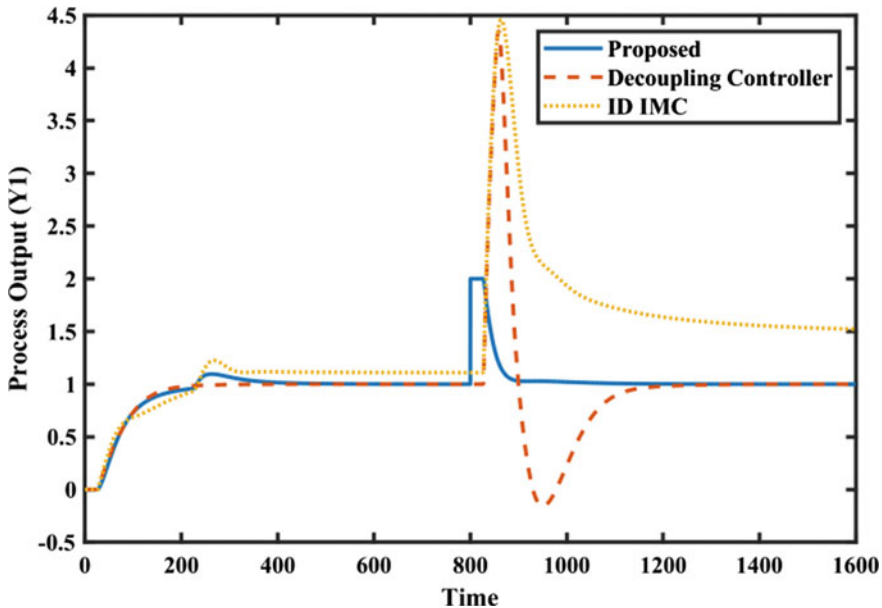
Fig. 7 Closed loop response of first output (Y_1) of HOP
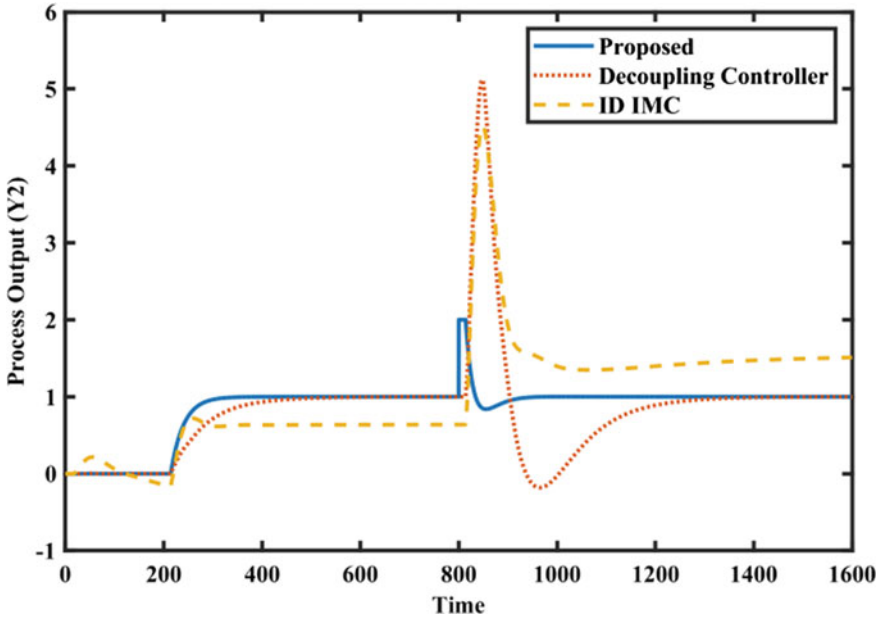
Fig. 8 Closed loop response of first output (Y_2) of HOP
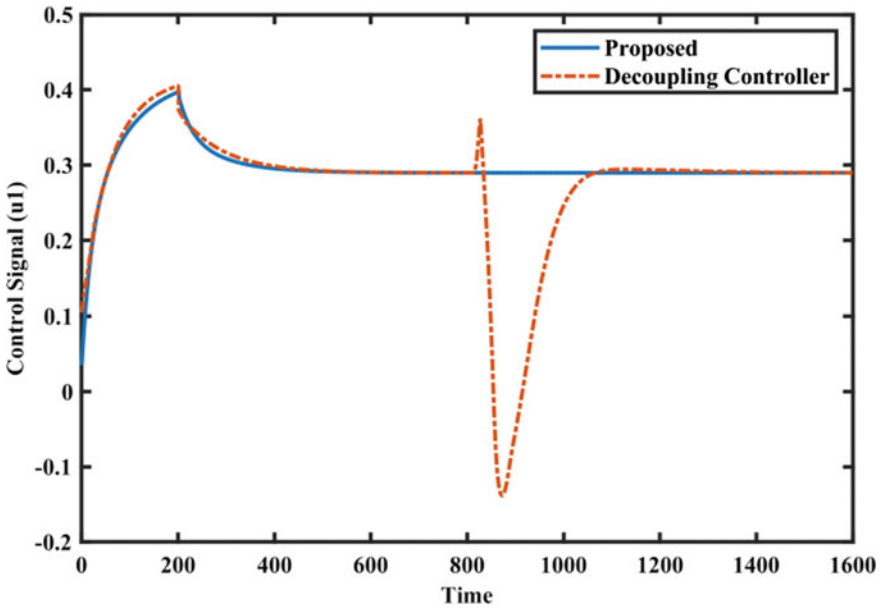
Fig. 9 Control signal of first loop ($Y_{s1} - U_1$) of HOP
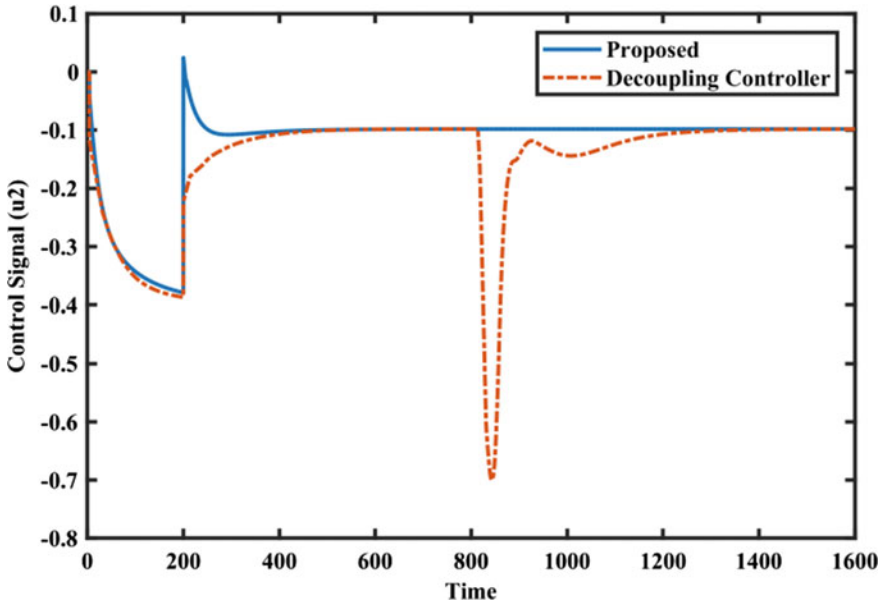
Fig. 10 Control signal of second loop ($Y_{s2} - U_2$) of HOP

Table 3 Performance indices of HOP

Controller method	Control loop	Set Point IAE	Disturbance IAE
Proposed Method	$Y_1 - U_1$ $Y_2 - U_2$	31.159 29.357	135.226 198.498
Decoupling Controller	$Y_1 - U_1$ $Y_2 - U_2$	40.741 31.148	203.237 215.139
ID IMC Method	$Y_1 - U_1$ $Y_2 - U_2$	47.280 41.659	267.481 443.77

The proposed method shows IAE for set point is 31.159 and 29.357 respectively for two channels. Similarly IAE for disturbance is 135.226 and 198.498 for the two channels respectively.

5   Conclusions

The present paper discusses the design of controller for multi variable process. The proposed method improves the set point response as well as disturbance response of the processes. The pairing of multivariable systems has been decided by relative gains of corresponding transfer function models of the given multivariable systems. The

performance of the system has improved by diagonal controllers and combinability among the different variables of the processes has been reduced by off diagonal controllers. The proposed method consists of diagonal and off diagonal controllers. The controllers of main diagonal are designed by using frequency response approach. Off diagonal are designed by inverted decoupled concept. The simulation results of proposed method have been compared with two similar approaches namely decoupler with disturbance observer and Shen et al. methods. The results are shows that the present approach improves the performance for both set point and disturbance effects. The proposed method is straight forward approach and it is simple to use in practical environment.

References

1. Hao Y et al (2020) Decoupling controller with disturbance observer for LTI MIMO systems. Asian J Control 22(2): 831–840
2. Wood RK, Berry MW (1973) Terminal composition control of a binary distillation column. Chem Eng Sci 28(9):1707–1717
3. Zhang W (2011) Quantitative process control theory, vol 45. CRC Press, Boca Raton
4. Xiong Q, Cai W-J, He M-J (2005) A practical loop pairing criterion for multivariable processes. J Process Control 15:741–747
5. Wang Q-G, Huang B, Guo X (2000) Auto-tuning of TITO decoupling controllers from step tests. ISA Trans 39(4): 407–418
6. Vu TNL, Lee M (2009) Independent design of Multiloop PI/PID controllers for Multi-delay processes, World Academy Science, Engineering and Technology (60):703–708
7. Tavakoli S, Griffin I, Fleming PJ (2006) Tuning of decentralised PI (PID) controllers for TITO processes. Control Eng Pract 14(9):1069–1080
8. Seborg DE et al (2010) Process dynamics and control. Wiley
9. Hanuma NR, Ashok Kumar DV, Anjaneyulu KSR (2014) Control configuration selection and controller design for multivariable processes using normalized gain. World Acad Sci Eng Technol Int J Electr Comput Electron Commun Eng 8(10)
10. Shen Y, Sun Y, Li S (2012) Adjoint transfer matrix based decoupling control for multivariable processes. Ind Eng Chem Res 51(50):16419–16426

Tuning of PID Controller Using SIMC Method for Systems with Time Delay and RHP Poles

P. Keerthi, R. Kiranmayi, K. Nagabhushanam, and K. Firdose Kowser Ahamadia

Abstract To control a class of system with RHP poles and time delay, a revised cascade control model is proposed in this work. This revised cascade control model aims to enhance both regulatory and servo performance of the system. Structurally, this proposed revised cascade control scheme has three controllers. To minimize the set point variations, an auxiliary controller is also used. An analytical method based on IMC design principle is used in designing the controller for set point tracing. It is developed using H_2 optimization and Internal Model Control theory. Disturbance rejecter is developed as P + I + D controller. To make the set point tracking controller physically realizable, a low pass filter is cascaded with it. Simplified Internal Modal Control (SIMC) with only one adjustable parameter 'T_c' is used for tuning the controllers. The simulation studies for proposed approach are done using MATLAB/SIMULINK to demonstrate its effectiveness. This proposed revised cascade control scheme depicts 3-Degree of Freedom Control Structure.

Keywords Time delay · Unstable processes · Cascade control · 3-DOF structure · P + I + D controller

1 Introduction

Processes with RHP poles are unstable in nature and are very difficult to control when compared to stable systems. This RHP pole may lead to large setting time and increased overshoot. In many systems, this unstable pole may also create an imbalance between the input and the output due to some external disturbance [1]. Therefore the study of the processes with RHP poles and time delay are carried out by some authors in past. In the initial stage, studies were focused mainly on tuning methods [2]. Chidambaram along with his co-workers proposed several techniques for systems having delay and RHP poles. They developed P + I controller using Ziegler-Nichols tuning formulae [3]. A formula for P and P + I controllers using

P. Keerthi (✉) · R. Kiranmayi · K. Nagabhushanam · K. Firdose Kowser Ahamadia
Department of EEE, JNTUA College of Engineering (Autonomous), Ananthapuramu, Andhra Pradesh, India

IMC principle is explained in [4]. To control a class of linear and non-linear systems, an Internal Model Controller cascaded with first order filter was suggested [5]. Anvil [6] developed a P + I + D controller using Internal Model Control theory and H_2 optimization technique for process with RHP poles. In [7, 8] revised form of SP structure based on 2-DOF control scheme is shown. But in general, one loop control method offers a satisfied output for systems with delay in few cases. The cascade control technique has an ability to reject the load disturbances very fast which enhances the closed loop performance of the system. Generally, there are two control loops in a cascade structure, one is secondary and the other is primary loop. Reducing the disturbances in the inner loop itself and not letting them enter into the primary loop is the main concept behind the cascade control technique [9]. A general cascade control scheme which was derived from Internal Model Control P + I + D method was proposed by Lee and Oh [10].

Cascade control structure alone might not give an effective regulatory control for systems with RHP poles whose control quality should be held at a peek level. Smith predictor (SP) can compensate the adverse effects of time delay. So combinations of cascade control scheme and SP have the ability to achieve a better control performance [11]. A cascade control scheme is derived from optimal Internal Model Control technique for control of systems with RHP poles with combination of controllers and filter was proposed by Dasari et al. [12]. In [13] the disturbance rejecter and set point tracking were developed using Internal Model Control. The main motive of the present study is to enhance both disturbance rejection and set point tracking. To achieve this we are using SIMC tuning method with only one adjustable parameter 'T_c'. This SIMC tuning method reduces recovery time and enhances the regulatory performance of the system.

2 Revised Cascade Control Model

The proposed revised cascade control scheme for systems consisting RHP poles is displayed in Fig. 1. In Fig. 1, P_{2m} and P_{1m} are the process models of the primary unstable systems and intermediate stable system, P_1 and P_2 respectively. P_{1mo} and P_{2mo} are the perfect transfer function models of P_{1m} and P_{2m} respectively without any time delay; i.e., $P_{1m}=P_{1mo}e^{-\theta_1 mS}$ and $P_{2m}=P_{2mo}e^{-\theta_2 mS}$

Proposed revised cascade control model has three controllers namely, the set point tracker C, inner loop load disturbance rejector F_1, and the outer loop load disturbance rejector F_2. To minimize the set point output variations an auxiliary controller set P_c is installed. F_1 rejects the disturbances that are introduced in the intermediate system and F_2 rejects the disturbances that are introduced in the primary systems. Note that, the auxiliary controller P_c has no effect on the controllers C, F_1 and F_2. This revised cascade control technique depicts the 3-DOF control scheme. This method offers a better control in the model parameter error case.

The intermediate system is stable in nature,
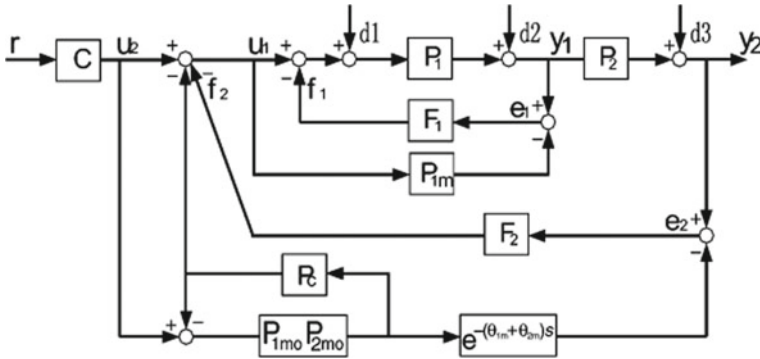
Fig. 1 Revised cascade control model

$$P_{1(s)} = K_1 e^{-\theta_{1s}}/(T_{1s}+1) \tag{1}$$

The primary system is unstable in nature,

$$P_{2(s)} = K_2 e^{-\theta_{2s}}/(T_{2s}-1) \tag{2}$$

3 Procedure for Controller Design

3.1 Stabilizing Controller P_c

The set point transfer function is derived from Fig. 1,

$$H_r = \frac{C P_1 P_2}{1 + P_C P_{1mo} P_{2mo}} * \frac{\left(1 + P_{1mo} P_{2mo} F_2 e^{-(\theta_{1m}+\theta_{2m})S}\right)\left(1 + P_{1m} F_1\right)}{1 + P_1 F_1 + P_1 P_2 F_2 + P_1 P_2 P_{1m} F_1 F_2} \tag{3}$$

Considering the ideal case, i.e., P_{1m} and P_{2m} are exact process models without any time delay. Therefore, set point transfer function is simplified

$$H_r = \frac{C P_1 P_2}{1 + P_c P_{1mo} P_{2mo}}$$

$$= \frac{K_1 K_2 C e^{-(\theta_1+\theta_2)S}}{T_1 T_2 S^2 + (T_2 - T_1)S + K_1 K_2 P_C - 1} \tag{4}$$

If auxiliary controller (P_c) wasn't introduced i.e., $P_c = 0$ then Eq. 4 is stable. Therefore, by using RH criterion for stability of its C.E, a realistic form of P_c is obtained. The selection of P_c can be done in two ways.

Case 1: $T_1 < T_2$, consider $P_c = K_c$ and $(K_c > 1/K_1 K_2)$ Hence, C.E of set point transfer function becomes $T_1 T_2 S^2 + (T_2 - T_1)S + K_1 K_2 K_c = 0$, is clearly stable as per RH Criterion for stability.

Case 2: $T_1 \;\; >= \;\; T_2$, consider $P_c \;\; = \;\; K_c + K_d S$, $(K_c > 1/K_1 K_2)\, K_d \;\; > \;\; (T_1 - T_2)/(K_1 K_2)$. Hence, C.E of set point transfer function becomes $T_1 T_2 S^2 + (K_1 K_2 K_d + T_2 - T_1)S + K_1 K_2 K_c - 1 = 0$ which is stable as per RH criterion for stability.

Note that, P_c can designed as a regular $P + I + D$ controller but it may result in complex tuning procedure. So, it is not suggestible.

3.2   Set Point Tracker C

If the set point tracker C is developed as stable and rational then there exist zero dead time element in the denominator of Eq. 4, which may yield a better set point response. We use H_2 performance measure min $\|e\|_2^2$ i.e., the integral square error (ISE) in designing optimal set point tracker C, which should meet the min $\|W(1 - H_r)\|_2^2$ condition.

Here W represents set point input weight function.

Using Pade's approximation for delay term in Eq. 4 gives

$$e^{-(\theta_1 + \theta_2)S} = Q_{nn}[-(\theta_1 + \theta_2)S]/Q_{nn}[(\theta_1 + \theta_2)S]$$

where,

$$Q_{nn}[(\theta_1 + \theta_2)S] = \sum_{j=0}^{n} \frac{(2n - j)!n!}{(2n)!j!(n - j)!} * [(\theta_1 + \theta_2)S]^j$$

There are 2 cases to perform the controller design procedure:

Case 1: $T_1 < T_2$ consider $P_c = K_c$ mentioned in above. Therefore, from Eq. 4

$$\|W(1 - H_r)\|_2^2 = \left\| \frac{1}{S}\left(1 - \frac{K_1 K_2 C(S) Q_{nn}[-(\theta_1 + \theta_2)S]}{\left[T_1 T_2 S^2 + (T_2 - T_1)S + K_1 K_2 P_c - 1\right]Q_{nn}[(\theta_1 + \theta_2)S]}\right)\right\|_2^2$$

$$= \left\| \frac{Q_{nn}[(\theta_1 + \theta_2)S]}{S Q_{nn}[-(\theta_1 + \theta_2)S]} - \frac{K_1 K_2 C(S)}{S\left[T_1 T_2 S^2 + (T_2 - T_1)S + K_1 K_2 P_c - 1\right]}\right\|_2^2$$

As $Q_{nn}(0)$ = 1 and all zeroes in $Q_{nn}[-(\theta_1 + \theta_2)S]$ term are in complex RHP, using orthogonality property in H_2 norm

$$\|W(1 - H_r)\|_2^2 = \left\| \frac{Q_{nn}[(\theta_1 + \theta_2)S] - Q_{nn}[-(\theta_1 + \theta_2)S]}{SQ_{nn}[-(\theta_1 + \theta_2)S]} \right\|_2^2$$

$$+ \left\| \frac{T_1T_2S^2 + (T_2 - T_1)S + K_1K_2K_C - 1 - K_1K_2C(S)}{S[T_1T_2S^2 + (T_2 - T_1)S + K_1K_2K_C - 1]} \right\|_2^{2.}$$

In order to design the ideal optimal controller, let its second term be zero.

$$C_{1m(s)} = \frac{T_1T_2S^2 + (T_2 - T_1)S + K_1K_2K_c - 1}{K_1K_2}$$

Since, $C_{1m(s)}$ is improper function. To make it physically realizable, low-pass filter $I_c(S) = \frac{1}{(T_cs+1)^2}$ is cascaded with it. Therefore, a practical optimal controller is derived as

$$C(S) = \frac{T_1T_2S^2 + (T_2 - T_1)S + K_1K_2K_C - 1}{K_1K_2(T_{CS} + 1)^2} \tag{5}$$

where T_c—adjustable term.

C attains the optimality only when T_c is adjusted to zero.

Case 2: $T_1 \geq T_2$ follows a similar deriving method gives,

$$C(S) = \frac{T_1T_2S^2 + (K_1K_2K_d + T_2 - T_1)S + K_1K_2K_c - 1}{K_1K_2(T_{cS} + 1)^2} \tag{6}$$

Remark 1: Substitute Eq. 5 or 6, (if $T_1 >= T_2$) in Eq. 4, then practical form of set point transfer function is derived as,

$$H_r(S) = \frac{1}{(T_{cs} + 1)^2} e^{-(\theta_1 + \theta_2)S} \tag{7}$$

Applying Inverse Laplace Transform,

$$Y_r(t) = \begin{cases} 0 & , t \leq \theta_1 + \theta_2 \\ 1 - \left(1 + \frac{t}{T_c}\right)e^{-(t-\theta_1-\theta_2)/T_c} & , t > \theta_1 + \theta_2 \end{cases} \tag{8}$$

There exist zero overshoot in ideal set point output and by tuning T_c quantitative time domain specifications are achieved.

From Eq. 8, tuning formula can be figured as $t_r = 3.8897T_c + \theta_1 + \theta_2$ where t_r is the rising time.

From Eqs. 7 and 8, it is clear and obvious that P_c does not affect the set point tracker C. In ideal case, i.e., when T_c is adjusted to zero, the set point response $H_r(S) =$

$e^{-(\theta_1+\theta_2)S}$ i.e., system output reaches the desired value for set point immediately after the overall systems' delay time. This single adjustable parameter T_c gives the trade off among the ideal performance of set point output and the output capacity of set point tracker C. Therefore, when T_c is adjusted to small value, the setpoint tracing becomes faster and the output of C is larger resulting in more aggressive dynamic behaviour. When T_c is adjusted to a higher value, the set point tracing becomes slow, and the output of C is small resulting in less aggressive dynamic behaviour. Therefore, it is suggested to adjust 'T_C' to a value which approximately equals the system delay initially. If tuning is unsatisfactory, then decrease/increase T_c to achieve required set point response.

3.3   Disturbance Rejector F_1

In order to make the design procedure simpler, leave the primary outer loop and consider the inner loop load disturbance output independently. This results in degradation of actual system performance, due to load disturbance response oscillation may seep into the inner loop.

From Fig. 1, the ideal load disturbance process models of inner loop are given as

$$H_{d1}(S) = \frac{y_1}{d_1} = \frac{P_1(S)}{1 + F_1(S)P_1(S)} \tag{9}$$

$$H_{d2}(S) = \frac{y_1}{d_2} = \frac{1}{1 + F_1(S)P_1(S)} \tag{10}$$

The inner loop complementary sensitivity function is termed as,

$$T_{d-inner} = \frac{f_1}{d_1} = \frac{F_1(S)P_1(S)}{1 + F_1(S)P_1(S)} \tag{11}$$

The required ideal complementary sensitivity function is $T_{d-inner} = e^{(-\theta_{1s})}$.

It means when a load disturbance d_1 is continously introduced in P_1, the load disturbance rejector F_1 must observe the resultant P_1 response error immediately after system time delay θ_1 and produce equal and opposite signal f_1 to oppose the injected load disturbance.

The practical required complementary sensitivity function relied on combined H_2 optimal performance measure of Internal Model Control theory is given

$$T_{d-inner}(S) = \frac{1}{T_{f_1s}+1}e^{(-\theta_{1s})} \tag{12}$$

where, T_{f_1} is as adjustable parameter.

From Eqs. 11 and 12,

$$F_1(S) = \frac{T_{d-inner}(S)}{1 - T_{d-inner}(S)} \frac{1}{P_1(S)} \tag{13}$$

From Eq. 13,

If T_{f_1} is tuned to be small, the load disturbance rejection in the inner loop becomes faster, the output of F is larger and vice-versa. Therefore, T_{f_1} aims the trade-off among load disturbance rejection capability and output capacity of F.

3.4 Disturbance Rejector F_2

From Fig. 1, the load disturbance model of outer loop is given as

$$H_{d3}(S) = \frac{y_2}{d_3} = \frac{1}{1 + F_2(S) P_1(S) P_2(S)} \tag{14}$$

The outer loop complementary sensitivity function is derived as,

$$T_{d-outer}(S) = \frac{f_2}{d_1} = \frac{F_2(S) P_1(S) P_2(S)}{1 + F_2(S) P_1(S) P_2(S)} \tag{15}$$

Similar analysis as in previous subsection is carried out to find required closed-loop complementary sensitivity function in practical form.

$$T_{d-\text{outer}}(S) = \frac{aS + 1}{\left(T_{f_2} S + 1\right)^3} e^{-(\theta_1 + \theta_2) S} \tag{16}$$

where T_{f_2} is adjustable parameter and a is obtained by an asymptotic constraint.

$$\lim_{S \to \frac{1}{T_2}} H_{d3}(S) = 0$$

To reject the step load disturbances, the following asymptotic constraint is required

$$\lim_{S \to \frac{1}{T_2}} (1 - T_{d-\text{outer}}(S)) = 0 \tag{17}$$

Substitute Eq. 16 in Eq. 17,

$$\lim_{S \to \frac{1}{T_2}} \left(1 - \frac{aS + 1}{\left(T_{f_2} S + 1\right)^3} e^{-(\theta_1 + \theta_2) S} \right) = 0$$

On solving above equation,

$$a = T_2 \left[\left(\frac{T_{f_2}}{T_2} + 1 \right)^3 e^{(\theta_1 + \theta_2)/T_2} - 1 \right] \tag{18}$$

from Eqs. 15, 16 and 18, the required load disturbance is obtained as,

$$F_2(S) = \frac{1}{K_1 K_2} \frac{(T_1 S + 1)(T_2 S - 1)(aS + 1)}{\left(T_{f_2} + 1 \right)^3 - (aS + 1)e^{-(\theta_1 + \theta_2)S}} \tag{19}$$

At $S = 1/T_2$, Eq. 19 has a zero-pole cancellation on Right Hand Plane, this may result in unreliable working of F_2.

So, Maclaurin expansion formulae are used to regenerate the ideal disturbance rejector.

Say $F_2(S) = M(S)/S$,

$$F_2(S) = \frac{1}{m} \left[M(0) + M'(0)S + \frac{M''(0)}{2!} S^2 + . + \frac{M^i(0)}{i!} S^i \right] \tag{20}$$

Considering the first three terms of Eq. 20 resemble a standard $P + I + D$ controller

$$F_{2-PID}(S) = K_F + \frac{1}{T_1 S} + T_D S \tag{21}$$

where $K_F = M'(0)$, $T_1 = 1/M(0)$, $T_D = M'(0)/2$.

First order low pass filter is cascaded with F_2 in Eq. 21 to implement it physically.

In [14], third order approximation controller using Pade series expansion is given as,

$$F_{2-3/3}(S) = \frac{d_3 S^2 + d_2 S + d_1}{C_2 S^2 + C_1 S + 1} + \frac{d_0}{S(C_2 S^2 + C_1 S + 1)} \tag{22}$$

The 1st part is a lead-lag controller of second order, whereas the 2nd part represents an integrator cascaded with second order low pass filter.

4 Robust Stability Analysis

Considering the formulae developed in above section, system stability is determined.

Practically, the frequently occurred process uncertainties are systems parameters perturbation, the response uncertainties of actuator and measurement errors of the sensors.

According to small gain theorem, a system is robust stable only if,

$$\| \Delta m(S) T_d(S) \|_\infty < 1 \tag{23}$$

Δm—Multiplicative system uncertainty bound,
T_d—Complementary sensitivity function.

To find the robust stability constraint of inner loop, substitution Eq. 12 in Eq. 23 yield the robust stability constraint.

$$\left\| \frac{\Delta m_1(S)}{T_{f_1} S + 1} \right\|_\infty < 1 \tag{24}$$

$\Delta m_1(S)$ multiplicative uncertainty bound of P_1.

Say $\Delta m_1(S) = \Delta K_1 / K_1$, the robust stability constraint of T_{f_1} is

$$\sqrt{T_{f_1}^2 w^2 + 1} > \frac{|\Delta K_1|}{K_1}, \forall w > 0 \tag{25}$$

$$\Delta m_1(S) = e^{-(\Delta \theta_1)S} - 1$$

$\Delta \theta_1$ Time delay uncertainty of P_1 To be robust stability, tune T_{f_1} as below

$$\sqrt{T_{f_1}^2 w^2 + 1} > \left| e^{-(j \Delta Q w)} - 1 \right|, \forall w > 0 \tag{26}$$

Substitute Eqs. 16 and 18 in Eq. 23 which gives the robust stability constraint.

$$\left\| \frac{\left[T_2 \left(\frac{T_2}{T_2} \right) + 1 \right]^3 e^{(\theta_1 + \theta_2)/(T_2 - 1)S} + 1}{\left(T_{f_2} S + 1 \right)^3} \right\|_\infty < \frac{1}{\| \Delta m_2(S) \|_\infty} \tag{27}$$

Δm_2—Multiplicative uncertainty bound of P_1, P_2.

Increasing T_{f_2} strengthens the robust stability but reduces performance of disturbance rejection and vice-versa. It is advised to initially adjust T_{f_2} is almost equal to overall process time delay and later decrease/increase T_{f_2} until the desired response and robust stability are achieved.

5 Simulation Examples

Three unstable cascade processes are simulated. For numerical comparison, performance measure Total Variation (TV), Integral Absolute Error (IAE) and Peak Value (PV) of the system output are considered.

$$IAE = \int_0^\infty |e(t)| dt$$

$$\text{TV} = \sum_{n=0}^{N} |u_{n+1} - u_n|$$

Example1: Consider an example studied by Cheng-Qiang Yin in [13]. The primary process is $P_2 = \frac{e^{-3S}}{10S-1}$ and the intermediate system is $P_1 = \frac{2e^{-2S}}{S+1}$ Method suggested by Cheng-Qiang Yin [13] is considered for comparison. For the Proposed method, $F_1 = S + 1/2S + 2, F_2 = 0.8228 + (0.8228/46.522S) + 0.8228 * 1.4649S$, $P_C = 0.6$ and $C = \frac{100S^2+4S+1}{160S^2+80S+10}$.

Considering the above controller values, the performance of the two techniques are simulated by enabling unit step change at t = 1 s in the set point and applying negative step change at t = 150 s in disturbance d_1 and negative step change at t = 350 s in disturbance d_2 respectively. System output and control signals are depicted in Figs. 2 and 3 for perfect model case.

To analyse the system robustness, assume a 20% raise in intermediate process gain along with 20% raise in intermediate delay time. The system output is given in Fig. 4 and control action response is given in Fig. 5. On observing Figs. 4 and 5, the
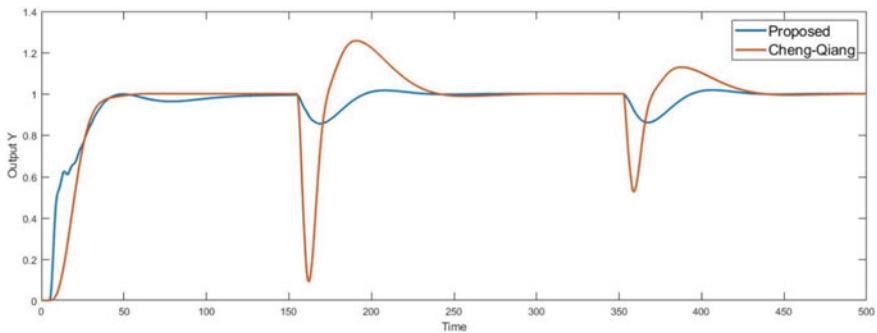
Fig. 2 System output for Example 1 (perfect)
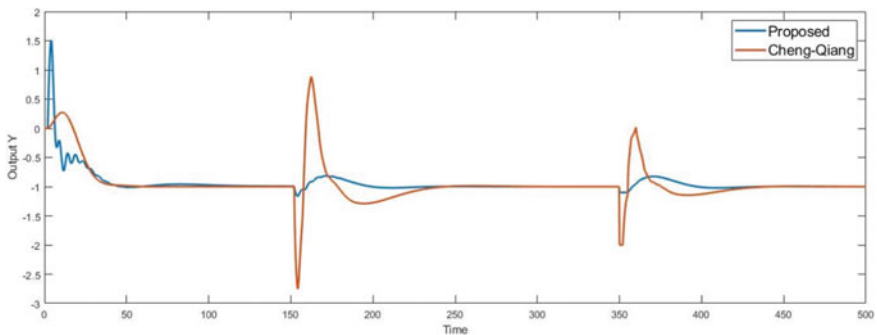
Fig. 3 Control signals of Example 1 (perfect)
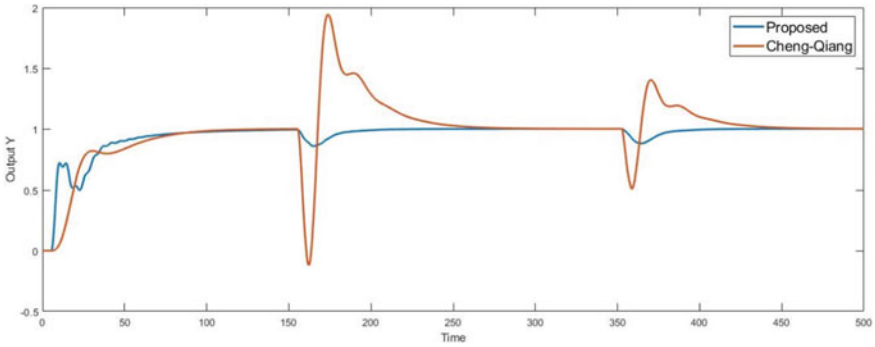
Fig. 4 System output for Example 1 (error).
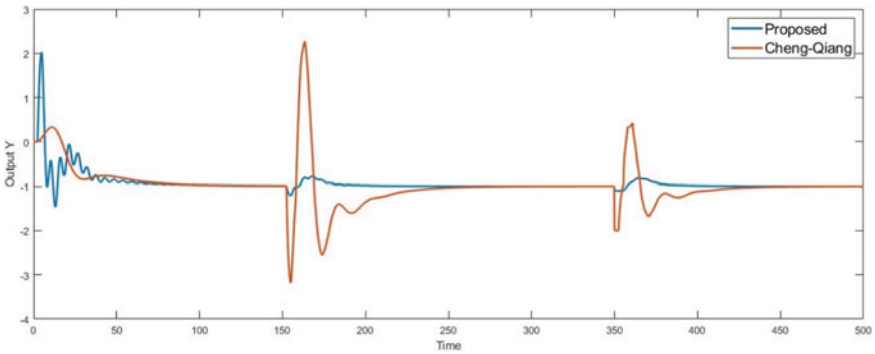
Fig. 5 Primary control signals of Example 1 (error)

proposed technique provides gives better closed loop response and the control action undergoes several oscillations at first which later becomes smoother obviously. The performance measures for proposed method and other method are tabulated below for both perfect and perturbed cases. From Table 1 the proposed method has lower IAE and TV than Cheng-Qiang technique. Therefore, the proposed technique gives an enhanced performance than the Cheng-Qiang technique.

Table 1 Performance measures of Example 1

Method	IAE	PV	TV
Perfect model parameter			
Cheng-Qiang	47.4591	1.01	39.65
Proposed	25.1821	1.26	31.45
Error model parameter			
Cheng-Qiang	55.81	1.52	62.74
Proposed	27.6376	1.0	59.28

Example 2: Another example studied by Cheng-Qiang Yin in [13] is considered whose primary process is $P_2 = \frac{e^{-4S}}{20S-1}$ and the intermediate process is $P_1 = \frac{2e^{-2S}}{20S+1}$ In proposed technique, $F_1 = \frac{20S+1}{m_S+2} F_2 = \frac{35.1847S^2+24.8678S+1.5707}{2.56S^2+0.8655S+1} + \frac{0.0208}{S(2.56S^2+0.8655S+1)}$, $P_C = 1$, and $C = \frac{400S^2+2S+1}{72S^2+24S+2}$.

Considering above controller parameters, the performance of two techniques are simulated by enabling a unit step change at t = 1 s in the set point and applying a negative unit step change at t = 150 s in disturbance d_1 and negative unit step change at t = 350 s in disturbance d_2 respectively. The system output and control action response are given in Figs. 6 and 7 for perfect model case. Therefore proposed method gives a better performance than Cheng-Qiang method for disturbance rejection.

To analyse robust stability of the system, assume a +10% error in intermediate process gain and delay and +20% error in primary process gain and delay. Its outputs are shown in Figs. 8 and 9. On observing Figs. 8 and 9, oscillations exists in both system responses and control actions.

The performance characteristics for two methods in both perfect and perturbed case are tabulated in Table 2.
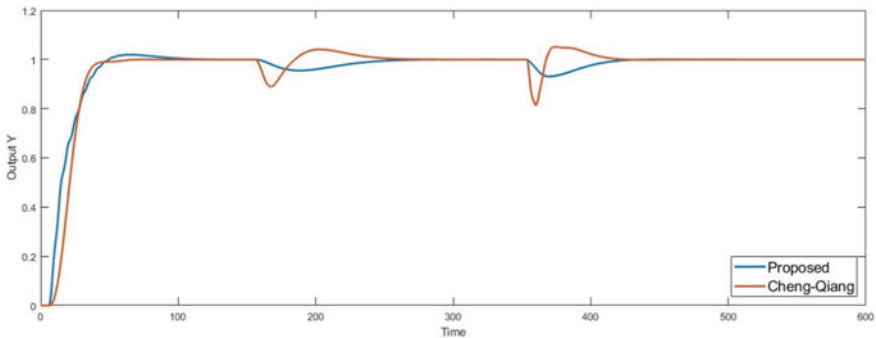
Fig. 6 System output for Example 2 (perfect)
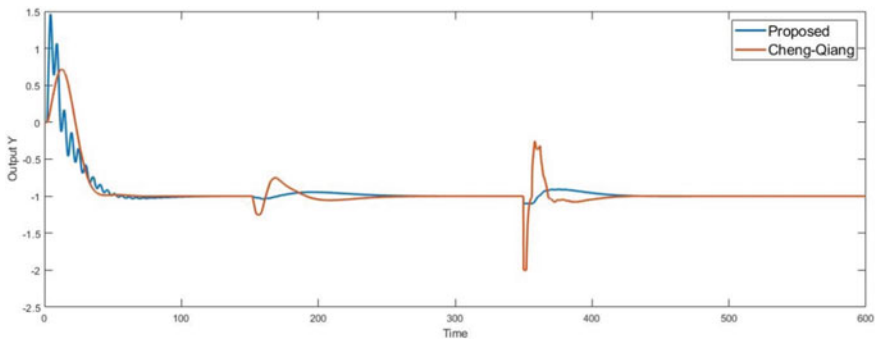
Fig. 7 Control signals of Example 2 (perfect)
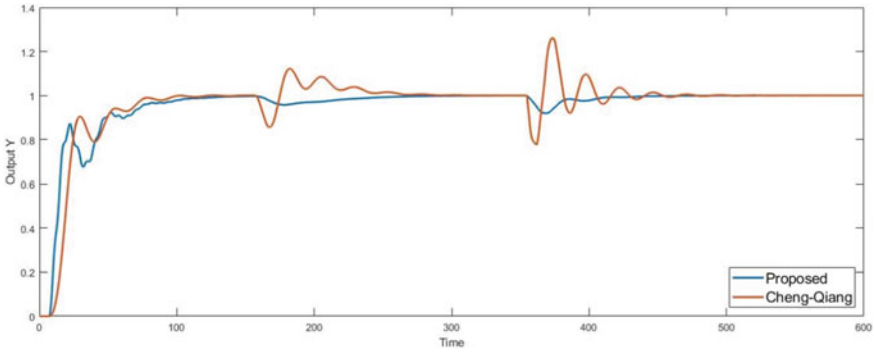
Fig. 8 System output for Example 2 (error)
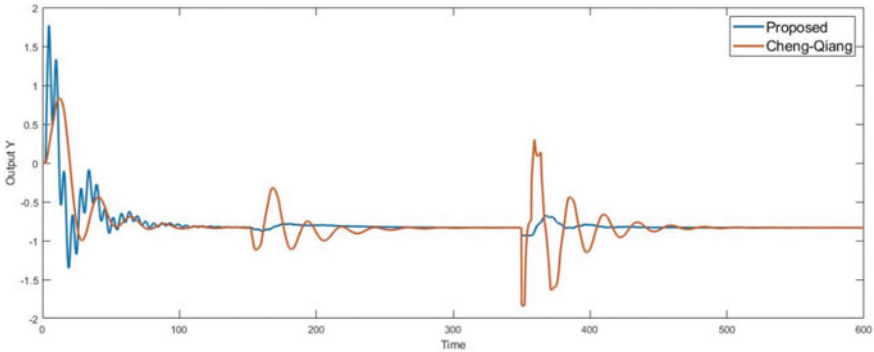
Fig. 9 Control signals of Example 2 (error)

Table 2 Performance characteristics of Example 2

Technique	IAE	PV	TV
Perfect model parameter			
Cheng-Qiang	28.3552	1.05	23.91
Proposed	24.2949	1.06	18.26
Error model parameter			
Cheng-Qiang	31.10	1.16	33.53
Proposed	28.3028	1.0	27.83

Example 3 Consider another example studied by Cheng-Qiang in [13] whose primary process is $P_2 = \frac{e^{-0.399S}}{5S-1}$ and the intermediate process is $P_1 = \frac{e^{-0.6S}}{2.07S+1}$. For the proposed method, $F_1 = \frac{2.07S+1}{0.6S+2}$, $F_2 = 4.4091 + \frac{1}{1.9636S} + 7.1874S$, $P_C = 2$, and $C = \frac{10.35S^2+2.938S+1}{S^2+2S+1}$.

Considering above controller parameters, the performance of these methods are simulated by enabling a unit step change at t = 1 s at the set point and applying a negative unit step change at t = 30 s in disturbance d_1 and negative step change at t = 60 s in disturbance d_2 respectively. The resulting system outputs are displayed in Fig. 10. The proposed method offers enhanced performance than other method.

To analyse robust stability, suppose a +20% error in intermediate process and −20% error in process delay and also +20% error in primary process gain along with +30% error in process delay. The corresponding outputs are displayed in Fig. 11. Initially proposed method has oscillations, but it results in better disturbance rejection.

The performance measures for the proposed and other method under perfect and perturbed cases are tabulated in Table 3.
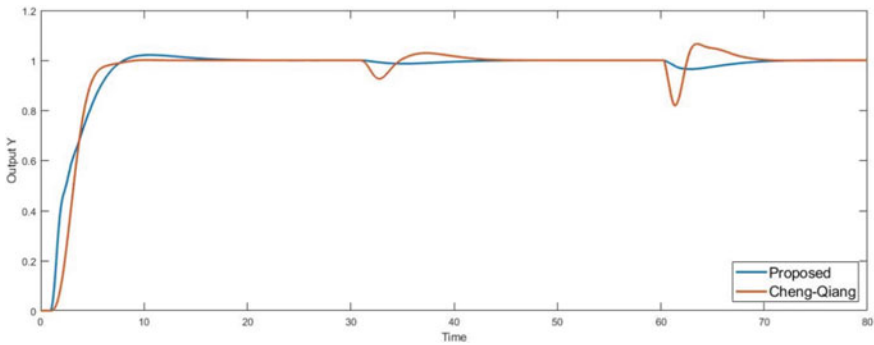
Fig. 10 Control signals of Example 3 (perfect)
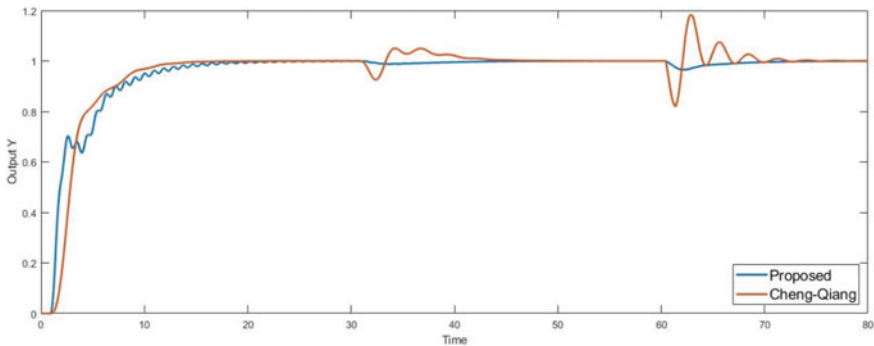
Fig. 11 System output for Example 3 (error)

Table 3 Performance characteristics of Example 3

Technique	IAE	PV	TV
Perfect model parameter			
Cheng-Qiang	4.1129	1.05	50.9
Proposed	3.4976	1.0	43.73
Error model parameter			
Cheng-Qiang	4.14	1.05	76.23
Proposed	3.7929	1.0	64.154

6 Conclusion

This paper work proposes a revised cascade control method for systems with RHP poles and time delay. The set point tracker C is developed derived from H_2 optimization and Internal Model Control theory. The disturbance rejector F_2 is developed as a $P + I + D$ controller. A filter is cascaded in series with controller C to make it physically realizable. SIMC tuning method is used to tune the controllers. SIMC has 'T_c' as variable measure which gives the trade-off among robustness and performance. MATLAB simulated results depicts that suggested technique gives enhanced outputs for both rejection of disturbance and also set point tracing. Thus, both regulatory and servo performances are enhanced. This method also yields lower IAE, PV and TV values.

References

1. Liu T, Gao FR (2011) Enhanced IMC design of load disturbance rejection for integrating and unstable processes with slow dynamics. ISA Trans 50(2):239–248
2. Chidambaram M (1997) Control of unstable systems: a review. J Energy Heat Mass Transf 19:49–56
3. Kavdia M, Chidambaram M (1996) On-line controller tuning for unstable systems. Comput Chem Eng 20(3):301–305
4. Jacob EF, Chidambaram M (1996) Design of controllers for unstable first-order plus time delay systems. Comput Chem Eng 20(5):579–584
5. Saxena S, Hote YV (2017) Internal model control based P+I+D tuning using first-order filter. Int J Control Autom Syst 15(1):149–159
6. Vanavil B, Anusha AVNL, Perumalsay M, RaoAS (2014) Enhanced IMC-P+I+D controller design with lead lag filter for unstable and integrating processes with time delay. Chem Eng Commun 201(11):1468–1496
7. Zhang WD, Gu DY, Wang W, Xu XM (2004) Quantitative performance design of a revised Smith predictor for unstable processes with time delay. Ind Eng Chem Res 43(43):56–62
8. Rao AS, Rao VSR, Chidambaram M (2007) Simple analytical design of revised Smith predictor with enhanced performance for unstable first-order plus time delay (FOPTD) processes. Ind Eng Chem Res 46(13):4561–4571
9. Vijaya Krishna B, Venkata Prasanth B, Sujatha P (2018) Design and implementation of DPFC for multi-bus Power system. Int J Eng Technol 7(2.8):18–29
10. Lee YH, Oh SG (2002) Enhanced control with a general cascade control structure. Ind Eng Chem Res 41(11):2679–2688

11. Vijaya Krishna B, Venkata Prasanth B, Sujatha P (2018) Enhancement of power quality in distribution system using hybrid seven Level H-bridge inverter based Dpfc. J Electrical Eng 2:61–71
12. Vijaya Krishna B, Venkata Prasanth B, Sujatha P (2018) MATLAB/Simulink study of multi-level inverter topologies using minimized quantity of switches. Int J Eng Technol 7(1.5):209–216
13. Yin C-Q, Wang H-T, Sun Q, Zhao L. Enhanced cascade control system for a class of unstable processes with time delay
14. Liu T, Zhang W, Gu D (2005) IMC-based control strategy for open-loop unstable cascade processes

Independent Controller Design for Non-minimum Phase Two Variable Process with Time Delay

C. Venkatasuresh, M. Rathaiah, R. Kiranmayi, and K. Nagabhushanam

Abstract Non-minimum phase (NMP) second order with time delay Process (SOPTDP) are stabilized by tuning process. Certain restrictions on the attainable performance when system have right hand side zeros of s-plane. The PID controller had been designed by Enhanced IMC method. The proposed Proportional Integral and Derivative controller (PID) are implemented to non-minimum phase systems like Fermenter, Isothermal CSTR. The performance was analysed under model uncertainty and considering perturbations in one model parameter. The performance of the system under recommend method compare with the existing method in the literature. From the performance detected as present Approach effectively stabilize and advances the performance of process with disturbances. From simulated result of three examples demonstrates present work offers better-quality presentation of the set-point track and Hload rejection than existing method through better time domain specifications.

Keywords Enhanced IMC · NMP · Second Order plus Time Delay systems (SOPTDS) · Fermenter and Isothermal CSTR

1 Introduction

PID controller was a naturally uses controller amongst industries. Valuation of PID controllers justifies conversation now. Normal PID controllers were exchanged by electronic controller. These offers various improvements than earlier controller. Thereafter, electronic controller is substituted with computer programs. presently PID controller are software programs in place of physical controllers. Dealing unstable processes were at all times difficulty with comparison to the stable processes. Regrettably, various chemical processes are unstable processes. Some samples were bottom level control of distillation column, chemical reactors, bioreactors, etc. The difficulty controlling becomes higher though the processes have time delay. Time

C. Venkatasuresh (✉) · M. Rathaiah · R. Kiranmayi · K. Nagabhushanam
JNTUA Collage of Engineering (Autonomous), Ananthapuramu, Andhra Pradesh, India

delay may be due to process lag. The goal has to develop a simple controller for stable/unstable SOPTD systems [1–5]. A tuning parameter selection depends on maximum sensitivity (Ms) function. Set point filter was used to decrease overshoots in servo problem. "Many Approaches are used to construction of PI/PID controller for stable NMP Processes are stated in literature as Internal Model Control approach (IMC) [6, 7], "PM and GM" [8], "optimization approach" [9, 10], control structure approach [11–13] etc.

Here numerous approaches existing in literature to construct PI/PID controllers to NMP FOPTD systems. But a very few approaches were offered to NMP SOPTDS processes.

2 Proposed Method

Therefore, in this study, a SOPDS process model $G_p(s)$ given by Eq. 1 is considered for the construct of PID controller $G_c(s)$

$$G_p(s) = \frac{k_p(u - ts)}{ps^2 + qs + r}e^{-ds} \tag{1}$$

Closed loop transfer function obtained as

$$\frac{C(s)}{R(s)} = \frac{G_p(s)G_c(s)}{1 + G_p(s)G_c(s)} \tag{2}$$

Wanted closed loop gain of the system is as follows

$$\frac{C(s)}{R(s)} = Z(s) = \frac{e^{-ds}}{(1 + \alpha s)^2} \tag{3}$$

From the closed loop transfer function controller transfer function derived as

$$G_c(s) = \frac{1}{G_p(s)} \frac{Z(s)}{1 - Z(s)} \tag{4}$$

By substituting above functions $G_c(s)$ is obtained as

$$G_c(s) = \frac{ps^2 + qs + r}{k_p(1 - ts)} \frac{e^{-ds}}{(1 + \alpha s)^2 - e^{-ds}} \tag{5}$$

By using Taylor series expansion $G_c(s)$ Approximated as follows

$$G_c(s) = \frac{1}{s}\left[\frac{ps^2 + qs + re^{-ds}}{k_p\left\{t(\alpha^2 - \frac{d^2}{2})s^2 + \left(u(\alpha^2 - \frac{d^2}{2}) + t(2\alpha + d)\right)s + u(2\alpha + d)\right\}}\right] \tag{6}$$

$$A(s) = \left[\frac{ps^2 + qs + re^{-ds}}{k_p\left\{t(\alpha^2 - \frac{d^2}{2})s^2 + \left(u(\alpha^2 - \frac{d^2}{2}) + t(2\alpha + d)\right)s + u(2\alpha + d)\right\}}\right] \tag{7}$$

Therefore, controller transfer function is written as

$$G_c(s) = \frac{A(s)}{s} \tag{8}$$

For the ideal form of PID apply Maclaurin series expansion theorem then

$$G_c(s) = \frac{1}{s}\left\{A(0) + A(0)'s + \frac{A(0)''}{2}s^2 + \ldots\right\} \tag{9}$$

The ideal form of PID as given below

$$G_c(s) = k_c\left\{1 + \frac{1}{\tau_I s} + \tau_D s\right\} \tag{10}$$

The controller parameters are assessed by comparing the constants of 's' of. The following terms for controller parameters are gotten and written as

$$k_c = A(0)', \tau_I = \frac{k_c}{A(0)}, \tau_D = \frac{A(0)''}{2k_c} \tag{11}$$

3   Selection of α

For the controller design the tuning parameter α is required. A suitable tuning parameter α is chosen by the use of maximum sensitivity function. The chosen tuning parameter for attained controller should offer better performance and robustness. For smaller α value of gives a faster response and displays improved results in the case of load disturbance for the processes. Ms is defined as

$$Ms = \left|\frac{1}{1 + G_p(s)G_c(s)}\right| \tag{12}$$

Ms value uses to determine a range for the Gain Margin (GM) and Phase Margin (PM) and relation between Ms and GM and Ms and PM is given below.

$$GM \geq \frac{Ms}{Ms - 1}; PM \geq 2\sin^{-1}\left(\frac{1}{2Ms}\right) \tag{13}$$

If the Ms value Increases then *GM* and *PM* decreases. The smaller values of *GM* is 1.6° and 34° have been suggested *PM* for stable processes. There are certain boundaries of the controller for controlling process and to attain the wanted robust closed loop performance. In this work, the α selected to reach *Ms* value in the range of 1.1–1.7.

4 Simulation Studies

Three cases of NMP SOPDT Process were taken to the Valuation of present approach. Simulation approach has been made to achieve the best value of α. For finding optimum PID values, the original PID values were essentially varied frequently using α through computer simulation till the lowest *MS* index and the required closed loop performance reached. Effectiveness of the controllers stands measure in the form of errors, total variation (TV) of operated variables and time domain specification.

Case study-1 (fermenter).

Fermenter process exact model has given below. In this the microbial growth was supposed to following Monod kinetics has specified by the following equations.

$$\frac{dZ}{dt} = (\mu - K_d)Z - Zu \tag{14}$$

$$\frac{dS}{dt} = (\frac{\mu}{Y} - m)Z - (S - S_t)u \tag{15}$$

$$\mu = \frac{\mu_{max}S}{k_m + S} \tag{16}$$

From solving above equations. Resulted as follows

$$\frac{X(S)}{U(s)} = \frac{0.2803(1 - 3.4063s)}{3.1638s^2 + 5.7382s + 1}e^{-0.3s} \tag{17}$$

In this present approach, PID values gotten are $k_c = 3.0725$, $\tau_i = 5.7382$, $\tau_d = 0.5514$ through choosing $\alpha = 0.1514$, $\beta = -5.475\alpha$, $\gamma = 1.925\alpha$ correspondingly. The performance has assessed for closed loop processes by unit step change in the set point as exposed in Fig. 1 also unit step change in load as exposed in Fig. 2. From figures analysed that the characteristics of a system like rise time (tr), settling
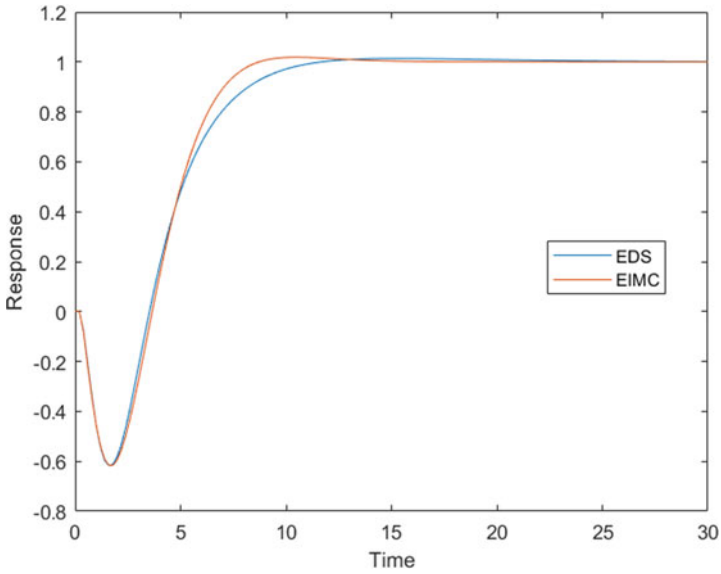
Fig. 1 Servo performance of process with the controller parameters obtained from the enhanced IMC, proposed method
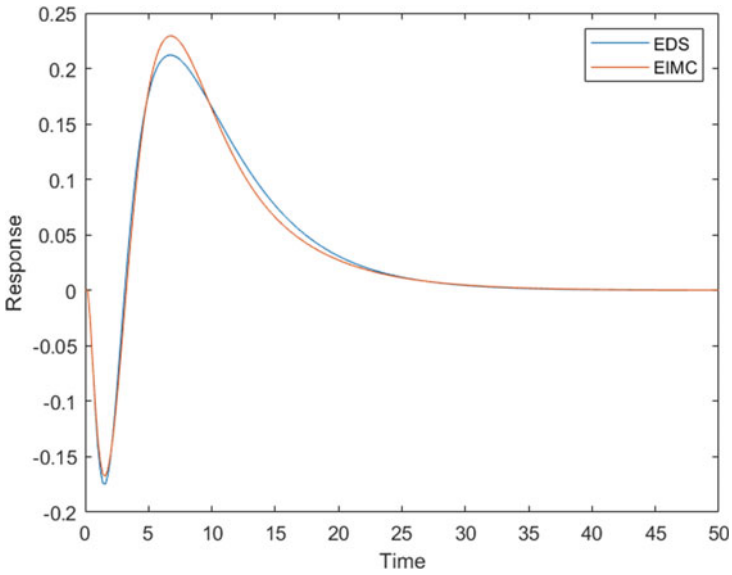
Fig. 2 Regulatory performance to fermenter process with controller parameters attained through enhanced IMC, proposed Approach

time (ts) were less than the available approach and these are recorded in Table 1. From both responses of current and existing method observes the current method accomplishes a smoother undershoot than the existing method. The current method expresses important development in the controller through reduce of ITAE 11% than enhanced IMC method. The controller actions of both servo and regulatory problems specified in Figs. 3 and 4 correspondingly.

Case study 2 (isothermal CSTR).

Van De Vesse isothermal CSTR [13] is taken now.

$$X \xrightarrow{P_1} 3Y \xrightarrow{P_2} Z$$

$$2X \xrightarrow{P_3} W$$
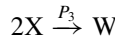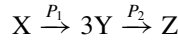
The mathematical model of the system is given below

$$\frac{dC_X}{dt} = \frac{F}{V}\left(C_{X,f} - C_X\right) - P_1 C_X - P_3 C_X^2 \tag{18}$$

$$\frac{dC_X}{dt} = -\frac{F}{V}C_B + P_1 C_X - P_2 C_Y \tag{19}$$

From solving above equations resulted function is as below

$$\frac{X(S)}{U(s)} = \frac{-1.117s + 3.1472}{s^2 + 4.6429s + 5.3821}e^{-0.1s} \tag{20}$$

In this present approach, PID values gotten are $k_c = 1.4588$, $\tau_i = 0.8627$, $\tau_d = 0.2154$ through choosing $\alpha = 0.1577$, $\beta = -1.009\alpha$, $\gamma = 2.586\alpha$ correspondingly. The performance has assessed for closed loop processes by unit step change in the set point as exposed in Fig. 5 also unit step change in load as exposed in Fig. 6. From figures analysed that the characteristics of a system like tr, ts were lower than the available approach and these are recorded in Table 1. From both responses of current and existing method observes the current method accomplishes a smoother undershoot than the existing method. The current method expresses important development in the controller through reduce of ITAE 25% than enhanced IMC method. The controller actions of both servo and regulatory problems specified in Figs. 7 and 8 correspondingly.

Case study-3 (fermenter).

The linearized form of case study 1 transfer function is as follows

$$\frac{X(S)}{U(s)} = \frac{0.0291(1 - 28.7121s)}{2.2024s^2 + 9.0155s + 1}e^{-0.3s} \tag{21}$$

Table 1 Comparisons of Performance indices of all case studies

Case study	Design method	Servo problem		Regulatory problem		TV		Rise time (s)	Settling time (s)
		IAE	ITAE	IAE	ITAE	Servo	Regulatory		
1	Enhanced IMC	6.372	16.306	2.398	22.384	131.23	2.987	0.839	1.63
	Proposed	6.042	15.019	2.064	21.31	109.58	2.489	0.802	1.108
2	Enhanced IMC	1.068	0.533	0.651	1.103	26.654	1.748	3.58	8.24
	Proposed	0.911	0.496	0.626	1.001	23.256	1.018	3.46	6.289
3	Enhanced IMC	36.423	267.67	3.575	57.261	148.81	6.347	11.3	26.7
	Proposed	35.7	225.36	2.92	50.215	139	5.983	10.86	25.9
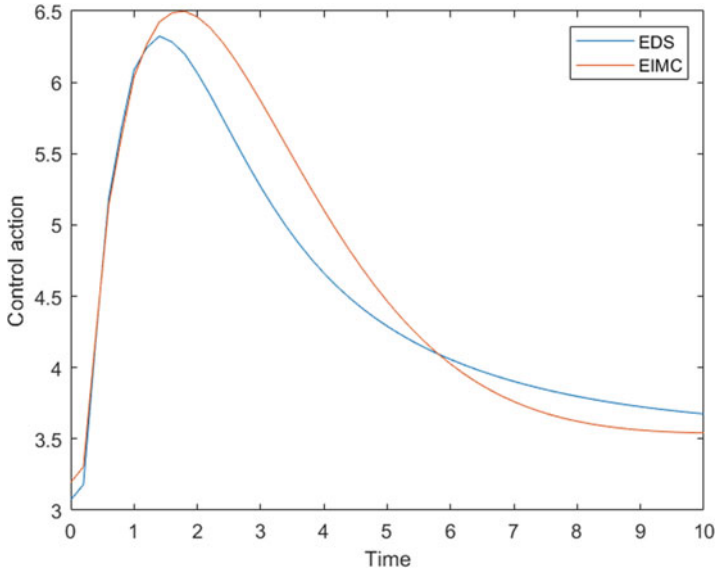
Fig. 3 Control action attained through servo response with PID parameters for enhanced IMC, proposed Approach
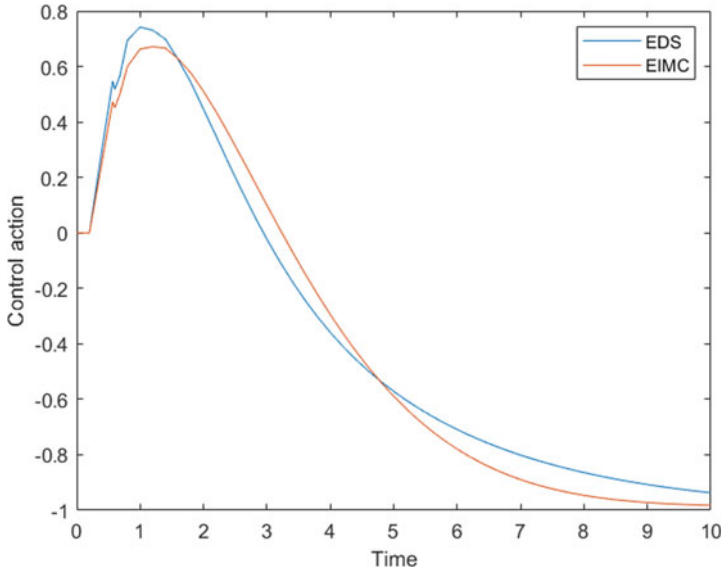
Fig. 4 Control action attained through regulatory response with PID parameters for enhanced IMC, proposed Approach
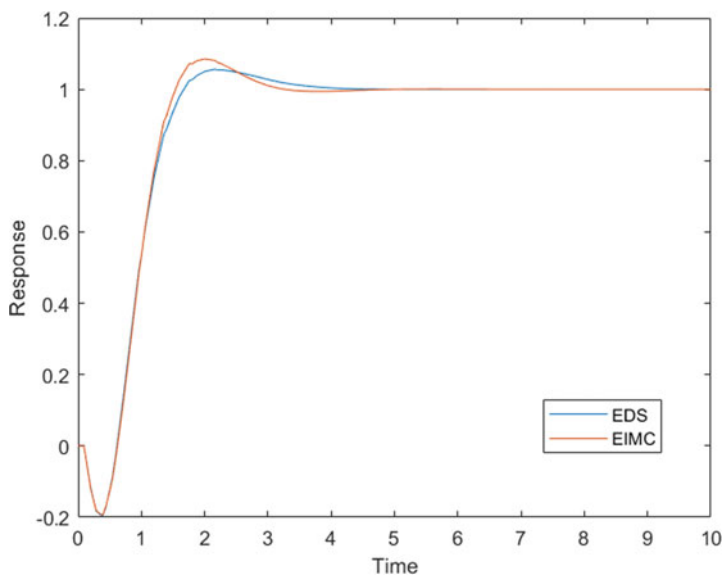
Fig. 5 Servo performance of isothermal process with the controller parameters attained from the enhanced IMC, proposed approach
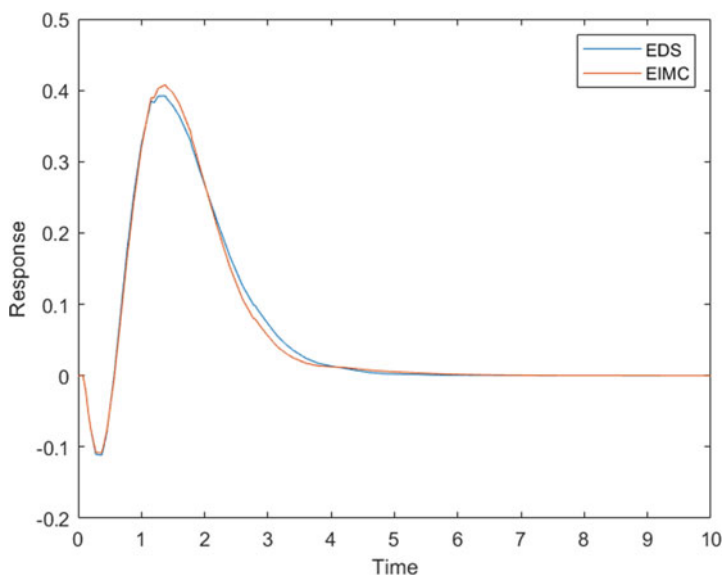
Fig. 6 Regulatory performance to isothermal process with controller parameters attained through enhanced IMC, proposed Approach
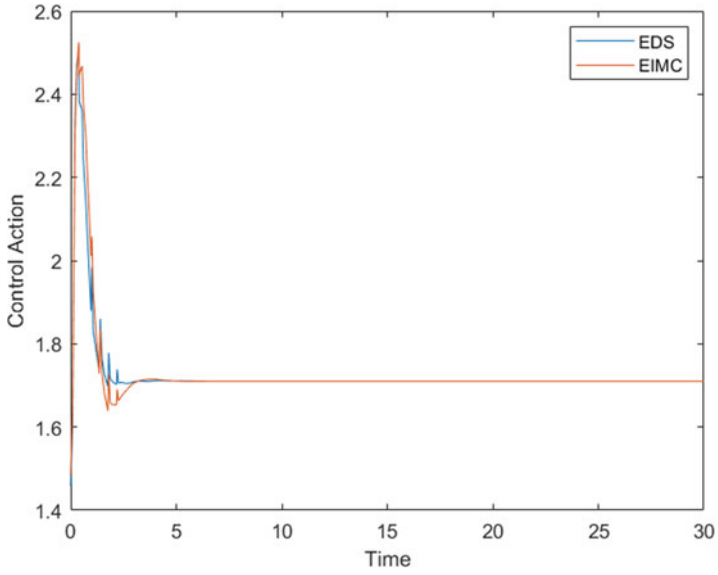
Fig. 7 Control action attained through servo response with PID parameters for enhanced IMC, proposed Approach
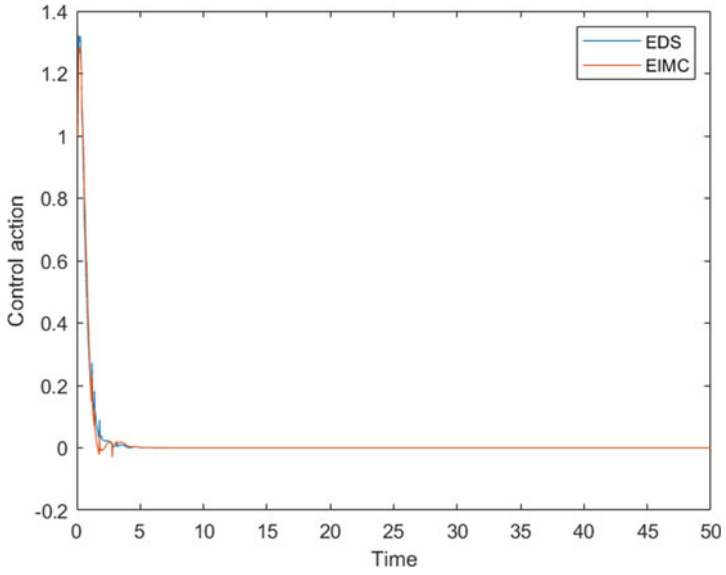
Fig. 8 Control action attained through regulatory response with PID parameters for enhanced IMC, proposed Approach
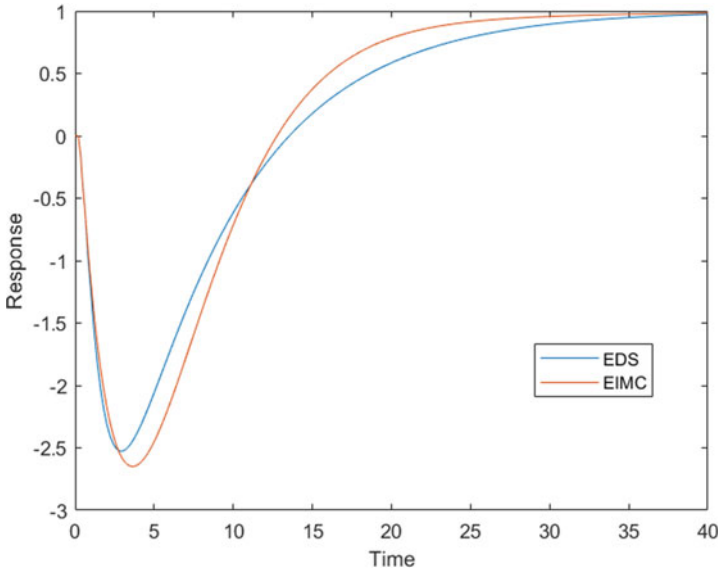
Fig. 9 Servo performance of fermenter process with the controller parameters attained from the enhanced IMC, proposed approach

For the present method, controller parameters attained are $k_c = 8.2075$, $\tau_i = 9.0155$, $\tau_d = 0.2443$ by choosing$\alpha = -0.0362$, $\beta = 87.78\alpha$, $\gamma = 0.0362\alpha$ correspondingly. The controller parameters for enhanced IMC method are $k_c = 1.4832$, $\tau_i = 0.8627$ and $\tau_d = 0.1723$. The performance has assessed for closed loop processes by unit step change in the set point as exposed in Fig. 9 also unit step change in load as exposed in Fig. 10. From figures analysed that the characteristics of a system like rise time tr, ts were lower than the available approach and these are recorded in Table 1. From both responses of current and existing method observes the current method accomplishes a smoother undershoot than the existing method. The current method expresses important development in the controller through reduce of ITAE 15% than enhanced IMC method. The controller actions of both servo and regulatory problems specified in Figs. 11 and 12 correspondingly.

5 Conclusion

A tuning process was used to develop the PID for the NMP SOPDT systems. The proposed method was associated with the method existing method in literature such as Enhanced IMC. Analysed that from simulation results, proposed method gives important development in the closed loop performance has detected in three cases related to the current method in literature. The current method expresses improved performance indices IAE and the ITAE. smother control action was attained by the
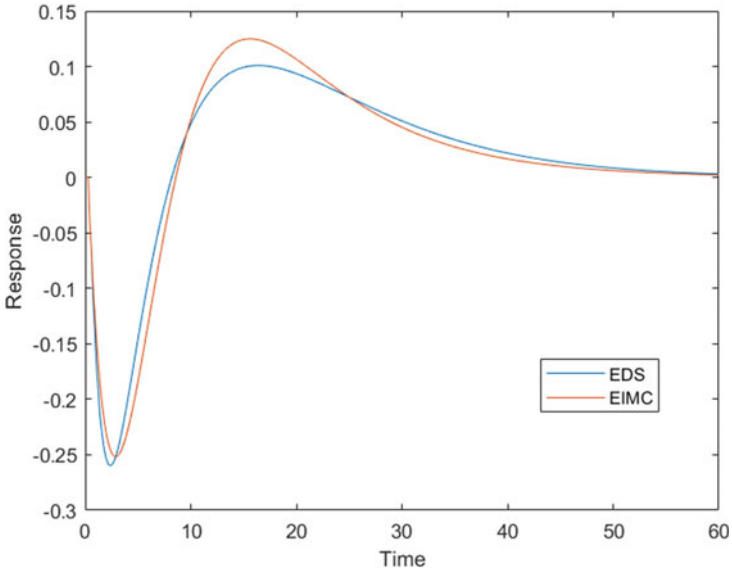
Fig. 10 Regulatory performance to fermenter process with controller parameters attained through enhanced IMC, proposed Approach
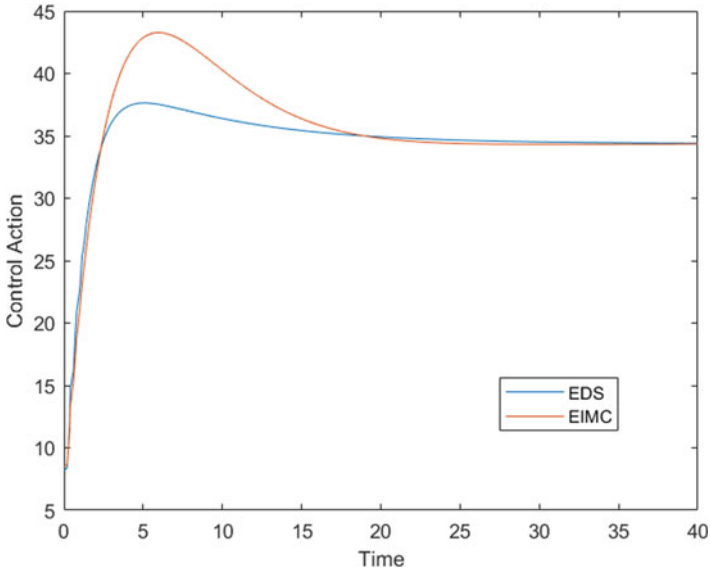
Fig. 11 Control action attained through servo response with PID parameters for enhanced IMC, proposed Approach
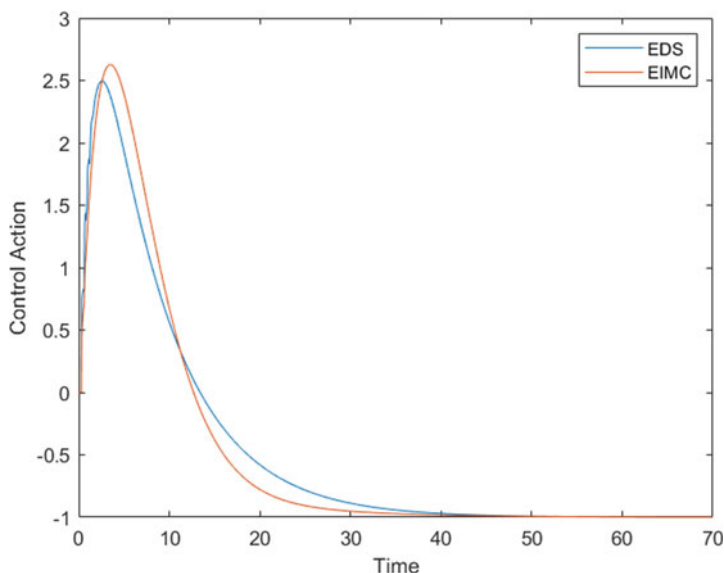
Fig. 12 Control action attained through regulatory response with PID parameters for enhanced IMC, proposed Approach

present Approach. The performance was enhanced in present approach than Existing approach. This could be exposed that the present PID is able to control NMP SOPTD systems.

References

1. Sankar Rao C (2019) Enhanced PID controller for non-minimum phase second order plus time delay system. In: Chemical product and process modeling, 20180059
2. Normey-Rico JE, Camacho EF (2008) Simple robust dead-time compensator for first-order plus dead-time unstable processes. Ind Eng Chem Res 47:4784–4790
3. Uma S, Rao AS (2016) Enhanced modified Smith predictor for second-order non-minimum phase unstable processes. Int J Syst Sci 47:966–981
4. Begum KG, Rao AS, Radhakrishnan TK (2017) Enhanced IMC based PID controller design for non-minimum phase (NMP) integrating processes with time delays. ISA T 68:223–234
5. Hauksdottir A, Sigurosson S (2011) The matching coefficients PID controller. In: American control conference, San Francisco, CA, USA
6. Kuhlmann A, Bogle ID (2001) Controllability evaluation for non-minimum phase-processes with multiplicity. AIChE J 47:2627–2632
7. Seaborg DE, Edgar TF, Mellichamp DA (2004) Process dynamics and control. Wiley, USA
8. Skogestad S (2003) Simple analytic rules for model reduction and PID controller tuning. J Process Contr 13:291–309
9. Simi S, Chidambaram M (2013) A simple method of tuning series cascade controllers for unstable systems. J Control Theory Appl 11:661
10. Sri RP, Chidambaram M (2003) A simple method of tuning PID controllers for integrator/dead-time processes. Comput Chem Eng 27:211–215

11. Vijaya Krishna B (2018) MATLAB/simulink study of multi-level inverter topologies using minimized quantity of switches. Int J Eng Technol 7:201–216
12. Vijaya Krishna B (2018) Design and implementation of DPFC for multi-bus power system. Int J Eng Technol 7:18–29
13. Vijaya Krishna B (2018) Enhancement of power quality in distribution system using hybrid seven level H-bridge inverter based DPFC. J Electr Eng 2:61–71

Control of DC Link Voltage and Load Voltage Variations in a Pitch Angle Controlled PMSG Based Wind Energy Conversion System

Check for updates

A. Nagarajan, K. Ramash Kumar, and G. Sureshkumaar

Abstract The world is looking for a better alternative energy source to generate electricity. Wind Energy Conversion Systems stands first in the choice of researchers in terms of renewable energy. Permanent Magnet Synchronous Generator (PMSG) plays an important role in wind energy generation system design. They exhibit a character of generating electrical energy even at a low speed of operation. Also, the control strategies of PMSG based WECS are quite easy compared to other generators used in wind energy generators. A pitch angle control technique along with load voltage variation control is proposed in this article. A MATLAB/Simulink based simulation work has been proposed here and the test results are quite satisfactory.

Keywords PMSG · DC link · Wind Energy Conversion System (WECS) · Pitch angle control · PIC controller

1   Introduction

Wind power conversion system is one of the cleanest sources of electrical power generation systems. The power produced per unit area is quite impressive compared to all other power generation techniques. Variable speed WECS are most popular in wind power technologies. They can deliver maximum power point tracking as well. Their efficiency is also compared to fixed speed and limited variable speed WECS.

The first choice of electrical generator now a days is doubly fed induction generator (DFIG). In this article PMSG based WECS have been proposed. This generator improves the reliability of the entire system. The features of PMSG are self-excitation

A. Nagarajan · G. Sureshkumaar
Karpagam College of Engineering, Coimbatore, India
e-mail: nagarajan.a@kce.ac.in

G. Sureshkumaar
e-mail: sureshkumar.g@kce.ac.in

K. Ramash Kumar (✉)
Dr. N.G.P Institute of Technology, Coimbatore, India
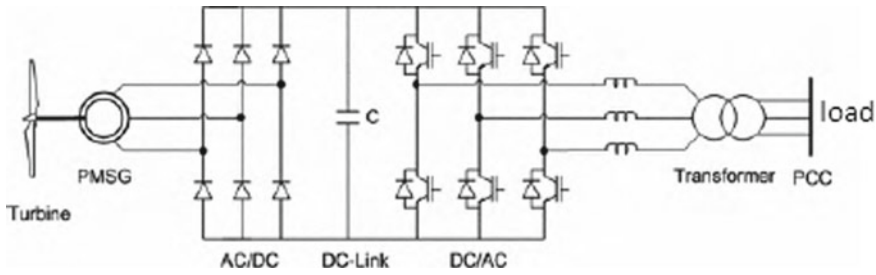
Fig. 1 PMSG based wind turbine

capability, leading to a high power factor and high efficient operation. A standalone PMSG based WECS can be used as a prime source of energy to the places where grid integration is impossible. The reason is PMSG based WECS can generate electricity even at a low speed of wind flow.

The main challenge of standalone power system is the feeding the end users with a balanced power supply with proper control of power converters. In wind energy conversion systems some power quality issues like voltage imbalalnce, generation of harmonics, flickering and load variation will occur very often. The change in voltage occurs often due to variation in the load. The oscillating voltage of WECS is mainly due to change in their wind speed. Also, the harmonics are produced due the high frequency operating power converters of WECS. Those power quality problem could not tolerated by the consumers and so, require mitigation methods [1]. Controllers for power converters were detailed in [2–9].

Figure 1 shows the schematic diagram of permanent magnet synchronous generator based wind energy conversion system. When load varies with in a short period, that causes a huge variation in voltage at common point of coupling. This will result in torque ripple sin the shaft which may affect the life of the machine,

A standalone small scale PMSG based WECS has been explained in this article. The major considerations are,

1. Implementation of pitch angle control under heavy wind turbulences.
2. Maintaining the constant voltage in the DC link to minimize ripple in the DC input to the generator side converter.

2 Mathematical Equations Analyis of Wind Turbine

The energy extracted from the wind is directly proportional to velocity of the wind. The power traced from the wind energy is expressed as below,

$$P = 0.5\rho A v_w^3 \tag{1}$$

where
ρ—air density (kg/m^3);

A—the area swept by blades (m^2); and.

V*w* —wind speed (m/s).

Betz limit explains the fraction at which the wind turbine can extract power from the wind flow. It is limited to 59% as per the Betz theory. Power coefficient C$_p$ is related to pitch angle of the blade and TSR (Tip Speed Ratio). This power coefficient describes the betz limit.

The power extracted from the wind by the turbine is expressed as,

$$P_w = 0.5C_p(\beta, \lambda)\rho A v_w^3 \tag{2}$$

where

Cp—power coefficient,

β—pitch angle in degrees.

λ—tip speed ratio.

The tip speed ratio is defined as the ratio of the speed at which the tip of blade rotates to the speed of the wind (wind velocity) **V*w*,**

$$\lambda = \frac{\Omega R}{v_w} \tag{3}$$

where

Ω—speed of the turbine rotor.

R—radius of turbine rotor tip swept area.

Figure 2 shows the relationship between speed of the turbine rotor and the mechanical power developed in the turbine. It is clear from the picture that is the mechanical power developed in the rotor is directly proportional to the speed of rotor also the developed power reaches maximum when at a particular point for various wind speeds.

The typical speed vs power extraction plot is shown in Fig. 3. It shows that all the WECS starts to produce electricity at a particular speed of wind. Till that the wind generator will be acting as a motor that is receiving power from the system. Once the

Fig. 2 Turbine rotor speed versus mechanical power
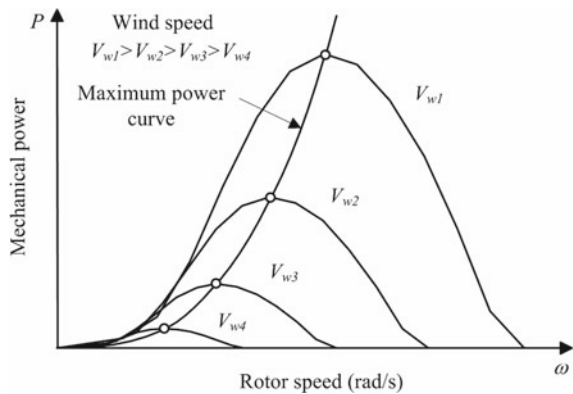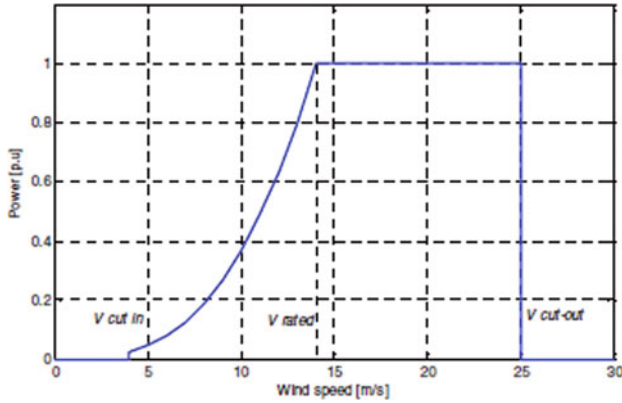
Fig. 3 Simulated responses of wind speed versus power extracted
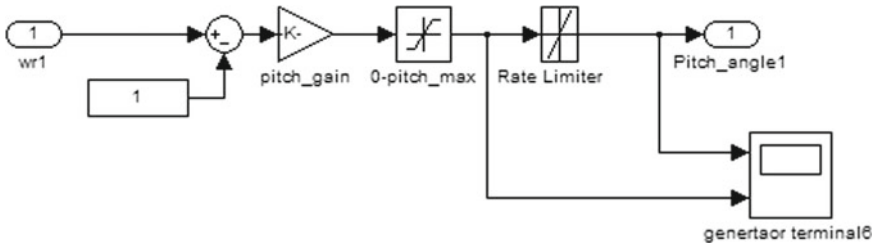
Fig. 4 MATLAB/Simulink model of control of pitch angle

cut-in velocity is attained the machine starts to generate electricity that is feeding the power to the power system. One major advantage of using PMSG base WECS is that it can produce electricity from a minimum wind velocity compared to all other types of wind energy conversion systems. From the Fig. 4 it is evident that that extracted power is maximum from the speed 12–25 m/s. if speed exceeds beyond this limit it can be controlled either mechanically by using pitch angle control technique or yaw mechanism or electrically using the power converters.

3 PAC of Wind Turbine

Pitch Angle Control (PAC) is defined as the angle at which the wind strikes the turbine blades with respect to the angle of rotation. When the speed of wind exceeds the wind blades experiences more torque as shown in the above figure. The figure says that the power generated beyond the cut off velocity is zero. That is when the speed of wind exceeding the maximum value of rated velocity the mechanical control comes into action.

There are two ways to control the wind turbine mechanically which are

1. Pitch angle control
2. Yaw mechanism

Since the power extracted from the wind is depending mainly on pitch angle and tip speed ratio the adjustment of pitch angle should be optimum below the rated speed of the wind. The need of pitch angle control is as below;

(1) Maximizing the obtained power of the wind turbine, below the cut off speed the angle must be set to derive maximum power from the system.

(2) Regulating the power generation within a safe value. Ensuring the machine operates with in the SOA (safe operating area).

4 Simulink Model of PAC

Following is the simulation model of the pitch angle control of wind energy conversion system. Here the actual wind speed is measured and compared with the reference value of wind flow. The error is amplified with help of a gain. After the gain point a rate limiter is introduced which limits the speed of the rotor within a safe value. In practical case a servo mechanism is employed to adjust the pitch angle. When the wind velocity exceeds a particular value the servo motor turns the wind blade to a particular angle. The feedback is given via the sensors which collect the speed of the rotor of the turbine.

5 Influence on Voltage at PCC and Its Compensation

It is not necessary that the load is to be constant always. Due to variation in load connected to the inverter the current in all the phases will not be uniform. In distribution network this is a major issue. This sudden change in load may lead to unbalance voltage at the point of common coupling. There for compensation of the load voltage variation is more important in the design of converters. A voltage control technique using the PWM inverter is employed in this proposal. PI controllers have been used to control the pulse generation. Here the actual value is compared with the reference value, based on the variation between actual and reference current magnitudes the PWM pulses are generated with different modulation index values (Fig. 5).
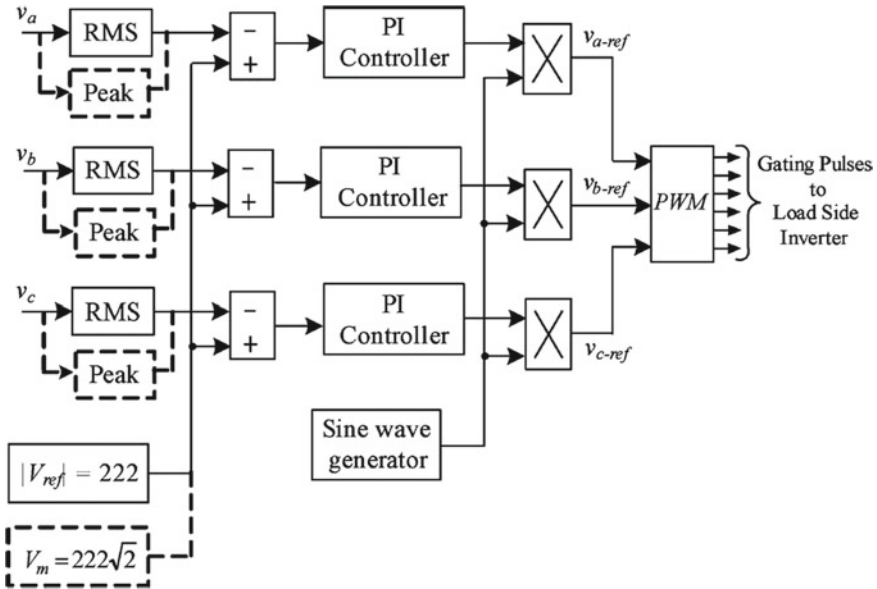
Fig. 5 MATLAB/Simulink model of PWM inverter controller for unbalanced load compensation

6 Simulation Responses and Its Discussions

The following figure shows the simulation model of the power converter used in the proposed system. The results are compared with and without the pitch angle control strategy (Fig. 6).

The speed of wind is set to 12 m per second for first 3 s and it is set to 25 m/s up to 7 s the again 12 m/s up to 10 s. in this the base speed is set to 12 m/s. From Fig. 7 the generator speed crosses more than 1.5 p.u which is 2550 rpm. In Fig. 8 it is shown that the generator speed is controlled to 1.2p.u at the excess wind speed (25 m/s). And parameters like generator torque and output voltage are also controlled.

Figure 9 shows the DC link voltage control. The disturbance occurs at 0.25 s of the simulation period. But the system is providing a constant voltage over the entire simulation period because of the control of the rectifier as explained in Sect. 5.

The load is varied at 0.25 s by using breaker in the Simulink model. In the above Fig. 10, the voltage output is kept constant throughout the run time. The disturbance is tolerated by the pulse generation. The magnitude of the modulating signal is shown in below plot.
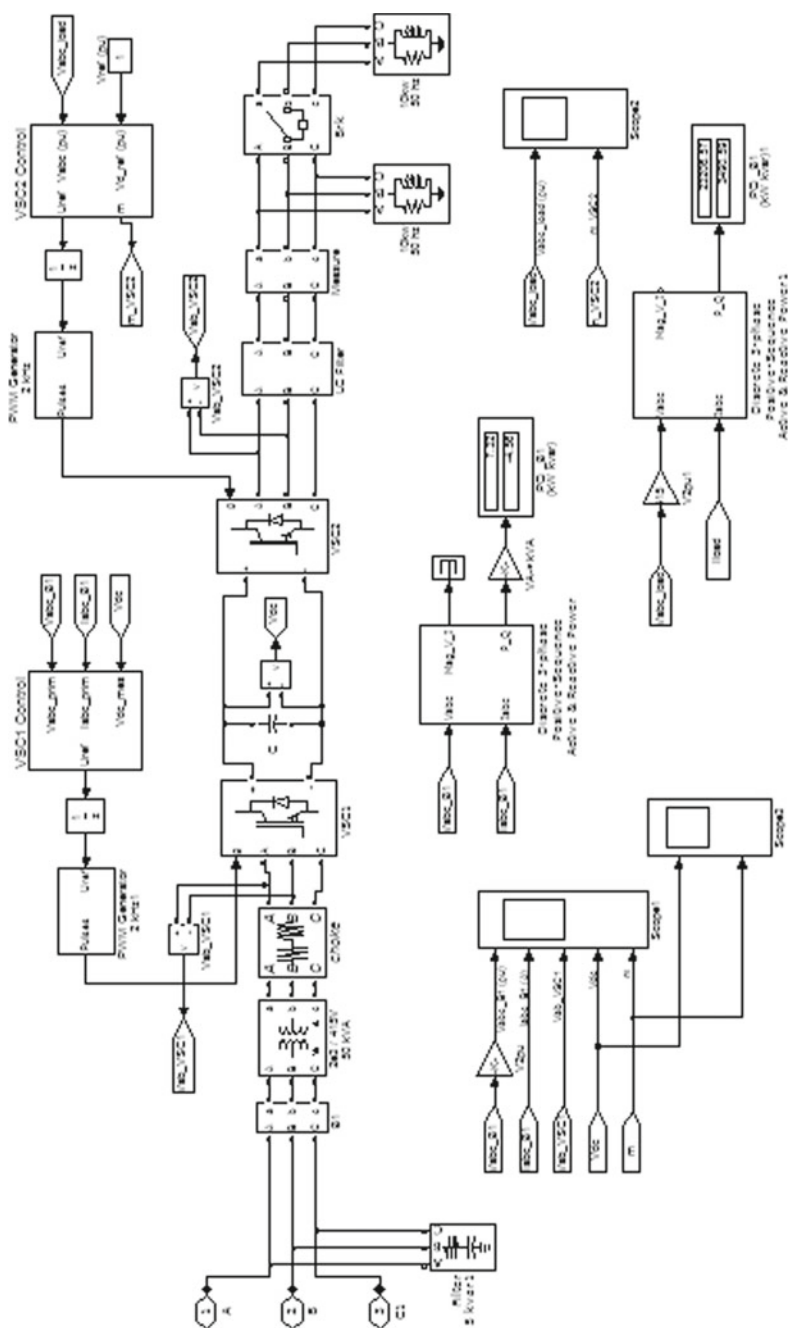
Fig. 6 MATLAB/Simulink model of power converters
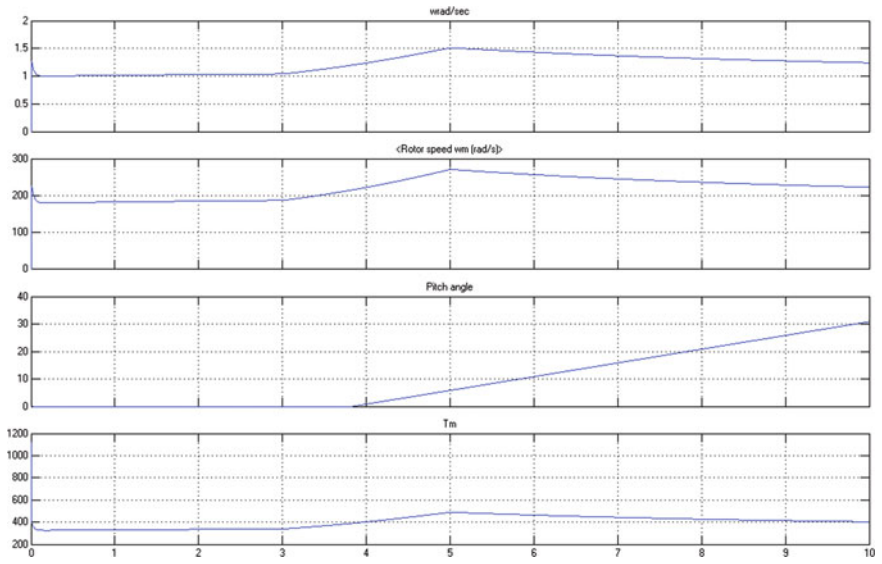
Fig. 7 Simulated responses of WECS without pitch control
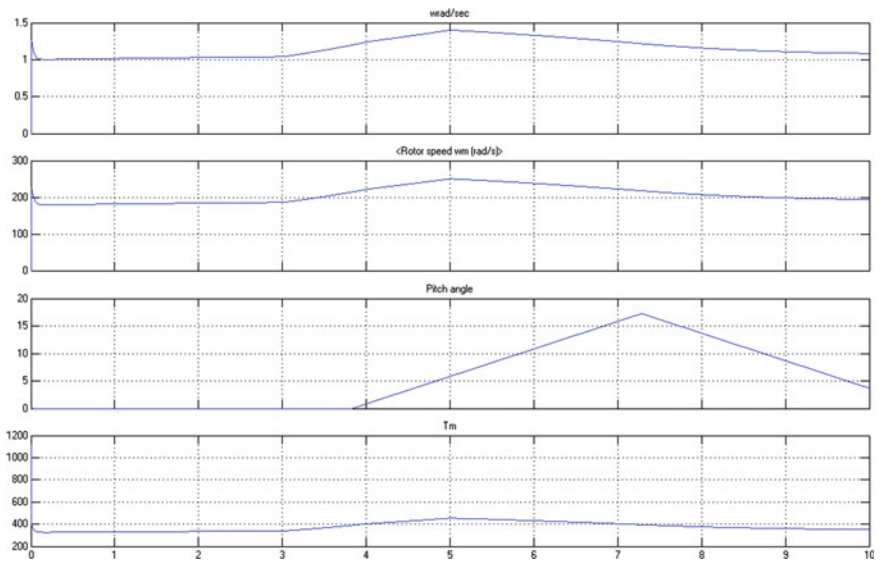
Fig. 8 Simulated responses of WECS with pitch control
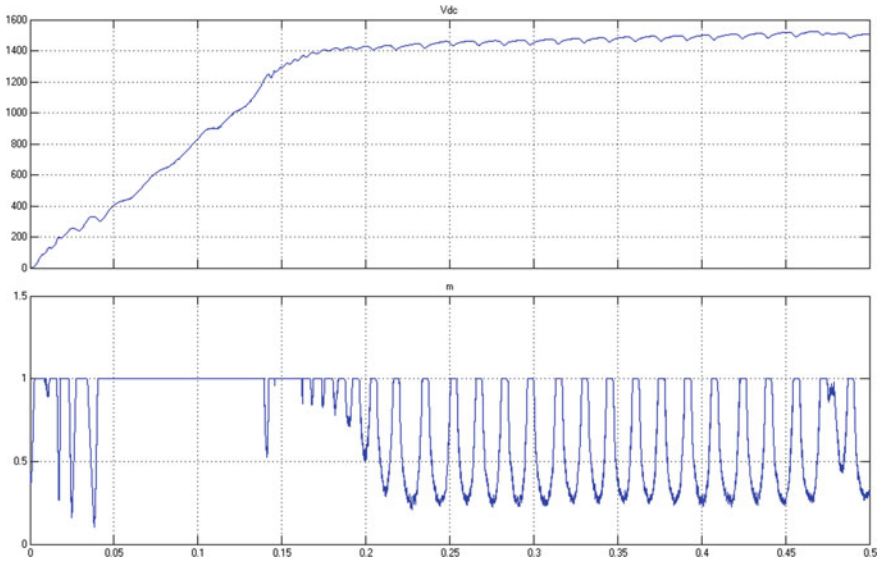
Fig. 9 Simulated responses of dc link voltage and magnitude of reference signal for PWM generation of rectifier
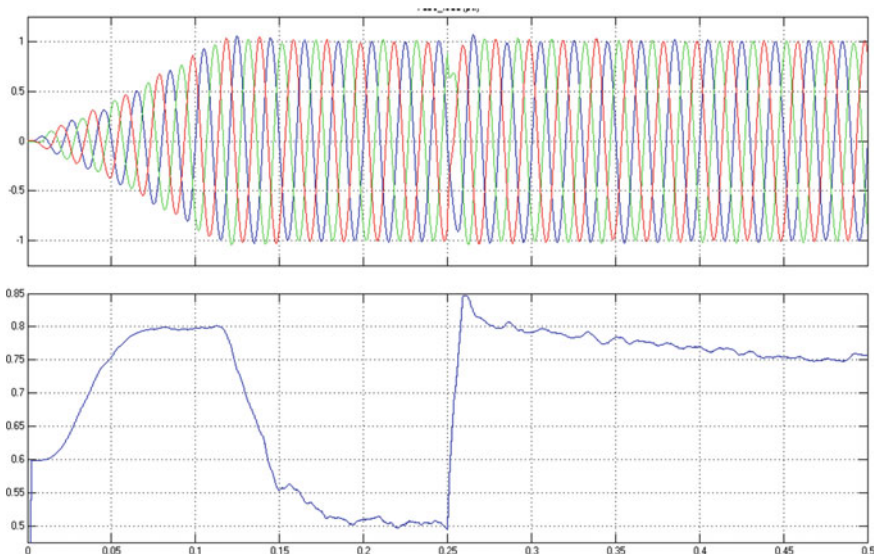
Fig. 10 Simulated responses of output voltage of inverter and magnitude of reference signal for pwm generation of inverter

Table 1 Generator ratings

Stator phase resistance	0.085 Ω
Rotor speed	1700 rpm
Line voltage	2500 V
Torque constant	16.54 Nm
No of pole pairs	4

7   Conclusion

It is evident that under mechanical discrepancy conditions like higher wind speed and multi directional wind power mechanical control of the systems works better compared to electrical control instantly. i.e. pitch angle control result is quick as compared to the electrical control. Also controlling the DC link voltage with in a safe limit saves the generator as well as the load. PI controller based control is giving better results as shown in the result. Modern controllers may give better result.

Appendix

See Table 1.

References

1. Bhende CN, Mishra S, Malla SG (2011) Permanent magnet synchronous generator-based standalone wind energy supply system. IEEE Trans Sustain Energy 2(4)
2. Rolan A, Luna A, Vazquez G, Aguilar D, Azevedo G (2009) Modeling of a variable speed wind turbine with a permanent magnet synchronous generator. In: IEEE International Symposium on Industrial Electronics Seoul Olympic Parktel, Seoul, Korea July pp 5–8
3. Gupta J, Ashwani Kumar. Fixed pitch wind turbine based permanent magnet synchronous machine model for wind energy conversion, pdf. (Available www.onlinejet.net)
4. Zhang J, Cheng M, Chen Z, Fu X (2008) Pitch angle control for variable speed wind. In: Third international conference on electric utility deregulation and restructuring and power technologies, Nanjing, China, pp 6–9
5. Ramash Kumar K (2019) FLC and PWM SMC for KY boost converter. J Circuits Syst Comput. https://doi.org/10.1142/S0218126619501846
6. Ramash Kumar K (2016) Implementation of sliding mode controller plus proportional integral controller for negative output elementary boost converter. Alexandria Eng J (Elsevier) 55(2):1429–1445
7. Ramash Kumar K, Jeevananthan S (2012) Analysis, design and implementation of hysteresis modulation sliding mode controller for negative output elementary boost converter. J Electric Power Components Syst 40(3):292–311

8. Gayathri Monicka J, Guna Sekhar NO, Ramash Kumar K (2011) Performance evaluation of membership functions on fuzzy logic controlled AC voltage controller for speed control of induction motor drive. Int J Comput Appl 13(5)
9. Ramash Kumar K, Arunkumar N, Sivakumaran TS (2019) Implementation of non-linear controller for contemporary DC-DC converters. Int J Electrical Eng Informatics 11(4):622–637

A Novel Approach of Wind MPPT Using Fuzzy Logic

Nesar Uddin, Md Saiful Islam, Jewel Sikder Joy, and Ishaque Mozumder

Abstract The key motive of this article is to carry out the optimum power as well as supreme power from wind turbine using fuzzy control technique. The wind power system largely relies on wind velocity and pitch angle, so Fuzzy control technique is used to extract supreme power in an efficient method which minimizes the dislocation in the system reaction. For developing the fuzzy-based wind MPPT system, a DC/DC step-up power converter is constituted to trigger the gate of its MOS-FET switch by applying Fuzzy rules for regulating rotor acceleration of wind generator. The fuzzy controller provides a signal as well as duty cycle as output through pulse width modulation (PWM) that switches the gate of DC/DC converter. Simulation results show that the developed fuzzy-based wind MPPT structure can perform nonlinear functions of random difficulty and cover an extensive range of working conditions.

Keywords Fuzzy Logic Controller (FLC) · Maximum power point tracking (MPPT) · Wind Energy Conversion System (WECS)

1 Introduction

In the mint condition, electricity is fundamental imperious of humankind, socioeconomic, artificial development of several countries [1]. Yet again, most of the regions did not entree to the light of electricity. The significant source of energy generation in the creation is fossil fuel that raises some gases that influence the human body [2].

N. Uddin (✉)
Department of EEE, Chittagong University of Engineering and Technology, Chattogram 4349, Bangladesh

M. Saiful Islam
Department of ETE, Chittagong University of Engineering and Technology, Chattogram 4349, Bangladesh

J. Sikder Joy · I. Mozumder
Department of IET, Chittagong University of Engineering and Technology, Chattogram 4349, Bangladesh

In 2019, be close to 19 TW fossil fuels have been used [3]. It also increases air pollution and damages the environment. To surpass the problem, the knowledge of new technology of sustainable energy from solar irradiation, hydro source, wind resource, geothermal store and biogas supply, are a more effective and reliable method in the creation [4].

Several regions in Bangladesh are also probable for wind resource as well as wind power generation via Patenga, Cox's Bazar, around Teknaf area, Kuakhata and Kutubdiya beach [5]. Generally, a wind turbine cannot generate extreme power by traditional control strategy [6]. To surpass these difficulties, MPPT methods are planned to extract supreme power by applying FLC. Howsoever, Fuzzy control technique is an expert intelligence thinking method that has beaten fruition in numerous cases to take effective decisions like human beings by adjusting themselves for present and future situations automatically [7]. Therefore, to extract supreme power from the wind turbine, the fuzzy control technique has been used in this paper. The recommended Fuzzy control technique can run an input/output planning by depending on the human sense which is a more effectual and optimum way.

2   Methodology

2.1   Block Diagram for Wind MPPT Structure

Maximum power point tracking (MPPT) is a method that is used to bring out supreme power from renewable foundations. The suggested wind MPPT is designed for extracting efficient power that stands on wind turbine, DC/DC step-up converter, fuzzy based control method and AC load as well as the electric appliance. The generated electricity from wind turbine relies on the atmospheric situation as well as wind speed and direction of pitch that is distributed through DC/DC step-up converter. The step-up converter is coupled with the fuzzy controller as well as the control unit. The control division acts according to fuzzy rules which give duty cycle as output. Subsequently, this output signal switches the gate of MOS-FET device through PWM incessantly to get the extreme power. Control units have double inputs viz; error signal and change in the error signal. The error signal is denoted as 'e' and change in the error signal is denoted as 'de'. Again 'e' and 'de' relies on a ratio of change in power and change in voltage. Change in power is denoted as 'ΔP' and change in voltage is denoted as 'ΔV'. Overall block diagram has been presented in Fig. 1.

2.2   Fuzzy Based Efficient Wind MMPT

Generally, wind power conversion method, the output power of the wind turbine is not linear because of unstable weather conditions. The fuzzy-logic-based controller
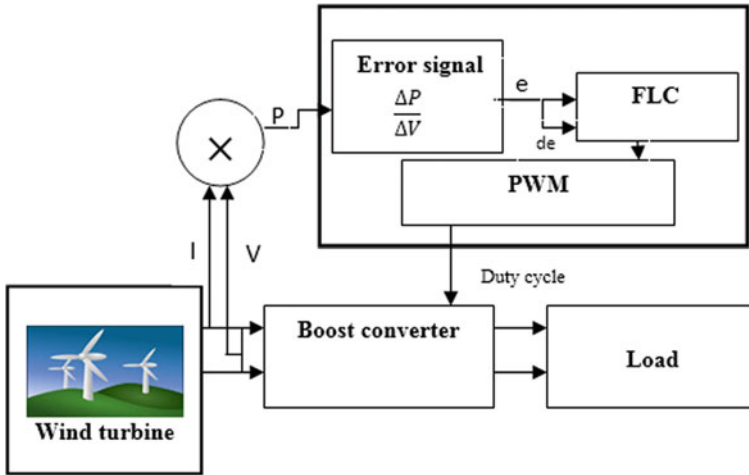
Fig. 1 Proposed Fuzzy based wind MPPT structure

helps to get better performance in such a situation. This arrangement is made up of two inputs and a single output that is illustrated below (Fig. 2). Error signal (e) and change in error signal (de) are the inputs. Five triangular memberships have been introduced for error signal. Those are Very Negative, Negative, Zero' side, Positive and Very Positive (denoted by VN, N, Z, P, VP), given below in Fig. 3. Three triangular memberships have been considered for changing the error signal. Those are Negative, Zero's side and Positive (denoted by N, Z, P), exposed in Fig. 4. Duty cycle represents the result as output membership function which has three triangular memberships. These functions are Very Negative, Zero and Very Positive (denoted by VN, Z, VP). They are illustrated in Fig. 5.

In support of MPPT method, we want e = (ΔP ÷ ΔV) is equal zero. Here, ΔP indicates power variation and ΔV indicates voltage variation. In condition, e = (ΔP ÷ ΔV) greater than zero, we supposed to raise the voltage to tap the MPPT. In condition, e = (ΔP ÷ ΔV) less than zero, it requires to lesser the voltage to tap the maximum power [8].
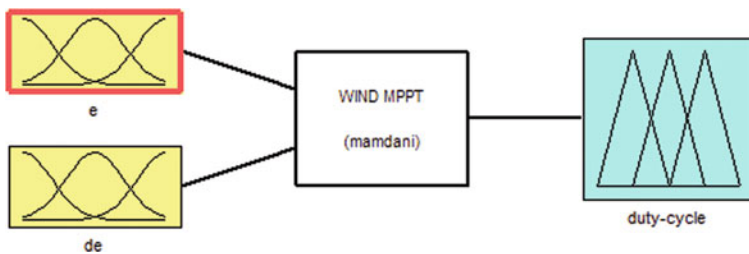
Fig. 2 Overall block diagram of wind MPPT
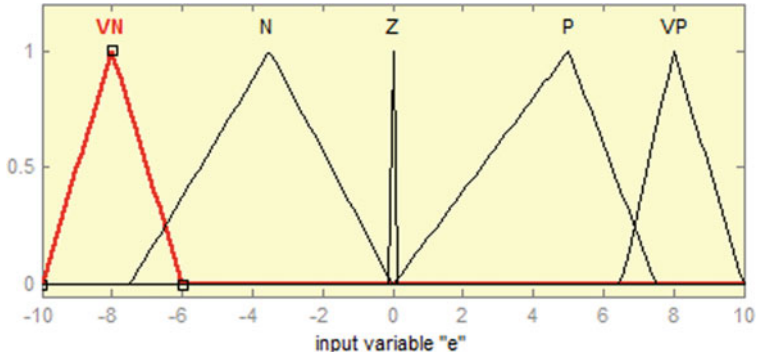
Fig. 3 Membership function of error signal (e)
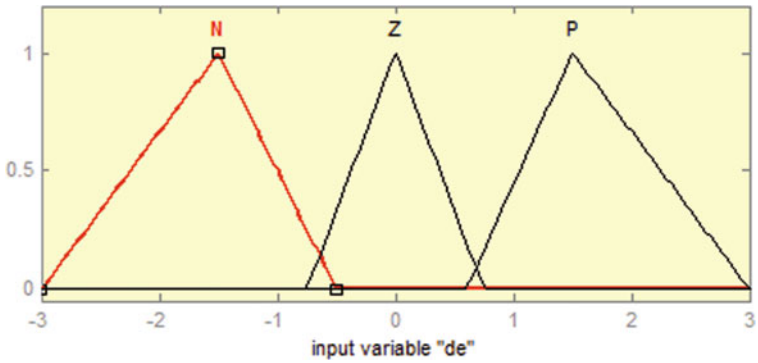
Fig. 4 Membership function of change in error signal (de)
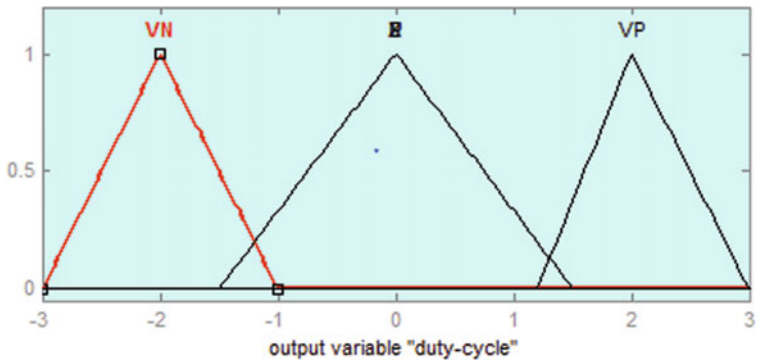
Fig. 5 Membership function of duty cycle

Table 1 Proposed fuzzy rules for MPPT system

	IF		THEN
Number of Fuzzy rules	Error signal, e	Change in error signal, de	Output signal
1–3	VN/VN/VN	N/Z/P	VP/VP/VP
4–6	N/N/N	N/Z/P	VP/Z/Z
7–9	Z/Z/Z	N/Z/P	Z/Z/Z
10–12	P/P/P	N/Z/P	Z/Z/VN
13–15	VP/VP/VP	N/Z/P	VN/VN/VN
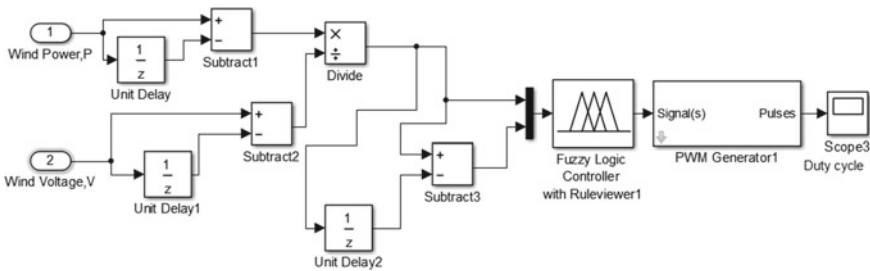
Fig. 6 Control unit of MPPT structure

The rotor velocity is the main dependence of the voltage of the wind turbine. As a result, the FLC controller is introduced here to be in command of rotor speed for higher power generation. The fuzzy formulas are précised in Table 1. The maximum output is picked up using fuzzy control technique. DC/DC step-up converter is applied here to position the duty ratio in the course of PWM. The error signal is a ratio of power deviation and voltage deviation with respect to time. A mathematical model of the control unit of MPPT system has been given in Fig. 6.

$$e = \frac{\Delta P}{\Delta V} = \frac{P(t) - P(t-1)}{V(t) - V(t-1)} \tag{1}$$

The FLC is associated with double input (e and de) and single output (Duty Cycle) that rely on IF–THEN affirmations which is named as fuzzy rules where IF indicates the input and THEN indicates the outputs parts. Overall, 15 rules have been set to produce supreme power from the scheme where various working conditions are described that is given in Table 1.

2.3   Simulink Structural Design of Suggested Wind MPPT

The ultimate schematic illustration of MPPT system is exposed in Fig. 7 where the arrangement is constructed with a wind turbine, DC/DC step-up converter, electric
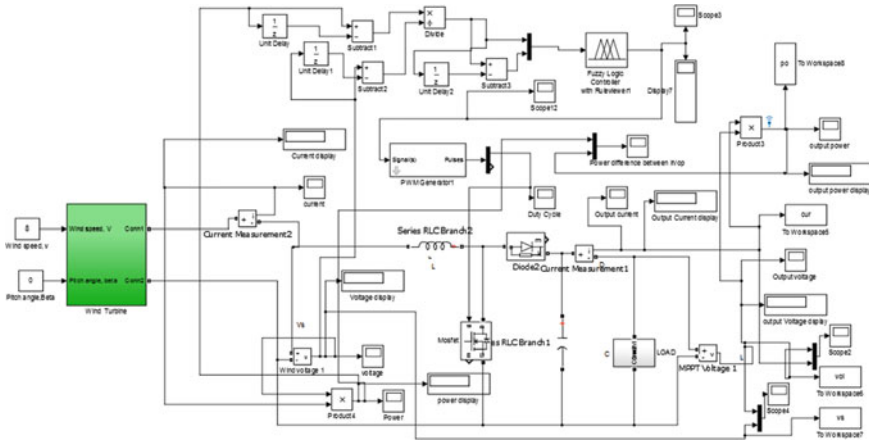
Fig. 7 Overall Simulink of suggested wind MPPT system

load and other system's controlling unit. The outputs have been culled by varying the wind speed from the range of (3–8) meter per second.

3 Evaluation of Wind MPPT

The 3-D plane inspection of FLC is represented in Fig. 8 where X-axis is representing the error signal (e) and Y-axis is indicating a change in error signal (de). The value of 'e' is measured from negative 10 to positive 10 that rely on the fluctuation of power and voltage. The value of 'de' is calculated as -3 to 3 which is précised for the interruption period to bring the utmost power. The outcome signal of the FLC is the duty cycle that varies with reference to 'e' and 'de'.

Fig. 8 Surface view of duty cycle of wind MPPT model
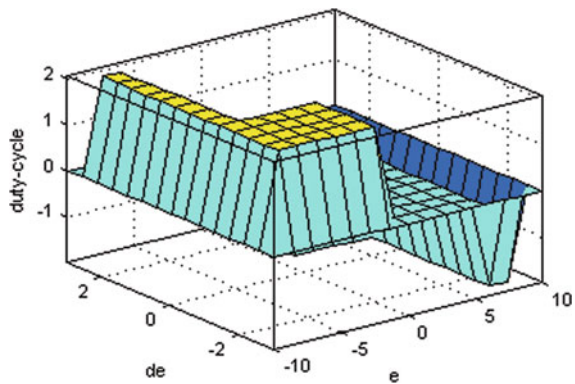
The value of the error signal has been considered negative ten to positive ten because of standard power as well as voltage of wind turbine. The standard power and voltage of WES5 tulipo wind turbine are 2.5 kW and 400 V. The standard voltage is converted to 220 V by 1-phase rectifier that is shown in Fig. 7. Therefore, we can get a ratio between power and voltage which is approximately 10. This value can vary negative ten to positive ten by depending on wind velocity. But the range of change in the error signal does not change gradually. So, this value has been considered negative three to positive three which is lower than the value of the error signal.

Case study—A

According to Fig. 9, when the value of the error signal is very negative and the value of change in the error signal is negative, then the duty cycle is very positive to get supreme power. Generally, it is not possible to get supreme power when the value of the error signal is very negative and the value of change in the error signal is negative. At any time, this situation may occur because of unavailable wind velocity. Therefore, it is impossible to get maximum power from the turbine. So, the duty cycle has been increased to bring out supreme power.

Case study—B

In this situations, when error signal is negative and change in error signal is considered as the value of zero' side, then duty cycle is assumed zero's side. That means the variation of change in error signal is not high by following error signal. So, it is not necessary to change the duty cycle (Fig. 10).

Case study—C

In this case, the value of the error signal and change in error signal are considered as positive and negative. So, the value of the duty cycle is assumed zero's side because of its operating condition. Error signal positive means the high operating condition
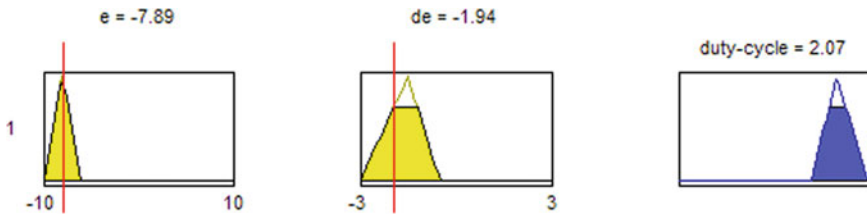
Fig. 9 When e-VN, de-N then Duty cycle-VP
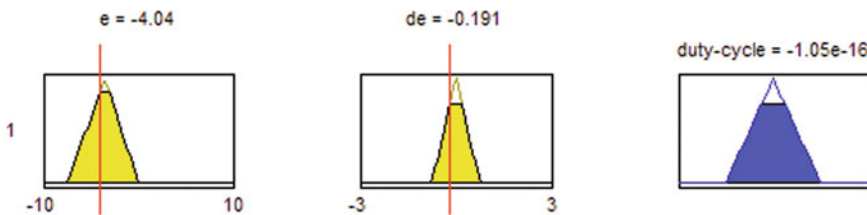
Fig. 10 When e-N, de-Z then Duty cycle- Z
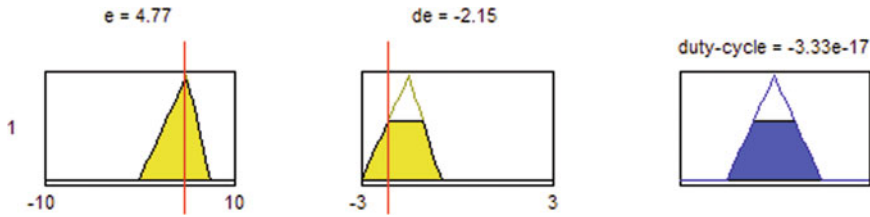
Fig. 11 When e-P, de-N then Duty cycle- Z

Table 2 Power comparison of proposed MPPT system with conventional system

Speed of wind in meter per second	Direction of Pitch in degree	Power-without MPPT in Watt	Power-Fuzzy Based Wind MPPT in Watt
3	0	28.22	30.95
3	1	28.22	30.95
4	0	81.62	89.79
4	1	81.62	89.79
5	0	233	252
5	1	233	252.7
6	0	438.8	482.9
6	1	438.8	482.9
7	0	1008	1086
7	1	1008	1086
8	0	1799	1988
8	1	1799	1988

of wind turbine and change in error signal negative means it does not influence on the operating system. So, the duty cycle has remained beside zero (Fig. 11).

Power difference of Fuzzy based wind MPPT system and without MPPT system is given in Table 2. From the Table, it is clear that Fuzzy based wind MPPT system is more efficient than a conventional system.

4 Conclusion

The suggested fuzzy based wind MPPT system that has diverse benefits over previously accessible systems. Fuzzy system is used which depends on nonlinear functions which is very modern and reliable. The petition of Fuzzy technique is swiftly mounting due to its strong performance regardless of complex situations and elasticity. It is possible to increase the efficiency by adapting supplementary intellectual fuzzy rules. A problem of the proposed structure is their remittent environment of

wind speed that is dependent on weather variations. Fuzzy technique has been applied to extract extreme power by two input parameters via Wind velocity & pitch angle. The results of the Fuzzy based wind MPPT model has been compared with the conventional system under a similar climate situation. Simulation result provides a grand performance of Fuzzy based wind MPPT model and less error. The application of renewable sources minimizes various gas discharges in the system. Besides, the system is environment kindly, favorable. Developed Fuzzy based wind MPPT structure more efficient than any other operable system as well as a grid-connected system that must reduce power loss.

References

1. Soedibyo T, Hadi A, Putra HP, Suryoatmojo H, Ashari M (2016) Integration of hybrid PV/wind generation system using fuzzy MPPT in grid connected system for remote area. In: MATEC web of conferences, International Conference on Frontiers of Sensors Technologies (ICFST 2016), vol 59, pp 1–2
2. Gowri M, Vinothkumar V, Punitha K (2015) ANFIS based MPPT and droop controller for power sharing in interline power system. Int J Sci Res Dev 3(04)
3. Khalil GA, Lee DC, Seok jk (2014) Variable speed, wind power generation system based on fuzzy logic control for maximum output power tracking. In: 35th annual IEEE power electronics specialists conference Aachen, Germany, pp 5–6
4. Uddin N, Islam MS (2018) Optimal fuzzy logic based smart energy management system for real time application integrating RES, grid and battery. In: 4th international conference on Electrical Engineering and Information & Communication Technology, IEEE
5. Islam MR, Rahman MN, Mannan MA (2016) Study of wind power in different parts of Bangladesh. Int Res J Eng Technol (IRJET) 03(09): 1290. ISSN: 2395-0056
6. Uddin N, Islam MS (2018) Optimization of PV energy generation based on ANFIS. In: International conference on Innovations in Science, Engineering & Technology, IEEE
7. Hailemariam MM, Mekonnen T, Sudheendra H (2015) Novel approach to fuzzy logic controller based hybrid solar/micro hydro/bio-mass generation, A real time analysis (Barsoma Village, Ethiopia). Int J Innov Res Eng Multidisciplinary Phys Sci (IJIRMPS) 3(1). ISSN: 2349-7300 ISO 9001
8. Minh HQ, Cuong NC, Chau TN (2014) A fuzzy-logic based MPPT method for stand-alone wind turbine system. Am J Eng Res (AJER) 3(9): 177–184. e-ISSN: 2320-0847 p-ISSN: 2320-0936

Shale Gas Productive Volume Optimization

Adamu Umar Ibrahim and Berihun Mamo Negash

Abstract Recently, a modified mathematical model was presented in order to determine the productivity of multi-fractured shale gas wells. However, the considerable challenge of the model was assuming an average fracture width for the determination of the practical fracture geometry of the stimulated reservoir volume. This challenge led to over prediction of the gas well. Therefore, this paper presents a simple and accurate method using a pseudo-3D model to determine the fracture width in shale gas reservoirs. The method utilized the hydraulic fracturing propagation capacity of the MATLAB software to visualize the fracture geometry, which includes the fracture height, fracture length, and the fracture width. So, the obtained fracture width can be incorporated with analytical models to predict long term productivity for multistage fractured shale gas wells. An accurate result for the productivity will assist in determining the ultimate gas recovery, propped volume, optimal fracture length, fracture spacing and predict the future performance of the well.

Keywords Shale gas · Hydraulic fracturing · Fracture geometry

1 Introduction

The combination of horizontal drilling and hydraulic fracturing technologies has already boosted the exploitation of shale reservoirs by many countries around the globe [1]. This unconventional resource compliment the hydrocarbon demand all over the world. Therefore, these technologies are revolutionary and will remain vital in shale gas production [2].

One of the significant challenges in shale gas production optimization is the forecasting of long-term productivity of shale gas wells. Currently, three types of models

A. Umar Ibrahim (✉)
Petroleum Engineering Department, Universiti Teknologi PETRONAS, Perak, Malaysia
e-mail: adamu_18000297@utp.edu.my

B. Mamo Negash
Shale Gas Research Center (SGRC), Universiti Teknologi PETRONAS, Perak, Malaysia
e-mail: bmamo.negash@utp.edu.my

are applied to predict productivity. These models are basically: analytical transient, numerical and empirical models [3, 4]. Recently, a simple model was presented, which successfully predict productivity with less than 3% over prediction. The analytical model is applied in pseudo-steady state flow conditions. However, the fracture width is assumed and computed as an average in the model [4]. Therefore, there is a need to find a more reliable way of obtaining a fracture width that will close the overprediction gap of the model. An accurate fracture width is vital to proper field performance prediction, fracture completion design and adequate placement of proppants in shale gas wells [5]. Also, this will give rise to designing a fracturing treatment that will propagate in a more optimal direction and avoid unnecessary damage to the subsurface environment.

This paper presents a simple and more accurate method using a pseudo-3D model to determine the fracture width in shale gas reservoirs. The model gives the visual result of the fracture geometry using MATLAB software. So, the obtained fracture width can be incorporated with the latest analytical model to forecast the long-term productivity of multistage fractured shale gas wells.

2   Methods and Procedures

Multistage hydraulic fracture in horizontal wells initiates many fractures from each perforation cluster. These change the formation stress field and affect the propagation of hydraulic fracture. Also, the low-viscosity fracturing fluid will leak off from the surface of hydraulic fracture into the reservoir, therefore, increasing its pressure. Consequently, the natural fractures will be stimulated by the induced stress. If the shear stress is greater than the shear strength, shear failure occurs [6]. Likewise, if the natural fractures inner pressure is greater than the tensile strength, tensile failure happens [7, 8]. Therefore, these increase the shale reservoir permeability. The fracture geometry, which has the fracture height, length, and width, directly control most of the productivity of the well [9, 10]. Proper design of these parameters aid in gas production optimization during fracturing treatment design.

Furthermore, it is good to visualize the fracture geometry using the pseudo-3D model in MATLAB software to determine the fracture width. The P3D model is compared with the well-known KGD and PKN models (Table 1).

Where, Qg = production rate of the gas (MScf/D); nf = number of fractures; μ = viscosity of the fluid (cp); T = temperature of the formation (°R); km = permeability of matrix (md); h = reservoir thickness (ft); p = average reservoir pressure (psia); pw = wellbore pressure (psia); e = exponential function; xf = fracture half-length (ft); kf = permeability of fracture (md); w = average fracture width (in); Sf = fracture spacing (ft).

Table 1 KGD, PKN & P3D models comparison

MODEL: KGD/2D	Assumptions
$L = 0.48(\frac{8GQ^3}{(1-v)\mu})^{1/6}.t^{2/3}$ $W_0 = 1.32(\frac{8(1-v)Q^3\mu}{G})^{1/6}.t^{1/3}$ $Pw = \sigma_{min} + 0.96(\frac{2G^3 Q\mu}{(1-v)^3 L^2})^{1/4}$	• Height is constant • A plane strain fracture in the horizontal plane • Viscous and Newtonian fracturing fluid • Elasticity theory used • No leak off into the formation
MODEL: PKN/2D	
$L = \left(\frac{Q}{\pi c_1 h}\right)t^{1/2}$ $W_0 = 4(\frac{2(1-v)Q^2\mu}{(1-v)\mu})^{1/4}.t^{1/8}$ $Pw = 2.5(\frac{G^4 Q^2\mu}{(1-v)^4 h^6})^{1/5}.t^{1/5}$	• Plane strain fracture • The fracture toughness does not influence the fracture geometry • No leak off into the formation
MODEL: PSEUDO-3D	
$Penetration depth(R) =$ $0.548(\frac{GQ^3}{G})^{1/9}.t^{1/9}$ $W_0 = 1.32(\frac{8(1-v)Q^3\mu}{G})^{1/6}.t^{1/3}$ $Pw = \sigma_{min} - \frac{5}{4\pi}\frac{GW_0}{R}ln\frac{(r_w)}{R}$ G = shear modulus (psi); Q = pumping rate (bbl/min); W_0 = maximum width (ft); v = poisson ratio; L = fracture length (ft); r_w = wellbore radius (ft); σ_{min} = minimum stress (psi); μ = fluid viscosity (cp)	• The vertical stress distribution is uniform • The shape of fracture is elliptical • No leak off into the formation [11]
MULTIFRACTURED SHALE GAS WELLS PRODUCTIVITY	$\dfrac{5.87\times10^{-5}n_f k_m h(\dot{P}^2 - P_w^2)}{\mu T S_f \sqrt{c}\left[\dfrac{1}{\left(1 - e^{-\sqrt{cx_f}}\right)} - \dfrac{1}{(3x_f\sqrt{c})}\right]}$ $c = \dfrac{96k_m}{k_f w S_f}$

3 Results and Conclusions

The single hydraulic fracture propagation is first shown to compare a P3D model with the conventional KGD and PKN model. The visualized fracture width provides the quantitative result of the width, which makes it easy to be applied in the productivity model.

4 Discussion

From Figs. 1, 2, 3, 4, 5, 6, 7 and 8, it is shown that the length of the fracture improves with greater rate than the fracture width of the fracture, which increases

Fig. 1 Fracture length
against time (KGD)
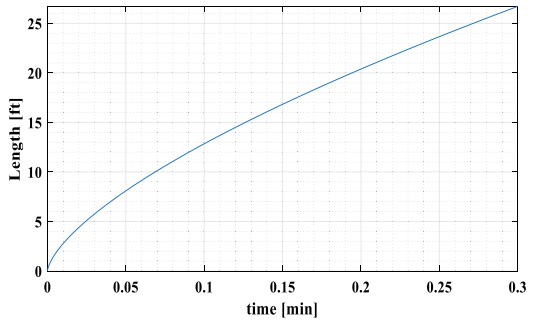
Fig. 2 Max width against
time (KGD)
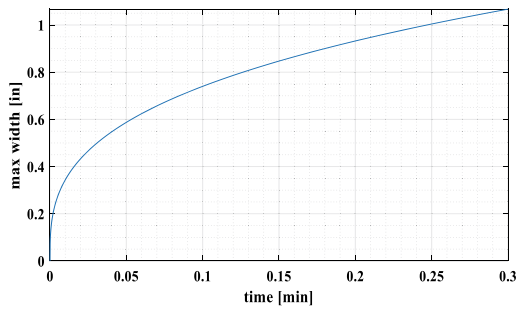
Fig. 3 Max width against
length (KGD)
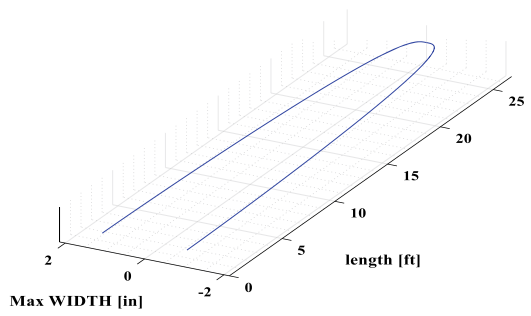
Fig. 4 KGD in 3D
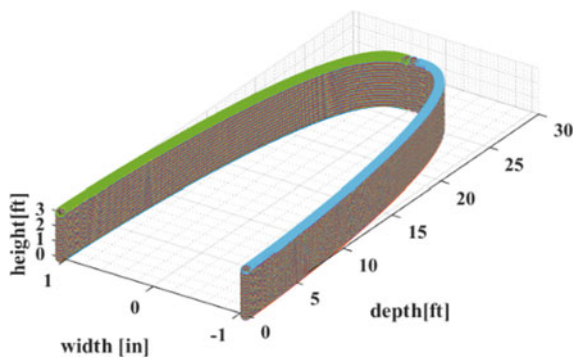
Fig. 5 Fracture width
against length (PKN)
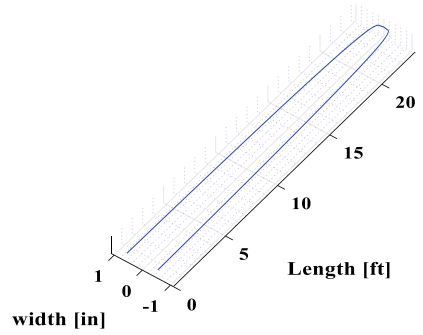
Fig. 6 Fracture width
against height (PKN)
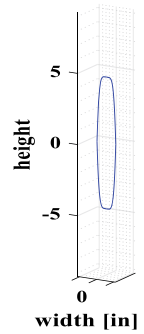
Fig. 7 Fracture height
against width (P3D)

$$z^2/2.250000 + y^2/3.438693 - 1 = 0$$

Fig. 8 Pseudo-3D model

$$z^2/2.250000 + x^2/38.259228 - 1 = 0$$
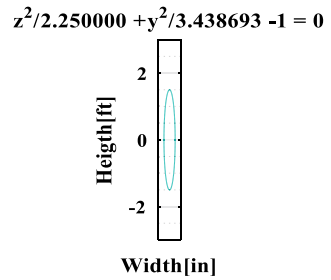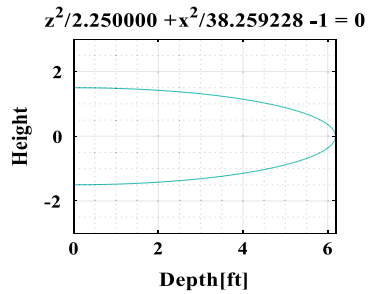
in proportional to t1/3. These models show a curvy tip which indicates that in the KGD model, the effect of the tip of the fracture is not large. Both KGD and PKN showed an approximation the width of the fracture and length. The pseudo-3D model forecasts a complete ellipsoid in 3D space at a time which can be seen in Figs. 9, 10 and 11. Therefore, the P3D model is good for fracture propagation modelling. The model can display the variation of the fracture height with length as the fracture propagates.

Fig. 9 Fracture length against time (PKN)
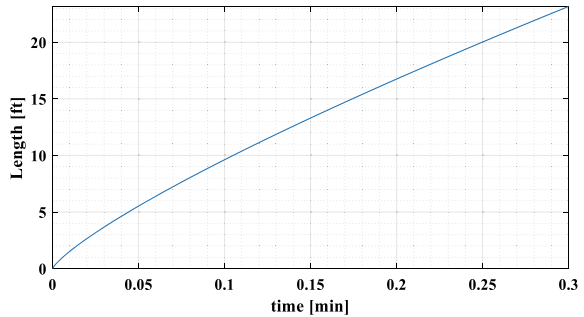
Fig. 10 Max. width against time (PKN)
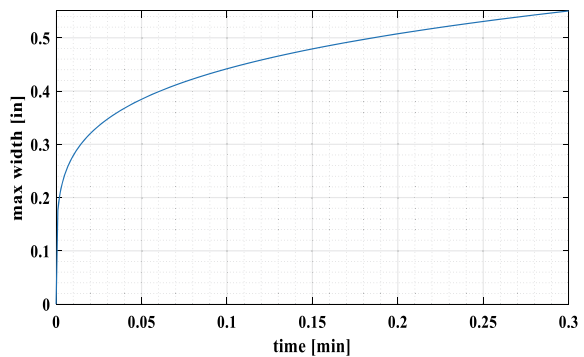
Fig. 11 P3D predicted shape
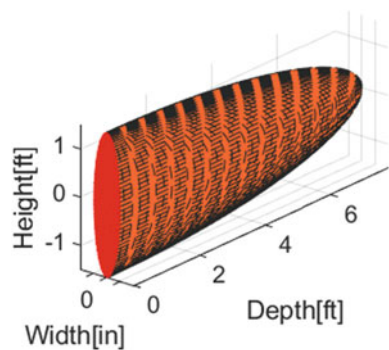
The fracture width is directly taken from the visualized fracture geometry and subsequently inputted in the multi fractured shale gas wells. This method is more dependable to determine the fracture width of a fractured shale gas reservoir to determine the productivity of gas wells. Also, it can be applied in order to optimize production by adequate fracture spacing and proppant placement.

References

1. Abdelaziz A et al (2019) Unconventional shale hydraulic fracturing under true triaxial laboratory conditions, the value of understanding your reservoir. In: SPE annual technical conference and exhibition, Society of Petroleum Engineers
2. Quainoo AK, Negash BM, Bavoh CB, Ganat TO, Tackie-Otoo BN (2020) A perspective on the potential application of bio-inhibitors for shale stabilization during drilling and hydraulic fracturing processes (in English). J Natural Gas Sci Eng 79:103380
3. Mahmoodi S, Abbasi M, Sharifi M (2019) New fluid flow model for hydraulic fractured wells with non-uniform fracture geometry and permeability. J Natural Gas Sci Eng, Article 68, Art no 102914
4. Li G, Guo B, Li J, Wang MJSD, and Completion (2019) A mathematical model for predicting long-term productivity of modern multifractured shale-gas/oil wells, vol 34, no 02, pp 114–127
5. Belyadi H, Fathi E, Belyadi F (2019) Hydraulic fracturing in unconventional reservoirs: theories, operations, and economic analysis. Gulf Professional Publishing
6. Wu Y-S (2018) Hydraulic fracture modeling. Gulf Professional Publishing
7. Ren L, Lin R, Zhao JZ (2018) Stimulated reservoir volume estimation and analysis of hydraulic fracturing in shale gas reservoir, (in English). Arab J Sci Eng 43(11):6429–6444
8. Ren L, Lin R, Zhao JZ, Rasouli V, Zhao JY, Yang H (2018) Stimulated reservoir volume estimation for shale gas fracturing: mechanism and modeling approach, (in English). J Petrol Sci Eng 166:290–304
9. Ahn CH, Dilmore R, Wang JY (2017) Modeling of hydraulic fracture propagation in shale gas reservoirs: a three-dimensional, two-phase model. J Energy Resources Technology 139(1):012903
10. Berawala DS, Andersen PO, Ursin JR (2019) Controlling parameters during continuum flow in shale-gas production: a fracture/matrix-modeling approach, (in English). SPE J Conf Paper 24(3):1378–1394
11. Yew CH, Weng X (2104) Mechanics of hydraulic fracturing. Gulf Professional Publishing

A Comparative Study and Validation of Kinematic Analysis of a Crank Rocker Engine Prototype Using MATLAB and ADAMS

A. M. Albaghdadi , **M. B. Baharom, and S. A. Sulaiman**

Abstract This work presents an analytical study of a suggested mechanism as part of a design process of a Crank-Rocker (CR) engine. The basic geometry of this engine is based on the principal of a four-bar crank rocker mechanism, which requires optimization for effective operation. Kinematic study of this mechanism is achieved using two analysis approaches, analytical and graphical model simulation approach. To achieve the objective, two softwares, namely MATLAB and ADAMS, were used to conduct this analysis and to validate mechanism functionality. Under certain constraints and operational conditions, the simulation results by both applications were compared and found to be almost identical, this gives an indication about mechanism operational behaviour. Moreover, this study provides clear idea about using different approaches and applications for system kinematic study purposes.

Keywords Crank-Rocker engine · Kinematic study · Mechanism synthesis · Model simulation

1 Introduction

A new engine called Crank-Rocker engine (CR) was patented in 2018 by engineers of Universiti Teknologi Petronas, Malaysia. Despite showing promising results, the engine is still considered belonging to early development stage. Of these studies which has been performed on crank-rocker engine and used as base reference for this paper and any further improvements on this engine, work in researches [1–4].

A. M. Albaghdadi (✉) · M. B. Baharom · S. A. Sulaiman
Department of Mechanical Engineering, Universiti Teknologi Petronas (UTP), 32610 Bandar Seri Iskandar, Perak, Malaysia
e-mail: anwr_18002826@utp.edu.my

M. B. Baharom
e-mail: masrib@utp.edu.my

S. A. Sulaiman
e-mail: shaharin@utp.edu.my

It is desired to enhance the crank rocker (CR) engine performance by upgrading engine components through geometry design optimization. An initial mechanism configuration of the crank rocker engine is suggested using algorithm optimization method. The next task is to perform kinematic and dynamic study of the configured as part of the design process.

The study of system kinematics can be done using analytical method, graphical method or iterative numerical methods [5]. Nowadays, it is significant to use rapid and effective method that is capable of solving real life problems, rather than using conventional methods which might consume designer time and effort. Hence, many applications and softwares are being used to accomplish design objectives effectively, with more reliable outcome in considerable time. Of these software, the ones used to perform kinematic analysis and simulation are namely MATLAB, MSC ADAMS, SAM, GIM, Universal Mechanisms, MMtool, and Linkage [5]. On one hand, for coded softwares such as MATLAB, the analyses are performed by constructing models and solving systems of mathematical equations. The implementation of such software is considered more efficient than solving these equations manually. However, the previous process might be very tedious due to the listing of all equations in a logical sequence that must be followed and understood by software compiler. On the other hand, application software such as ADAMS, is a virtual modeling tool, meaning that it only requires mechanism geometry and initial operation condition to generate full body analysis and animate corresponding actions [6].

In this paper, a kinematic analysis study of a particular crank rocker engine mechanism analyzed using two softwares, namely MATLAB and ADAMS. Position, velocity and acceleration of this mechanism are studied under specified driving motion and system constraints. Moreover, to validate mechanism kinematics, a comparison between both simulation methods is conducted using the outcome results of both applications.

2 Analytical Analysis of Crank Rocker Mechanism in MATLAB

Since a MATLAB code is utilized to perform analytical study on this system [7–9], it is required to include related mathematical equations that represent system physics. Hence this section introduces kinematic fundamentals which are involved in analyzing the system. Once the mechanism geometry and other conditions such as Grashof's law and initial conditions are identified, mathematical equations are constructed. This study adopted the usage of vector-loop method for a crank-rocker mechanism kinematic analysis. Figure 1, shows the crank rocker basic geometry which consists of four main linkages namely ground R_1, Crank R_2, Coupler R_3 and Rocker R_4. Also, each respected link angular position with reference to the ground is set to θ_1, θ_2, θ_3 and θ_4, which are (ground linkage angle, crank angle, coupler

Fig. 1 Crank-Rocker basic geometry
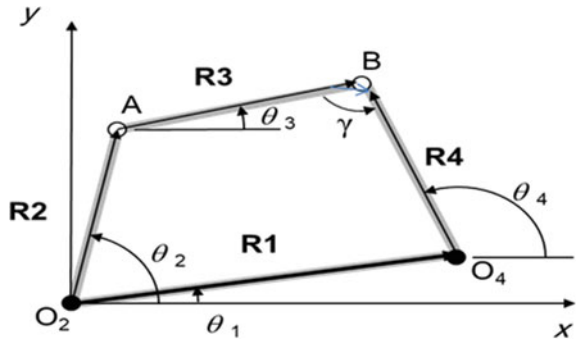
angle and rocker angle) respectively. The angle between the coupler and rocker is called the transmission angle γ. Another angle of concern in this design process is the throw angle φ, which represents maximum rocker displacement, in other words it is obtained by subtracting maximum and minimum crank rocker angle (θ_4).

In this case, the vector loop equation is represented in Eq. (1):

$$R_2 + R_3 = R_1 + R_4 \tag{1}$$

By arranging this relation, we have

$$R_2 + R_3 - R_1 - R_4 = 0 \tag{2}$$

Arbitrary link position relative to the X–Y axis, for each components of Eq. (2) can be rewritten as:

$$R_2 \cos \theta_2 + R_3 \cos \theta_3 - R_4 \cos \theta_4 - R_1 \cos \theta_1 = 0 \tag{3}$$

$$R_2 \sin \theta_2 + R_3 \sin \theta_3 - R_4 \sin \theta_4 - R_1 \sin \theta_1 = 0 \tag{4}$$

Equation (3) represents system component position in the X-axis direction, while Eq. (4) represents system component position in the Y-axis direction. When applying first time derivative on Eqs. (3) and (4), also since is ground link is fixed ($\omega_1 = 0$), we get angular velocity Eqs. (5) and (6):

$$-R_2 \sin \theta_2 \omega_2 - R_3 \sin \theta_3 \omega_3 + R_4 \sin \theta_4 \omega_4 = 0 \tag{5}$$

$$R_2 \cos \theta_2 \omega_2 + R_3 \cos \theta_3 \omega_3 - R_4 \cos \theta_4 \omega_4 = 0 \tag{6}$$

In the above Equations ω_2, ω_3 and ω_4 are angular velocities of links R_2, R_3, and R_4 respectively. Similarly, when applying second time derivative on both Eqs. (3) and (4), they yield angular acceleration Eqs. (7) and (8):

$$- R_2 \sin \theta_2 \alpha_2 - R_2 \cos \theta_2 \omega_2^2 - R_3 \sin \theta_3 \alpha_3 - R_3 \cos \theta_3 \omega_3^2$$
$$+ R_4 \sin \theta_4 \alpha_4 + R_4 \cos \theta_4 \omega_4^2 = 0 \tag{7}$$

$$R_2 \cos \theta_2 \alpha_2 - R_2 \sin \theta_2 \omega_2^2 + R_3 \cos \theta_3 \alpha_3 - R_3 \sin \theta_3 \omega_3^2$$
$$- R_4 \cos \theta_4 \alpha_4 + R_4 \sin \theta_4 \omega_4^2 = 0 \tag{8}$$

Also, α_2, α_3, and α_4 above represent the angular accelerations of links R2, R3, and R4 respectively. Linear velocity for joints between linkages, in this case point A and B as illustrated in Fig. 1, basically can be found by Eq. (9):

$$\upsilon = \omega \times R \tag{9}$$

where, R is linkage length and ω is the angular speed. Likewise, linear acceleration for both points can be found by Eq. (10):

$$a = a_t + a_n \tag{10}$$

where, a_t is the tangential acceleration of this point and a_n is the normal acceleration of the same point, and both can be found by:

$$a_t = \alpha \times R \tag{11}$$

$$a_n = \frac{v^2}{R} \tag{12}$$

Here, α is the angular acceleration of link R, and v is the linear velocity at the measured point. Using all of the above equations with known crank position ($0° \leq \theta_2 \leq 360°$) and constant rotational speed $\omega_2 = C$, MATLAB codes are programmed for system analysis. All remaining parameters are calculated and translated into system simulation analysis.

3   Modelling and Simulation in ADAMS

Unlike programming with softwares such as MATLAB [6, 10, 11], ADAMS has the advantage of less need to write mathematical equations to identify system physics. The required task is only to represent the geometry of this system into the application interface, and then ADAMS has the ability to translate this system into a model for simulation and analysis. ADAMS also allows users to perform static, kinematic and dynamic study on different system conditions, in addition to the ability to simulate and predict the behavior of each case. However, other procedures are required to
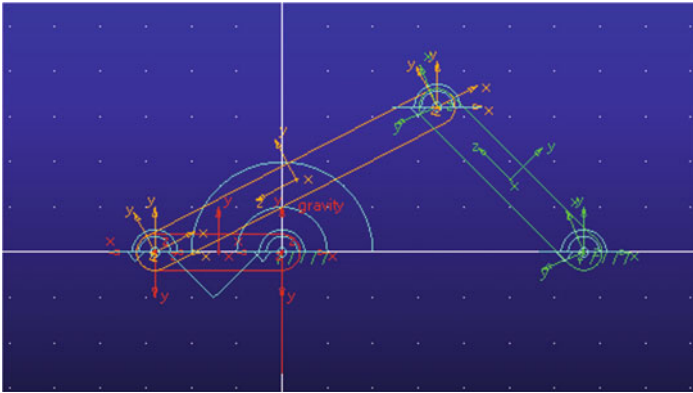
Fig. 2 Crank-rocker model in ADAMS

accomplish valid simulation such as identifying all links and joints dimensions. Also, it is desired to create proper measurements and variables of each object to be included and calculated during the simulation process. After model set-up is finished, the simulation process can be performed, and a full evaluation of system characteristics is created. Figure 2, shows ADAMS workspace environment, the crank rocker mechanism is drawn with corresponding system geometry. Also, joints between linkages are identified with crank joint having a motion function for system initiation.

4 Results and Discussion

Selection of mechanism dimensions was previously conducted using algorithm optimization method [12]. Table 1, shows the geometry specifications of the crank rocker mechanism.

The above data were then supplied to MATLAB and ADAMS as a system configuration. Also, a constant crank rotation speed of 200 rad/sec was applied at the crank link. The ground linkage angle θ_1 is assumed to be zero so that there is no offset between the crank and rocker linkages. The path motion of the mechanism is illustrated in Fig. 3, by MATLAB. In ADAMS, an animated simulation of the system is firstly performed to validate this mechanism then carry on system analysis.

Table 1 Crank-rocker mechanism dimensions

R_1 /mm	R_2 /mm	R_3 /mm	R_4 /mm	θ_2 /degree	γ /degree
33.23	14.03	35	21.07	180	84.34
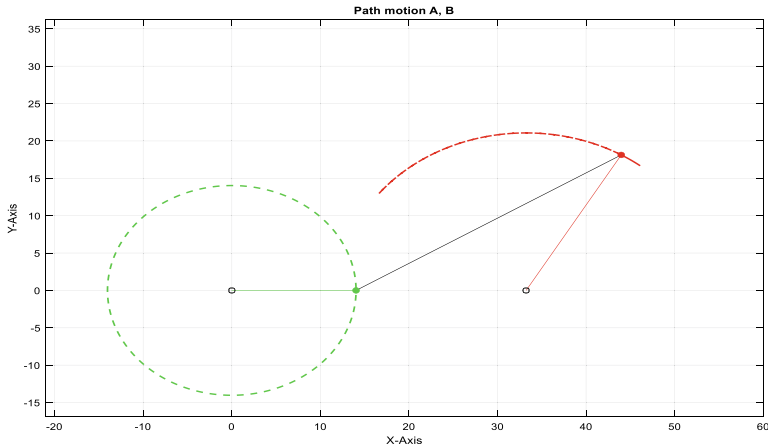
Fig. 3 Crank-Rocker mechanism path motion in MATLAB

Figure 3, shows the complete rotation of the crank link (green), and the corresponding rocker path motion (red). After analyzing this system path behavior, a simulation was conducted in both softwares for studying the kinematic performance of this mechanism. The outcome results were plotted in overlayed graphs for better illustration and verification purpose. The plots of angular position, velocity and acceleration versus crank position (θ_2), obtained from MATLAB and ADAMS simulations are compared and shown in Figs. 4, 5, 6, 7 and 8.

Figures 4, 5, 6, 7 and 8, shows overlaying graphs of the different kinematic analysis for the crank-rocker mechanism by MATLAB and ADAMS. These data are representing mechanism various positions, angular and linear velocities, angular
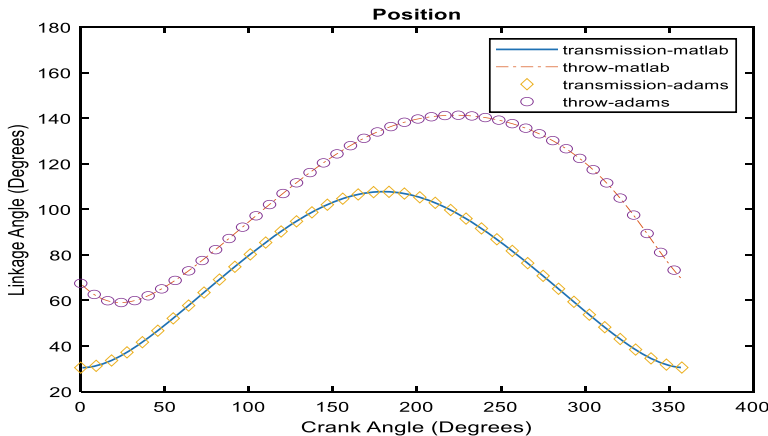
Fig. 4 Kinematic simulation comparison results between MATLAB and ADAMS, Mechanism angular position

Fig. 5 Kinematic simulation comparison results between MATLAB and ADAMS, Mechanism linear velocity
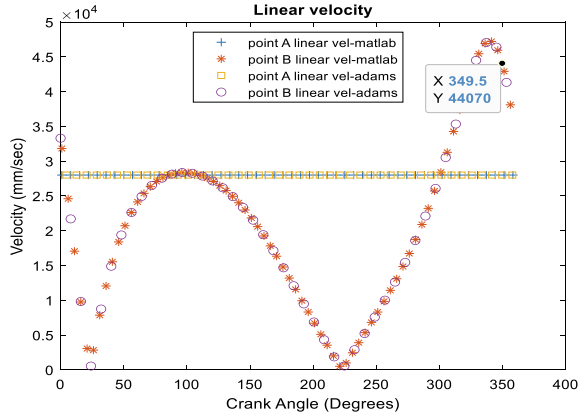
Fig. 6 Kinematic simulation comparison results between MATLAB and ADAMS, Mechanism angular velocity
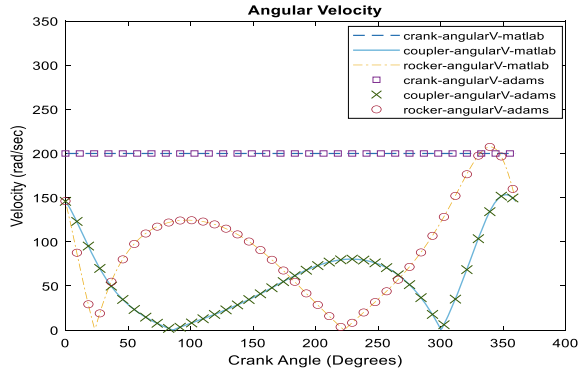
Fig. 7 Kinematic simulation comparison results between MATLAB and ADAMS, Mechanism linear acceleration
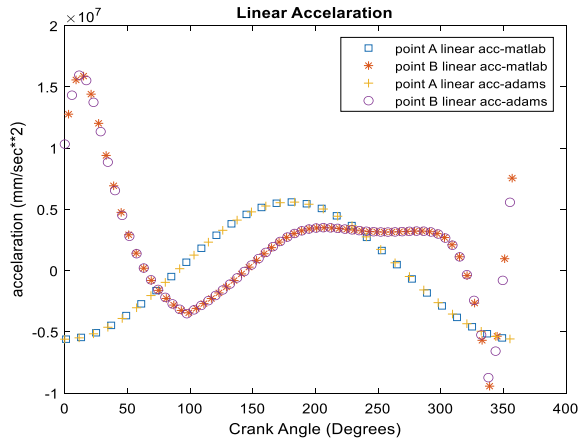
Fig. 8 Kinematic simulation comparison results between MATLAB and ADAMS, Mechanism angular acceleration
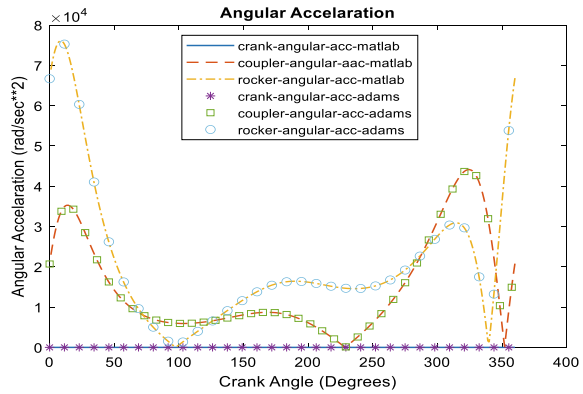
and linear accelerations with respect to the crank position. When examining and comparing results by both softwares, it can be noticed that both models are showing almost same readings, which gives an indication that both results are verifying each other. Hence, these results can be considered and applied for further system analysis and enhancement.

5 Conclusion

This paper studied the kinematic characteristics of a possible mechanism design of the crank rocker engine. Two different approaches were implemented for this goal, analytical coding using MATLAB and animated model using ADAMS. Both methods simulated this system for determined crank rotation speed and results show this mechanism kinematic behavior in term of position, velocity and acceleration. A comparison between outcome results by both softwares validated each other, where all plots were almost identical. These two methods considered practical and useful for kinematic analysis of such system.

References

1. Mohammed SE, Baharom MB, Rashid A, Aziz A (2016) Thermodynamic analysis of a single cylider crank-rocker engine. ARPN J Eng Appl Sci 11:12239–12246
2. Mohammed SE, Baharom MB, Aziz ARA (2017) Performance and combustion characteristics of a novel crank-rocker engine. J Mech Sci Technol 31:3563–3571
3. Hassan SEM (2018) Development, performance and emission assessment of a single-cylinder air-cooled gasoline Crank-Rocker engine. Universiti Teknologi Petronas
4. Mohammed SE, Baharom MB (2019) Modelling of combustion characteristics of a single curved-cylinder spark-ignition crank-rocker engine. Energies 1–14

5. Verma SK, Kumar R, Chittawadigi RG, Saha SK (2016) Kinematic analysis of mechanisms using velocity and acceleration diagrams (VAD) module in mechanalyzer software. In: The proceedings of the Asian conference on multibody dynamics, p. 01_771082
6. Hroncová D, Delyová I, Frankovský P (2014) Kinematic analysis of mechanisms using MSC Adams. Appl Mech Mater 611:83–89
7. Norton RL (2012) Design of machinery: an introduction to the synthesis and analysis of mechanisms and machines, McGraw-Hill, New York
8. Mohammadzadeh A (2007) Analytical synthesis and analysis of mechanisms using MATLAB and SIMULINK. Am Soc Eng Educ 12(242):1
9. Myszka DH, Celik A (2012) Machines and Mechanism. Prentice Hall
10. Hroncová D, Binda M, Šarga P, Kičák F (2012) Kinematical analysis of crank slider mechanism using MSC ADAMS/View. Procedia Eng 48:213–22
11. Hroncová D, Frankovský P, Virgala I, Delyová I (2014) Kinematic analysis of the press mechanism using MSC Adams. Am J Mech Eng 2:312–315
12. Albaghdadi AM, Baharom MB, bin Sulaiman SA (2021) Parameter design optimization of the crank-rocker engine using the FMINCON function in MATLAB. IOP Conference Series: Materials Science and Engineering 1088(1):012072. https://doi.org/10.1088/1757-899X/1088/1/012072

Application of Machine Learning Models in Gas Hydrate Mitigation

Sachin Dev Suresh, Bhajan Lal [ID], Ali Qasim [ID], Khor Siak Foo, and Jega Divan Sundramoorthy

Abstract The production of oil and natural gas contributes to a significant amount of revenue generation in Malaysia thereby strengthen the country's economy. The flow assurance industry is faced with impediments during smooth operation of the transmission pipeline in which gas hydrate formation is the most important. It affects the normal operation of the pipeline by plugging it. Gas hydrate is a crystalline structure composed of a network of hydrogen bonds between host molecules of water and guest molecules of the incoming gases under high pressure and low temperature conditions. Industry uses different types of chemical inhibitors in pipeline to suppress hydrate formation. To overcome this problem, machine learning algorithm has been introduced as part of risk management strategies. The objective of this research is to evaluate the various types of machine learning models used to predict the gas hydrate formation where the input parameters are gas composition, pressure and concentration of inhibitor and the output parameter is hydrate deposition/formation temperature (HDFT). Three machine learning models are compared: Artificial Neural Network (ANN), Least Square version of Support Vector Machine (LSSVM), and Extremely Randomized Trees (Extra Trees). Comparison of the three different machine learning models is based on the correlation coefficient, R^2. The best choice of machine learning model that has highest R^2 is obtained by Extra Trees

S. D. Suresh · B. Lal (✉) · A. Qasim
Chemical Engineering Department, Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar Perak, Malaysia
e-mail: bhajan.lal@utp.edu.my

B. Lal · A. Qasim
CO2 Research Centre (CO2RES), Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar Perak, Malaysia

K. S. Foo
PTTEP-Malaysia, Petronas Twin Towers Kuala Lumpur City Centre, Level 26-30, Tower 2, 50088 Kuala Lumpur, Malaysia

J. D. Sundramoorthy
Federal Territory of Kuala Lumpur, Baker Hughes (M) Sdn. Bhd, 207 Jalan Tun Razak, 50400 Kuala Lumpur, Malaysia

model of 0.9991 compared to other two machine learning models which predicted R^2 value greater than 0.96.

Keywords Gas hydrate · Machine learning · Hydrate formation temperature

1 Introduction

In the 1930s, gas hydrate deposition was discovered in pipelines where several attempts were made to understand the clathrate hydrate in terms of compound, composition, equilibrium state and manner of handling them [1]. Gas hydrate, mostly known as clathrate hydrates exist when sufficient amount of water molecules in oil and hydrocarbon vapour form crystallize during transient flow condition [2, 3]. Gas hydrates are developed in the crude oil and natural gas industries' output, refining and transmission facilities. [4–8]. Depending on the type and size of guest molecules, the gas hydrates are produced in 3 structures, which are I, II and H [7, 9]. The unit cells of Structure I had eight cavities (two smalls and six large ones) and were formed by hydrogen-bound 46 water molecules. Structure II unit cells, by contrast, contained 136 molecules of water and were enclosed in 24 cavities, including 16 small and 8 wide cavities. Whereas structure H is a mixture of methane, propane, isobutene, CO_2, H_2S, N_2, and stabilises with one of the two hydrate structures forming micro cavities[10, 11]. According to Qin et al. [12] and Olajire [13], at high pressure and low temperatures as well as long distances, gas hydrates are produced, often contained in subsea oil and gas pipelines. Although gas hydrate production contributes to critical flow assurance problem, it is therefore important to address this issue by implementing effective preventive and predictive strategies [4].

There are several risk management strategies that has been used to prevent the growth of hydrates which are dehydration of natural gas, usage of inhibitors such as thermodynamic hydrate inhibitors (THIs) and low dosage hydrate inhibitors (LDHIs) [12–16]. Methanol (MeOH) and monoethylene glycol (MEG) are two commercially used Thermodynamic hydrate inhibition compounds [17]. The goal of these THIs is to interrupt the process in which hydrates form, often like water–ice antifreeze, to ensure safety under the most severe hydrate forming conditions. However, the downside of these THIs is that the hydrate condition gets extreme as the subcooling process is of better consistency, leading to more inhibitor requirements. This leads to a rate limit of methanol supply due to availability, storage and injection restrictions, leading to an adverse effect on capital costs, operating costs, non-optimum performance, as well as an increased risk of hydrate plug formation [1].

In contrast with THIs, LDHIs play a crucial position as an alternative to THIs in terms of lower doses of inhibitors used and dosage levels that cause lower loss of inhibitors due to evaporation. Other than that, LDHIs would decrease the chemical storage and injection needs, which will have a beneficial impact by minimising CAPEX and OPEX [1]. Two kinds of LDHIs, such as kinetic hydrate inhibitors (KHI), are used to slow hydrate crystallisation and anti-agglomerants (AAs), which facilitate

hydrate molecule distortion [12]. For simulation purposes, computer intelligence can be used, such as machine learning (ML) models, to test the accuracy of the inhibitor used, composition of gas molecules, gas hydrate temperature and pressure regulation [4].

Usually, ML is called a universal toolbox, ready to be used for classification issues, defining an appropriate group for a new collection of findings, and regression tasks that approximate the relationship between groups of data provided. ML is fundamentally a numerical representation of a phenomenon, provided a certain value and based on a certain setting, aimed at performing a task. ML techniques may be classified based on learning activity priorities that define patterns for detection and estimation, action learning, or inductive methods of learning. The algorithms can be further grouped into three distinct families of learning, known as supervised learning, unsupervised learning and reinforcement learning [18]. The main objective of this paper is to compare numerous forms of ML models that will be used, such as Aritifical Neural Network (ANN), Least Square Support Vector Machine (LSSVM) and Extremely Randomized Trees (Extra Trees), which are used to help create advanced gas hydrate control techniques [12]. In this article, theory and method employed for different ML models have been discussed. Afterwards, a comparison among the models has been performed and the their effectiveness has been assesed and identified. Finally, on these basis, some important conclusion have been drawn and discussed.

2   Methods and Theory

Authors like Rebai et al. [19], stated that machine learning techniques such as ANN, LSSVM and Extra Trees are considered eligible to reproduce any form a function and have been successful in predicting the required output of the gas hydrate formation which is hydrate formation/deposition temperature (HDT/HFT). The input parameters that has been used in this 3 ML models were concentration of the inhibitors, gas composition and pressure. Each ML models theory and method will be explained below in details.

2.1   *Least Square Version of Support Vector Machine (LSSVM)*

LSSVM development is aimed at reducing the complexity of the Support Vector Machine (SVM) and to eliminate complexity in the optimization process, linear formulas are deciphered by linear encoding rather than solving quadratic encoding equations to improve their speed. LSSVM algorithm's output can be overcome by fixed linear equality rather than inequality constraints [4]. General formulation that being used in approximating a given data set $\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$ into
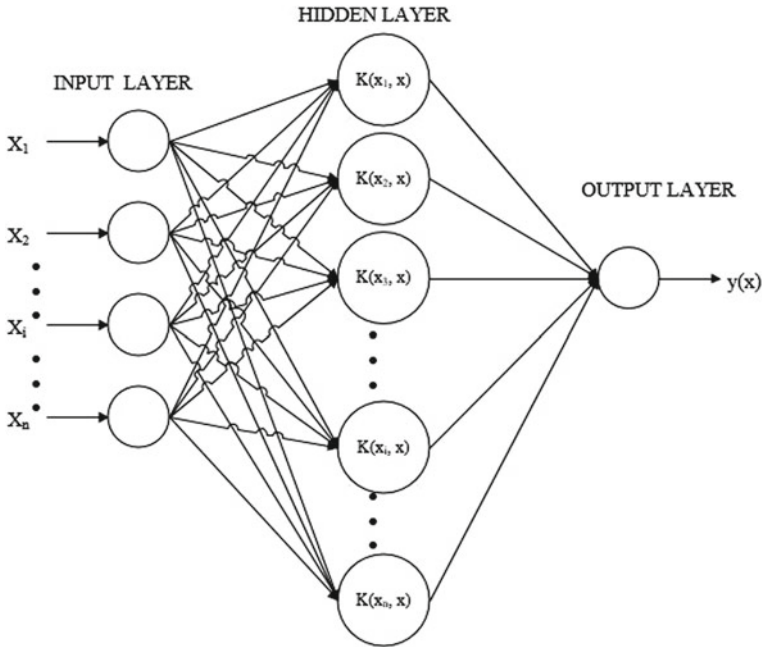
Fig. 1 LSSVM model architecture [20]

consideration by employing following nonlinear function [14]:

$$f(x) = \langle w, \Phi(\mathrm{x}) \rangle + b \tag{1}$$

where $\langle ., . \rangle$ represent dot product; $\Phi(\mathrm{x})$ denotes the nonlinear function that performs regression; b and w are bias terms and weight vector, respectively. The LSSVM model structure can be seen in Fig. 1 where the K (x_i, x_j) is defined as kernel function [20].

2.2 Extremely Randomized Trees (Extra Trees)

Tree models were identified by the assembly method named as Extra Trees, in which this process produces multiple trees from the complete testing examples in terms of regression or decision trees. Extra Trees is a top-down procedure where it forms a joint of untrimmed choice or regression tress where the act of splitting nodes by selecting action points randomly completes and uses the entire experiment model to grow these trees, distinguishing it from other tree-based ensemble approaches [4].

The growth of the tree is in the top-down order where the numbering is from left to right, as can be seen from Fig. 2. The value of a parameter is compared with the
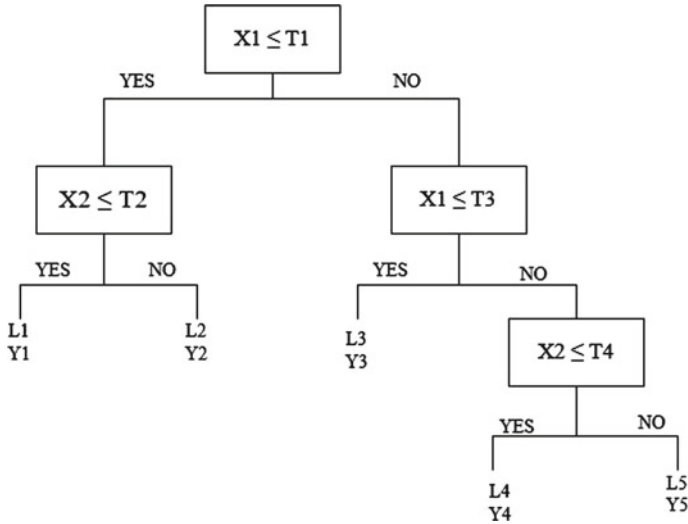
Fig. 2 Study of Extra Trees framework [4]

threshold value of each internal node. If the shown inequality above each node is satisfied, the path is either the left branch or the path is not the right branch [4].

2.3 Artificial Neural Network (ANN)

Artificial neural networks (ANN) are a method of calculation based on how the brain conducts calculations [20, 21]. According to Samsudin et al. [20], this model has the potential to carry out extensive input–output mapping and can create a network that approximates nonlinear functionality. Figure 3 shows ANN feed-forward backward propagation consists of three layers; data transmitted is the input (first layer), interpreted data is hidden layer, and result provided is output (last layer) [20].

Output of the ANN assuming a linear output neuron j, a single hidden layer with h sigmoid hidden nodes and the output variable (x_t) is given by [20]:

$$x_t = g\left(\sum_{j=1}^{h} w_j f\left(s_j\right) + b_k\right) \quad (2)$$

where $g(.)$ is the linear transfer function of the output neuron k and b_k is its bias, w_j is the connection weights between hidden layers and output units, $f(.)$ is the transfer function of the hidden layer.

Below Fig. 4, shows the exact prediction of the distribution of temperature in the gas-dominated system, with due regard to the exothermic effect of the composition
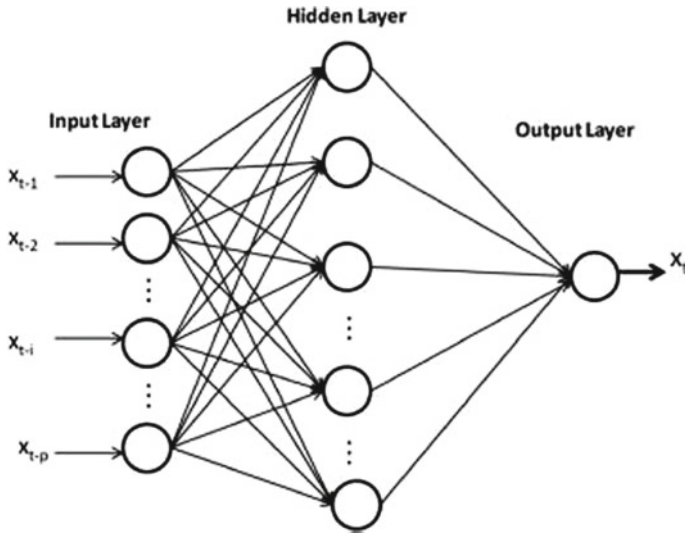
Fig. 3 Sample Structure of ANN [20]

Fig. 4 The interdependency in the gas-dominated environment between hydrate formation, hydrate deposition and temperature variation
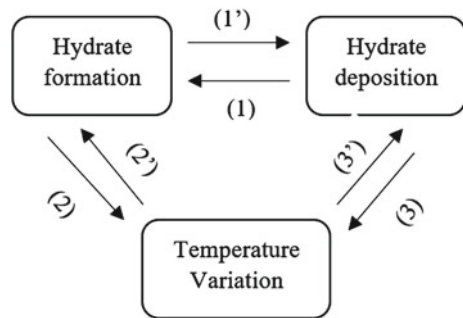
of hydrate and the resistive effect of the hydrate deposition layer on the rate of heat transfer in the pipeline. The number on the Fig. 4 is decoded by (1) supplying hydrate forming materials for deposition, (1') reflecting the indirect effect by temperature, pressure, and effective inner diameter variations. Whereas (2) signifies heat release from hydrate formation by interaction of free water and gas; (2') indicates the effect of temperature differences on the hydrate formation sub-cooling. The heat tolerance effect denoted by (3) of the hydrate layer on the heat transfer while (3') indicates the indirect effect influencing the rate of hydrate formation [13].

3 Results and Discussion

It is assumed that the accuracy and credibility of every statistical method is linked to the universality and validity of the database used for its foundation. To compare the results based on the several factors such as correlation coefficient, R^2, comparison of the three ML models must be represented in terms of hydrate dissociation/ formation temperature (HDT/HFT). All the data and information taken and referred from [4, 22] in order to obtain effective comparison and accuracy.

Based on article Yarveicy and Ghiasi [4], in the first step,1800 experimental data points of different gas systems for liquid water-hydrate-vapor (LHV) or ice-hydrate-vapor (IHV) balance were collected and divided into two classes called "training set" and "test set" for Extra Trees and LSSVM models. 90% were used as training procedures from the data point and the remaining 10% were used as testing procedures. The primary goal is to approximate the HDT / HDFT of gases where electrolysis and alcohol(s) are present in pure water or aqueous solutions, where it is correlated as function of concentration of inhibitor(s) ($C_{additive}$), pressure (*P),* and the gas composition (Z_i):

$$HDFT = f(C_{additive}P, Z_i) \qquad (3)$$

The composition of gas (Z_i) consisting of C_1, C_3, i-C_4, i-C_5, n-C_4 n-C_6, N_2, and CO_2 in (mol %). Consequently, LSSVM model used radial basis function (RBF) kernel function in its methodology with optimization of Coupled Simulating Annealing (CSA). Two considerations were crucial to obtaining the HDT / HFT Extra Tree algorithm, which was the sum of tress generated for calculating significance and maximum depth of each tree using the test and error method. From trial and error method, optimum number of tress obtained was 5 as well as maximum depth for the Extra Trees algorithm changed from 3 to 40. By monitoring the error, best performance can be seen at maximum depth value of 30 [4].

The objective of the ANN model was the same as LSSVM and Extra Trees, which was to predict the temperature of the gas mixture (HFT / HDT), where 279 sweet and sour gas experimental data points were collected. For the training procedure, 201 data points were used, 42 data points were used for testing set. The input was a gas composition composed of H_2S, CO_2, and CH_4 in moles percentage (*mol%*) and pressure (*P*) in mPa, respectively [22]. For the three models [14], a comparative study between the predicted hydrate formation temperature (HFT / HDT) values and the experimental data for the train, validation and test data was developed. Below seen Table 1, shows the results obtained from ANN, LSSVM and Extra Trees models in terms of R^2 and AARD in percentage (%).

Below mentioned are the formulas used to calculate the values placed in Table 1 which are coefficient of determination percent ((R^2%), and average absolute relative deviation percent (*AARD%*) defined by Eqs. (3) and (4) in respective order, are mathematical parameter used to estimate the HDT / HDF of gas hydrate as well as

Table 1 Error evaluation outcomes for the presented Extra Trees, LSSVM and ANN models

Parameter/model	Extra trees			LSSVM			ANN		
	Train	Test	Total	Train	Test	Total	Train	Test	Total
R^2%	99.91	97.46	99.7	97.54	92.4	96.99	99.47	98.82	99.28
AARD%	0.01	0.35	0.04	0.28	0.61	0.32	0.1331	0.2137	0.1509

determining the reliability and efficiency of the built Extra Trees, LSSVM and ANN models.

$$R^2\% = 100\left(1 - \frac{\sum_i^n (o_i - t_i)^2}{\sum_i^n (o_i - average(t_i))^2}\right) \tag{4}$$

$$AARD\% = 100 \sum_{i=1}^n \left(\frac{o_i - t_i}{t_i}\right)/n \tag{5}$$

where n denotes the number of samples; o_i and t_i are the predictions of the model and corresponding targets, respectively.

The error analysis results for the tabulated in Table 1 shows an excellent compliance with the experimental hydrate deposition / forming temperature (HDT / HDF) for the three distinct types of ML models that are acceptable outcomes [4]. Nevertheless, as the AARD percent is 0.04 relative to the other models with AARD percent of 0.32 and 0.1509 for LSSVM and ANN, respectively, the suggested tree-based model has stronger prediction.

4 Conclusion

Interaction in this paper provides the significant comparison of ML based models such as Extra Trees, LSSVM and ANN where their respective methodologies were employed for phase equilibrium modelling of clathrate hydrates. From the coefficient of determination percent, R^2 is found to be greater than 96% for all three ML models. Based on the analysis of the error study, it can be argued that the Extra Trees algorithm, which was used for the first time in the supremacy of gas hydrate phenomenon experimentation, is extremely capable of demonstrating the formation of gas hydrate depending on the temperature, resulting in the Extra Trees providing more precision than the LSSVM and ANN models in terms of AARD percent of 0.04, which is the lowest compared to LSSVM and ANN models.

References

1. Bavoh CB, Lal B, Osei H, Sabil KM, Mukhtar H (2019) A review on the role of amino acids in gas hydrate inhibition, CO2 capture and sequestration, and natural gas storage. J Nat Gas Sci Eng 64:52–71

2. Bavoh CB, Partoon B, Lal B, Kok Keong L (2017) Methane hydrate-liquid-vapour-equilibrium phase condition measurements in the presence of natural amino acids. J Nat Gas Sci Eng 37:425–434

3. Bomba J, Chin D, Kak A, Meng W (2018) Flow assurance engineering in deepwater offshore-past, present, and future. In: Offshore technology conference

4. Ghiasi MM, Yarveicy H, Arabloo M, Mohammadi AH, Behbahani RM (2016) Modeling of stability conditions of natural gas clathrate hydrates using least squares support vector machine approach 223:1081–1092

5. Kazemeini M, Freidoonian F, Fattahi M (2012) Developing a mathematical model for hydrate formation in a spray batch reactor 2:244–247

6. Khamehchi E, Shamohammadi E, Yousefi SH (2013) Predicting the hydrate formation temperature by a new correlation and neural network

7. Khan MS, Bavoh CB, Partoon B, Nashed O, Lal B, Mellon NB (2018) Impacts of ammonium based ionic liquids alkyl chain on thermodynamic hydrate inhibition for carbon dioxide rich binary gas. J Mol Liq 261:283–290

8. Khan MS, Yaqub S, Manner N, Karthwathi NA, Qasim A, Mellon NB, Lal B (2018) Experimental equipment validation for methane (CH4) and carbon dioxide (CO2) hydrates. IOP Conf Ser Mater Sci Eng

9. Nashed O, Dadebayev D, Khan MS, Bavoh CB, Lal B, Shariff AM (2018) Experimental and modelling studies on thermodynamic methane hydrate inhibition in the presence of ionic liquids. J Mol Liq 249:886–891

10. Nasir Q, Lau KK, Lal B, Sabil KM (2014) Hydrate dissociation condition measurement of CO2-rich mixed gas in the presence of methanol/ethylene glycol and mixed methanol/ethylene glycol + electrolyte aqueous solution. J Chem Eng Data 59(11):3920–3926

11. Olajire A (2020) Flow assurance issues in deep-water gas well testing and mitigation strategies with respect to gas hydrates deposition in flowlines—a review. J Mol Liq 114203

12. Partoon B, Sahith SJK, Lal B, Maulud ASB (2020) Gas hydrate models. In: Chemical additives for gas hydrates. Springer, Cham, pp 67–85

13. Qasim A, Khan MS, Lal B, Abdullah MZ, Maulud AS, Simulation of hydrate phase boundary for natural gas mixture with high CO2 content through simulation

14. Qasim A, Khan MS, Lal B, Shariff AM (2019) A perspective on dual purpose gas hydrate and corrosion inhibitors for flow assurance. J Pet Sci Eng 183:106418

15. Qasim A, Lal B, Shariff AM, Ismail MC (2020) Role of ionic liquid-based multipurpose gas hydrate and corrosion inhibitors in gas transmission pipeline. In; Nanotechnology-based industrial applications of ionic liquids. Springer, pp 221–244

16. Qin H, Srivastava V, Wang H, Zerpa LE, Koh CA (2019) Machine learning models to predict gas hydrate plugging risks using flowloop and field data. In: Offshore technology conference

17. Rafique D, Velasco L (2018) Machine learning for network automation: overview, architecture, and applications [Invited Tutorial]. 10(10):D126–D143

18. Rebai N, Hadjadj A, Benmounah A, Berrouk AS, Boualleg SM (2019) Prediction of natural gas hydrates formation using a combination of thermodynamic and neural network modeling. 182:106270

19. Sahith SAK, Pedapati SR, Lal B (2019) Application of artificial neural networks on measurement of gas hydrates in pipelines

20. Samsudin R, Saad P, Shabri AJH, Sciences ES (2011) River flow time series using least squares support vector machines 15(6)

21. Soroush E, Mesbah M, Shokrollahi A, Rozyn J, Lee M, Kashiwao T, Bahadori A (2015) Evolving a robust modeling tool for prediction of natural gas hydrate formation conditions. 12:45–55

22. Yarveicy H, Ghiasi MM (2017) Modeling of gas hydrate phase equilibria: extremely randomized trees and LSSVM approaches. J Mol Liq 243:533–541

Analysis of Signal Sensing with Adaptive Threshold for Energy Detector in Cognitive Radio Systems

Biji Rose and B. Aruna Devi

Abstract Overviewing the past the past period of 20 years, it is evident that in certain geographical regions, the spectrum, lay unutilized whereas other areas face spectrum crunch. To boost the spectral efficiency, unlicensed users (secondary users) will be allowed to access unutilized parts of the spectrum and adapt their communications. It assists to utilize them while minimizing the interference on licensed users. Such ability is the predominant feature of cognitive radio (CR) nodes, which needs, algorithms and protocols for spectrum sensing, sharing, and management. The rationale is to facilitate optimization of the spectrum and make a conducive environment between licensed and unlicensed users. The seamless method of spectrum sensing is Energy Detection method, since it does not require to have priory knowledge of Primary User (PU).In this paper, sequential sensing event at the Energy Detector is used to decide on the optimal adaptive threshold value over Rayleigh fading channel.

Keywords Cognitive radio · Energy detector

1 Introduction

The year 2020 which is not yet over, went through health and economic crisis worldwide due to Covid-19, faced lockdowns of months. All people during this pandemic situation were connected and jointly fight the Covid-19 situation using digital technologies, which completely depends on mobile wireless network. At the outburst of pandemic situation, an emergency meeting was called in starting week of April and ITU emphasized on the importance of communication technologies in disseminating timely critical information, supporting e-learning for more than 1.5 billion

B. Rose (✉)
Department of IT, Dr. N.G.P. Institute of Technology, Coimbatore, Tamil Nadu, India
e-mail: bijirose@drngpit.ac.in

B. A. Devi
Department of ECE, Dr. N.G.P. Institute of Technology, Coimbatore, Tamil Nadu, India

students, training workers by digital means to improve productivity, and promoting e-businesses [1].

The service mobile wireless network increased tremendously in 2020 due to pandemic situation. Already in 2018 it was reported: the need of smartphones, tablets, laptops, and other wireless devices that allow users to browse the Web, use email, and download videos, multimedia, and applications, by 2023 will grow. Thus by 2023, the average 5G connection speed will reach 575 Mbps, the average Wi-Fi speeds will grow to 92 Mbps, and there will be 3.6 networked devices/connections per person and nearly 10 devices and connections per household [2].

As a result, the growing demand for mobile wireless broadband is expected to surpass the usable radio spectrum. Prior highlighting the concepts of the performance analysis, standards of spectrum sensing include IEEE 802.22 WRAN (wireless regional area network) and its amendments, IEEE 802.11af for wireless LANs, IEEE 1900.x series, and licensed shared access (LSA) for LTE mobile operators [3]. Spectrum sensing technique achieves detection of channel state by (1) Energy detection (ED), (2) Cyclostationary detection, (3) Matched filter-based detection, (4) Waveform-based detection, (5) Covariance absolute value detection (CAV), and (6) Eigen-values based detection (EVD). The spectrum sensing problem is solved as a binary hypothesis test. The main attribute of interest is decision statistic, detection, and false-alarm probabilities and the decision threshold. The Central Limit Theorem (CLT) is used to achieve energy detection with prescribed performance level.

Studies of spectrum usage have shown that this spectrum allocation paradigm often results in severe underutilization of the spectrum [4, 5]. In opportunistic or dynamic spectrum access, a band of licensed spectrum can be used by unlicensed users whenever it is not being used by licensed users. The licensed user is often referred to as the primary user, whereas the unlicensed user is called the secondary user. In addition, the secondary users must be capable of sensing the radio environment to determine the spectrum hole opportunities for dynamic spectrum access. Such capabilities are realized in cognitive radio technologies [6].

In a cognitive radio system, a Markovian model of the primary user can be used for state prediction, which in turn can prevent unintended collisions [7].A popular edge detection technique uses the continuous wavelet transform (CWT) to decompose the edge detector into multiple resolutions and multiplies the resolutions together, which has a beneficial effect of reducing the noise [8]. While the novel Discrete Wavelet Transform (DWT) based hybrid edge detection algorithm, where we use conventional DWT technique in low noise variance scenario and DWT-Multiscale-Sum (DWTMS) technique in high noise variance scenario depending on the Threshold Value. The DWT Hybrid Edge Detection gives higher sensing accuracy, and more noise reluctance for the secondary user in the cognitive radio system compared to CWT [9].

Thus, an adaptive spectrum sensing algorithm in which the threshold is made adaptive to the fading environment is the need of the hour [10].

2 Related Works

In the case of relatively low SNR,when the difference between noise and signal energy is not distinct, single-threshold detection methods are easy to cause SUs to perceive the existence of a PU incorrectly, which will reduce the detection probability of the entire system. To improve the detection the single threshold is replaced over the years by adaptive or dynamic threshold.

There are methods, named three-event ED, which makes decision considering the event immediately before and the one immediately after current sensing event. The primary users (PU) activity duty cycle value was exploited for tracking the changes of the PU state [11–13].

Furthermore, an adaptive double threshold based on history energy detection is proposed for Cooperative Spectrum Sensing. The weighting coefficient of thresholds according to the SNR of all cognitive nodes is calculated during sensing period; thus adjusting the upper and lower thresholds, if the average history energy is still between two thresholds, the single-threshold method will be used for the end decision. Finally, the fusion center aggregates the detection results of each node and obtains the final cooperative conclusion through "or" criteria [14, 15].

However, the double threshold methods mentioned above are all no-decision when the energy is between the high and low thresholds. The multilevel quantization method is adopted in the uncertain region of the traditional double-threshold energy detection to solve the problem that the sensing information is ignored between the two determined thresholds [16, 17].

In this method, an adaptive threshold method to overcome sensing failure at very low SNR with uncertain noise power using a check parameter and double threshold concept is proposed. It was noted that this method improved the detection probability at a low SNR [18].

Nowadays, many Machine Learning algorithms are put forward for Optimization of Adaptive Threshold in energy detection. The machine learning algorithm looks for the optimum decision threshold, which is the most important parameter to decide the presence or absence of the primary user, using historical detection data [19].

3 Methodology

Let the complex signal received at time t by the secondary user in the desired band be [20]

$$y(t) = \begin{cases} n(t): & H_0 \\ h(t)s(t) + n(t): & H_1 \end{cases} \tag{1}$$

where n(t) is an additive complex white Gaussian noise process, h(t) represents a fading process (e.g., nonfading means h(t) = 1), and s(t) is a signal transmitted by

Fig. 1 Hypothesis model
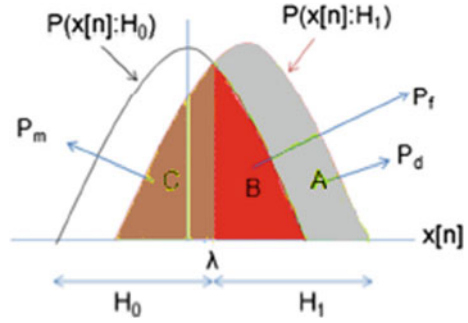
the primary node [20]. The sensing decisions are made using N >> 1 samples of y(t). A decision statistic T(y) is used for hypothesis testing, given by

$$\begin{cases} T(y) < \lambda accept : H_0 \\ T(y) > \lambda accept : H_1 \end{cases}$$

where $0 < \lambda < \infty$ is called the decision threshold. The reliability associated with the decision rule (2) can be characterized by Probability of Detection P_d and Probability of False Alarm P_f is shown in Fig. 1. Mathematically, it is a conditional probability given as below:

$$P_d = \Pr(T(y) > \lambda | H1) \tag{2}$$

$$P_f = \Pr(T(y) > \lambda | H0) \tag{3}$$

The exact values of P_d and P_f depend on how T(y) is constructed using received samples, channel estimates, propagation characteristics, the choice of the threshold, and other information. The false-alarm and detection probabilities may be approximated using CLT for large number of samples (N), as

$$P_f = Q\left(\frac{\lambda - 2N}{2\sqrt{N}}\right) \tag{4}$$

$$P_d = Q\left(\frac{\lambda - 2N(1+\gamma)}{2(1+\gamma)\sqrt{N}}\right) \tag{5}$$

where γ is the SNR and Q is the standard Gaussian upper tail probability function given as $Q(u) = \frac{1}{2\pi} \int_u^\infty \exp\left(\frac{-x^2}{2}\right) dx$.

The probability of missed detection is given as P_m

$$P_m = 1 - P_d \tag{6}$$

Therefore, the error decision probability P_e is written as

$$P_e = P_f + P_m \tag{7}$$

Substituting Eqs. (6) in (7), results in

$$P_e = P_f + (1 - P_d) \tag{8}$$

From Eqs. (4) and (5) suitable sample size N can be calculated from P_f and P_d as,

$$N = \left[\gamma^{-1} Q^{-1}(P_f) - \left(1 + \gamma^{-1}\right) Q^{-1}(P_d) \right]^2 \tag{9}$$

For constant detection probability (CDR) the normalized threshold λ_d (optimum threshold) can be calculated as,

$$\lambda_d = 2\sqrt{N} \left(Q^{-1}(P_f) + \sqrt{N} \right) \tag{10}$$

For constant false alarm rate (CFAR) the normalized threshold λ_f (optimum threshold) can be calculated as,

$$\lambda_f = 2(1 + \gamma)\sqrt{N} \left(Q^{-1}(P_f) + \sqrt{N} \right) \tag{11}$$

Optimum Threshold is chosen to make P_d large ($P_d \geq 0.9$) and P_f small ($P_f \leq 0.1$).CFAR and CDR is used for the selection of threshold [21]. It has been noted that CFAR improves the throughput [22], but CFAR fails to give enough protection to the PU as compared to that of the CDR approach. To improvise the overall throughput at low SNR, the combination of CFAR and CDR is utilized to select the optimum threshold [23]. In this paper, considering all of the above aspects, the optimal threshold is selected for ED which jointly achieve the targeted P_f and P_d at low SNR.

3.1 Optimum Threshold Condition

The error decision probability Pe is given by

$$\lambda = (1 - \delta)P_f + \delta(1 - P_d) \tag{12}$$

where δ $(0 \leq \delta \leq 1)$ is PU's spectrum usage ratio and $(1-\delta)$ is PU's channel is vacant & noise occupied, $(1-P_d)$ gives probability of missed detection.

The adaptive threshold to minimize the error decision probability is

$$\lambda = \sigma_n^2 \cdot \frac{(1+\gamma)}{(2+\gamma)} \cdot \left\{ 1 + \sqrt{1 + \frac{4(2+\gamma)ln.\left[(1+\gamma).\frac{(1-\delta)}{\delta}\right]}{N\gamma}} \right\} \qquad (13)$$

where $\sigma_n^2 =$ noise variance. Thus, optimum threshold setting should consider both the probability of false alarm and the probability of mis-detection in order to benefit both PU and SU.

4 Conclusion

The overall throughput at low SNR is improved by the combinational use of the optimum threshold. Thus, the optimal threshold is selected for ED which jointly achieve the targeted P_f and P_d at low SNR. While the SNR increases P_d is also increases.

References

1. Special emergency session of the broadband commission pushes for action to extend internet access and boost capacity to fight COVID19. https://www.itu.int/en/mediacentre/Pages/PR05-2020 Broadband Commission-emergency-session-internet-COVID-19.aspx, online. Accessed 20 Apr 2020
2. Cisco visual networking index Global mobile data traffic forecast update, 2018–2023. https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html
3. Gavrilovska L, Denkovski D, Rakovic V, Angjelichinoski M (2014) Medium access control protocols in cognitive radio networks: overview and general classification. IEEE Commun Surveys Tutor 16(4):2092–2124
4. FCC (2002) Spectrum policy task force. Technical report 02–135, Rep. ET Docket, Federal Communications Commission
5. Shared Spectrum Company (2010) General survey of radio frequency bands: 30MHz to 3 GHz. Technical report
6. Haykin S (2005) Cognitive radio: brain-empowered wireless communications. IEEE J Sel Areas Commun 23(2):201–220
7. Liu Y, Jun LIANG, Xiao N, Hu Y, Hu M (2016) Dynamic double threshold energy detection based on Markova model in cognitive radio. J Electron Inf Technol 38(10):2590–259
8. Tian Z, Giannakis GB (2006) A wavelet approach to wideband spectrum sensing for cognitive radios. In: Proceedings of 1st international conference on cognitive radio oriented wireless networks and communications (CROWNCOM), pp 1–5
9. Kumar A, Saha S (2019) Discrete wavelet transform based hybrid edge detection technique for wideband spectrum sensing in CRNs. In: IEEE region 10 symposium (TENSYMP) pp 529–534
10. Verma P (2020) Adaptive threshold based energy detection over Rayleigh fading channel. Wirel Person Commun 113(1):299–311

11. Vladeanu C, Nastase CV, Martian A (2016) Energy detection algorithm for spectrum sensing using three consecutive sensing events. IEEE Wirel Commun Lett 5(3):284–287
12. Martian A, Al Sammarraie MJ, Vlădeanu C, Popescu DC (2020) Three-event energy detection with adaptive threshold for spectrum sensing in cognitive radio systems. MDPI J Sens (Basel) 20(13):3614
13. Nasrallah A, Hamza A, Boukaba T, Baudoin G, Messani A (2018) Energy detection with adaptive threshold for cognitive radio. In: IEEE international conference on communications and electrical engineering, pp 1–5
14. Yu S, Liu J, Wang J, Ullah I (2020) Adaptive double-threshold cooperative spectrum sensing algorithm based on history energy detection. Hindwai J Wirel Commun Mob Comput
15. Yang H, Fu Y, Lei J (2018) A new cooperative spectrum sensing algorithm based on double threshold. In: IEEE 2nd advanced IMCEC conference, pp 588–592
16. Hu L, Cao N, Mao M, Liao H, Shi R, Wang S (2019) Dynamic adaptive double-threshold cooperative spectrum sensing with multi-level quantization. In: IEEE 3rd advanced IMCEC conference, pp 1381–1385
17. Sarala B, Devi SR, Sheela JJ (2020) Spectrum energy detection in cognitive radio networks based on a novel adaptive threshold energy detection method. Elsevier J Comput Commun 152:1–7
18. Mahendru G, Shukla AK, Banerjee P, Patnaik LM (2019) Adaptive double threshold based spectrum sensing to overcome sensing failure in presence of noise uncertainty. In: IEEE 6th international conference on SPIN, pp 466–471
19. Kockaya K, Develi İ (2020) Spectrum sensing in cognitive radio networks: threshold optimization and analysis: research square
20. Urkowitz H (1967) Energy detection of unknown deterministic signals. Proc IEEE 55(4):523–531
21. Verma G, Sahu OP (2016) Intelligent selection of threshold in cognitive radio system. Telecommun Syst 63(4):547–556
22. Koley S, Mirza V, Islam S, Mitra D (2015) Gradient based real-time spectrum sensing at low SNR. IEEE Commun Lett 19(3):391–394
23. Verma G, Sahu OP (2016) Opportunistic selection of threshold in cognitive radio networks. Wirel Person Commun 92(2):711–726

Collaborative Design in Concurrent Engineering of Industry 4.0 Through Virtual Reality Simulation in Achieving Accelerated Time-To-Market During COVID-19 (Coronavirus) Pandemic Outbreak

A. F. Fudzin, M. Amin, and A. Mokhtar

Abstract In the era of industry 4.0, it is important for businesses to conduct design and engineering activities via data exchange in cloud computing. Collaborative design in concurrent engineering through virtual reality simulation using digital mock up (DMU) presented in this paper is to construct a simulation in digital environment in order to reduce development time, cost and improve quality of product which will impact to accelerated time to market. However, due to current COVID-19 (Coronavirus) pandemic outbreak, the activities could be interrupted. The virtual reality design collaboration seems to be as a solution with the support of cloud storage of product design and satisfy design issue in concurrent engineering. It is also overcome the geographical limitation and social distancing constraint in "face to face" collaboration. Thus, stakeholders' activities are carried out as usual without interruption in the challenging period during the pandemic. Finally, virtual simulation of collaborative design in concurrent engineering of previous related works conducted by the author are briefly presented.

Keywords Industry 4.0 · Concurrent engineering · Collaborative engineering · Digital manufacturing · Simulation · COVID-19

1 Introduction

Competitive market has forced industry to improve new product development and improve the speed of product launch through digitalization. The basic formulation and solution of accelerated time to market in new product launch is a vision for the

A. F. Fudzin (✉) · M. Amin · A. Mokhtar
Faculty of Mechanical, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Perak, Malaysia
e-mail: ahmad_g02810@utp.edu.my

M. Amin
e-mail: mamin_amajid@utp.edu.my

realization of the product development digitalization process. In the global competition of new product introduction, the time required for the realization of products as well as production is an absolute prerequisite in product development. For this purpose, the virtual reality simulation of product development is an absolute requirement to achieve accelerated time to market and reduce cost [1]. On the other hand, many considerations in product development is the involvement of various stakeholder. Single stakeholder alone in development team will not be capable of meeting the challenge. For this purpose, extensive cooperation among numerous stakeholder will be necessary. Despite the fact that these issues have been brought to the attention of many scholars and industrialists since the early stages of digitalization, the current situation of global pandemic outbreaks has also come to global attention in recent literature.

In current situation of COVID-19 (Coronavirus) global pandemic outbreak, the face to face stakeholder collaboration is prohibited in many countries despite the challenge of accelerated new product introduction [2]. In the event of catastrophe such as COVID-19 pandemic, the industry was badly affected by forcing the workforce to follow public health guidelines such as movement control order and work from home (WFH) [3]. Therefore, it is essential for stakeholders or citizen to adhere to the social distancing initiative by having virtual meeting from their location using digital technology [4]. The launch of product need to be realized in timely manner as planned ahead before the pandemic. Thus, this is the benefit of industry 4.0 by having the cloud data storage which all the stakeholders be able to communicate and retrieve the data for analysis. Subsequently, product development can still be performing and interruption of time to market is less affected.

Nowadays, the digitalization reaching beyond boundary. Cloud-based digitalization was introduced [5]. The emerging of technologies such as Industry 4.0 is known as the fourth industrial revolution, which developed in Germany plays an important major role in the design and development [6–8]. The ninth pillars of industry 4.0 which consists of big data, autonomous robot, simulation, system integration, internet of things, cloud computation, additive manufacturing, augmented reality and cyber security can help to achieve accelerated time to market [8, 9]. Virtual simulation was used to perform an analysis such as clash interference check, tools accessibility, workability, ergonomics and serviceability in the early stage of product development. In the virtual simulation world, virtual commissioning may also be carried out [10]. The advancement in new digital technology provides the advantage of how industry reacts to the latest phenomena.

Before this pandemic shock all industrialist and engineering-related activities normally getting in touch with stakeholders is a requisite. In order to have a common emphasis on reducing product design and production time and costs and increasing product quality, a lot of discussion with all stakeholders through effective collaboration are needed [11]. As the purpose of the study in achieving uninterrupted product launch and accelerated time to market in the current pandemic situation, virtual people's collaboration is highly essential to the success of concurrent engineering in the Industry 4.0 platform. This work proposed a method to support the execution of

collaborative design in concurrent engineering in industry especially in automotive environment.

2   Collaborative Design in Concurrent Engineering

In collaborative design, a design communication and involvement of several stakeholders such as marketing team, designer, process engineer, tools engineer and sourcing team need to established in the early stage of development process. Design coordination can be enhanced across multiple knowledge channels in multiple ways compared to conventional settings as a one-way mechanism [5]. Nowadays, it is also important for designers to co-develop components with various stakeholder in different geographical locations [12]. Globalization of industry into several parts of the world makes it difficult to gather all the stakeholder in one location, because of geographical constraints. This initiative is a blessing in disguise by having virtual collaborative design because of the COVID 19 (Coronavirus) pandemic problem. Therefore, this phenomenon is a push factor for web-based virtual communication method is taking place in rapid pace [13].

Concurrent engineering is defined an attempt to implement a parallelism in product development. The realization of concurrent engineering through utilizing DMU analysis in digital method helps to minimize the time taken from start of project until start of series production. Figure 1 illustrate the DMU's target process
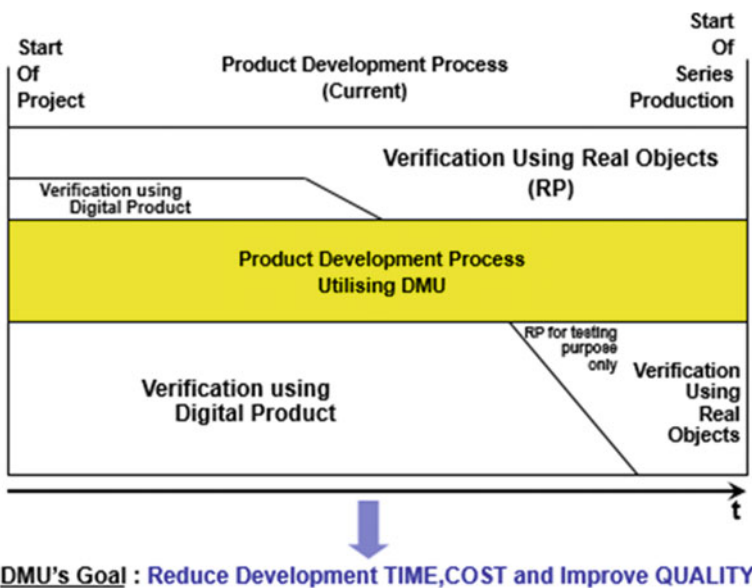
Fig. 1   Product development process utilizing DMU

concept of verification using digital product and verification using real objects [14]. The DMU process lead to significant reduction in the number of real object verification and just limited for testing purpose only [15]. As the DMU's goal are to reduce development time, cost and improve quality, this will impact to accelerated time to market of product. Furthermore, the number of design change due to assembly problems and the amount of prototype validation needed for production during the product validation stage will be reduced. The advantage of virtual simulation in the design process is a key enabler for manufacturers to have a faster and more flexible approach to solving complex issues in the collaborative design environment.

3 Methodology—Virtual Reality Simulation

In the early stage of design, face to face communication during collaboration design of stakeholders will support faster feedback between designer and process engineer. Figure 2 demonstrates the process of collaboration design in concurrent engineering. The designer will issue design to process engineer for review process, so that, product change request can immediately be considered for the change. At this point of time, the industry 4.0 through cloud computing which shared platforms that serve to multiple users (Shared Data Format) is very essential to support engineer to review the same data and same format [9, 16]. Faster feedback can be achieved from several parties in respond to the design issue aroused. It can be visualized as
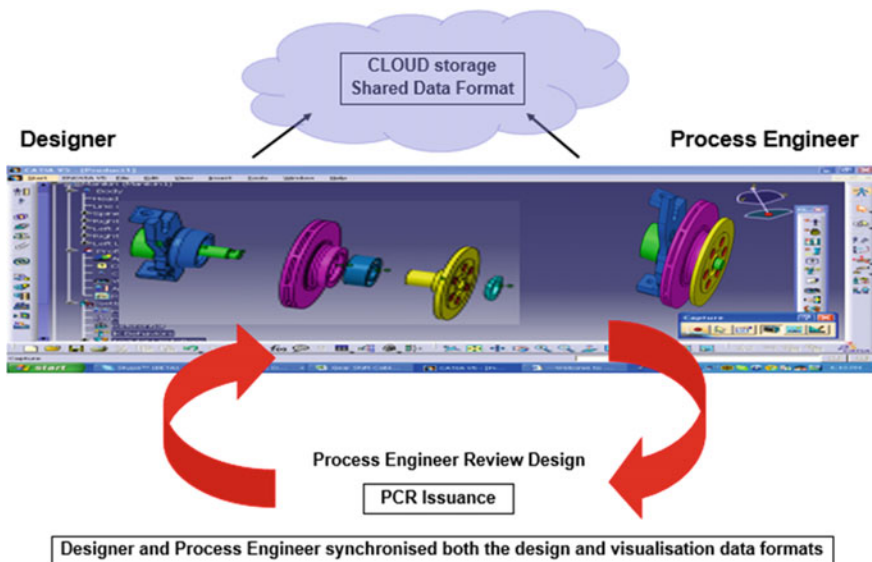
Fig. 2 Collaboration design over the cloud storage

designer and process engineer synchronized both the design and visualization data formats on the same computer screen, eventhough they are from different part of the world. This will help to reduce design-cycle time and cuts cost.

On design stage of product development, virtual computing analysis of manufacturing processes can be applied using 3D CAD/CAM modeling using CATIA V5 (Dassault Systemes) [17]. It is a key enabler in the technology of Industry 4.0 and integrated process from Design for Manufacturing (DFM) [18] to visual and augmented reality simulation can be achieved using seamless data transfer using cloud computing [19]. The presented case study in automotive simulation analysis is used in connection to 3D modeling. Three (3) stages of analysis investigated using virtual simulation in development phase which are as followings;

3.1 Stage 1, Tools Accessibility Analysis

Tools accessibility analysis is an approach in assessing the tools require in production assembly line [20]. This to ensure the correct tools to be purchased by the procurement team. Thus, eliminate the use of incorrect tools for production.

3.2 Stage 2, Investigate Serviceability Using Human Manikin Virtual Simulation

This investigation is to provide a visual simulation on how humans and objects function together in term of serviceability. All potential problems of serviceability are analyzed; human movements will work effectively with the proposed tools at an early stage [21].

3.3 Stage 3, Tools Clash and Interference Check Using Human Manikin Virtual Simulation

At this stage clash and interference check is using human manikin virtual simulation. This investigation is to provide a visual simulation on how the inference of tools clash with components during assembly and disassembly. It is very important for automotive maintenance and service operation [15]. All possible clash and interference are analyzed.

4 Case Study—Virtual Reality Simulation

The described approach has been implemented in a practical application, which supports the tools accessibility analysis, investigate serviceability using human manikin virtual reality and tools clash and interference check using human manikin virtual simulation.

4.1 Tools Accessibility Analysis

The case study of tools accessibility analysis is to verify assembly tools' accessibility applied during the assembly process during assembly process. Particularly, tool placement constraints during tool applications are verified into the assembly product. Tools accessibility analysis is an efficient approach in assessing the tools require in production assembly line. Prior to the purchase of tools by the procurement team, this activity may eliminate the use of incorrect tools for production. The analysis utilized CAD data which retrieved from cloud data storage. Further, DMU CAD models of the assembly product are imported into a virtual reality visualization tools accessibility analysis.

The designer provided the components design to the cloud data storage and process engineer as well as tool engineer have to analyze the tool accessibility to ensure that the assembly process does not pose accessibility related problems. Prompt response to designer in the case of accessibility issue can be rectified by the designer immediately. In this case study, a virtual reality simulation is being utilized for the analysis as shown in Fig. 3. This approach would detect any problem earlier in design phase.
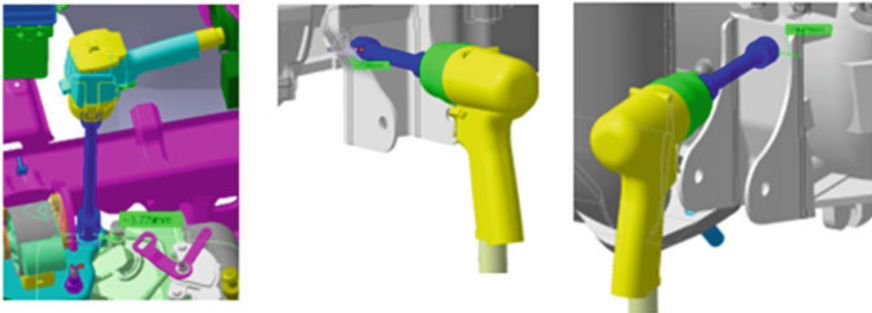
Fig. 3 Virtual simulation of tools accessibility
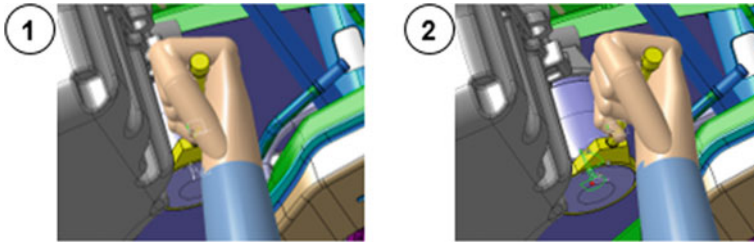
Fig. 4 Virtual simulation of human hand movements position 1 to positon 2 (Diassembly process)

4.2 Investigate Serviceability Using Human Manikin Virtual Simulation

This investigation is adopted DMU is to provide a virtual simulation on how humans and objects function together in term of serviceability. All potential difficulties during serviceability are analyzed. Such a criteria during the procedure of disassembly operation sequence is essential. Figure 4 shows virtually how human hand with tool movements from position 1 to position 2 in disassembly process of oil filter from the engine. The movement for disassembly of the oil filter can be preliminary identified so it can be easily disassembly using dedicated tool. Others issues such as difficulties, interference and clearance analysis with tools, human and product is analyzed. A suitable tools need to be purchased, and in the case unavailability of tools, special tool need to designed to ease the serviceability process.

4.3 Tools Clash and Interference Check Using Human Manikin Virtual Simulation

At this stage clash and interference check is using human manikin virtual simulation. The verification of the reach-ability of tools to perform a task is very crucial. This investigation is to provide a visual simulation on how the reach-ability and inference of tools clash with components. In the design phase all possible class and interference are analyzed virtually. In Fig. 5, tool clash with the surrounding can be identified. Feedback from process engineer to designer can immediately improve in the first review session. The improvement of product design is validated in the second review before proceed to the next phase. In order to support accelerated time to market, this rapid and accurate assessment is needed.

Another example of reach-ability is the interference analysis for disassembly of oil filter from the engine. As explained in previous section the tool used to disassembly by turning the tool from position 1 to position 2. In continuation of that, human hand is used to completely disassemble the oil filter. Figure 6 shows the virtual simulation interference study during the disassembly of oil filter using hand. During
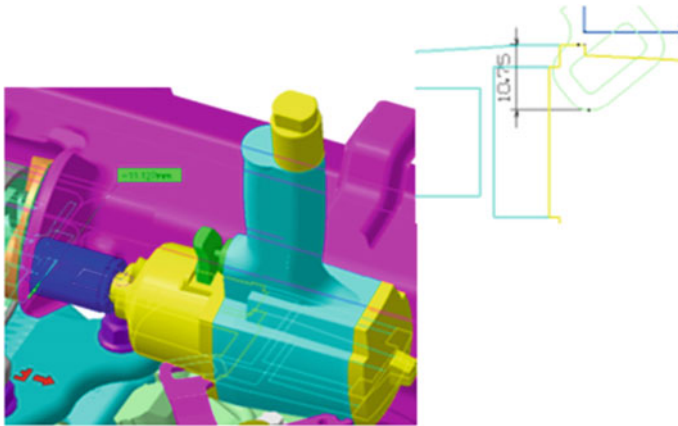
Fig. 5 Virtual clash tool with product
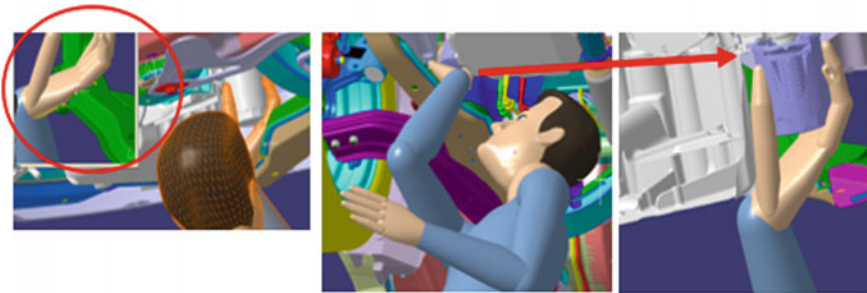
Fig. 6 Virtual simulation interference study

the disassembly proceed we found the interference occurred between hand's manikin and sub frame. The red circle in the diagram clearly visualize the interference. With help of simulation software this occurrence can immediately feedback to designer in the form of product change request. It is proven that early detection during DMU virtual simulation check will eliminate any cost during prototype stage. Figure 7 shows the proven real serviceability check where the interference was eliminated in design phase.

5 The Role of Industry 4.0 in COVID-19 Pandemic

In automotive manufacturing environment, activities such as simultaneous engineering, concurrent engineering and process engineering design, the role of Industry 4.0 is an important factor in fostering accelerated time to market. The deployment of Industry 4.0 platform in the COVID-19 pandemic outbreak, helps the engineer

Fig. 7 Real serviceability check

to continue moving forward using virtual reality simulation supported the design review activities through the data exchange in cloud computing. It is a good tool in order to increase social distancing and to adhere to the restriction on face to face collaboration beside making simulation in economical ways. Based on the case study presented in the previous section, stakeholders have been shown to be able to work in same space of virtual environment without face to face collaboration as new norm in working culture.

Collaborative design in concurrent engineering using DMU approach is part of successful factor in virtual reality simulation where product introduction is uninterrupted due to pandemic. In the area of automotive design, the such operation presented in the case study, the operations were checked on a physical vehicle must now be checked by means virtual verification tools. The development of the virtual verification is opens new approach of the analyses performed on the DMU. The application not only allows designers to simulate the tools accessibility, serviceability check, tools clash and interference check, but also offers a set of tools useful for analyzing the DMU to detect potential modelling errors which could jeopardize product functionality.

As virtual reality simulation has been presented during product development cycle of the automotive industry, which it is known as an important tool in virtual verification. The advantage of virtual verification in development design phase, cost reduction is achievable due to reduce and eliminate most of physical prototype. The availability of Industry 4.0 as a key for future business is leading to these two goals, as previously addressed in achieving uninterrupted product launch and accelerated time to market in the current pandemic situation.

6 Conclusions

Today, automotive industries are required to transfer most of the activities in virtual environment to meet the product launch time line in competitive market. In the availability of industry 4.0 platform, it is an important for businesses to conduct design and engineering activities of the data exchange in cloud computing. The deployment

of Industry 4.0 pillars in the COVID-19 pandemic outbreak, helps the engineer to continue moving forward using virtual reality simulation supported the design review activities through the data exchange in cloud computing. The Internet is becoming an important element for concurrent engineering applications. The presented approach supported a collaborative design in concurrent engineering through the virtual reality simulation using DMU to construct a simulation in the digital environment in order to reduce development time, cost and improve quality, which will impact to accelerated time to market of product. The stakeholders' activities are carried out as usual without interruption in the challenging period during the COVID-19 pandemic outbreak. In this way, the virtual simulation design including cloud storage supports product design and satisfy design issue in concurrent engineering. It is also overcome the geographical limitation and social distancing constraint in "face to face" collaboration in order to avoid possible infections among stakeholders. Therefore, web-based virtual communication method is taking place in rapid pace. Finally, it should be mentioned that the virtual simulation of collaborative design in concurrent engineering of the previous related works has been implemented into the development process of automotive manufacturer and finds its application in design phase. In conclusion, with the presence of the latest Industry 4.0 application, there is no significant impact on the disruption due to the COVID-19 pandemic in the virtual reality simulation phase.

References

1. Fragner A, Kreis A, Hirz M (2020) Virtual tools to support design and production engineering: early detection of stone chips to optimize production processes. In 2020 IEEE 7th international conference on industrial engineering and applications (ICIEA), 2020, pp 399–403
2. Avdiu B, Nayyar G (2020) When face-to-face interactions become an occupational hazard: jobs in the time of COVID-19, ed: The World Bank
3. del Rio-Chanona RM, Mealy P, Pichler A, Lafond F, Farmer D (2020) Supply and demand shocks in the COVID-19 pandemic: An industry and occupation perspective. arXiv preprint arXiv:2004.06759, 2020
4. Katapally TR (202) A global digital citizen science policy to tackle pandemics like COVID-19. J Med Int Res 22:e19357
5. Wu D, Rosen DW, Wang L, Schaefer D (2015) Cloud-based design and manufacturing: a new paradigm in digital manufacturing and design innovation. Comput Aided Des 59:1–14
6. Zhou J, Li P, Zhou Y, Wang B, Zang J, Meng L (2018) Toward new-generation intelligent manufacturing. Engineering 4:11–20
7. Xu LD, Xu EL, Li L (2018) Industry 4.0: state of the art and future trends. Int J Prod Res 56:2941–2962
8. Hermann M, Pentek T, Otto B (2016) Design principles for industrie 4.0 scenarios. In: 2016 49th Hawaii international conference on system sciences (HICSS), 2016, pp 3928–3937
9. Erboz G (2017) How to define industry 4. 0: the main pillars of industry 4. 0. In: 7th international conference on management (ICoM 2017), At Nitra, Slovakia, 2017, pp 1–2

10. Ružarovský R, Holubek R, Sobrino DRD, Janíček M (2018) The simulation of conveyor control system using the virtual commissioning and virtual reality. Adv Sci Technol Res J 12
11. Vila C, Ugarte D, Ríos J, Abellán J (2017) Project-based collaborative engineering learning to develop Industry 4.0 skills within a PLM framework. Procedia Manuf 13:1269–1276
12. Fuh JY, Li W (2005) Advances in collaborative CAD: the-state-of-the art. Comput Aided Des 37:571–581
13. Sułkowski Ł (2020) Covid-19 pandemic; recession, virtual revolution leading to deglobalization? J Intercult Manage 12:1–11
14. Döllner G, Kellner P, Tegel O (2000) Digital mock-up and rapid prototyping in automotive product development. J Integr Des Process Sci 4:55–66
15. Gynn M, Steele J (2015) Virtual automotive maintenance and service confirmation. In: SAE technical paper 0148–7191
16. Zheng P, Sang Z, Zhong RY, Liu Y, Liu C, Mubarok K et al (2018) Smart manufacturing systems for Industry 4.0: conceptual framework, scenarios, and future perspectives. Front Mech Eng 13:137–150
17. Ghionea IG, Devedžić G, Ćuković S (2015) Parametric modeling of surfaces using CATIA v5 environment. In: Applied mechanics and materials, 2015, pp 93–98
18. Myung S, Song K, Lee J (2002) Integration of DFM and virtual NC manufacturing process. In: CIRP ISMS, pp 175–180
19. Posada J, Toro C, Barandiaran I, Oyarzun D, Stricker D, de Amicis R et al (2015) Visual computing as a key enabling technology for industrie 4.0 and industrial internet. IEEE Comput Graphics Appl 35:26–40
20. Kang X, Peng Q (2007) Analysis of tool accessibility in fixture setup planning. In: International design engineering technical conferences and computers and information in engineering conference, 2007, pp 1007–1016
21. Bruno M, Milite A, Monacelli G, Pina M, Sessa F (2006) Serviceability analyses in virtual environment for the automotive industry. In: DS 36: proceedings design 2006, the 9th international design conference, Dubrovnik, Croatia, 2006

Stability Analysis of Semi-Markovian Discrete-Time Neural Networks with Time-Varying Leakage Delays

K. Maheswari, S. N. Shivapriya, and C. Ramkumar

Abstract This paper studies the stabilization of discrete-time semi-Markovian neural networks with coupling time varying delays with unreliable communication links with sensor non-linearity in communication channels. There occurs the time varying leakage delays with random parameters which obeys the uncorrelated Bernoulli-distributed sequences. By the proper construction of the Lyapunov Krasovskii functional (LKF), sufficient conditions for stability analysis has been attained that are derived by the means of linear matrix inequalities (LMI's). Numerical example with simulation is illustrated for the effectiveness of the proposed method.

Keywords Discrete-time NNs · Leakage delays · Stability analysis · Semi-Markovian neural networks

1 Introduction

Devoting much attention to the study of neural networks (NNs) in last few years has tremendously impacted the application prospect in the real world scenario. NN's are extensively applied in varied areas of research in large data and artificial intelligence which includes applications in pattern recognition, robotics, optimization etc. Significant number of researches has also been carried out in dynamic analysis of various NNs. Stability analysis via state estimation procedure is one of the hottest research topic in the current trends of digital technology using discrete-time NNs.

In the areas of control and signal processing application problem procedures, discrete-time NNs' play a vital role along with time delays. There are a lot of

K. Maheswari (✉)
Department of Mathematics, Kumaraguru College of Technology, Coimbatore, India

S. N. Shivapriya
Department of Electronics and Communications Engineering, Kumaraguru College of Technology, Coimbatore, India

C. Ramkumar
Department of Biomedical Engineering, Dr. N. G. P. Institue of Technology, Coimbatore, India

approaches for the state estimation procedure in discrete time NNs that holds its consideration with time delays several times and leakage delays at fewer number of times. These time delays in the state estimation problems has drawn increasing research attention in today's digital world. This state estimation problem along with Markovian jump parameters in discrete-time NN's has been discussed in various dynamical problems and the references therein [1–3]. Due to the strong modeling ability of Markovian jump neural networks (MJNNs), greater attention has been drawn in the field of hybrid dynamical systems like in the case of biology, medicine, industry production, in aerospace Engineering.

But as is well known, in MJNNs, the transition matrix is converted into an invariant matrix caused due to exponential distribution and obeys the stochastic process and is not relevant with the past modes. This is one of the disadvantages of MJNNs. Therefore, in order to overcome this shortcoming, semi-Markovian Jump Neural Networks (SMJNNs) has been used where the probability density function matrix takes the key role. Some of the recent researches have emphasized the synchronization controller using SMJNNs given in [4] and the robust stochastic stability conditions have also been incorporated in [5]. Inspired by all these works, in this paper, SMJNNs has been applied to the model that has time varying coupling delays with randomly occurring uncertainties [6]. By employing Lyapunov Karsovskii functional (LKF) with different decision variables, sufficient conditions to attain the stability is achieved.

On the basis of the these discussions, the aim of this paper is to investigate the asymptoticity for SMJNNs with delays that are mixed and time varying in nature. By constructing suitable LKF's and proceeding with the LMI technique, we derive the feasibility of the stability conditions by the addressed NNs.

Notations:

Throughout the paper T and -1 represent the transpose and the inverse of the matrix respectively. The n-dimensional Euclidean space is given by R^n for the set of m x n matrices in R^{mxn} and $\lambda_{\max(\min)}(A)$ is the corresponding eigenvalue of the spectral norm matrix H.

2   Formulation of the Problem

For $\breve{\Omega}$ being the space in the probability P measured on F in the space is given by $(\breve{\Omega}, F, P)$. Let (r(t), t > 0) be the semi-Markov process in the discrete-time state given by $\mathbf{S} = \{1, 2,\ldots, N\}$ with the generator $\Gamma \cong \gamma_{ij}, i, j \in S$

$$\Pr ob\{r(t + \hat{h}) = j/r(t) = i\} = \begin{cases} \gamma_{ij}(\hat{h})\hat{h} + o(\hat{h}), & i \neq j \\ 1 + \gamma_{ij}(\hat{h})\hat{h} + o(\hat{h}), & i = j \end{cases}$$

With $\lim_{\hat{h} \to 0} \frac{o(\hat{h})}{\hat{h}} = 0$ with $\gamma_{ij}(\hat{h}) \geq 0$ for $i \neq j$ when it moves from i at time t to mode j at time c with

$$\gamma_{ii}(\hat{h}) = - \sum_{j \in S, j \neq i} \gamma_{ij}(\hat{h}) \tag{1}$$

Discrete-time delayed NNs with time-varying leakage delay as follows:

$$\varsigma(t+1) = B(\varsigma(t - \psi(t))) + x(t) + C\breve{f}(\varsigma(t)) + D\breve{g}(\varsigma(t - d(t))) + J' \tag{2}$$

where $\varsigma(t) = [\varsigma_1(t), \varsigma_2(t), \ldots, \varsigma_n(t)]^T \in R^n$ is the state vector and the activation functions are given as $\breve{f}(\circ) = [\breve{f}(\varsigma_1(t)), \breve{f}(\varsigma_2(t)), \ldots, \breve{f}(\varsigma_n(t))]^T \in R^n$, $\breve{g}(\varsigma(t)) = [\breve{g}(\varsigma_1(t)), \breve{g}(\varsigma_2(t)), \ldots, \breve{g}(\varsigma_n(t))]^T \in R^n$ and $J' = [J'_1, J'_2, \ldots, J'_n]^T$ is the input external vector. Each of $\psi(t)$ in the system represents the delay due to leakage which satisfies $0 < \psi_m \leq \psi(t) \leq \psi_M$ where ψ_m is the lower bound and ψ_M is the upper bound of $\psi(t)$ and $d(t)$ represents the delay in the transmission satisfying $0 < d_m \leq d(t) \leq d_M$ with the positive integers d_m representing the lower bound and d_M representing the upper bound of $d(t)$.

Assumption 1

For any $c_1, c_2 \in R, c_1 \neq c_2$, the activation functions $\breve{f}_i(\circ)$ and $\breve{g}_i(\circ)$ satisfy,

$$M_i^- \leq \frac{\breve{f}_i(c_1) - \breve{f}_i(c_2)}{c_1 - c_2} \leq M_i^+,$$

$$N_i^- \leq \frac{\breve{g}_i(c_1) - \breve{g}_i(c_2)}{c_1 - c_2} \leq N_i^+, \quad i = 1, 2, \ldots, n$$

where $M_i^-, M_i^+, N_i^-, N_i^+$ are the known constants.

Assuming $\varsigma^* = [\varsigma_1^*, \varsigma_2^*, \ldots, \varsigma_n^*]^T$ equation in (2) and letting

$$y_i(t) = \varsigma_i(t) - \varsigma_i^*, \ f_i y_i(t) = \breve{f}_i(\varsigma_i(t))$$

$$- \breve{f}_i(\varsigma_i^*,), \ g_i(\varsigma_i(t - d(t)))$$

$$= \breve{g}_i(\varsigma_i(t - d(t))) - \breve{g}_i(\varsigma_i^*)$$

Then, the NN system Eq. (1) can be written as,

$$y(t+1) = B(y(t - \psi(t))) + Cf(\varsigma(t)) + Dg(\varsigma(t - d(t))) \tag{3}$$

where $y(t) = [y_1(t), y_2(t), \ldots, y_n(t)]^T$,

$$y(t - \psi(t)) = [y_1(t - \psi(t)), y_2(t - \psi(t)), \ldots, y_n(t - \psi(t))]^T,$$

$$f(y(t)) = [f(y_1(t)), f(y_2(t)), \ldots, f(y_n(t))]^T,$$
$$g(y(t - \psi(t))) = [g(y_1(t - \psi(t))), g$$
$$(y_2(t - \psi(t))), \ldots, g(y_n(t - \psi(t)))]^T$$

By the above assumption given, it can be verified that for every $f_i(\circ), g_i(\circ), i = 1, 2, \ldots, n$ satisfy $\begin{matrix} M_i^- \le (f_i(c_1) - f_i(c_2))/(c_1 - c_2) \le M_i^+, \\ N_i^- \le (g_i(c_1) - g_i(c_2))/(c_1 - c_2) \le N_i^+ \end{matrix}$ for any $c_1 \neq c_2$ and $f_i(0) = g_i(0) = 0$.

The model is associated with the initial condition as

$$y(c) \le \theta(c), c = -\delta, -\delta + 1, \ldots, 0 \tag{4}$$

where $\delta = \max\{\psi_M, d_M\}$.

Lemma 1 In an open set A of R^m, let $f_1, f_2, \ldots, f_N : R^m \to R$ then f_i over D satisfies.

$$\min_{\{\beta_i | \beta_i > 0, \sum \beta_i = 1\}} \sum_i \frac{1}{\beta_i} f_i(l) = \sum_i f_i(l) + \max_{g_{i,j}(l)} \sum_{i \neq j} g_{i,j}(l) \tag{5}$$

given by

$$g_{j,i}(l) \equiv g_{i,j}(l), \begin{bmatrix} f_i(l) & g_{i,j}(l) \\ g_{i,j}(l) & f_{ij}l) \end{bmatrix} \ge 0. \tag{6}$$

Lemma 2 For any matrices $\wp_1, \wp_2$ and $\wp_3$ where $\wp_1^T = \wp_1$ and $\wp_2^T = \wp_2 > 0$, then $\wp_1 + \wp_3^T \wp_2^{-1} \wp_3 < 0$ with $\begin{bmatrix} \wp_1 & \wp_3^T \\ * & -\wp_2 \end{bmatrix} < 0$.

3   Main Results

In this section, we apply the semi-Markovian and investigate the delay dependent sufficient conditions of LKF.

Theorem 3.1 The NNs system (3) is globally asymptotically stable, under the assumption (1), if there exists the symmetric matrices $P_i > 0, i = 1, \ldots, 6, Q_r > 0, r = 1, \ldots, 4, R_1 > 0, R_2 > 0, T_j > 0$ with the diagonal matrices $\Xi_1 > 0, \Xi_2 > 0$ and matrices $U_j, j = 1$ to 4 with weights W_0, W_1 of appropriate dimensions, such that following LMI holds with.

$$\begin{bmatrix} T_1 & T_2 \\ T_3 & T_4 \end{bmatrix} > 0 \tag{7}$$

where $\tilde{P}_i = \sum_{j \in S} \gamma_{ij}(h) P_j$,

$$\begin{aligned}
\Omega_1 =\ & \tilde{e}_8 \tilde{P}_1 \tilde{e}_8 + 2\tilde{e}_8 \tilde{P}_1 \tilde{e}_1 + 2\tilde{e}_8 \tilde{P}_3 (\tilde{e}_5 - \tilde{e}_7) \\
& + 2\tilde{e}_8 \tilde{P}_2 (\tilde{e}_2 - \tilde{e}_4) + 2\tilde{e}_8 \tilde{P}_3 (\tilde{e}_{11} - \tilde{e}_{12}) \\
& + 2\tilde{e}_1 \tilde{P}_2 (\tilde{e}_2 - \tilde{e}_4) + 2\tilde{e}_1 \tilde{P}_3 (\tilde{e}_5 - \tilde{e}_7) \\
& + (\tilde{e}_2 - \tilde{e}_4)^T \tilde{P}_4 (\tilde{e}_2 - \tilde{e}_4) + 2(\tilde{e}_2 - \tilde{e}_4)^T 2\tilde{e}_1 \tilde{P}_2 (\tilde{e}_2 - \tilde{e}_4)
\end{aligned}$$

$$\begin{aligned}
& 2(\tilde{e}_2 - \tilde{e}_4)^T \tilde{P}_5 (\tilde{e}_5 - \tilde{e}_7) + 2(\tilde{e}_2 - \tilde{e}_4)^T \tilde{P}_5 (\tilde{e}_{11} - \tilde{e}_{12}) \\
& + 2(\tilde{e}_5 - \tilde{e}_7)^T \tilde{P}_6 (\tilde{e}_5 - \tilde{e}_7) 2(\tilde{e}_5 - \tilde{e}_7)^T \tilde{P}_6 (\tilde{e}_{11} - \tilde{e}_{12}).
\end{aligned}$$

$$\begin{aligned}
\Omega_2 =\ & (\psi_1 + 1)\tilde{e}_1^T Q_1 \tilde{e}_1 - \tilde{e}_3^T Q_1 \tilde{e}_3, \Omega_3 \\
=\ & (d_1 + 1)\tilde{e}_1^T Q_2 \tilde{e}_1 - \tilde{e}_6^T Q_2 \tilde{e}_6, \\
\Omega_4 =\ & d_1^2 \tilde{e}_1^T R_1 \tilde{e}_1 + d_1 \tilde{e}_5^T U_4 \tilde{e}_5 + d_1 \tilde{e}_6^T (U_3 - U_4)\tilde{e}_6 \\
& - d_1 \tilde{e}_7^T U_3 \tilde{e}_7 + d_1^2 \tilde{e}_8^T R_2 \tilde{e}_8, \\
\Omega_5 =\ & -(\tilde{e}_{11}^T R_1 \tilde{e}_{11} + 2\tilde{e}_{11}^T U_3 (\tilde{e}_6 - \tilde{e}_7)) \\
& + (\tilde{e}_6 - \tilde{e}_7)^T (R_2 + U_3)(\tilde{e}_6 - \tilde{e}_7) + \tilde{e}_{12}^T R_1 \tilde{e}_{12} \\
& + 2\tilde{e}_{12}^T U_4 (\tilde{e}_5 - \tilde{e}_6) + (\tilde{e}_5 - \tilde{e}_6)(R_2 + U_4)(\tilde{e}_5 - \tilde{e}_6)), \\
\Omega_6 =\ & -\tilde{e}_1^T (M_1 \Xi_1)\tilde{e}_6 + 2\tilde{e}_1^T (M_2 \Xi_1) e_9 \\
& - \tilde{e}_9^T \Xi_1 \tilde{e}_9, \Omega_7 = -\tilde{e}_6^T (N_1 \Xi_2)\tilde{e}_6 \\
& + 2\tilde{e}_6^T (N_2 \Xi_2) e_{10} - \tilde{e}_{10}^T \Xi_2 \tilde{e}_{10}, \Omega_8 \\
=\ & 2(\tilde{e}_8 W_0^T + \tilde{e}_1 W_1^T)(B\tilde{e}_3 + C\tilde{e}_9 + D\tilde{e}_{10} - \tilde{e}_8^T - \tilde{e}_1).
\end{aligned}$$

Proof The LKF is.

$$V_i(t) = \begin{bmatrix} y(t) \\ \sum_{m=t-\psi_M}^{t-\psi_m-1} y(m) \\ \sum_{m=t-d_M}^{t-d_m-1} y(m) \end{bmatrix} \begin{bmatrix} P_1 & P_2 & P_3 \\ * & P_4 & P_5 \\ * & * & P_6 \end{bmatrix} \begin{bmatrix} y(t) \\ \sum_{m=t-\psi_M}^{t-\psi_m-1} y(m) \\ \sum_{m=t-d_M}^{t-d_m-1} y(m) \end{bmatrix} \tag{8}$$

$$V_2(t) = \sum_{s=t-\psi(t)}^{t-1} y^T(s) Q_1 y(s) + \sum_{j=-\psi_M+1}^{-\psi_m} \sum_{s=t+j}^{t-1} y^T(s) Q_2 y(s) \tag{9}$$

$$V_3(t) = \sum_{s=t-d(t)}^{t-1} y^T(s)Q_3 y(s) + \sum_{j=-d_M+1}^{-d_m} \sum_{s=t+j}^{t-1} y^T(s)Q_4 y(s) \tag{10}$$

$$V_4(t) = d_1 \sum_{j=-d_M+1}^{-d_m} \sum_{s=t+j}^{t-1} \begin{bmatrix} y(s) \\ \eta(s) \end{bmatrix}^T \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \begin{bmatrix} y(s) \\ \eta(s) \end{bmatrix} \tag{11}$$

where $\eta(s) = y(s+1) - y(s)$, $d_1 = d_M - d_m$.

Calculating the difference of $V_a(t)$, $a = 1, 2, 3, 4$,

$$E\{\Delta V_1(t)\} = E\{V(t+1) - V(t)\}$$

$$\Delta V_1(t) = \begin{bmatrix} \eta(t) + y(t) \\ \sum_{m=t-\psi_M}^{t-\psi_m-1} (\eta(m) + y(m)) \\ \sum_{m=t-d_M}^{t-d_m-1} (\eta(m) + y(m)) \end{bmatrix} \begin{bmatrix} \tilde{P}_1 & \tilde{P}_2 & \tilde{P}_3 \\ * & \tilde{P}_4 & \tilde{P}_5 \\ * & * & \tilde{P}_6 \end{bmatrix} \begin{bmatrix} \eta(t) + y(t) \\ \sum_{m=t-\psi_M}^{t-\psi_m-1} (\eta(m) + y(m)) \\ \sum_{m=t-d_M}^{t-d_m-1} (\eta(m) + y(m)) \end{bmatrix}$$

$$- \begin{bmatrix} y(t) \\ \sum_{m=t-\psi_M}^{t-\psi_m-1} y(m) \\ \sum_{m=t-d_M}^{t-d_m-1} y(m)) \end{bmatrix} \begin{bmatrix} \tilde{P}_1 & \tilde{P}_2 & \tilde{P}_3 \\ * & \tilde{P}_4 & \tilde{P}_5 \\ * & * & \tilde{P}_6 \end{bmatrix} \begin{bmatrix} y(t) \\ \sum_{m=t-\psi_M}^{t-\psi_m-1} y(m) \\ \sum_{m=t-d_M}^{t-d_m-1} y(m)) \end{bmatrix},$$

$$\Delta V_1(t) = \begin{bmatrix} \eta(\hat{t}) \\ y(\hat{t}) \\ \sum_{m=t-\psi_M}^{t-\psi_m-1} \eta(m) \\ \sum_{m=t-\psi_M}^{t-\psi_m-1} y(m) \\ \sum_{m=t-d_M}^{t-d_m-1} \eta(m) \\ \sum_{m=t-d_M}^{t-d_m-1} y(m) \end{bmatrix} \begin{bmatrix} \tilde{P}_1 & \tilde{P}_1 & \tilde{P}_2 & \tilde{P}_2 & \tilde{P}_3 & \tilde{P}_3 \\ \tilde{P}_1 & 0 & \tilde{P}_2 & 0 & \tilde{P}_3 & 0 \\ \tilde{P}_2^T & \tilde{P}_2^T & \tilde{P}_4 & \tilde{P}_4 & \tilde{P}_5 & \tilde{P}_5 \\ \tilde{P}_3^T & \tilde{P}_3^T & \tilde{P}_5^T & \tilde{P}_5^T & \tilde{P}_6 & \tilde{P}_6 \\ \tilde{P}_3^T & 0 & \tilde{P}_5^T & 0 & \tilde{P}_6 & 0 \end{bmatrix} \begin{bmatrix} \eta(\hat{t}) \\ y(\hat{t}) \\ \sum_{m=t-\psi_M}^{t-\psi_m-1} \eta(m) \\ \sum_{m=t-\psi_M}^{t-\psi_m-1} y(m) \\ \sum_{m=t-d_M}^{t-d_m-1} \eta(m) \\ \sum_{m=t-d_M}^{t-d_m-1} y(m) \end{bmatrix}$$

$$\Delta V_1(t) = \begin{bmatrix} \eta(\hat{t}) \\ y(\hat{t}) \\ \sum\limits_{m=t-\psi_M}^{t-\psi_m-1} \eta(m) \\ \sum\limits_{m=t-\psi_M}^{t-\psi_m-1} y(m) \\ \sum\limits_{m=t-d_M}^{t-d_m-1} \eta(m) \\ \sum\limits_{m=t-d_M}^{t-d_m-1} y(m) \end{bmatrix} \begin{bmatrix} \tilde{P}_1 & \tilde{P}_1 & \tilde{P}_2 & \tilde{P}_2 & \tilde{P}_3 & \tilde{P}_3 \\ \tilde{P}_1 & 0 & \tilde{P}_2 & 0 & \tilde{P}_3 & 0 \\ \tilde{P}_2^T & \tilde{P}_2^T & \tilde{P}_4 & \tilde{P}_4 & \tilde{P}_5 & \tilde{P}_5 \\ \tilde{P}_3^T & \tilde{P}_3^T & \tilde{P}_5^T & \tilde{P}_5^T & \tilde{P}_6 & \tilde{P}_6 \\ \tilde{P}_3^T & 0 & \tilde{P}_5^T & 0 & \tilde{P}_6 & 0 \end{bmatrix}$$

$$\begin{bmatrix} \eta(\hat{t}) \\ y(\hat{t}) \\ \sum\limits_{m=t-\psi_M}^{t-\psi_m-1} \eta(m) \\ \sum\limits_{m=t-\psi_M}^{t-\psi_m-1} y(m) \\ \sum\limits_{m=t-d_M}^{t-d_m-1} \eta(m) \\ \sum\limits_{m=t-d_M}^{t-d_m-1} y(m) \end{bmatrix} = \varphi^T(t)\Omega_1\varphi(t) \tag{12}$$

$$\Delta V_2(t) = (\psi_1 + 1)\hat{y}^T(t)Q_1\hat{y}(t) - \hat{y}^T(t - \psi(t))Q_2\hat{y}^T(t - \psi(t))$$
$$= \Phi^T(t)\Omega_2\Phi(t) \tag{13}$$

$$\Delta V_3(t) = (d_1 + 1)\hat{y}^T(t)Q_3\hat{y}(t) - \hat{y}^T(t - d(t))Q_4\hat{y}^T(t - \psi(t))$$
$$= \Phi^T(t)\Omega_2\Phi(t) \tag{14}$$

$$\Delta V_4(t) = d_1^2 \begin{bmatrix} \hat{y}(t) \\ \hat{\eta}(t) \end{bmatrix}^T \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \begin{bmatrix} \hat{y}(t) \\ \hat{\eta}(t) \end{bmatrix}$$
$$- d_1 \sum_{s=t-d_M}^{t-d_m-1} \begin{bmatrix} y(t) \\ \eta(t) \end{bmatrix}^T \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \begin{bmatrix} y(t) \\ \eta(t) \end{bmatrix} \tag{15}$$

where $\psi_1 = \psi_M - \psi_m$ and $d_1 = d_M - d_m$.
Also, by zero initial conditions, it is seen that,

$$d_1\big[y^T(t-d(t))U_1y^T(t-d(t)) - y^T(t-d_M)U_1y^T(t-d_M)\big]$$

$$-\sum_{s=t-d_M}^{t-d(t-1)} \eta^T(s)U_1(\eta(s)+2y(t)) = 0,$$

$$d_1\big[\hat{y}^T(t-d_m)U_2\hat{y}^T(t-d_m) - \hat{y}^T(t-d(t))U_2\hat{y}^T(t-d(t))\big]$$

$$-\sum_{s=t-d(t)}^{t-d_m-1} \eta^T(s)U_2(\eta(s)+2y(t)) = 0 \tag{16}$$

Also,

$$\Phi^T(t)$$

$$= \begin{bmatrix} y^T(t),\, y^T(t-\psi_m),\, y^T(t-\psi(t)),\, y^T(t-\psi_M),\, y^T(t-d_m),\, y^T(t-d(t)),\, y^T(t-d_M) \\[4pt] \eta(t) \quad f^T(y(t)) \quad g^T(x(t-d(t))) \quad \displaystyle\sum_{s=t-d_M}^{t-d_m-1}\eta(s) \quad \displaystyle\sum_{s=t-d_M}^{t-d_m-1} y(s) \end{bmatrix},$$

By differencing the above inequalities based on the expectation, we get,

$$E\{\Delta V_4(t)\} = d_1^2\Big\{ \begin{bmatrix} \hat{y}(t) \\ \eta(t) \end{bmatrix}^T \begin{bmatrix} R_1 & R_2 \\ R_3 & R_4 \end{bmatrix} \begin{bmatrix} \hat{y}(t) \\ \eta(t) \end{bmatrix}$$

$$- d_1 \sum_{s=t-d_M}^{t-d(t)-1} \begin{bmatrix} \hat{y}(t) \\ \eta(t) \end{bmatrix}^T \begin{bmatrix} R_1 & R_2 \\ R_3 & R_4 \end{bmatrix} \begin{bmatrix} \hat{y}(t) \\ \eta(t) \end{bmatrix}$$

$$- d_1 \sum_{s=t-d(t)}^{t-d_m-1} \begin{bmatrix} \hat{y}(t) \\ \eta(t) \end{bmatrix}^T \begin{bmatrix} R_1 & R_2 \\ R_3 & R_4 \end{bmatrix} \begin{bmatrix} \hat{y}(t) \\ \eta(t) \end{bmatrix}$$

$$+ d_1\big[\hat{y}^T(t-d(k))U_1\hat{y}^T(t-d(k))\big]$$

$$- \hat{y}^T(t-d_M)U_1\hat{y}^T(t-d_M)$$

$$+ \hat{y}^T(t-d_m)U_2\hat{y}(t-d_m)$$

$$- \hat{y}^T(t-d(t))U_2\hat{y}^T(t-d(t))$$

$$- \sum_{s=t-d_M}^{t-d(t)-1} \eta^T(s)U_1\eta(s) + 2\hat{y}(s)$$

$$- \sum_{s=t-d(t)}^{t-d_m-1} \eta^T(s)U_2\eta(s) + 2\hat{y}(s)\Big\}$$

$$< \Phi^T(t)\Omega_4\Phi(t) - \frac{d_1}{d_M-d(t)}\left(\sum_{t=t-d_M}^{t-d(t)-1} \begin{bmatrix} \hat{y}(t) \\ \eta(t) \end{bmatrix}\right)^T$$

$$
\begin{bmatrix} R_1 & U_1 \\ * & R_2 + U_1 \end{bmatrix} \left(\sum_{t=t-d_M}^{t-d(t)-1} \begin{bmatrix} \hat{y}(t) \\ \eta(t) \end{bmatrix} \right)
$$

$$
- \frac{d_1}{d(t) - d_m} \left(\sum_{t=t-d(t)}^{t-d_m-1} \begin{bmatrix} \hat{y}(t) \\ \eta(t) \end{bmatrix} \right)^{TT}
$$

$$
\begin{bmatrix} R_1 & U_2 \\ * & R_2 + U_2 \end{bmatrix} \left(\sum_{t=t-d(t)}^{t-d_m-1} \begin{bmatrix} \hat{y}(t) \\ \eta(t) \end{bmatrix} \right)
$$

Letting, $\partial_1 = \frac{d_M - d(t)}{d_1}$, $\partial_2 = \frac{d_M - d(t)}{d_1}$ and since $\partial_1 + \partial_2 = 1$, by Lemma 2, there exists matrix $\begin{bmatrix} T_1 & T_2 \\ T_3 & T_4 \end{bmatrix} > 0$, such that

$$
\Delta V_4(t) \le \Phi^T(t) \Omega_4 \Phi(t) - \left[\begin{pmatrix} \sum_{t=t-d_M}^{t-d(t)-1} \begin{bmatrix} \hat{y}(t) \\ \eta(t) \end{bmatrix} \end{pmatrix} \\ \begin{pmatrix} \sum_{t=t-d(t)}^{t-d_m-1} \begin{bmatrix} \hat{y}(t) \\ \eta(t) \end{bmatrix} \end{pmatrix} \right]^T
$$

$$
\begin{bmatrix} R_1 & U_1 & T_1 & T_2 \\ * & R_1 + U_1 & T_3 & T_4 \\ * & * & R_1 & U_2 \\ * & * & * & R_2 + U_2 \end{bmatrix}
$$

$$
\begin{bmatrix} \begin{pmatrix} \sum_{t=t-d_M}^{t-d(t)-1} \begin{bmatrix} \hat{y}(t) \\ \eta(t) \end{bmatrix} \end{pmatrix} \\ \begin{pmatrix} \sum_{t=t-d(t)}^{t-d_m-1} \begin{bmatrix} \hat{y}(t) \\ \eta(t) \end{bmatrix} \end{pmatrix} \end{bmatrix}
$$

$$
\Delta V_4(t) \le \Phi^T(t) \Omega_4 \Phi(t)
$$

$$
- \begin{bmatrix} \sum_{t=t-d_M}^{t-d_m-1} \eta(t) \\ \hat{y}(t - d(t) - \hat{y}(t - d_M)) \\ \sum_{t=t-d_M}^{t-d_m-1} y(t) \\ \hat{y}(t - d_m) - \hat{y}(t - d(t))) \end{bmatrix}^T \begin{bmatrix} R_1 & U_1 & T_1 & T_2 \\ * & R_1 + U_1 & T_3 & T_4 \\ * & * & R_1 & U_2 \\ * & * & * & R_2 + U_2 \end{bmatrix}
$$

$$\times \begin{bmatrix} \sum\limits_{t=t-d_M}^{t-d_m-1} \eta(t) \\ \hat{y}(t-d(t)) - \hat{y}(t-d_M)) \\ \sum\limits_{t=t-d_M}^{t-d_m-1} y(t) \\ \hat{y}(t-d_m) - \hat{y}(t-d(t))) \end{bmatrix}$$

$$\leq \Phi^T(t)(\Omega_4 + \Omega_5)\Phi(t) \tag{17}$$

Considering the activation functions from the assumptions

$$\begin{bmatrix} \hat{y}(t) \\ f(\hat{y}(t)) \end{bmatrix}^T \begin{bmatrix} M_1 \Xi_1 & -M_2 \Xi_1 \\ -M_2 \Xi_1 & \Xi_1 \end{bmatrix} \begin{bmatrix} \hat{y}(t) \\ f(\hat{y}(t)) \end{bmatrix} \leq 0,$$

$$\begin{bmatrix} \hat{y}(t-d(t)) \\ g(\hat{y}(t-d(t))) \end{bmatrix}^T \begin{bmatrix} N_1 \Xi_2 & -N_2 \Xi_2 \\ -N_2 \Xi_2 & \Xi_2 \end{bmatrix} \begin{bmatrix} \hat{y}(t-d(t)) \\ g(\hat{y}(t-d(t))) \end{bmatrix} \leq 0 \tag{18}$$

where

$$\Xi_1 = diag\{\tilde{\alpha}_{11}, \tilde{\alpha}_{12}, \ldots, \tilde{\alpha}_{1n}\},$$

$$\Xi_2 = diag\{\tilde{\alpha}_{21}, \tilde{\alpha}_{22}, \ldots, \tilde{\alpha}_{2n}\},$$

$$M_1 = diag\{\tilde{M}_1^- \tilde{M}_1^+, \tilde{M}_2^- \tilde{M}_2^+, \ldots, \tilde{M}_n^- \tilde{M}_n^+\},$$

$$M_2 = diag\left\{ \frac{\tilde{M}_1^- + \tilde{M}_1^+}{2}, \frac{\tilde{M}_2^- + \tilde{M}_2^+}{2}, \ldots, \frac{\tilde{M}_n^- + \tilde{M}_n^+}{2}, \right\},$$

$$N_1 = diag\{\tilde{N}_1^- \tilde{N}_1^+, \tilde{N}_2^- \tilde{N}_2^+, \ldots, \tilde{N}_n^- \tilde{N}_n^+\},$$

$$N_2 = diag\left\{ \frac{\tilde{N}_1^- + \tilde{N}_1^+}{2}, \frac{\tilde{N}_2^- + \tilde{N}_2^+}{2}, \ldots, \frac{\tilde{N}_n^- + \tilde{N}_n^+}{2}, \right\}.$$

We obtain the zero equation, by introducing the relaxation matrix

$$2\big(\eta^T(t)W_0^T + \hat{y}^T(t)W_1^T\big)[B\hat{y}(t-\psi(t))$$
$$+ Cf(\hat{y}(t)) + Dg(\hat{y}(t-d(k))) - \hat{y}(t) - \eta(t)] = 0 \tag{19}$$

Combining all the equations,

$$\Delta V(t) \leq \Phi^T(t)\left(\sum_{s=1}^{8} \Omega_s\right)\Phi(t) < 0 \tag{20}$$

If $\tilde{\lambda} = \lambda_{\max}(\Lambda) < 0, \quad \Lambda = \sum\limits_{s=1}^{8} \Omega_s$ with

$$\Delta V(t) \le \tilde{\lambda} \|\hat{y}(t)\|^2, \ \forall \hat{y}(t) \ne 0. \tag{21}$$

This implies that system (3) is globally asymptotically stable.

Hence, the proof is complete.

Remark These results can also be applied to the system of randomly occurring parameter uncertainties, where the stability conditions does not depend upon the transmission delays caused due to delay bounds but also depends upon the time-varying leakage delays.

4 Numerical Example

Example 1 Let us take the uncertain NNs with the subsystem as:

$$B = \begin{bmatrix} 0.8 & 0 \\ 0 & 0.7 \end{bmatrix}, C = \begin{bmatrix} -2.3 & 0 \\ 1.7 & 0.4 \end{bmatrix}, D = \begin{bmatrix} 1.3 & 0.2 \\ 0.1 & -0.02 \end{bmatrix},$$

$$N_a = \begin{bmatrix} 0.04 & 0.02 \\ 0 & -0.03 \end{bmatrix}, N_b = \begin{bmatrix} 0.07 & 0 \\ 0.04 & 0.03 \end{bmatrix}, N_c = \begin{bmatrix} 0.03 & 0.02 \\ 0.01 & -0.04 \end{bmatrix}$$

With the nonlinear activation functions given by,

$$f(\hat{y}(t)) = \begin{bmatrix} \tanh(0.6\hat{y}_1(t)) + 0.6 \\ \tanh(-0.4\hat{y}_2(t)) + 0.2 \end{bmatrix}, \quad g(y(t)) = \begin{bmatrix} \tanh(0.3\hat{y}_1(t)) \\ \tanh(-0.4)) \end{bmatrix}$$

The time-varying delay and the leakage delay are given by.

$d(t) = \frac{1}{2}\left(7 + 5\sin\left(\frac{t\pi}{2}\right)\right)$ and $\psi(t) = \frac{1}{2}\left(5 + 3\sin\left(\frac{t\pi}{2}\right)\right)$ with corresponding bounds given by $d_m = 3$, $d_M = 8$, $\psi_m = 1$, $\psi_M = 5$. By solving the LMIs, a solution is obtained.

Figure 1 showcases NNs converge smoothly in the absence of leakage delay and there is some kind of dynamic behavioral change in the network in the presence of leakage delay. Hence, the effectiveness of the given result is thus verified.

5 Conclusion

In this paper, the analysis of stability problem is discussed with leakage delays. Sufficient conditions are encountered to ensure the delay-dependent conditions and to ensure the robustness by using the LMI approach. Numerical example has also been for the derived theoretical result.
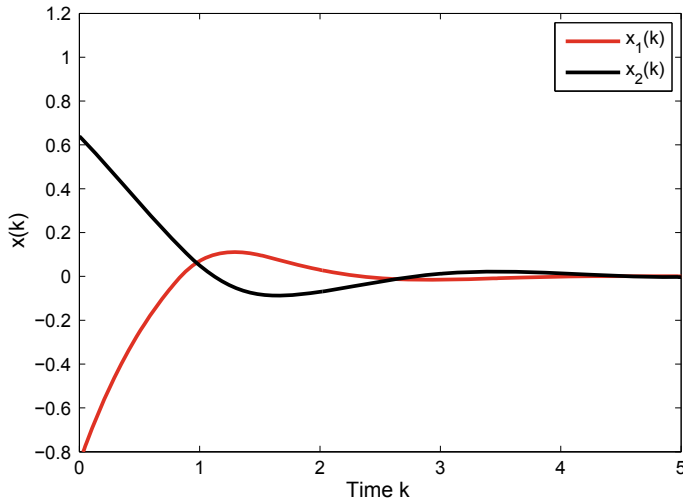
Fig. 1 State responses of the system

References

1. Das S, Moura JMF (2015) Distributed Kalman filtering with dynamic observations consensus. IEEE Trans Sign Process 63(17):4458—4473
2. Hu J, Wang Z, Gao H (2018) Joint state and fault estimation for time-varying nonlinear systems with randomly occurring faults and sensor saturations. Automatica 97(2018):150–160
3. Maheswari K (2016) H∞ state estimation for discrete-time neural networks with multiple missing measurements using second order reciprocal convex approach. Adv Intell Syst Comput 412:102–119
4. Rakkiyappan R, Maheswari K, Sivaranjani K (2018) Young Hoon Joo, Non-Fragile finite-time l2 − l∞ state estimation for discrete-time neural networks with semi-Markovian switching and random sensor delays based on Abel lemma approach. Nonlinear Anal Hybrid Syst 29:283–302
5. Rakkiyappan R, Maheswari K, Velmurugan G, Park JH (2018) Event-triggered H∞ state estimation for semi-Markov jumping discrete-time neural networks with quantization. Neural Netw 105:236–248
6. Nagamani G, Ramasamy S (2016) Dissipativity and passivity analysis for discrete-time T-S fuzzy stochastic neural networks with leakage time-varying delays based on the Abel lemma approach. J Franklin Inst 353:3313–3342

Forecasting PM10 Concentration Based on a Hybrid Fuzzy Time Series Model

Yousif Alyousifi and Mahmod Othman

Abstract Developing statistical models for air pollution forecasting is crucial for managing air quality. Nevertheless, many researchers have concentrated on improving the model's accuracy when applying for data with many fluctuations in the pollutant's concentration. Also, they have attempted to address the uncertainty analysis that might lead to inadequate outcomes. The fuzzy time series (FTS) is considered one of the powerful models that are commonly applied in predicting air pollution. However, most FTS models are not accurate in partitioning the universe of discourse. Therefore, a new hybrid model based on the FTS-based Markov chain and C-Means clustering technique with an optimal number of clusters is proposed in this study. This hybridization is contributed to produce an adequate partition and improve the model accuracy accordingly. The superiority of the proposed model is validated using three common statistical criteria. The PM_{10} concentration data collected from Melaka, Malaysia is used in this study. Results prove that the proposed model greatly improved the prediction accuracy, for which it outperformed several fuzzy time series models. Hence, we have concluded that the model proposed is a good option for forecasting air pollution and any type of random data.

Keywords Air pollution · C-Means clustering techniques · Fuzzy time series · Markov transition matrix · PM_{10}

Y. Alyousifi (✉) · M. Othman
Department of Fundamental and Applied Sciences, Universiti Teknologi PETRONAS, Bandar Siri Iskandar, 31750 Tronoh, Perak, Malaysia
e-mail: yalyousifi@tu.edu.ye

Y. Alyousifi
Department of Mathematics, Faculty of Applied Sciences, Thamar University, Thamar, Yemen

177

1 Introduction

Prediction of air pollution is a crucial environmental problem in most countries as it is beneficial to air pollution assessment where the result found can be used for managing the air quality. Also, predicting the concentration of air pollutants is vital for assessing the effects of air pollutants on human health [1, 2].

Many researchers have been applied the FTS models for forecasting air pollution [3]. For instance, Cagcag et al. [4] have predicted air pollution in Ankara based on a new seasonal FTS model. Koo et al. [5] have utilized some statistical models in order to forecast the air pollution events and made a comprising to determine the adequate model. Apart from that, the FTS model based on Fuzzy K-Medoids clustering is suggested for minimizing the sensitivity of outliers in order to produce adequate forecasts of air pollution [6]. Furthermore, in air pollution forecasting, an integrated model of the fuzzy theory and advanced optimization algorithm is introduced [7]. A hybrid FTS model is also suggested for predicting air pollution [8].

Since the hybrid models are performed better than their single models. For example, the hybrid models in conventional time series have shown an improvement in the model performance more than their single models such as the ARMA and ARIMA models, which are combined between the autoregressive (AR) and moving average (MA) models. These hybrid models have outperformed the singles models. Accordingly, many researchers have proposed a variety of hybrid FTS models to address some of these issues. For example, a hybrid fuzzy time series Markov chain (FTSMC) model has been proposed by Tsaur [9], which is used for establishing proper weights of the fuzzy relationships between the data points of the time series pattern. Aripin et al. [10] have applied the FTSMC model for air pollution forecasting. FTSMC is superior to many methods available in the literature in terms of model accuracy. Nevertheless, it has a drawback in utilizing an arbitrary partition of intervals. So, it could not determine the appropriate length of intervals, which is considered a deficiency in determining the effective length of the interval in this hybrid model. Apart from that, Alyousifi et al. [11], proposed the use of the FTSMC model based on the optimal grid partition method for modeling air pollution events in Malaysia.

Furthermore, the hybrid model of the fuzzy time series model and the C-Means clustering technique has outperformed the fuzzy time series models. For example, Van Tinh et al. [12] proposed a hybrid fuzzy time series and K-Means clustering for forecasting the enrolment of the University of Alabama. Likewise, Kai et al. [13] proposed a forecasting model of fuzzy time series based on K-Means clustering. However, it has a limitation in dealing with recurrences of the observations and the number of clusters that have been chosen randomly, which can be led to an insufficient length of intervals and insufficient forecasting accuracy. To address these two limitations of the existing hybrid models that were previously mentioned, this study proposed a new hybrid model (FTSMC-CMeans) by combining the fuzzy time series, Markov chain, and C-Means clustering technique. It is anticipated that the proposed model may overcome these limitations and produce a better model

forecasting and to produce better model accuracy. The proposed model is validated by comparison with several existing models.

2 Methodology

FTS is a forecasting model based on the fuzzy set theory and Fuzzy Logic proposed by Zadeh in 1965 [14]. FTS models are considered as advanced statistical time series analysis applying to fuzzy sets. due to their advantages. Song and Chissom [15, 16] are firstly introduced the first-order FTS model. To reach the high accuracy of the forecasting model, developing FTS models have been done by enhancing the three main stages of the FTS algorithm, which are fuzzification, fuzzy inference, and defuzzification [16]. For instance, Chen [17] has improved the FTS model proposed by Song and Chissom [15, 16] through the use of fuzzy logical relation tables for the purpose of reducing the computational complexity of the model. Also, Huarng [18, 19] have developed Chen's model [17] by handling the issue of the length of the interval.

2.1 C-Means Clustering Technique

The C-Means clustering technique [20, 21] is one of the famous unsupervised learning algorithms. It is a partitioning clustering algorithm, which partitions a given data set into a set of C clusters. The result of the C-Means clustering Technique depends on the number of clusters. Apart from that, the main matter in partitioning clustering is determining the optimal number of clusters in a data set. For instance, the C-Means clustering requires the user to specify the number of clusters k in order to be generated. Accordingly, in this paper, an optimal number of clusters is determined based on three different methods, which are elbow, average silhouette, and gap statistic methods [22], through using two functions in R, which are *fviz_nbclust( )* and *NbClust*() [23]. By using these functions, the optimal number of clusters determines the best partition selected. For further details about these methods see [24].

2.2 Hybrid Fuzzy Time Series Model

The algorithm of this study is adopted the arithmetic processes proposed by Tsaur [9] in order to conduct the proposed model. The steps of the model are as follows:

Step 1. Define U.
Step 2. Divide U into subintervals based on the C-Means Clustering technique.
Step 3. Define the fuzzy sets A_i for each time series observation on U.

Step 4. Fuzzify the actual values of the time series (observations) into fuzzy numbers based on the maximum membership value.

$$A_i = \frac{f_{A_i}(u_1)}{u_1} + \frac{f_{A_i}(u_2)}{u_2} + \cdots + \frac{f_{A_i}(u_n)}{u_n} \tag{1}$$

where f_{A_i} is the membership function of fuzzy set A_i; $f_{A_i} : U \to [0, 1]$. $f_{A_i}(u_r) \in [0, 1]$ and $1 \le r \le n$.

Step 5. Establish the fuzzy logical relationships (FLRs) and determine fuzzy logical relation groups (FLRGs). Then, create the Markov transition probability matrix. The matrix P is $P_{n \times n}$, and P_{ij} is transition probability from state A_i to state A_j. P_{ij} can be calculated as follows

$$P_{ij} = \frac{N_{ij}}{N_{i.}}, i, j = 1, 2, \ldots, n \tag{2}$$

where N_{ij} is the frequencies and $N_{i.} = \sum_{j=1}^{n} N_{ij}$ is the total frequencies.

Step 6. Calculate the forecasted values by considering the following cases:

Case 1. If the FLRG of A_i is one-to-one, i.e., $A_i \to A_k$, with $P_{ik} = 1$ and $P_{ij} = 0, j \ne k$), then the forecasting of $F(t)$ is m_k, the midpoint of u_k, $k = 1, 2, ..n$, which is determined by using Eq. (3)

$$F(t + 1) = m_k P_{ik} = m_k \tag{3}$$

Case 2. If the FLRG of A_i is one-to-many, i.e., $A_i \to A_1, A_2, ...A_n, i = 1, 2, .., n$). Thus, if the state is A_i for the actual value $Y(t)$ at time t, the forecasted value $F(t + 1)$ is calculated according to Eq. (4)

$$\begin{aligned} F(t + 1) &= m_1 p_{i1} + m_1 p_{12} + \cdots + m_{i-1} p_{i(i-1)} + Y(t) \, p_{ii} \\ &\quad + m_{i+1} p_{i(i+1)} + \cdots + m_n p_{in} \end{aligned} \tag{4}$$

where $m_1, m_2, \ldots, m_n$ are the midpoint of $u_1, u_2, \ldots, u_n$ and m_i replaced by $Y(t)$.

Step 7. Adjust the predicted values by adding the differences of actual values $Y(t)$. The adjusted forecasted values can be written by

$$\hat{F}(t + 1) = F(t + 1) + \text{diff}(Y(t)) \tag{5}$$

2.3 Model Evaluation

Three statistical criteria used in this study in order to validate the proposed model, which are MAPE, RMSE, and Thiels' U statistics, are given in Equations as follows:

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{Y_i - F_i}{Y_i} \right| \times 100, \tag{6}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - F_i)^2}{N}} \quad \text{and} \tag{7}$$

$$\text{Theil's U} = \frac{\sqrt{\sum_{i=1}^{N}(Y_i - F_i)^2}}{\sqrt{\sum_{i=1}^{N} Y_i^2} + \sqrt{\sum_{i=1}^{N} F_i^2}} \tag{8}$$

where Y_i means the actual data, F_i the predicted values and N is the total number of the actual data.

3 Results and Discussion

In this section, we will provide the results of the proposed model through the implementation and comparison of the model using two datasets. The algorithm of the proposed model is implemented for predicting the concentration of PM_{10}. To validate the proposed model, the datasets are used to assess the model performance and compare it with other models.

3.1 Air Pollution Forecasting

For validation of the proposed model, the algorithm of the proposed model is also applied for both training and testing datasets of PM_{10} concentrations are used to evaluate the performance of the proposed model and compare it with some of the existing FTS models. For implementing the Algorithm of the proposed model, we start by defining U from data and then apply the C-Means clustering method.

As can be seen from Table 1 that the proposed model has been performed very well with producing small errors. Figures 1 and 2 demonstrate the results of a comparison of the proposed model and some existing models using the testing data of PM_{10} concentrations. It observed that the model has predicted PM_{10} data well with producing the smallest error. Besides, the proposed model outperformed the other existing models. Also, the model shows its superiority as compared with the other

Table 1 A comparison of the hybrid model and some FTS models using PM10 data

	Model	Using training dataset			Using testing dataset		
		RMSE	MAPE	U statistic	RMSE	MAPE	U statistic
1	Chen's model [17]	18.04	14.67	1.32	9.12	14.26	0.80
2	Sliva2 et al.'s model [25]	16.10	17.79	1.18	10.23	18.68	0.89
3	Yu's model [26]	17.06	9.74	1.25	8.87	14.20	0.78
4	Cheng's model [27]	16.70	10.88	1.22	8.67	13.69	0.76
5	Severiano et al.'s [28]	16.10	17.79	1.18	10.23	18.68	0.89
6	Sliva et al.'s model [29]	15.67	9.28	1.14	8.90	13.92	0.78
7	Sadaei et al.'s [30]	17.06	9.82	1.25	8.73	13.99	0.77
8	Tsuar's model [9]	17.07	9.40	1.25	8.66	13.86	0.76
9	Hybrid Model	7.55	6.83	0.55	5.01	7.25	0.45
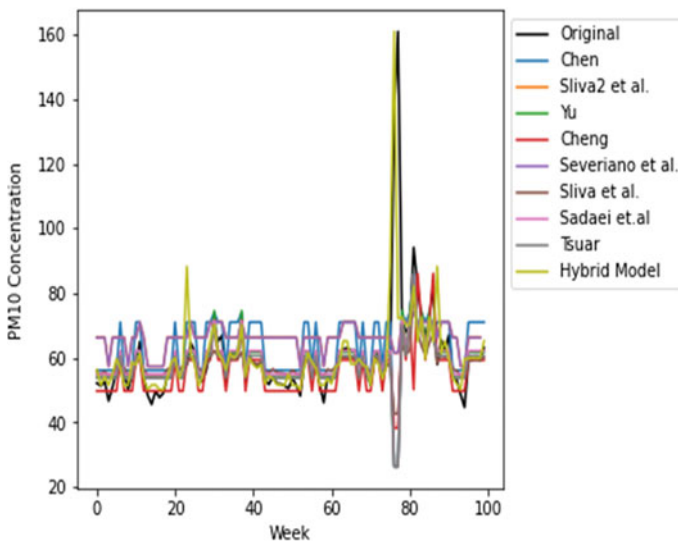
Fig. 1 A comparison of the hybrid model and some FTS models using the training PM_{10} Data

models. This implies the proposed model is an improved option for forecasting air pollution events.

4   Conclusion

This study proposed a new hybrid fuzzy time series model, which is implemented for predicting PM10 concentration. The model hybridization has been contributed in
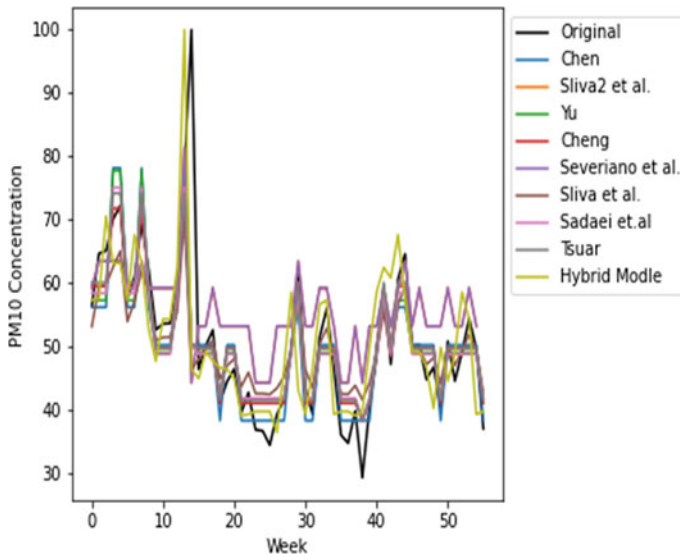
Fig. 2 A comparison of the hybrid model and some FTS models using the testing PM_{10} Data

producing adequate partitioning and improving the model accuracy. The proposed model was evaluated by comparison with eight FTS models, which are famous in the literature. A comparison has demonstrated the ability of the model in avoiding the arbitrary selection of intervals and dealing with recurrent observations, which greatly improves model accuracy. The predicted values found by the model display its flexibility in the FTS for predicting air pollution. Generally, the proposed model has the flexibility to apply for many types of time series data.

References

1. Yan Y, Li Y, Sun M, Wu Z (2019) Primary pollutants and air quality analysis for urban air in China: evidence from Shanghai. Sustainability 11(8):2319
2. Alyousifi Y, Othman M, Faye I, Sokkalingam R, Silva PC (2020) Markov weighted fuzzy time-series model based on an optimum partition method for forecasting air pollution. Int J Fuzzy Syst 22:1468–1486
3. Cheng CH, Huang SF, Teoh HJ (2011) Predicting daily ozone concentration maxima using fuzzy time series based on a two-stage linguistic partition method. Comput Math Appl 62(4):2016–2028
4. Cagcag O, Yolcu U, Egrioglu E, Aladag CA (2013) Novel seasonal fuzzy time series method to the forecasting of air pollution data in Ankara. American Journal of Intelligent Systems. 3(1):13–19
5. Koo JW, Wong SW, Selvachandran G, Long HV (2020) Prediction of Air Pollution Index in Kuala Lumpur using fuzzy time series and statistical models. Air Q Atmosp Health 1–12
6. Dincer NG, Akkuş Ö (2018) A new fuzzy time series model based on robust clustering for forecasting of air pollution. Eco Inform 43:157–164

7. Yang H, Zhu Z, Li C, Li R (2019) A novel combined forecasting system for air pollutants concentration based on fuzzy theory and optimization of aggregation weight. Appl Soft Comput 105972
8. Wang J, Li H, Lu H (2018) Application of a novel early warning system based on fuzzy time series in urban air quality forecasting in China. Appl Soft Comput 71:783–799
9. Tsaur RC (2012) A fuzzy time series-Markov chain model with an application to forecast the exchange rate between the Taiwan and US dollar. Int J Innov Comput Inf Control 8(7):4931–4942
10. Aripin A, Suryono S, Bayu S (2016) Web-based prediction of pollutant PM10 concentration using Ruey Chyn Tsaur fuzzy time series model. AIP Conf Proc 17(46):20–46
11. Alyousifi Y, Othman M, Sokkalingam R, Faye I, Silva PC (2020) Predicting daily air pollution index based on fuzzy time series Markov chain model. Symmetry 12(2):293
12. Van Tinh N, Vu VV, Linh TTN (2016) A new method for forecasting enrolments combining time-variant fuzzy logical relationship groups and k-means clustering. Int Res J Eng Technol 3(3):1–32
13. Kai C, Fang-Ping F, Wen-Gang C v A novel forecasting model of fuzzy time series based on k-means clustering. IWETCS, IEEE, 223–225
14. Zadeh LA (1965) Fuzzy sets. Inf Control 8(7):338–353
15. Song Q, Chissom BS (1993) Forecasting enrollments with fuzzy time series-Part I. Fuzzy Sets Syst 54:1–10
16. Song Q, Chissom BS (1994) Forecasting enrollments with fuzzy time series-Part II. Fuzzy Sets Syst 62(1):1–8
17. Chen SM (1996) Forecasting enrolments based on fuzzy time series. Fuzzy Sets Syst 81(3):311–319
18. Huarng K (2001) Effective lengths of intervals to improve forecasting in fuzzy time series. Fuzzy Sets Syst 123(3):387–394
19. Huarng K, Yu TH (2006) Ratio-based lengths of intervals to improve fuzzy time series forecasting. IEEE Trans Syst Man Cybern Part B Cyber 36:328–340
20. Hartigan JA (1979) A K-means clustering algorithm: algorithm AS 136. Appl Stat 28(11):126–130
21. Zhang Z, Zhu Q (2012) Fuzzy time series forecasting based on k-means clustering. Open J Appl Sci 2:100–103
22. Rousseeuw PJ, Kaufman L (1990) Finding groups in data. Wiley, Hoboken
23. Charrad M, Ghazzali N, Boiteau V, Niknafs A (2014) NbClust: An R package for determining the relevant number of clusters in a data set. J Stat Softw 61:1–36
24. Kaufman L, Rousseeuw PJ (2009) Finding groups in data: an introduction to cluster analysis. John
25. Silva PC, Sadaei HJ, Guimarães FG (2017) Interval forecasting with fuzzy time series. Conference: IEEE symposium series on computational intelligence
26. Yu HK (2005) Weighted fuzzy time series models for TAIEX forecasting. Physica A 349(34):609–624
27. Cheng CH, Chen TL, Chiang CH (2006) Trend-weighted fuzzy time-series model for TAIEX forecasting neural information processing. Spring Berlin/Heidelberg Lect Notes Comput Sci 42(34):469–477
28. Severiano, C. A., Silva, P. C., Sadaei, H. J., Guimarães, F. G.: Very short-term solar forecasting using fuzzy time series. In 2017 IEEE international conference on fuzzy systems (FUZZ-IEEE). 1–6 (2017).
29. Silva PC, Sadaei HJ, Guimarães FG (2019) Probabilistic forecasting with fuzzy time series. IEEE Trans Fuzzy Syst 99
30. Sadaei HJ, Enayatifar R, Abdullah AH, Gani A (2014) Short-term load forecasting using a hybrid model with a refined exponentially weighted fuzzy time series and an improved harmony search. Int J Electr Power Energy Syst 62:118–129

Carbonated Water Injection for EOR and CO₂ Storage: A Simulation Study

A. Abdulrahman and J. Foroozesh

Abstract CO$_2$ Enhanced Oil Recovery (EOR) techniques have gained massive attention by oil companies as they target the oil industry's two main concerns of CO$_2$ contribution to the climate change and the decline in oil production. Carbonated Water Injection (CWI) is a promising EOR technique that promotes safe and economic CO$_2$ storage, moreover, it mitigates the pitfalls of CO$_2$ injection which include low sweep efficiency, early CO$_2$ breakthrough, and the risk of CO$_2$ leakage in subsurface formations. Upon the contact of Carbonated Water (CW) with oil inside the reservoir, CO$_2$ molecules cross the interface between the two fluids moving towards the oil phase due to the concentration gradient and its higher solubility in hydrocarbons. Subsequently CO$_2$ molecules diffuse inside the oil until thermodynamic equilibrium is reached. CO$_2$ dissolution in oil causes it to swell and consequently leads to changes in its physical properties (viscosity and density). Such changes in oil properties lead to improved sweep and displacement efficiency and thus higher recovery factor. Several experimental studies have been reported in the literature, but little work has been done on the simulation of CWI due to the complex multi-physics nature of this process. In this paper, Schlumberger equilibrium-based compositional simulator (ECLIPSE300) has been used to investigate the oil recovery and CO$_2$ storage during CWI. The carbonated water has been simulated using two injector wells placed at the same location where they inject free CO$_2$ and water at a certain volumetric ratio to account for the mass fraction of dissolved CO$_2$ inside the carbonated water. CO$_2$SOL option has been used to account for CO$_2$ solubility inside the water phase and a 2D cartesian model (x, y) has been considered to avoid the effects of gravity which might reduce the amount of CO$_2$ dissolved inside the oil phase. A sensitivity analysis on CW injection rate, and the effect of CO$_2$ diffusion have been investigated. It was found that low injection rate promotes longer contact time thus more CO$_2$ molecules

A. Abdulrahman
Petroleum Engineering Department, Universiti Teknologi PETRONAS, Perak, Malaysia

J. Foroozesh (✉)
School of Energy and Electronic Engineering, University of Portsmouth, Portsmouth, United Kingdom
e-mail: jalal.foroozesh@port.ac.uk

will get transferred to the oil leading to high oil recovery and CO_2 storage. In addition, it was found that CO_2 diffusion between grid cells has minimal impact on oil recovery and CO_2 storage in this case study.

Keywords Carbonated water injection · Compositional simulation · Enhanced oil recovery · CO_2 storage · CO_2 diffusion · CO_2 solubility in water

1 Introduction

Carbonated water injection is a promising CO_2 based EOR technique that injectes CO_2 in an effective and efficient way. In this method, the CW is prepared by dissolving CO_2 inside the water at surface conditions using gas infusion generator [1], then the CW mixture is injected into the formation to increase oil recovery and store the dissolved CO_2 [2]. When CW contacts the oil phase inside the reservoir, CO_2 molecules gradually cross the interface between the two fluids due to the concentration gradient (cross phases diffusion) and higher solubility of carbon dioxide in hydrocarbons, and subsequently, CO_2 molecules diffuse over time within the resident oil until thermodynamic equilibrium is reached. CO_2 dissolution in oil causes it to swell and consequently leads to a change in its physical properties (density–viscosity) [2–6].

Several experimental studies have been carried out to evaluate the additional recovery factor during CWI, these experiments have been conducted using different setups such as coreflood rigs, sand-packs, and high-pressure transparent micromodels. It has been found that CWI in secondary mode (after primary depletion) yields a higher recovery factor compared to CWI in tertiary mode (after water injection). Sohrabi et al. [7] have performed coreflood experiment in a water-wet reservoir core (sandstone), in the first part of the experiment, CWI was injected in a secondary mode and the recorded increment in the recovery factor was 19%. In the second part of the experiment, CWI was injected in a tertiary mode, the recorded increment the recovery factor was determined to be 9%. Mosavat and Torabi [2] have conducted a flooding experiment in a sand-pack apparatus filled with sand made of 99.88% silicon, it was found that the injection of CO_2-saturated water resulted in improving the conventional water flooding oil recovery factor by about 19.0% and 12.5% of OOIP for the secondary and tertiary scenario, respectively. Riazi et al. [8] have injected CWI (tertiary mode) in a micro model made of two flat glass plates, the recorded recovery factor by water injection was 27% and the increment in the recovery factor due to CWI (post water flood) was found to be 6%.

Carbonated water injection EOR has the potential to store CO_2 safely and economically in the geological formations. Kechut et al. [9] have performed an experimental investigation of CO_2 storage in sandstone core samples. During secondary CWI, 45% of the dissolved CO_2 was stored while in tertiary CWI it was found that about 51% of injected CO_2 was stored. Tavakolian et al. [10] have performed series of coreflood experiments using sandstone core plugs and crude oil at real reservoir conditions. The

experiments were carried out to compare the performance of CWI and CO_2 flooding. CWI resulted in a less and more controllable CO_2 production profile compared to CO_2 flooding. Also, it was found that duringCO_2 flood experiments, CO_2 production exceeded that of CO_2 storage very quickly and after that large volume of CO_2 had to be reinjected (recycled), whereas CO_2 production during CWI was significantly less than CO_2 storage even after a large volume of CWI.

The additional oil recovery reported during CWI can be attributed to the effect of several mechanisms simultaneously. These recovery mechanisms include oil swelling, viscosity reduction, wettability alteration, IFT reduction, and CO_2 exsolution [3–6, 11–13]. Sohrabi et al. [4–6] have investigated oil swelling during CWI by conducting many flooding experiments in a micro model using a light oil (viscosity = 0.83 cp) and also a viscous oil (viscosity = 16.5 cp). They have found out that the swelling in the light oil was 105% while the swelling in the viscous oil was 23% which led to the conclusion of swelling degree is a function of oil viscosity. A series of experimental studies have been conducted by Miller and Jones [3] to investigate the change in oil viscosity during the CWI process. They have run the experiments at different values of pressures and temperatures, different oil samples with a viscosity of 10 and 17 API were used. It was found out that in all the cases the viscosity of the oil in the absence of CO_2 increased as the pressure increased, while the viscosities of the oil with CO_2 in the solution decreased as the pressure increased. Mosavat and Torabi [13] have performed several experiments to investigate the effect of CO_2 diffusion on wettability alternation. It was also concluded that once CO_2 molecules reach the rock surface (in contact with the oil droplet), the molecules started replacing the hydrocarbon molecules were adsorbed on the surface. This gradual replacement caused the rock surface to shift its wettability towards the water-wet condition. Yang et al. [11] have conducted lab experiment to measure dynamic interfacial tension of oil-brine systems compared to oil-brine CO_2 systems at different values of pressures and temperatures. They realized that in the crude oil-brine system, the dynamic interfacial tension reduction was primarily caused by the natural surfactant which was possibly generated due to chemical reactions between the crude oil and the reservoir brine. While For the crude oil-brine CO_2 system, the dissolution of CO_2 into crude oil was an additional factor to reduce the dynamic interfacial tension. Riazi [12] have conducted a fluid flow experiment on CWI using a high-pressure micro-model. They have noticed an interesting phenomenon of gas nucleation subsequent to depressurizing of the CW flooded. This pressure reduction appreciably affected the production process, leading to an additional oil recovery (due to CO_2 exsolution).

2 Simulation Studies of CWI and Challenges Encountered

An accurate reservoir model with the ability to reliably predict the performance of different recovery methods is vital for successful reservoir management. There are black oil and compositional simulation approaches available depending on type of reservoir fluids and recovery methods, Compositional simulation approach should

be considered when simulating carbonated water injection since significant cross-phase mass transfer of CO_2 molecules occurs between carbonated water and hydrocarbon leading to oil properties changes. In addition, hydrocarbon physical properties (viscosity and density) are a function of both pressure and CO_2 content. The following paragraphs will explain how CWI is simulated in ECLIPSE300, the complex physics required for successful simulation of CWI, the simulation studied that have been performed and the limitations encountered.

2.1 Simulation of CWI in ECLIPS300 Compositional Simulator

There is no unique option in ECLIPSE300 that for precise simulation of CWI. However, a few researchers and industry experts tried to simulate CWI using "CO_2SOL" which is a three-phase compositional option that accounts for CO_2 solubility in water and CO_2 dissolution in oil phase [14]. In this option, it is assumed that instantaneous equilibrium (instantaneous phase equilibrium) exists between oil, gas, and water phases. Furthermore, water is not allowed to dissolve in gas or oil phase. CO_2 component can be present in the three phases, while hydrocarbon components can only be present in gas and oil phases. "SOLUBILI" keyword is used in the PROPS section to provide experimental CO_2 in water solubility data. If the solubility data is not provided by the user, built in correlations will be used to generate the CO_2 in water solubility data [14–16].

ECLIPSE300 has two diffusion models that can be used to represent CO_2 diffusion within the oil or gas phase (diffusion between two adjacent cells containing the same phase); in the first model, diffusion is driven by the concentration gradient, while in the second model diffusion is driven by the gradient of the chemical potential. Diffusivity coefficient can be specified in two possible ways based on the mentioned diffusion models; normal diffusion coefficients Di defined by keywords DIFFCOIL and DIFFCGAS or activity corrected diffusion coefficient D_i^{α} defined by keywords DIFFAGAS and DIFFAOIL [15].

2.2 The Complex Physics Required for Successful Simulation of CWI

Kechut et al. [9, 17, 18] used compositional simulators to simulate a set of carbonated water coreflood experiments. The coreflood experiments had been performed using two types of core (a reservoir water-wet core and Clashach water-wet sandstone core) and two different oil samples: a light oil (n-decane) and a viscous stock-tank crude oil with known composition. Secondary water injection, secondary CWI and tertiary CWI were simulated using the ECLIPSE300 commercial reservoir simulators. First

water injection (WI) was simulated. The input water–oil relative permeability (Krw-o) curve into the simulator was obtained based on Corey correlations when production data was history matched. Next CWI was simulated. The ECLIPSE300 compositional simulator was used with a tuned EOS, which could acceptably predict the PVT properties of the oil-CO_2 system. The Krw-o obtained for WI tests were input into the ECLIPSE300 during CWI simulation. The simulation results by ECLIPSE300 showed an over prediction of oil recovery in comparison with experimental data for all coreflood tests. They explained that compositional simulators were developed based on the inherent assumptions of instantaneous equilibrium and complex mixing while during carbonated water flooding, CO_2 transferring process from water into the oil phase happens gradually. The resistance against CO_2 migration at the interface between water and oil phases plays an important role, which prevents immediate distribution of CO_2 between the phases. Consequently, the assumption of instantaneous equilibrium may not be realistic during CWI process. To support their explanation, they mentioned the micro-model observation reported by Sohrabi et al. [5], who used image processing technique during the conducted visualisation tests. Riazi et al. [8] have reported that in the micro-model system the swelling of oil blobs as a result of CO_2 diffusion happens gradually and it needs time to stabilise and reach equilibrium state (170 h to reach its final shape corresponding to maximum swelling value). Embid and Rivas [19] explained that this assumption of instantaneous equilibrium can lead to a large error in the cases where mass transfer resistances are large. They mentioned that this may happen when there are short contact times for the mass transfer process (laboratory displacement in cores) or large diffusion patterns are available for components to diffuse through (field scale), moreover, if there are slow diffusion velocities due to large viscosity. All these can lead to not having an instantaneous thermodynamic equilibrium state.

The effect of CO_2 diffusion within the oil phase is another challenge encountered during the simulation of CWI. Molecular diffusion is defined as the movement of molecules caused by Brownian motion or composition gradient in a fluid mixture that can facilitate the dissolution of CO_2 into the oil systems [20]. Injected CO_2 during CO_2 based injection projects could just pass through the reservoir, causing early breakthrough, if there is no molecular diffusion. This would lead to lower oil recovery prediction as some parts of the reservoir are not touched by the CO_2 causing no benefit from the CO_2 presence. This may lead to wrong prediction as diffusion mechanism can play important roles during CO_2 based injection projects [21, 22]. Alfarge et al. [23] reported that most of the studies in the literature identified molecular diffusion as an important mechanism for CO_2 to enhance the oil recovery in oil reservoirs. Grogan and Pinczewski [24] highlighted the importance of molecular diffusion during the oilfiled life. As such disregarding the diffusion mechanism in the model can lead to overestimation or underestimation of the recovered oil [22, 23].

2.3   Nonequilibrium-Based CWI Simulation

Foroozesh et al. [25, 26] have come up with a novel 1D compositional simulation approach whereby the assumption of instantaneous equilibrium was relaxed by adding a new term to the governing equation called mass transfer coefficient (MTC). This term used to capture the slow CO_2 mass transfer from CW to oil within the grid cell. Their kinetic based model was used to simulate a secondary CWI experiment reported by Sohrabi et al. [7] for validation purposes. MTC was tuned to match differential pressure (DP) and recovery factor (RF) data of the experiments and a good match was achieved. Additionally, Foroozesh and Jamiolahmady [25] investigated the role of rock wettability and wettability alteration in the performance of CWI process. They simulated CWI coreflood experiments in mixed wet sandstone cores. The relative permeability, capillary pressure and MTC values were tuned and a good matching was achieved. The also used their kinetic based model to map the result of CWI in core to reservoir (scale-up practice). Although, Foroozesh et al.'s simulation approach was a novel way of simulating CWI but they did not capture the effect of CO_2 diffusion within the phases.

3   Simulation Of CWI Using ECLIPSE300

In this study, a 2D two-phase compositional model has been constructed using ECLIPSE300 to investigate the oil recovery and CO_2 storage during CWI. The 2D model is in the direction of the cartesian coordinates (x, y) to avoid the effects of gravity (water underride and gas override) which might reduce the amount of CO_2 dissolution inside the oil phase. CO_2SOL option (keyword) was used for this simulation. It was not possible to set the concentration of CO_2 in injected water stream by using CO_2SOL. As such, the carbonated water has been simulated using two injector wells placed at the same location where the first well injects free CO_2 while the second well injects water. The volumetric ratio was adjusted between the injected water and CO_2 to represent a CW with a mass fraction of 5% dissolved CO_2. The following table and figure show the reservoir description, injection rate and relative permeability curves (Table 1; Fig. 1).

The oil composition consists of Decane (C10) with a density of 43.6288 Ib/ft3 and viscosity of 0.2780 cp at the reservoir initial conditions. The total simulation time was about 8.3 years or 3000 days and no free CO_2 was present inside the reservoir during the simulation. The total oil recovery at the end of this simulation (base case) was found to be 68.4% and all the amount of injected CO_2 was stored safely inside both oil and water phase as CO_2 did not appear in the production stream. The following figures show the oil saturation, free gas saturation and mass fraction of CO_2 inside oil phase at the end of simulation (Figs. 2, 3 and 4).

Table 1 Details of the created ECLIPSE300 model

Number of grid blocks	$50 \times 50 \times 1$
Gridblock dimensions, ft3	$100 \times 100 \times 1$
Permeability, md (kx = ky)	500
Porosity, fraction	0.25
Reservoir temperature, F	212
Initial reservoir depth, psi	3500
Initial water saturation, fraction	0
Water injection rate, STB/day	4750
CO$_2$ injection rate, MSCF/day	2.305
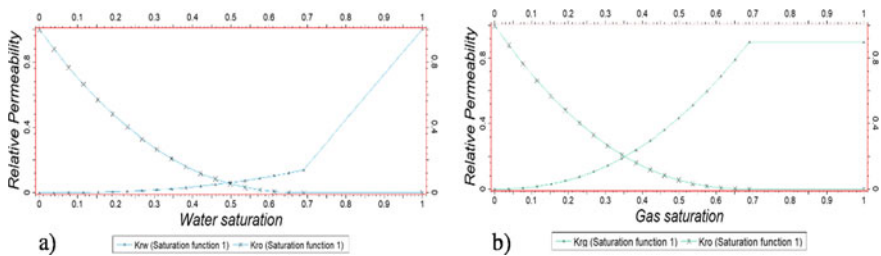
Fig. 1 a Oil water relative permeability and **b** gas oil relative permeability

Fig. 2 Oil saturation (C10) after 3000 days
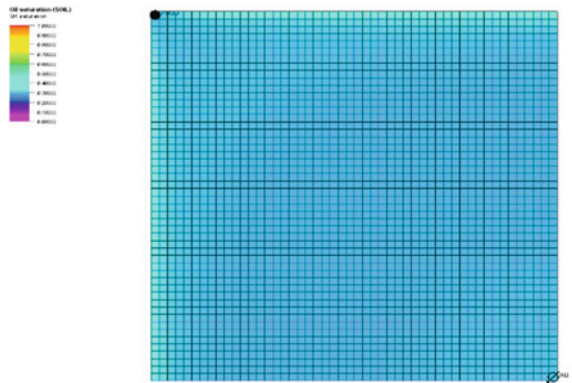
3.1 CW Injection Rate Sensitivity

A sensitivity analysis has been done on CW injection rate to investigate its impact on oil recovery, CO$_2$ storage and spatial spreading inside the reservoir. Three different injection rates have been studies and the mass fraction of dissolved CO$_2$ inside injected CW for each scenario was 5%. In the first scenario, a total of 2000 STB/day CW has been injected into the reservoir (1900 STB/day of water, 0.922 MSCF/day

Fig. 3 Free gas saturation after 3000 days
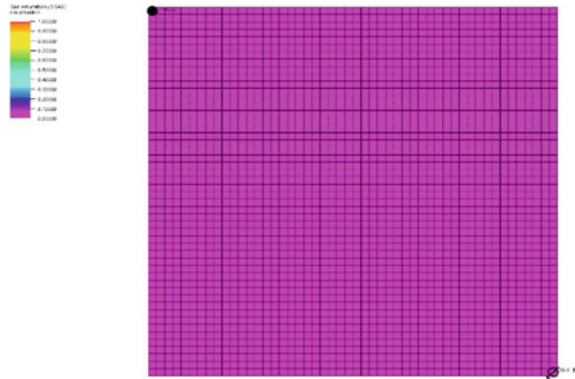
Fig. 4 Mass fraction of dissolved CO_2 inside the oil phase at the end of simulation
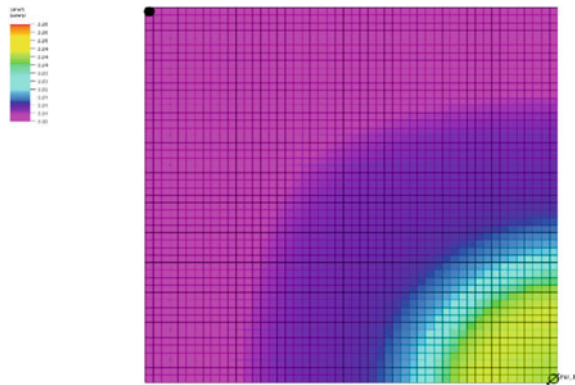
of free CO_2) for 7500 days, while in the second scenario a total of 5000 STB/day CW has been injected (4750 STB/day of water, 2.305 MSCF/day of free CO_2) for 3000 days, and in the last scenario 8000 STB/day of CW has been injected (7600 STB/day of water, 3.688 MSCF/day of free CO_2) for 1875 days.

It was found that the oil recovery factor for the injection rate of 2000 STB/day is the highest (68.6%) and the wide spatial spreading of CO_2 inside the reservoir led to early CO_2 breakthrough compared to the other scenarios thus less CO_2 storage (96%). This can be attributed to the long contact time between CW and oil phase which allows more CO_2 to be transfered based on the thermodynamic equilibrium calculation. While for scenario three, less oil recovery was found (58%) and the limited CO_2 spatial spreading inside the reservoir caused 100% CO_2 storage as CO_2 did not reach the producer well. The flowing figures show the oil recovery, CO_2 spatial spreading and CO_2 storage for the three scenarios (Figs. 5, 6 and 7).

Fig. 5 Oil recovery factor at the end of simulation for the three different injection rates of CW

■ 2000 STB/day   ■ 5000 STB/day   ■ 8000 STB/day

a)   b)   c)
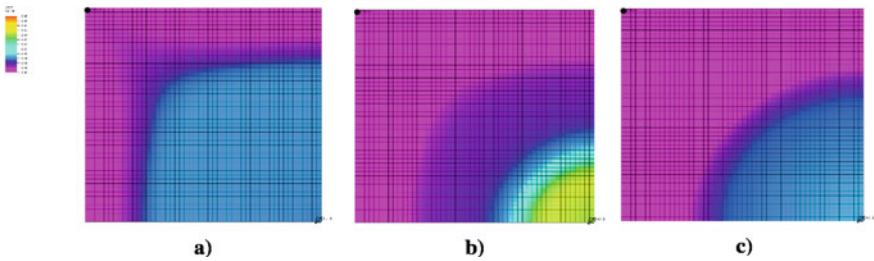
Fig. 6 Mass fraction of dissolved CO_2 inside the oil phase at the end of simulation for **a** 2000 STB/day, **b** 5000 STB/day, **c** 8000 STB/day
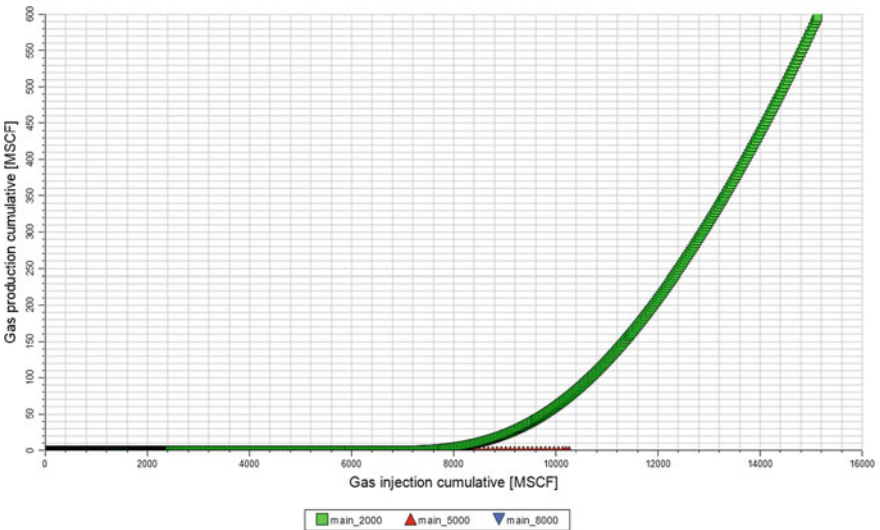
Fig. 7 Cumulative injected free CO_2 and cumulative produced CO_2 for the different scenarios of CW injection rate
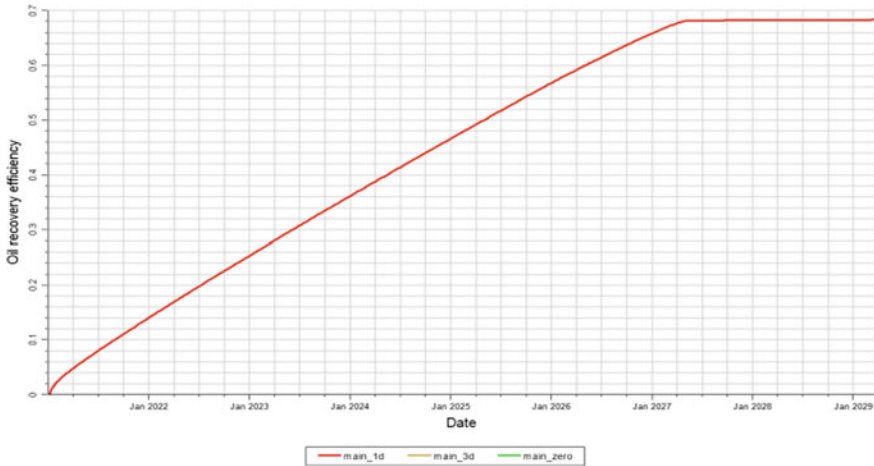
Fig. 8 Oil recovery vs time for the 3 scenarios of different CO_2-Oil diffusion coefficients

3.2 Effect of CO_2 Diffusion During CWI

As CO_2 gets transferred from CW to the oil phase, CO_2 molecules diffuse inside the oil due to the concentration gradient until the equilibrium state is reached. DIFFCOIL keyword added in PROPS section provides CO_2-Oil diffusion coefficient between adjacent grid cells. To study the effect of CO_2 diffusion, three scenarios have been simulated where in the first scenario the diffusion coefficient was set to zero while in the second and third scenario the diffusion coefficient was set to 0.007758103 ft2/day (1d) and 0.02327 ft2/day (3d) respectively. It was found that CO_2 diffusion has a minimal (negligible) impact on both oil recovery and CO_2 storage. The following figure shows the oil recovery for all the scenarios (68.4%) thus led to the same CO_2 storage (100% as CO_2 did not reach production well) (Fig. 8).

4 Conclusion

Carbonated water injection is a promising EOR technique that promotes high oil recovery and safe CO_2 storage. Upon the contact of Carbonated Water (CW) with oil inside the reservoir, CO_2 molecules cross the interface between the two fluids moving towards the oil phase due to the concentration gradient, and subsequently, CO_2 molecules diffuse inside the oil until thermodynamic equilibrium is reached. CWI was studied numerically using ECLIPSE300 where the CO_2SOL option was used to account for CO_2 dissolution inside the water phase. A sensitivity analysis on injection rate and CO_2 diffusion coefficient was done to understand their impact on oil recovery and CO_2 storage. It was found that low CW injection rates leads to

a higher oil recovery and a wider spatial spreading of CO_2 inside the oil thus a higher storage, while in the case of high injection rates, a less oil recovery was found and the limited spreading of CO_2 causes a lower CO_2 storage. Additionally, the effect of CO_2 diffusion was found to be minimal (negligible) on both oil recovery and CO_2 storage in this case study.

References

1. Mosavat N (2014) Utilization of Carbonated Water Injection (CWI) as a Means of Improved Oil Recovery in Light Oil Systems: Pore-Scale Mechanisms and Recovery Evaluation. University of Regina, Faculty of Graduate Studies and Research
2. Mosavat N, Torabi F (2014) Application of CO_2-saturated water flooding as a prospective safe CO_2 storage strategy. Energy Procedia 63:5408–5419
3. Miller, J.S. and R.A. Jones. A laboratory study to determine physical characteristics of heavy oil after CO_2 saturation. in SPE/DOE Enhanced Oil Recovery Symposium. 1981. Society of Petroleum Engineers.
4. Sohrabi M et al (2011) Safe storage of CO_2 together with improved oil recovery by CO_2-enriched water injection. Chem Eng Res Des 89(9):1865–1872
5. Sohrabi M et al. (2009) Mechanisms of oil recovery by carbonated water injection. in SCA annual meeting
6. Sohrabi M et al (2011) Carbonated water injection (CWI)–A productive way of using CO2 for oil recovery and CO2 storage. Energy Procedia 4:2192–2199
7. Sohrabi M et al (2012) Coreflooding studies to investigate the potential of carbonated water injection as an injection strategy for improved oil recovery and CO 2 storage. Transp Porous Media 91(1):101–121
8. Riazi M, Sohrabi M, Jamiolahmady M (2011) Experimental study of pore-scale mechanisms of carbonated water injection. Transp Porous Media 86(1):73–86
9. Kechut, N.I., M. Sohrabi, and M. Jamiolahmady. Experimental and numerical evaluation of carbonated water injection (CWI) for improved oil recovery and CO2 storage. in SPE EUROPEC/EAGE Annual Conference and Exhibition. 2011. Society of Petroleum Engineers.
10. Tavakolian, M., et al. Significant improvement in oil recovery and CO2 storage by carbonated water injection (CWI). in Third EAGE CO2 Geological Storage Workshop. 2012.
11. Yang D, Tontiwachwuthikul P, Gu Y (2005) Interfacial tensions of the crude oil+ reservoir brine+ CO2 systems at pressures up to 31 MPa and temperatures of 27 C and 58 C. J Chem Eng Data 50(4):1242–1249
12. Riazi, M., Pore scale mechanisms of carbonated water injection in oil reservoirs. 2011, Heriot-Watt University.
13. Mosavat N, Torabi F (2016) Micro-optical analysis of carbonated water injection in irregular and heterogeneous pore geometry. Fuel 175:191–201
14. Schlumberger, ECLIPSE Technical Description. 2014.
15. Schlumberger, ECLIPSE Reference Manual. 2014.
16. Whitson, C.H. and M.R. Brulé, Phase behavior. Vol. 20. 2000: Henry L. Doherty Memorial Fund of AIME, Society of Petroleum Engineers ….
17. Kechut NI, Jamiolahmady M, Sohrabi M (2011) Numerical simulation of experimental carbonated water injection (CWI) for improved oil recovery and CO2 storage. J Petrol Sci Eng 77(1):111–120

18. Kechut, N.I., et al. Tertiary oil recovery and CO2 sequestration by carbonated water injection (CWI). in SPE International Conference on CO2 Capture, Storage, and Utilization. 2010. Society of Petroleum Engineers.
19. Embid S, Rivas O (1994) Simulation of miscible displacement with interphase mass transfer resistance. SPE Advanced Technology Series 2(01):161–168
20. Mohebbinia, S. and T. Wong. Molecular diffusion calculations in simulation of gasfloods in fractured reservoirs. in SPE Reservoir Simulation Conference. 2017. Society of Petroleum Engineers.
21. Bear, J., Dynamics of fluids in porous media. 2013: Courier Corporation.
22. Peksa, A.E., K.-H.A. Wolf, and P.L. Zitha. Molecular Diffusion of CO From Carbonated Water (CW) Into the Oil-Experimental Observations. in SPE Asia Pacific Oil and Gas Conference and Exhibition. 2013. Society of Petroleum Engineers.
23. Alfarge D, Wei M, Bai B (2017) Factors affecting CO2-EOR in shale-oil reservoirs: numerical simulation study and pilot tests. Energy & Fsuels 31(8):8462–8480
24. Grogan A, Pinczewski W (1987) The role of molecular diffusion processes in tertiary CO2 flooding. J Petrol Technol 39(05):591–602
25. Foroozesh J, Jamiolahmady M (2016) Simulation of carbonated water injection coreflood experiments: an insight into the wettability effect. Fuel 184:581–589
26. Foroozesh J, Jamiolahmady M, Sohrabi M (2016) Mathematical modeling of carbonated water injection for EOR and CO2 storage with a focus on mass transfer kinetics. Fuel 174:325–332

Study of Efficient FIR Filter Architecture for Fixed and Reconfigurable Applications

B. Aruna Devi, V. Prasanna Devi, and S. Preethi

Abstract Transpose type Finite-Impulse Response (FIR) filters are implicitly pipelined and endorse the technique of Multifold Constant Multiplications (MCM). The studied model explores the possibility of understanding block FIR filters in the transpose form pattern for the efficient realization of broad order FIR filters for both fixed and reconfigurable applications in area-delay. A flow graph optimized registry complexity is built based on the computational study of transpose type configuration of the FIR filter. The proposed translate type block filter is an architecture based on a multiplier. Analysis is done towards the need of an optimum low complexity design using the MCM model for the block realization of fixed FIR filters. Since power dissipation is directly linked to the hardware, the use of MCM algorithms indirectly achieves a certain amount of power reduction. MCM is therefore largely found to be sufficient for the implementation of Broad Order design research. Using Data Flow Graph, which has been comprehensively iterated based on the ability to execute algorithms and their enhancement with a high degree of parallel and asynchronous operation, the computation process in MCM is likely to be well performed.

Keywords Reconfigurable architecture · FIR filter · Block processing · Finite-impulse response (FIR) filter · VLSI

1 Introduction

High Complexity in Field Programmable Gate Array (FPGA) is the major demerit for the program to execute in a running period which increases the delay duration, which makes the parameter of power to be high. The complexity in the number of slices used in Field Programmable Gate Array(FPGA) is increased in large numbers and gradually increases the storage memory space. Earlier, the duration period of high Look Up Table (LUT) were introduced, so that delay is increased in execution

B. Aruna Devi (✉) · V. P. Devi · S. Preethi
Department of Electronics and Communication Engineering, Dr. N.G.P Institute of Technology, Coimbatore, Tamil Nadu, India
e-mail: arunadevi@drngpit.ac.in

period of field arithmetic and one of the methods used is Distributed Arithmetic (DA). Here the multiplications are rearranged and for the explicit multiplications using Read Only Memory lookups, in analysis of FPGA based implementation, with advantages of reduction in memory in FPGA. The timing of the DA based FIR filter models is done using minimum clock periods depending on the delay of the final adder stage in the PSAT (Pipeline Shift-Add Tree).However, which depends on the critical path calculation of the adder. The throughput rate is the amount of output produced by the critical path over a period of time.

In this paper, the study of various models of Finite Impulse Response Filter with energy efficient architecture for reconfigurable applications is done. The paper presents a comparative study of optimum low complexity, reconfigurable digital FIR filter for low power with reduction in delay implementation.

2   Literature Survey

Many studies have been undergone towards FIR efficient implementation for various applications [1]. The implementation of filter aids in proposing various models for efficacy of power, delay, area and test bed analysis.

The key study deals with distributed FIR filter using the least mean block square algorithm using Look Up tables, reducing the number of flip flops and adder. It provides n/6 adders with reduction in area delay product, with block size tables less than four block size tables. The implementation of the parallel structure with weighted adaptation phrase, output filter computation and modified distributed arithmetic helps to reduce the logic and power consumption of the look up tables with less words to be updated per output. The algorithm includes decomposition into smaller vectors of both input and weight vector, and filtering analysis of the inner product computation [2].

Shah et al. [3] has implemented a competent FIR filter towards reconfigurable implementation for an coherent FIR filter for fixed and reconfigurable applications by implanting delay and area efficient addition based carry selector by optimizing the superfluous operations, by performing the carry process before the ultimate sum function analysis. The ability of the transpose form FIR filter implementation resulted in 26.6% of reduction in delay and area of 15% in reconfigurable models.

An architecture based on distributed arithmetic FIR filter for increased speed and less power consumption was proposed using block least mean square algorithm. The implementation design provided an efficiency of 17.3 times, lower power consumption of 17.3 times and 95% higher in terms of through output. The variable coefficient adaptive proposed filter units includes the following operations: Register, partial product computation, weight update and error generation, simultaneously with shift accumulator look-up tables, emulate multiply and accumulate functions. A systolic array architecture with parallel data processing units has been suggested for the architecture. The parallelism was found to be successful with the number of clock cycles reduced [4].

The least-square linear phase FIR filter (LLFE) is frequently used in the analysis of ECG signals. It aids in removal of low frequency mode using discrete wavelet transform.

Distributed Arithmetic (DA) is a technique that is bit-serial in nature. It can therefore appear to be slow. It turns out that when the number of elements in a vector is nearly the same as the word size, DA is faster in nature, reinstate the explicit multiplications overdone by read only memory look up table, by ROM look-ups and an effective model to realize and to implement on Field Programmable Gate Arrays (FPGAs). The coherent model is found effective for estimating of inner product or multiply and accumulate (MAC). In Digital Signal Processing algorithms, the MAC operation is popular. Multiplications are reordered and combined in DA so that the arithmetic becomes "distributed" rather than "lumped" through the structure. When DA is implemented in FPGAs, to execute the MAC operation, one can take advantage of memory in FPGAs. When DA was devised (in the 1970s), the comparisons given were in terms of number of TTL ICs required for mechanization of a certain type of filter. In particular, for an eighth order digital filter operating at a word rate close to 1 MHz, 72 ICs with a total power consumption of about 30 W was stated with the DA approach while 240 ICs with a power dissipation of 96 W was indicated for a multiplier-based solution [5].

3 Proposed Work

FIR filter transpose type block was suggested with decreased register complexity. A low-complexity design methodology that can be optimized for greater throughput in reconfigurable applications for the block realization of fixed FIR filters using the MCM scheme is proposed. It can be optimized for higher output in application that are reconfigurable. A low-complexity design approach for the block implementation of fixed FIR filters using the MCM scheme It can be configured for greater throughput in reconfigurable applications as shown in Fig. 1. Here presented an optimum low complexity design using the MCM scheme for the block implementation of fixed FIR filters. MCM technique minimizes the number of adders and subtractors. As a result the computation process is reduced. Since Power dissipation is directly related to the hardware, some amount of power reduction is indirectly achieved by using MCM
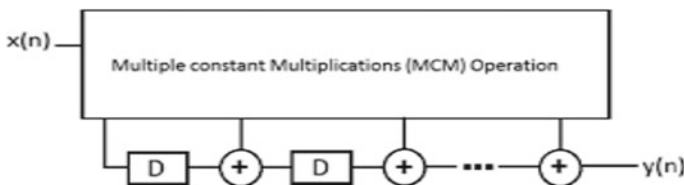
Fig. 1 Proposed model

algorithms. Therefore MCM is suitable for implementation of Large Order Filters with fixed coefficients. On simulation of serial in serial out register, merged data flow graph and the power analysis of the partial generator unit, the power consumption is 40.05 mW compared to traditional model to 42.8 mW. The area product delay is reduced around 13 % and energy per sample improvement is 12.8%. Hence in this method, the number of flip flops, adders and multipliers are reduced for large order filter which reduces the power and delay.

4 Conclusion

Finite Impulse Response (FIR) digital filter is extensively used in various applications ranging from biomedical applications, to speech processing in various domains of digital signal processing. Finite-Impulse Response (FIR) Reconfigurable Filter whose filter coefficients alter dynamically during runtime has an important role. The multiplier unit is the optical FIR filter's building structure in its transposed form. With the filter input realization, the multiplication filter coefficients can be introduced, having a major effect on the complexity and design efficiency. Hence calculation of inner product or multiply and accumulate can be done efficiently. Hence we have explored the possibility of realization of block FIR filter in transpose form configuration for area-delay efficient realization of large order FIR filters for fixed and reconfigurable applications.

References

1. Reddy KS, Suresh HN (2019) FPGA implementation of reconfigurable FIR filter using Vedic design with CLA adder. Int J Adv Sci Technol 28(1):144–161. Retrieved from http://sersc.org/journals/index.php/IJAST/article/view/228
2. Ahmed F (2019) The energy efficient fir filter for biomedical application using FPGA. Comput Rev J 3:236–244
3. Shah S, Rajula S (2019) Design of FIR filter architecture for fixed and reconfigurable applications using highly efficient carry select adder. Adv Intell Syst Comput 898:627–637 (2019)
4. Chowdari CP, Seventline JB (2020) Systolic architecture for adaptive block FIR filter for throughput using distributed arithmetic. Int J Speech Technol 23:549–557 (2020). https://doi.org/10.1007/s10772-020-09745-4
5. Mohanty BK, Kumar P (2013) A high-performance energy-efficient architecture for FIR adaptive filter based on new distributed arithmetic formulation of block LMS algorithm. IEEE Trans Signal Process 61(4):921–932

Improved Deep Learning Based Prediction of Crop Yield Using Bidirectional Long Short Term Memory

V. Saveetha, R. Kingsy Grace, S. Sophia, and S. Birundha

Abstract Yield estimation for crops is an important problem to be addressed with respect to agricultural planning. Deep learning and machine learning provide feasible solutions for the above issue. In the current research perspective yield prediction is vastly addressed and more techniques evolve in the objective to improve accuracy. This research work tries to analyze the predictive accuracy of machine learning methods along with deep learning methods. The models are analyzed by metrics such as the Mean absolute percentage error (MAPE), root-mean-square error (RMSE), mean absolute error (MAE), and accuracy. The proposed Bidirectional LSTM deep learning technique exhibits high accuracy and low error rates in order to predict rice yield in INDIA The results show that Bi-LSTM have better accuracy when compared to other models, Bi-LSTM and XGB obtain the lowest RMSE errors, the lowest MAPE error percentage is for Bi-LSTM and the lowest MAE errors is also for Bi-LSTM. Since Bi-LSTM exhibits crop yield model with better accuracy and lowest errors it is suitable massive crop yield prediction in agricultural planning.

Keywords Artificial Intelligence · Neural networks · Deep learning · Recurrent neural networks · Long short term memory

1 Introduction

Crop yield prediction (CYP), a challenging task, benefits farmers in order to take managerial decisions and financial stands [1]. CYP is challenging because of factors like environmental and field based features. Also, the food production of the world is

---

V. Saveetha (✉)
Dr. N.G.P. Institute of Technology, Coimbatore, India
e-mail: saveetha@drngpit.ac.in

R. Kingsy Grace · S. Birundha
Sri Ramakrishna Engineering College, Coimbatore, India

S. Sophia
Sri Krishna College of Engineering and Technology, Coimbatore, India

dependent on CYP for precise forecast which leads to timely decisions [2]. Machine learning (ML), that act has a stem of Artificial Intelligence, has been implemented for agricultural prediction enormously in literature [3] due to its behavior of treating outputs with respect to inherent function of the input features. Precision Agriculture, the field of CYP manages to monitor measure and produce optimized outputs with response to variability in crops [4]. Current research contributes vastly to estimate the CYP justifiably but always a better performance is preferred [5]. Recently ML techniques like Clustering, Classification, Regression and Neural Networks are employed for CYP. In our study data of crop yield and other factors including statistics and Deep Neural Networks (DNN) enables to produce effectual forecasting of yield. DNN with hidden layers act to be more powerful to reveal the fundamental nonlinear relationship between input and response variables [6]. In this paper Recurrent Neural Networks (RNN) [1] is used to predict and check crop yield. RNN works with multiple stacked layers that convert the raw data into meaningful representations. RNN are well suited for modeling functions for which the data is composed of vectors that involve a time dependency between the values like CYP [7]. It includes a feedback loop that it uses to learn from sequences, including sequences of varying lengths. The long short-term memory (LSTM) type of RNN has received enormous attention in the realm of sequence learning. LSTM a popular variant of RNN has internal memory to allow long-term dependencies to affect the output [8]. The LSTM is good for CYP because it is designed to confine to the time dependencies of environmental factors, it enables capability to predict crop yield without drop in accuracy and combined with the bi directional method, the model predicts effectively.

This research aims to evaluate methods to detect rice cultivation in India, using the LSTM and Bidirectional LSTM (Bi-LSTM) methods. The study compared the results based on LSTM and Bi-LSTM with traditional machine learning techniques: Support Vector Machines (SVM), Random Forest (RF) and Gradient Boost(XGB) Regression. The rest of the paper is structured as Sect. 2 describes related work of various models used for CYP. Section 3 presents the proposed methodology. Section 4 discusses the results and finally, concluding remarks is discussed in Sect. 5.

2   Review

Many methods exist in literature for CYP and every time an attempt is performed by researchers to enhance the performance of prediction. In this paper an ensemble based classification algorithm using Artificial Neural Networks (ANN), Decision Tree and SVM is designed and implemented to predict crop yield. They try to increase the accuracy of classification. The authors modeled an learning technique based on AdaBoost ensemble algorithm and enhancement is performed using genetic optimization technique [9]. Khaki and Wang suggested a DNN model for CYP across from 2008 to 2016 which was analysed along with methods like Lasso, shallow neural networks, and decsion tree [10]. You et al. designed a hybrid model with Convolution Neural Network(CNN) and RNN to predict soybean yield [11]. Kim

et al. suggested a DNN model for CYP using optimized input features from satellite products and meteorological datasets from 2006 to 2015 [12]. Wang et al. [13] developed a Deep Learning(DL) framework to predict soybean crop yields in Argentina and achieved good results with a transfer learning approach to predict Brazil soybean harvests. Yang et al. (2019) analyzed the working of CNN to predict rice grain yield on images and concluded that CNN model gives robust yield forecast through the ripening stage [14]. This study shares a model to perform soybean yield forecasts in Brazil using LSTM. They compare performance of linear regression, random forest and LSTM for forecasting yield and evaluate accuracy [15]. The authors propose a CNN-LSTM model for soybean yield prediction at the county-level. The model was evaluated by crop growth variables and environment variables weather data. They show the results of the experiment that performance of the proposed deep CNN-LSTM model outperforms the CNN or LSTM model [16]. The above review instigates that Deep learning can be used for CYP.

3   Methodology

Decision making in CYP require crop response models that make accurate estimations to predict yields [17]. The purpose of the study is to find out whether RNN can effectually predict crop yield for varying seasons, assess model performance and compare the same with other models. The machine learning algorithms already examined with respect to DNN are LSTM and CNN [8]. CNN is a specific class of neural networks mostly used in image recognition. In this work Bi directional LSTM is proposed for CYP in order to increase the performance accuracy. A comparison of machine learning models and feature levels was performed based on the yield estimation. Models were trained to predict yields from 1997 to 2015 for each state. Model performance was evaluated using Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and Accuracy. The input variables for the prediction model include the Crop yield year, Climatic data and Area Cultivated. The yield prediction is based on the equation y = f(x), where y is the yield predicted and x is the input data with respect to time.

Figure 1 shows the flow of the research performed. Initially Data Acquisition is performed and the dataset obtained from [18] is explored for information following which the data related to rice is extracted. Data preprocessing like encoding categorical variables, filling missing values and scaling are performed on the derived dataset. The machine learning model and deep learning model are implemented and the accuracy and error metrics for the same are analyzed.

LSTM works exclusively for prediction of sequence problems. The different architectures of LSTM are LSTM, CNN-LSTM, Encoder-Decoder LSTM, Generative LSTM and Bi-LSTM. LSTM a RNN has an internal state, models parallel input sequence, process variable-length input to generate variable-length output (Fig. 2). The memory cell is the computational unit of the LSTM and consists of weights and gates.
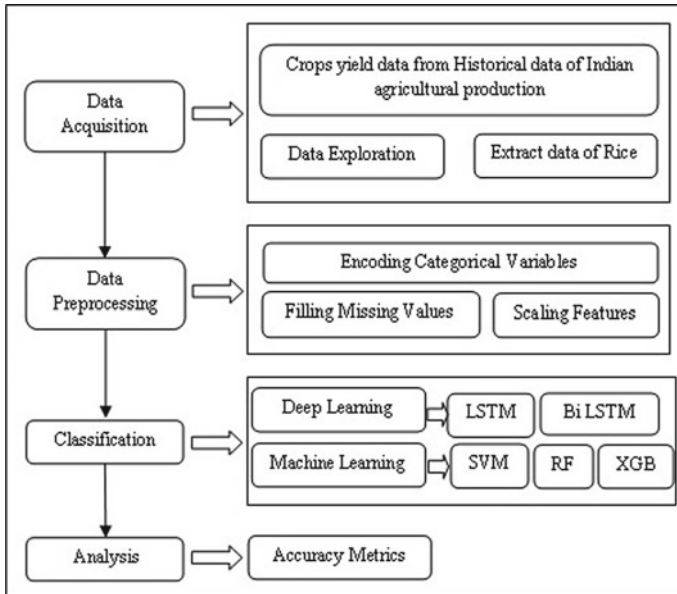
Fig. 1 Methodology used
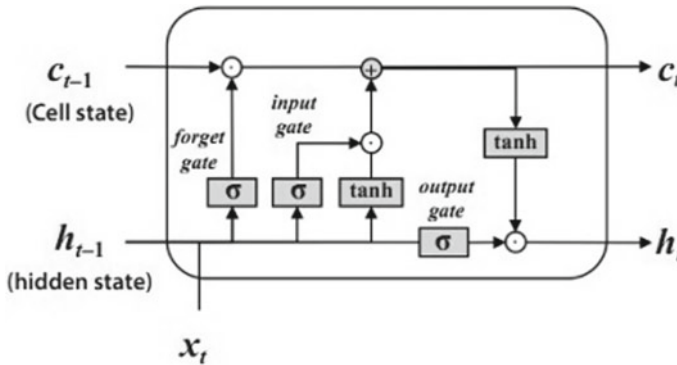
Fig. 2 LSTM architecture

Huang et al. [19] The objective of a Bidirectional LSTM is to make forward and backward passes on the features of CYP in order to predict the yield. The past and future features for a given time help to process the given sequence and thus provide efficient implementation of the BDLSTM model. This model is trained using back propagation through time (BPTT) [20] and the passes front and back are carried over by unfolding the hidden states for all time steps. In fields like phoneme classification, traffic prediction and speech recognition Bi-LSTM model are significantly better than unidirectional ones and so an effort is made to use it in CYP [21]. The Bi-LSTM
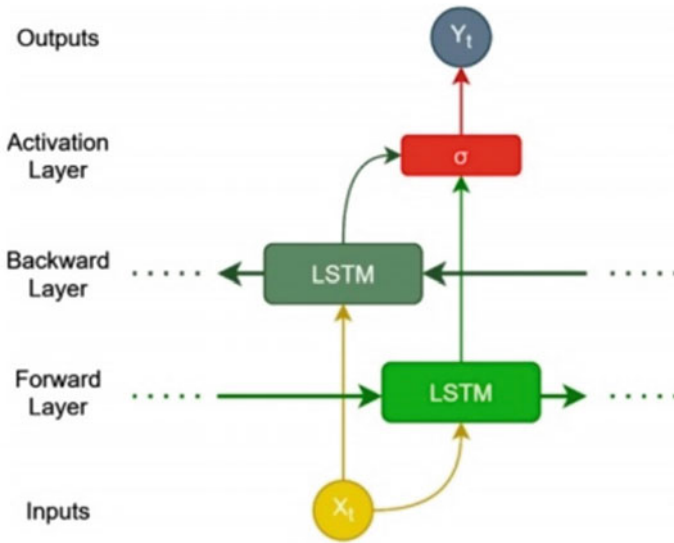
Fig. 3 Bidirectional LSTM

model (Fig. 3) generate an output Y_t which is calculated by $Y_t = \sigma(h_{ft}, h_{bt})$ where σ function combines the two output sequences from forward pass and backward pass.

4 Result Discussion

4.1 Data Set

Historical data of Indian agricultural production on various states and districts acquired from the Indian government web page available in [18] is used. This dataset is utilized to investigate and analyze production contribution to particular land area. Dataset used is a Comma Separated Values file. The dataset contained observed production for rice and other crops between 1997 and 2015 across 646 districts from 33 states.

4.2 Metrics Used

The following indexes are selected to evaluate the forecast model performance: Accuracy, Mean absolute percentage error (MAPE), root-mean-square error (RMSE) and mean absolute error (MAE) [9]. The MAPE is a measure of prediction accuracy in statistics and is the most common measure used to forecast error, and works best if

Fig. 4 Actual versus predicted trend
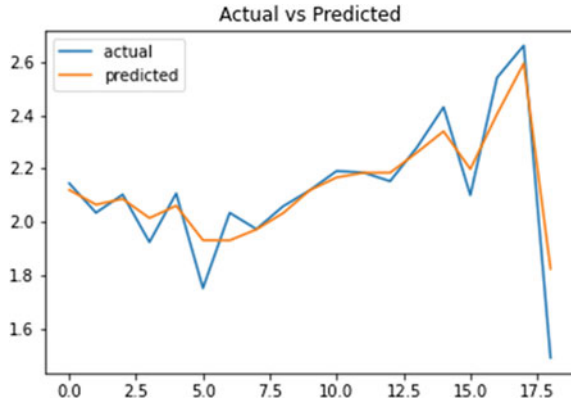
Table 1 Formula of performance metrics

Performance metrics	Formula
Accuracy	$Accuracy = \frac{P-A}{P}$
MAPE	$MAPE = \frac{1}{N} \sum\limits_{k=1}^{N} \left| \frac{A_k - P_k}{A_k} \right|$
RMSE	$RMSE = \sqrt{\frac{1}{N} \sum\limits_{k=1}^{N} (|A_k - P_k|)^2}$
MAE	$MAE = \frac{1}{N} \sum\limits_{k=1}^{N} |A_k - P_k|$

A Actual value, *P* Predicted value

there are no extremes to the data (and no zeros). RMSE is the standard deviation of the residuals errors. Residuals tells how concentrated the data is around the line of best fit. The RMSE is a measure of the differences between values predicted by a model and the values actually observed. It is used to compare forecasting errors of different models for a particular variable and not between variables, as it is scale-dependent. The values for the above metrics are obtained from Actual Values and Predicted Values of yield after training and testing the dataset. The sample Actual versus Predicted trend is shown in Fig. 4. Table 1 shows the formula employed to find the metrics.

4.3 Comparison

Table 2 shows the performance metrics values for the models which are compared with Bi-LSTM. The accuracy graph shows good percentage for Bi-LSTM and LSTM. The MAPE percentage error is minimum for Bi-LSTM. The RMSE error values

Table 2 Performance metrics

Model	Accuracy %	MAPE %	RMSE	MAE
SVM	88	10.6	0.4601	9.54
RF	90.8	9.02	0.3280	2.98
XGB	90.5	9.60	0.2500	2.65
LSTM	91.6	8.6	0.3350	2.42
Bi-LSTM	93.0	7.8	0.2450	2.11
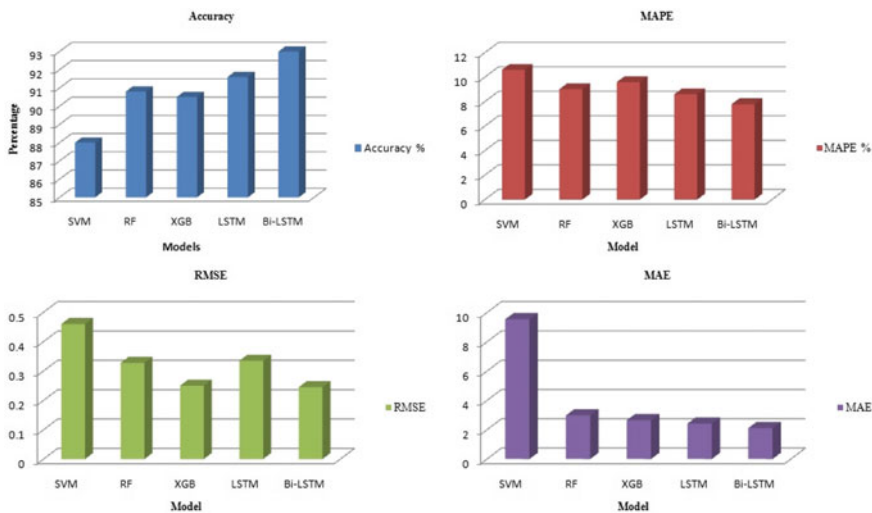
Fig. 5 Performance metrics

are good for both XGB and Bi-LSTM. The MAE values are good for Bi-LSTM, LSTM, XGB and RF. The results show that Bi-LSTM is better in predicting CYP when compared with other methods with respect to accuracy and minimized error metrics which is shown in Fig. 5.

5 Conclusion

This study analyzed the performance of crop yield prediction using different models like SVM, RF, XGB, LSTM and Bi-LSTM of which four are machine learning models and two are Deep Learning models. The yield prediction for test set is reasonable high in accuracy. The Bi-LSTM model outperforms other models both in accuracy and stability. The Bi-LSTM computationally scales better than SVM, RF and XGB. These results contribute to a comprehensive understanding of the response of rice

yield to variables and the comparative performance of crop yield forecasts across different types of machine learning algorithms. Further improvement of the yield prediction modeling could benefit from more explicit information on crop type. The proposed method shows great potential in improving the accuracy of yield prediction for other crops like corn, wheat, and potatoes at fine scales in the future.

References

1. Saeed K, Wang L (2019) Crop yield prediction using deep neural networks. Front Plant Sci 10
2. Horie T, Yajima M, Nakagawa H (1992) Yield forecasting. Agric Syst 40(1–3):211–236
3. Van Klompenburg T, Kassahun A, Catal C (2020) Crop yield prediction using machine learning: a systematic literature review. Comput Electron Agricult 177:105709
4. Patrício DI, Rieder R (2018) Computer vision and artificial intelligence in precision agriculture for grain crops: a systematic review. Comput Electron Agric 153:69–81
5. Filippi P, Jones EJ, Wimalathunge NS, Somarathna PD, Pozza LE, Ugbaje SU, Jephcott TG, Paterson SE, Whelan BM, Bishop TF (2019) An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. Precision Agric 20(5):1015–1029
6. Yang X, Sun M (2019) A survey on deep learning in crop planting. In: IOP conference series: materials science and engineering, vol 490, no 6. IOP Publishing, p 062053
7. Khaki S, Wang L, Archontoulis SV (2020) A cnn-rnn framework for crop yield prediction. Front Plant Sci 10:1750
8. Kang Y, Ozdogan M, Zhu X, Ye Z, Hain C, Anderson M (2020) Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest. Environ Res Lett 15(6):064005
9. Koduri SB, Gunisetti L, Ramesh CR, Mutyalu KV, Ganesh D (2019) Prediction of crop production using adaboost regression method. J Phys Conf Ser 1228(1):012005. IOP Publishing
10. Khaki S, Khalilzadeh Z, Wang L (2019) Classification of crop tolerance to heat and drought—a deep convolutional neural networks approach. Agronomy 9(12):833
11. You J, Li X, Low M, Lobell D, Ermon S (2017) Deep gaussian process for crop yield prediction based on remote sensing data. In: 31rst AAAI conference on artificial intelligence
12. Kim N, Ha KJ, Park NW, Cho J, Hong S, Lee YW (2019) A comparison between major artificial intelligence models for crop yield prediction: case study of the midwestern united states, 2006–2015. ISPRS Int J Geo-Inf 8(5):240
13. Wang AX, Tran C, Desai N, Lobell D, Ermon S (2018) Deep transfer learning for crop yield prediction with remote sensing data. In: Proceedings of the 1st ACM SIGCAS conference on computing and sustainable societies, pp 1–5
14. Yang Q, Shi L, Han J, Zha Y, Zhu P (2019) Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. Field Crop Res 235:142–153
15. Schwalbert RA, Amado T, Corassa G, Pott LP, Prasad PV, Ciampitti IA (2020) Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. Agricult Forest Meteorol 284:107886
16. Sun J, Di L, Sun Z, Shen Y, Lai Z (2019) County-level soybean yield prediction using deep CNN-LSTM model. Sensors 19(20):4363
17. Ji B, Sun Y, Yang S, Wan J (2007) Artificial neural networks for rice yield prediction in mountainous regions. J Agricult Sci 145(3):249
18. Indian Agricultural Production. https://data.gov.in
19. Huang Z, Xu W, Yu K (2015) Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv:1508.01991

20. Cui Z, Ke R, Pu Z, Wang Y (2020) Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values. arXiv preprint arXiv:2005.11627
21. Bolton DK, Friedl MA (2013) Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agric For Meteorol 173:74–84

Instrument Cluster IoT Enclosure Design and Production Implementation in Self Driven Vehicles

R. Thirumalai , M. Seenivasan, A. Sivakumar , M. Vivekraj, and V. Balaji

Abstract MNC's has introduced a Self-driven two-wheeler in the metropolitan cities, so that vehicle consists of numerous electronics and the major patented is self-control through smartphones. The most important electronic PCB board is embedded in an enclosure. The industry develops the enclosure design and production implementation with the help of Solid works, the 3D enclosure was designed into tow sub-assemblies, one is cover part and another one is the PCB carrier. The material for enclosure was selected as per IPC standards. (PA66 30%GF). The enclosure was also designed as self-locking even without screw. The fasteners (TWIN FAST SCEW) were selected as per the IPC standards. The metal to non-metal torque rate was analyzed and selected as per requirements. The mechanical endurance such as Hardness test and Salt Spray Test were conducted for the selection of screw. The enclosure was produced using Hot injection molding process.

R. Thirumalai (✉) · V. Balaji
Department of Mechanical Engineering, Dr N G P Institute of Technology, Coimbatore, Tamil Nadu, India
e-mail: thirumalai@drngpit.ac.in

M. Seenivasan
Department of Mechanical Engineering, KGISL Institute of Technology, Coimbatore, Tamil Nadu, India

A. Sivakumar
Department of Mechanical Engineering, Kongu Engineering College, Perundurai, Tamil Nadu, India

M. Vivekraj
Department of Mechanical Engineering, Research Scholar, Anna University, Chennai, Tamil Nadu, India

1 Introduction

The bounce vehicle principle is based on IoT, the major controls of the vehicle depends on the IoT. From the powering ON to Powering OFF everything is controlled through IoT. So the major essential electronic parts are being embedded in the instrument cluster enclosure, that is designed and produced by Squraesri technologies. IoT security: Review, blockchain solutions, and open challenges are discussed [1, 2]. Modeling of information processing in the internet of things at agricultural enterprises and solar powered buildings are more successful [3, 4]. Design and implementation of vehicle navigation system in urban environments, automatic garage doors and weather stations using internet of things (IoT) are becoming familiar nowadays [5–7]. Managing IoT devices using blockchain platform is developed in many countries [8, 9]. Secure integration of IoT and cloud computing is described with several case studies [10]. IoT middleware survey on issues and enabling technologies is demonstrated [11]. Internet of Things (IoT) is the network of physical objects embedded with sensors, and computers which will helpful in exchange of data and information with several systems. There are too many applications with IoT available such as; Automobiles, Heart monitoring implants, Bio-chips etc., IoT consists of web enabled devices to collect and send data across the environment using sensors and processors. A machine to machine communication acts on the information they get from each other. The lifecycle of IoT is to collect, communicate, analyze and act on the data information. The sensors are useful in collecting the information in the manufacturing plant and a cloud platform is created to send the data and events through a network used in the industry. The analyzing of the data includes filtering the data, building reports and visualizing the data. The final lifecycle of IoT—Act is useful in communication to the other machine and send successful notifications to the other systems effectively and efficiently. IoT works with RFID, Sensors, Smart technology and Nano Technology. The applications of IoT is wider in the areas of Medical and Healthcare systems, Manufacturing, Building and home automations, Transportation, Energy management, Environmental monitoring and so on. Also the IoT faces several challenges; technological standardization, complexity in the software, wireless communication, power supply, data volumes and interpretation and so on. IoT provides dynamic control of the industry and daily life, integrates human society and physical systems, improves the resource utilization ratio, flexible configuration, technological integrator.

2 Methodology

The electronic PCB board dimensions were taken into considerations with keeping all the electronic components embedded into it. After that the area in vehicle which is allocated for the installation of the enclosure is inspected. An 2D rough draft or sketch were drawn of the enclosure in simple drawing sheet with all necessary dimensions

that were taken before. The 3D design is done with keeping 2D sketch as a reference, in Solidworks software. When the deigning stage is over, the material for production is selected. Preceding the screw for fastener as well as the manufacturing method were selected as per the IPC standards. After all the above steps, the production implementation of enclosure is done by the hot injection moulding method. Then finally all the electronics are embedded into the enclosure and installed in the vehicle. The process flow chart methodology is shown in Fig. 1.
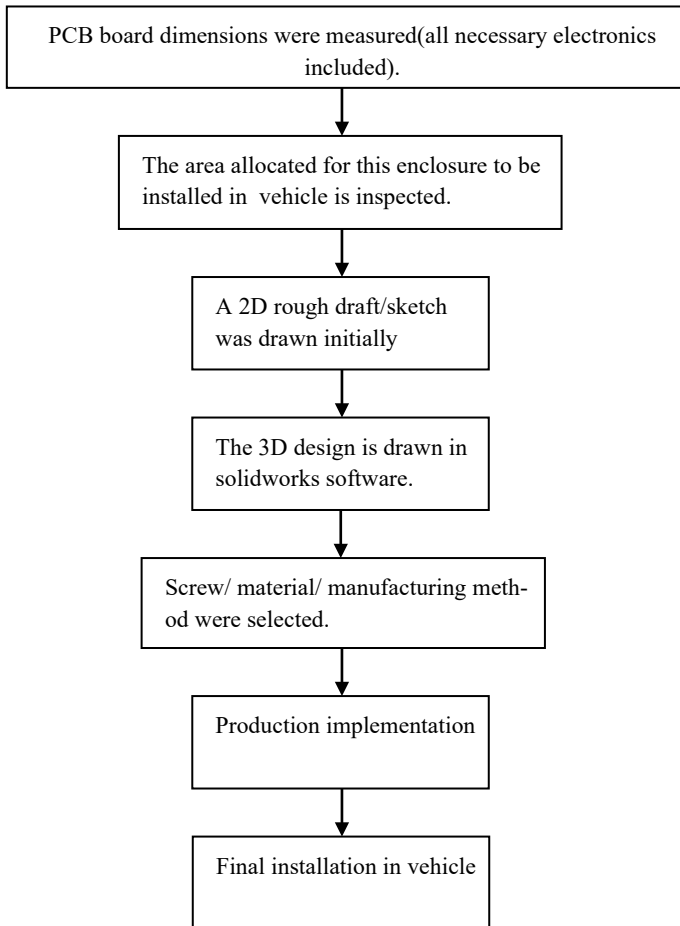
Fig. 1 Methodology

3 Overview of the Development of Enclosure in Solidworks Software

While designing the 3D sketch in solidworks software, the 2D rough draft is kept as the reference which was taken from various parameters while drawing. As earlier mentioned, the enclosure was divided into two parts, and each part was designed separately and then assembled together. So while designing in software several commands were used to design the enclosure. One of the important step in the design process is selection of the materials. This systematic selection of the materials will give the minimization of the cost during the production process. The design engineer should be aware of the properties of the material and behavior of the materials under several working conditions. The thermal properties, machinability properties, strength, hardness and other mechanical properties are to be well known the designer. The material used for the production of the enclosure is PA66 30% GF. PA66 GF30 is a 30% glass fibre reinforced PA 66(Poly Amide).Poly (hexamethyleneadipamide). Polyamide 66, or Nylon 66 (**PA66**) is an engineering-grade thermoplastic.

PA66 provides lower impact resistance when compared to PA6. High strength, creep strength, toughness are the mechanical properties which are the outstanding properties of PA66. PA66 is a glass fibre reinforced about 30 percentage to form PA66 GF30. Glass fibres exhibit high static load withstanding properties at elevated temperatures.

4 Results and Discussion

MNC's has introduced a Self-driven two-wheeler in the metropolitan cities, so that vehicle consists of numerous electronics and the major patented is self-control through smartphones. The most important electronic PCB board is embedded in an enclosure. The industry develops the enclosure design and production implementation with the help of Solid works, the 3D enclosure was designed into tow sub-assemblies, one is cover part and another one is the PCB carrier. The material for enclosure was selected as per IPC standards. (PA66 30%GF). The enclosure was also designed as self-locking even without screw. The fasteners (TWIN FAST SCEW) were selected as per the IPC standards. The metal to non-metal torque rate was analyzed and selected as per requirements. The mechanical endurance such as Hardness test and Salt Spray Test were conducted for the selection of screw. The enclosure was produced using Hot injection molding process. The material selected is tested for hardness and salt spray test.

4.1 Hardness Test

Hardness is the mechanical property which resists deformation which is plastic in nature when induced by mechanical abrasion. Materials are basically stronger and harder than plastics and wood. Strong intermolecular bonds are generally present in harder materials and therefore the behavior of solid material under force gives a tough behavior under variable loads. The material selected for the enclosure design provides high hardness and it is found satisfactory.

4.2 Salt Spray Test

For testing the corrosion resistance of the material, salt spray test is performed. The materials tested for salt spray test are metallic in nature. These materials are also finished with a surface coating for providing better test results. The material that is to be tested provides a corrosion protection to the undercovering metal. Corrosive attacks are produced in the Salt spray testing method. This salt spray test will evaluate the correctness of the coated material for further usage. The viscosity is very sensitive to temperature. Shrinkage is of the order of **0.01–0.02 mm/mm [1–2%]**. The addition of reinforcing glass fibers reduces the shrinkage to **0.2–1%.** Differential shrinkage in the flow and cross-flow directions is quite high. Mineral fillers yield more isotropic moldings. PA66 is resistant to most solvents, but not to strong acids or oxidizing agents. The material selected for the enclosure design provides sufficient success rate and hence it is found satisfied.

5 Conclusion

The IoT consist of four layers namely; sensor layer, Gateway and Network layer, Management service layer and application layer. Sensor layer collects and process the real time information. Gateway and network layer allows the organizations to share the information's through the robust networks available in the industry. Management service layer ensures security and privacy of data. Application layer provides a user interface for the IoT. The enclosure for electronics to be embedded in it has been successfully designed and produced by injection molding process. The enclosure has been installed in vehicle without causing any interruption to other parts and system in electric vehicle. Hence, the enclosure was designed initially from 2D sketch to 3D model by using Solidworks software, followed by selecting appropriate material for production and also the fasteners as well and finally fabricated using injection moulding process.

References

1. Khan MA, Salah K (2018) IoT security: review, blockchain solutions, and open challenges. Fut Gener Comput Syst 82:395–411
2. Bagas AD, Budi WA (2018) IoT-based integrated home security and monitoring system. In: IOP conferences series: journal of physics: conference series, vol 1140
3. Lvovich IY, Lvovich YE, Preobrazhenskiy AP, Preobrazhenskiy YP, Choporov ON (2009) Modeling of information processing in the internet of things at agricultural enterprises. IOP Conf Ser Earth Environ Sci 315(3):032029
4. Handayani TP, Hulukati SA, Jaya R, Tiandho Y, Abdullah R (2019) The prototype of solar-powered building lighting IoT. IOP Conf Ser Mater Sci Eng 486(1):012079
5. Godavarthi B, Nalajala P, Ganapuram V (2017) Design and implementation of vehicle navigation system in urban environments using internet of things (IoT). IOP Conf Ser Mater Sci Eng 225:012262
6. Jianyun C, Yunfan S, Chunyan L (2017) Research on application of automatic weather station based on internet of things. IOP Conf Ser Earth Environ Sci 104(1):012015
7. Uriawan W, Zulfikar WB, Sofa RM, Ramdhani MA (2018) Internet of things for automatic garage doors using {ESP} 8266 module. IOP Conf Ser Mater Sci Eng 434:12057
8. Huh S, Cho S, Kim S (2017) Managing IoT devices using blockchain platform. In: International conference on advanced communication technology (ICACT), IEEE, pp 464–467
9. Khan MA, Salah K (2018) IoT security: review, blockchain solutions, and open challenges. Futur Gener Comput Syst 82:395–411
10. Stergiou C, Psannis KE, Kim BG, Gupta B (2018) Secure integration of IoT and cloud computing. Futur Gener Comput Syst 78:964–975
11. Ngu AH, Gutierrez M, Metsis V, Nepal S, Sheng QZ (2016) IoT middleware: a survey on issues and enabling technologies. IEEE Internet Things J 4(1):1–20

General Adversarial Networks: A Tool to Detect the Novel Coronavirus from CT Scans

R. Shriram, T. R. Kaushek Kumar, V. Samuktha, and R. Karthika

Abstract Detection of the novel Corona virus in the early stages is crucial, since no known vaccines exist. Artificial Intelligence- aided prognosis using CT scans can be used as an effective method to identify symptoms of the virus and can thus significantly reduce the workload on the radiologists, who have to perform this task using their eyes. Among the most widely used deep learning convolutional neural networks, research shows that the Xception, Inception and the ResNet50 provide the best accuracy in detecting Covid-19. This paper proposes that using General Adversarial Network (GAN) as a data augmentation technique, in combination with these models will significantly improve the accuracy and thereby increase the chances of detecting the same. The paper also compares and contrasts how each of the three GANs namely DCGAN, LSGAN, CoGAN, perform in association with the afore-mentioned models. The main aim of this paper is to determine the most credible GAN network to carry out the task of data augmentation as well to prove that involving GANs would improve the existing accuracy of our model, paving way for an effective approach to train the model.

1 Introduction

The giant tide of SARS-CoV-2 virus, widely known as Covid-19 or novel corona virus has swept off all of mankind in one fell swoop and with no known vaccine,

---

R. Shriram · T. R. K. Kumar · V. Samuktha · R. Karthika (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: r_karthika@cb.amrita.edu

rapid detection and diagnosis of COVID-19 is of utmost importance in the efforts to improve the chances for a quick recovery. Detection usually involves taking samples from the nose, throat and lungs or carrying out a blood test. The standard test for infection is the Reverse transcription- Polymerase chain reaction (RT-PCR) test and takes up to two days for results to be received.

There are two major problems with this method-the finite number of detection kits available and valuable time lost to obtain the results.

Fortunately, tools aided by artificial intelligence (AI) can help in detecting novel corona virus outbreaks. Although these tools are being used in hospitals to detect mild cases, it is highly unlikely that these AI tools will replace RT-PCR as the primary diagnostic tool. A nucleic acid test can be carried out by taking a swab in any isolated location as per their convenience, while a chest scan has to be performed in an enclosed space with staff nearby. Also, most of these tools offer an accuracy of around 96% [1] which, given the sheer amount of COVID positive cases, is not good enough. In a bid to improve the accuracy, we can incorporate more such deep learning tools and techniques.

One such technique is the use of GAN's (General Adversarial Networks) to better train state-of-the-art models like Xception [2], Inception [3] and ResNet-50, culminating in better accuracy. GAN is a class of deep learning framework that incorporates neural networks named generator and discriminator, contesting with each other in a game-like situation. to add images to the dataset which are as superficially authentic as possible. This leads to improvement in accuracy. GANs will be dealt with in detail in Sect. 21.1. Section 2 deals with related work that have been carried out, Sect. 2 elaborates on the different GAN architectures dealt with in this paper, followed by the experiment performed and then the paper is concluded with the result and analysis.

2 Related Work

A plethora of Deep Neural Networks have been designed to aid in the process of detection since every new model may train a specific dataset in a better way. Some of them include using a light CNN architecture based on SqueezeNet [4] for classification of CT scans that averages a computation time of just 7.81 s. Abdul Waheed et al. [5] proposed an ACGAN (Auxiliary Classifier Generative Adversarial Network) based GAN, called CovidGAN to generate synthetic Chest X-Ray images for data augmentation the training sets along with a CNN model for better detecting COVID-19. Another Deep Learning framework called the CovNet was proposed to identify Covid-19 and also distinguish it from other Community-acquired pneumonia (CAP) and non-pneumonic lung diseases that it can be confused with [6].

Various regression models including Lasso Regression model, Exponential Smoothing model, etc. have been used to predict the number of COVID patients to be anticipated in the future [7]. But these tools perform well only when they have sufficient usable information, accumulating which is not easy. CNN models have been used not only in classifying CT scans, but are also used in other domains such

as labelled Waste management [8] and detecting Breast cancer from mammographic images by using the retrieved similar cases as reference [9].

Not only CNN's but also VGGNet (which is a significantly more accurate CNN [10]) has been used for classification purposes. Since 4D imaging is significantly better in detecting diseases in motion-reliant organs like heart, lungs, etc. than 3D, a variety of synchronization mechanisms including wavelet de-noising is used to reconstruct the said 4D image[11].

3 GAN Architectures

All GAN architectures have a unique generator portion or a discriminator portion, connected in that order.

3.1 GAN

GANs are basically neural networks that are trained in an adversarial manner to generate a probability distribution of data that mimics the original probability distribution. It consists of two models namely, the generator and the discriminator. Both these models compete with each other with an objective of mimicking the original distribution. The discriminator model performs the classification of whether a given image is fake (labelled 0) or real (labelled 1). The generator generates images to fool the discriminator by synthesizing artificial images that have a close resemblance to the real ones, making it difficult for the discriminator to distinguish the fake (the one generated from generator) from the real which is sampled from the true dataset. In essence, the generative network is trained to maximize the final classification error (between true and generated data), while the discriminative network is trained to minimize it. Equilibrium is reached when the generator produces samples that follow the original probability distribution and the discriminator predicts fake or not-fake with equal probability. It is highly important that both networks learn equally during training and converge together. So, during the training of the discriminator, we ignore the generator loss and focus on the discriminator loss, which penalizes the discriminator for misclassifying real images as fake or vice-versa. Generator's weights are not updated. During the generator training, we use the generator loss, which penalizes the generator for failing to fool the discriminator and generating an image that the discriminator classifies as fake. The discriminator is frozen during generator training and only generator's weights are updated through back propagation.

Consider the training dataset as uniform distribution, $p(X)$, a real image drawn from it is taken as x and a random image in R^d to be z. Let $G(z)$ and $F(z)$ be the generative and discriminative models, respectively. The generative model generates an image as its output, $G(z)$. Let $p(G)$ be the probability distribution of $G(z)$. The discriminative model is tasked with computing the probability that an input image

drawn is real. Ideally, *F(x) = 1* if *x* is real and *F(x) = 0* if *x* is generated. The GAN framework corresponds to a mini-max, with the objective function given by the equation 1 [12].

$$\max_{g} \min_{f} V(f, g) = E_{x-px}\big[-\log f(x)\big] + E_{x-px}\big[-\log(1 - f(g(z)))\big] \quad (1)$$

$E(x)$ is the Expected value (or First moment) of a function. f and g are the respective functions of the generator and the discriminator.

However, in practice this is solved by alternating these two gradient update steps:

$$\text{step } 1 : \theta_f^{t+1} = \theta_f^t - \lambda^t \nabla_{\theta f} V\big(f^t, g^t\big) \quad (2)$$

$$\text{step } 2 : \theta_g^{t+1} = \theta_g^t - \lambda^t \nabla_{\theta g} V\big(f^{t+1}, g^t\big) \quad (3)$$

where λ is the learning rate and θ_f and θ_g are the parameters of *F* and *G*. Goodfellow et al. [12] show that the distribution, *p(G)*, converges to *p(X)*.

3.2 DCGAN

Deep Convolutional GANs are essentially tailor-made GANs that have certain architectural constraints that help in stabilizing the learning process of standard GANs by learning a hierarchy of representations from object parts to scenes in both the generator and discriminator. Standard GANs are known for being unstable to train.

DCGAN utilizes logical steps towards preventing this instability such as incorporating the all convolutional net which replaces spatial pooling functions (like Max Pooling), the benefits of which are explained by Springenberg et al. [13] and allowing the generator to determine the down-sampling on its own by providing strided convolutional layers.

It also includes removing fully connected layers and applying Batch Normalization (technique by which each input is transformed to have unit variance and zero mean) to all but the input layer of the discriminator and output layer of the generator. Another unique feature is that generators employ the ReLu activation layer whereas the discriminator uses the Leaky ReLu activation function. Generator and Discriminator of a DCGAN architecture is shown in Figs. 1 and 2.
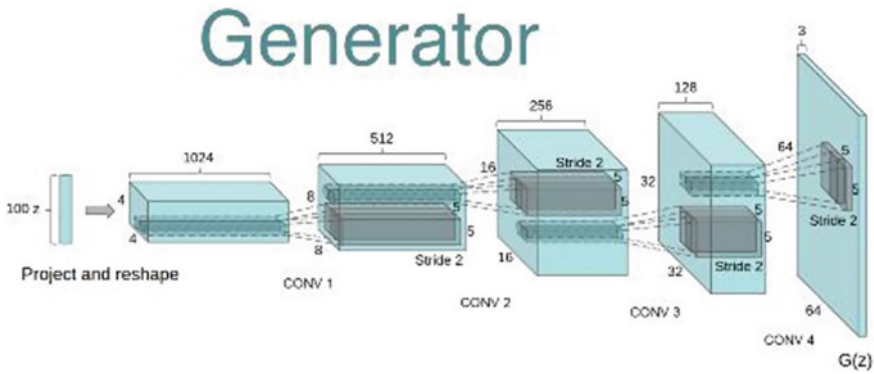
Fig. 1 Generator of a DCGAN [14] with strided convolutional layers

Fig. 2 Discriminator of a DCGAN with strided convolutional layers with the same dimensions

3.3 CoGAN

Coupled GAN (CoGAN) was proposed as an alternative approach to learn a joint distribution of multi-domain images from the data. It is a function that assigns probabilistic densities to every joint occurrence of an image in different domains like images portraying the same scenery but using different modalities or images of the same face portraying varying emotions like happy, sad, etc. Once this joint distribution is obtained, novel tuples of images can be generated by the GAN. Traditional approach dictates that to obtain this joint distribution, each set of multi-domain images must be fed individually as tuples of images. Building such a dataset consisting of tuples of corresponding images can be quite cumbersome, which greatly diminishes the application and usability of this concept. CoGAN comes to the rescue since it does not have this correspondence constraint. Rather it just requires a set of images to be drawn from the marginal distributions of each domain of images. The architecture of a CoGAN is based on that of a GAN. CoGAN incorporates
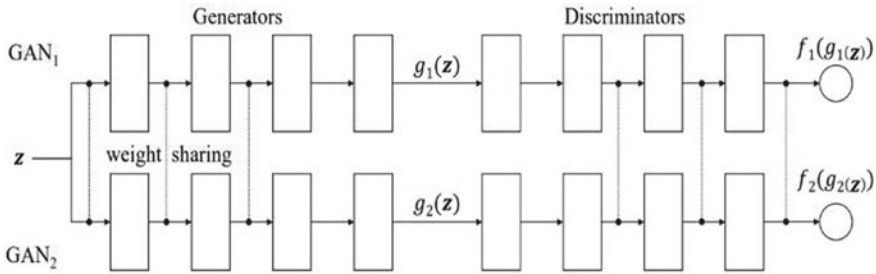
Fig. 3 Weight sharing in CoGAN [15]

a tuple of GANs for each individual image domain, which under normal circumstances produces a product of marginal distributions instead of a joint distribution. This inconvenience is overcome by using a weight-sharing constraint that favors a joint distribution result over a product of marginal distributions result. This forces the individual GANs, which decode the high- level semantics of the image to share the weights with the other GANs. Then the layers of these GANs responsible for decoding the low-level features translate the shared representation exclusively to fool their corresponding discriminators. The structure of a CoGAN having two image domains is shown in Fig. 3 of CoGAN proposal [14]. Although only two domains are shown, it can be easily extrapolated to as many domains as required. As discussed, the first few layers of the generators, which are responsible for extracting high-level semantics, are forced to share the parameters from a common subset.

Also, the last few layers of the discriminators, which are responsible for encoding the high-level semantics, are enforced to share the weights. Note that the flow of information is in opposite directions. Thus, CoGANs effectively synthesizes pairs of images without the need for correspondence in the input dataset and thereby aiding in better training the models.

Here the forced weight sharing is represented using a dotted line.As mentioned earlier, the weight sharing occurs in the first few layers of the generator and the last few layers of the discriminator.

3.4 LSGAN

The Least Squares GAN is a slightly modified variant of the normal GAN. Standard GAN employs the 'Sigmoid Cross-entropy' loss function in its discriminators while classifying the synthesized images from the real ones. This loss function may lead to this undesirable vanishing gradient problem. This problem arises in gradient based learning methods that update the weights based on the partial derivative (gradient) of the error function, where in some cases, the gradient becomes vanishingly small and thereby effectively stops the training process. LSGAN provides a solution by changing the loss function to the least squares loss function. Study [16] shows that
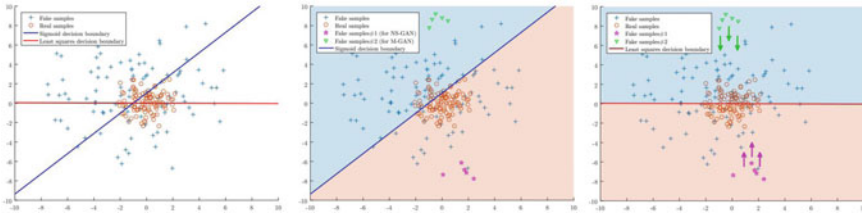
Fig. 4 **a** The decision boundaries of the two loss functions, **b** Fake images that may lead to vanishing gradients, **c** Effect of LSGAN [15]

this approach is better than the standard GAN in two aspects-generating higher quality images and learning in a more stable fashion.Least squares loss focuses on the distance of the image from the discriminator boundary and thus penalizes generated samples that, although appear on the right side of the boundary, are too far away from the real data and subsequently the boundary and thus may lead to vanishing gradients. Figure 4 shows how the LSGAN forces the generated samples (in magenta) to move closer to the decision boundary.

4 Experiment

For this paper we have used an open source dataset of SARS-CoV-2 CT scan [17], containing 2482 CT scans in total (1252 positive CT scans and 1230 negative CT scans). This data has been collected from hospitals in the locality of Sao Paulo, Brazil.

All the three different GANs were implemented with the help of pyTorch, an open source ML library. Each scenario consists of a validation phase and testing phase. In the validation phase, 15% of images were used whereas in the testing phase, 25% of images were used. All images have a dimension of 75*75*3. A sample for the fake images generated by the DCGAN is shown in Fig. 5 and their corresponding generator and discriminator losses with respect to the iterations are shown in Fig. 6.

These images were generated after 5000 iterations. Better quality can be obtained by increasing the number of iterations.

In addition to the deep transfer models, we added 3 dense layers with sigmoid activation function in the last layer to classify outputs accordingly. Number of epochs used to train is 100.

As a first step of approach, we trained all three deep transfer models using the actual dataset (without GAN). Later, we trained them using the dataset obtained by concatenating the actual dataset with the datasets obtained from various GANs (artificial images) individually.
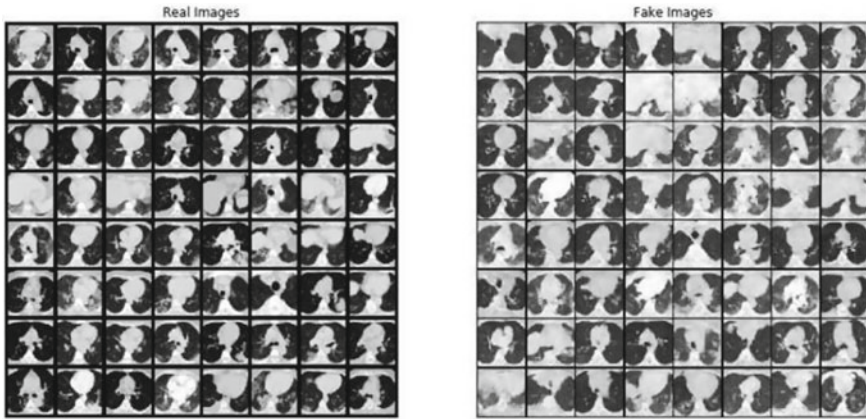
Fig. 5 Real versus generated images for DCGAN
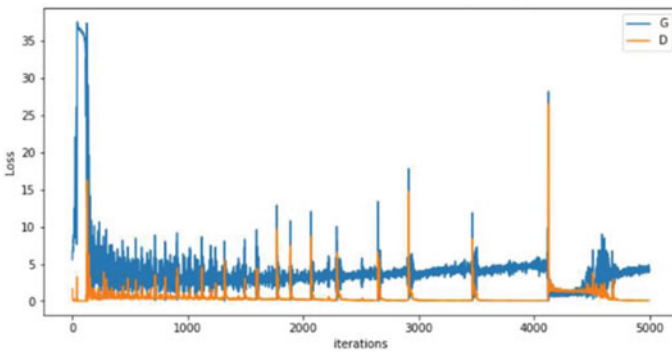
Fig. 6 Discriminator and generator loss versus iterations (DC-GAN)

5   Results and Discussion

CT scan is a computerized imaging technique to produce cross-sectional images using X-rays directed at a patient and interpreting the output signal. There are some visible attributes in the lungs of patients with COVID- 19 signs, such as ground-glass "opacities" (darkened and blurry patches in the lung) and zones of greater lung density called consolidation which are more recurrent, wide-spread and often with a rounded appearance and peripheral lung distribution when compared to a healthy person.

For evaluation, we used the confusion matrix which can be obtained explicitly using inbuilt libraries. The details of the confusion matrix are shown in Table 1.

The below mentioned parameters (4), (5), (6) are used for evaluation along with accuracy.

Table 1 Confusion matrix output for various GANs

GAN\Models	ResNetSO	InceptionV3	Xception
DC-GAN	[[532 34] [27 528]]	[[541 5] [31 544]]	[[533 13] [6 569]]
CoGAN	[[536 10] [20 555]]	[[531 15] [12 563]]	[[541 5] [12 563]]
LS GAN	[[536 10] [5 570]]	[[536 10] [15 560]]	[[541 5] [11 564]]

$$\text{Precision} = \frac{\text{True Positives}}{\text{True positives} + \text{False Negatives}} \tag{4}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True positives} + \text{False Negatives}} \tag{5}$$

$$\text{F1 Score} = \frac{\text{Precision*Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

Here a True Positive refers to a Covid patient being correctly identified as one whereas a True Negative refers to a healthy person identified correctly as one. A False Positive is a healthy person incorrectly identified as a Covid patient whereas a False Negative is a Covid patient was not identified as one. The plots of Validation Accuracy Vs epoch and Validation Loss Versus epoch for all three of the deep transfer models in each case (without GAN as well with each GAN) have been provided. Xception has the most deviation among the models considered but it also has the best validation accuracy among them. The series of tables drafted below shows the brief analysis of the results obtained (Fig. 7).

Table 2 shows the results obtained by testing the deep transfer models trained without GAN implementation. As we can say, Xception tops the list with its accuracy being 97.26% which is a considerably good result.

Because of Xception's complex network architecture, which simultaneously trains 3–4 small networks and concatenates all the results together for further processing,
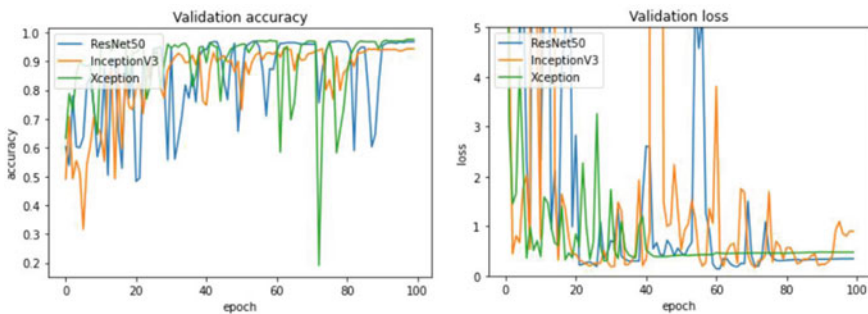
Fig. 7 Validation accuracy and validation loss plot for deep transfer models without GAN

Table 2 Test results of 3 deep transfer models without using GAN architecture

Model / Results	ResNetSO	InceptfonV3	Xception
Testing accuracy	94.52	96.3	97.26
Precision	0.928	0.948	0.96
Recall	0.981	0.981	0.987
FI score	0.948	0.965	0.974

it can learn more details than ResNet50 consisting of 50 layers, whose architecture is more of a repetitive one. Same analogy goes with InceptionV3 as its architecture is less complex when compared to Xception's but more than ResNet50's.

Table 3 shows the results obtained by testing the deep transfer models trained with the LS-GAN dataset along with the real dataset. This significant improvement in each deep transfer model's results attests the use of augmentation, especially using GAN. Here, ResNet50 tops the list with an accuracy of 98.66%. It has improved its accuracy to 4.14% which is to be noted. Due to the increase in dataset size, ResNet50 has a greater exposure resulting in its improvement. Xception performed exceptionally well with its accuracy being 98.57% which is much comparable with ResNet50, with an improvement of 1.31%. InceptionV3 also showed improvement in its accuracy with a difference of 1.47% which is more than Xception's (Fig. 8).

Table 4 shows the results obtained by testing the deep transfer models trained with CoGAN implementation. Here, Xception tops the list in this analysis with its accuracy being 98.48%, followed by InceptionV3 and ResNet50. All the deep transfer models
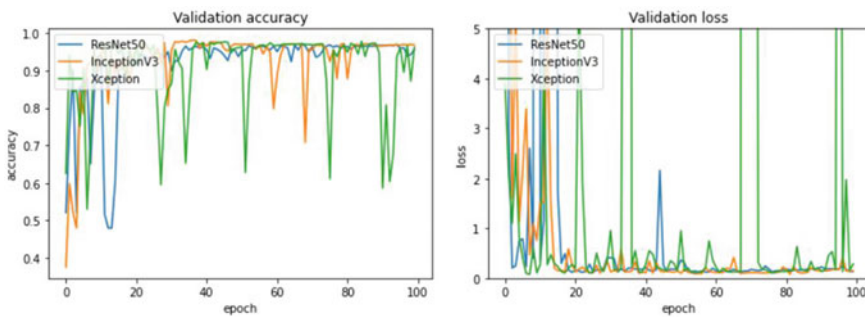
Fig. 8 Validation accuracy and validation loss plot for deep transfer models with LSGAN

Table 3 Test results of 3 deep transfer models using LS-GAN architecture

Model / Results	ResNet50	InceptionV3	Xception
Testing accuracy	98.66	97.77	98.57
Precision	0.991	0.973	0.98
Recall	0.982	0.982	0.991
FI score	0.986	0.977	0.985

showed improvement in its scores with ResNet50 being the highest with an increase of 2.8% (Fig. 9).

Table 4 Test results of 3 deep transfer models using CoGAN architecture

Model / Results	ResNetSO	InceptionV3	Xception
Testing accuracy	97.32	97.59	98.48
Precision	0.964	0.978	0.978
Recall	0.982	0.973	0.991
FI score	0.973	0.975	0.9845
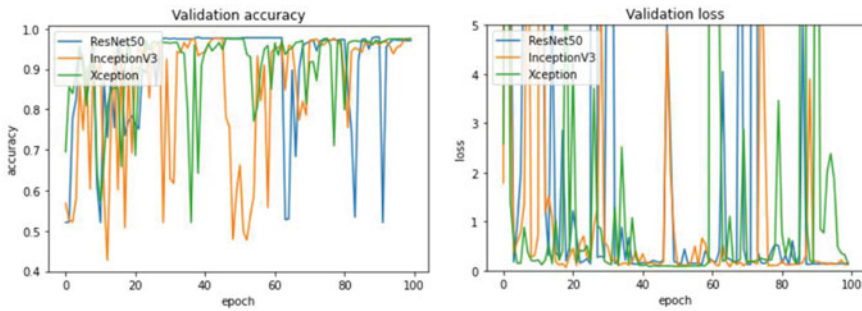
Fig. 9 Validation accuracy and validation loss plot for deep transfer models with CoGAN

Table 5 Test results of 3 deep transfer models using DC-GAN architecture

Model / Results	ResNetSO	InceptionV3	Xception
Testing accnracy	96.52	96.79	98.31
Precision	0.97	0.946	0.989
Recall	0.958	0.991	0.976
FI score	0.964	0.968	0.9825
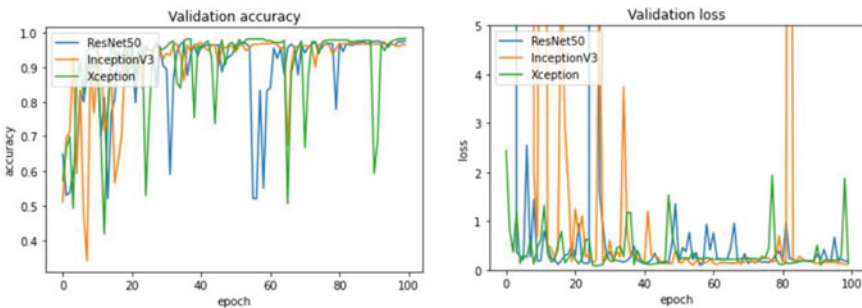
Fig. 10 Validation accuracy and validation loss plot for deep transfer models with DC-GAN
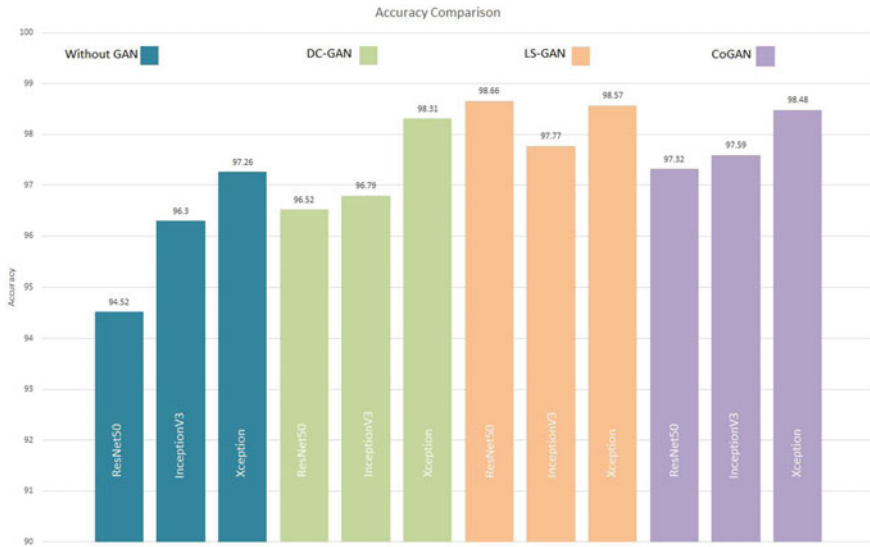
Fig. 11 Accuracy comparison

Table 5 shows the results obtained by testing the deep transfer models trained with DCGAN implementation. Here, Xception again tops the list in this analysis with its accuracy being 98.31%, followed by InceptionV3 and ResNet50. Here, DCGAN has the least incremented results among the GANs taken into consideration, yet shows a significant improvement when compared to results obtained without GAN implementation (Fig. 10).

From the above analysis, we can say that models trained with LSGAN augmented images proved to be the best among the GANs considered in this proposal. This is due to the 'least squared' loss function which solves the vanishing gradient problem confronted by the other GANs here, resulting in a greater stability. LSGAN works with the advantage of penalizing those samples that have no gradients though they have been classified rightly (Fig. 11).

6 Conclusion

In this paper, the detection of the novel coronavirus proved to be more efficient by augmenting the already available dataset [16]. The fabrication of artificial images was primarily based on the adversarial network GAN. The main aim of this paper is to determine the most credible GAN network to carry out the task of data augmentation as well to prove that involving GANs would improve the existing accuracy of our model, paving way for an effective approach to train the model. From our analysis, ResNet50 (98.66%) and Xception (98.57%) trained with LS-GAN augmented

images provides the best testing accuracy. The best improvement between architecture implemented without GAN and with GAN can be observed in ResNet50 implementation with LS- GAN (4.14%). The pre-processing phase involved here solves the issue of limited datasets available in the biomedical areas of research and also aids in training the Deep Learning models more efficiently. The work carried out in this paper could be further experimented by venturing into other recent and robust versions of GANs.

References

1. Ardakani A, Kanafi A, Acharya A, Rajendra U, Nazanin K, Afshin M (2020) Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: results of 10 convolutional neural networks. Comput Biol Med. https://doi.org/10.1016/j.compbiomed.2020.103795
2. Chollet F (2017) Xception: deep learning with depthwise separable convolutions. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), Honolulu, HI, pp 1800–1807. https://doi.org/10.1109/CVPR.2017.195
3. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), Boston, MA, pp 1–9. https://doi.org/10.1109/CVPR.2015.7298594
4. Polsinelli M, Cinque L, Placidi G (2020) A light CNN for detecting COVID-19 from CT scans of the chest. https://doi.org/10.1016/j.patrec.2020.10.001
5. Waheed A, Goyal M, Gupta D, Khanna A, Al-Turjman F, Pinheiro PR (2020) CovidGAN: data augmentation using auxiliary classifier GAN for improved Covid-19 detection. IEEE Access 8:91916–91923. https://doi.org/10.1109/ACCESS.2020.2994762
6. Li L, Qin L, Xu Z et al (2020) Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary ct: evaluation of the diagnostic accuracy. Radiology 296(2):E65–E71. https://doi.org/10.1148/radiol.2020200905
7. Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, Cai M, Yang J, Li Y, Meng X, Xu B (2020) Deep learning algorithm using CT images to screen for Corona Virus 2 disease (COVID19). medRxiv 2020.02.14.20023028. https://doi.org/10.1101/2020.02.14.20023028
8. Sidharth R, Rohit P, Vishagan S, Karthika R, Ganesan M (2020) Deep learning based smart garbage classifier for effective waste management. In: 2020 5th international conference on communication and electronics systems (ICCES), IEEE, pp 1086–1089. https://doi.org/10.1109/ICCES48766.2020.9137938
9. Kiruthika K, Vijayan D, Lavanya R (2019) Retrieval driven classification for mammographic masses. In: 2019 International conference on communication and signal processing (ICCSP), pp 0725–0729. https://doi.org/10.1109/ICCSP.2019.8698044
10. Saiharsha B, Diwakar B, Karthika R, Ganesan M (2020) Evaluating performance of deep learning architectures for image classification. In: 2020 5th International conference on communication and electronics systems (ICCES), pp 917–922. https://doi.org/10.1109/ICCES48766.2020.9137884
11. Ennesai S, Narayanankutty KA, Ganesan M (2010) Temporal alignment of non-gated image sequences for 4D cardiac imaging using wavelets. In: 2010 International conference on computer and communication technology (ICCCT), pp 198–200. https://doi.org/10.1109/ICCCT.2010.5640532
12. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. NIPS

13. Springenberg JT, Dosovitskiy A, Brox T, Riedmiller M (2014) Striving for simplicity: the all convolutional net. arXiv preprint arXiv:1412.6806
14. Radford A, Metz L, Chintala S (2016) Unsupervised representation learning with deep convolutional generative adversarial networks
15. Liu M-Y, Tuzel O (2016) Coupled generative adversarial networks. Mitsubishi Electric Research Labs (MERL)
16. Mao X, Li Q, Xie H, Lau RYK, Wang Z, Smolley SP (2017) Least squares generative adversarial networks. In: 2017 IEEE international conference on computer vision (ICCV), Venice, 2017, pp 2813–2821. https://doi.org/10.1109/ICCV.2017.304
17. Soares E, Angelov P, Biaso S, Froes MH, Abe D K (2020) SARS-CoV-2 CT-scan dataset: a large dataset of real patients CT scans for SARS-CoV-2 identification. medRxiv. https://doi.org/10.1101/2020.04.24.20078584

VLSI Implementation of Multipliers for Artificial Intelligence Applications: A Survey

R. Karthick, S. Shanmuga Raju, G. Abinaya, G. Ashwanth, and G. Kaviya

Abstract As the competition among different multiplier models keeps rising we have explored a 2-speed, radix-4, serial-parallel (SP) multiplier for stimulated applications like Machine Learning and Artificial Intelligence. The designed multiplier is a variant of the modified radix-4 with serial-parallel (SP) Booth multiplier that skips over the zero operations and adds only the ones in Booth encodings, making the latency dependent critical paths are taken for process so that throughput and latency are invoked higher for a subnet of multiplier values. We show that for bit widths of 32 and 64, our optimizations can result in a 1.42 times to 3.36 times improvement with the standard parallel (P) Booth multiplier in terms of time—area depending on the incoming set.

Keywords Multiplier · Speed optimized · Dynamic CMOS · Booth multiplier

1 Introduction

Manipulating data on multiplication is the most important milestone in applications like Artificial Intelligence(AI), digital signal processing (DSP) and machine learning (ML), detailing the delay, area, power and overall performance of parallel(P) implementations. An extensive work on optimizing the multiplication circuits, however, the problem is that Wallace or dadda tree in combination with higher radixes on modified Booth algorithms. Parallel-Parallel (P-P), Serial-Parallel(S-P) and serial-serial(S–S) in which any one of the multiplication that is generally performed in digital circuits. Modified Booth algorithm, in which we target serial-parallel(SP) two-speed multiplier (TSM) that rigidly adds only the ones encoded parts of the multiplication and bounce over the zero encoded portions.

In AI, DSP and ML implementations, the improvement in performance of a design depends on the turned down precision representations, to gain computational accuracy we strive for the bottom-most possible bit width. Exactness is usually fixed at

R. Karthick (✉) · S. Shanmuga Raju · G. Abinaya · G. Ashwanth · G. Kaviya
Department of Electronics and Communication Engineering, Dr.N.G.P Institute of Technology, Coimbatore, India

design time, and therefore, any changes in the requirements entail that further modification involves reconfiguration of the implementation. In cases where a tiny bit span would be sufficient, the design runs at bottom efficiency since least mandatory computation is taken on. To lighten this, mixed-accuracy algorithms intended to be used in a lower bit span some partition of time, and a substantial bit space when necessary. Two data paths that make the operation at different precisions which is normally implemented in the circuit.

The target of the paper is to remove a portion of the computation completely and significantly during the runtime which is a dynamic control structure. The goal is accomplished using modified serial Booth multiplier, which is independent of location and skips over encoded entire-zeroes or entire-ones computation. A parallel bit arrangement of all operands both multiplicand and multiplier that is designed to give a smart block which can be injected into existing AI bots, DSPs and ML. For certain intake sets, the multiplier hits considerable striking improvements in computational system performance. The main benefits of this method are as mentioned below.

(1) The data path of the booth multiplier has two sub-circuits. These circuits operate at different critical paths.
(2) Implementing two critical paths in booth multiplier provides enhancement in bit-pattern calculation and increases the throughput. Therefore its results in reduced time and improved latency than conventional multipliers.
(3) The FPGA implementation of the proposed multiplier model for evaluating the performance and comparison of throughput with conventional multipliers (Fig. 1).

Fig. 1 Multiplier architecture
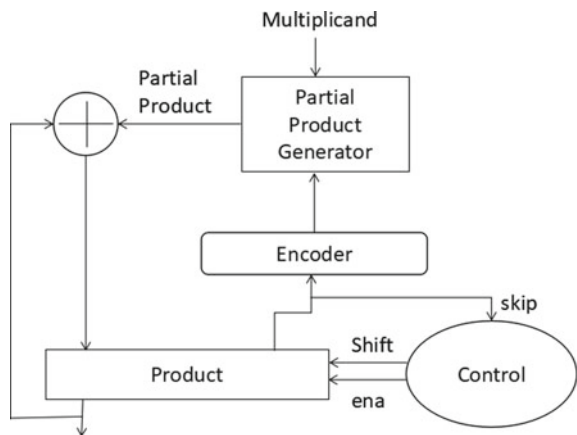
2 Related Works on Multipliers for AI

2.1 Booth Multiplier, Dadda Multiplier, Wallace Multiplier

Vasudeva et al. [4] proposed a "Study of 8 Bits Fast Multipliers for Low Power Application". In VLSI the design of multipliers is the most complex architecture block. Designing the multiplier may face challenges in optimizing power, delay and area. When the number of stages is increased the width of the multiplier is also increased which leads to complexity in Areas. For accuracy and high sampling rate wide bit width is required. So the requirement for designing and developing an IP for high speed applications. The design is proposed utilizing the basic multiplier building blocks. In this paper we analyze the different types of 8-bit multiplier and compare their performance. We are comparing three multiplies namely modified Booth, Wallace and Dadda.

Modified Booth's Multiplier:
    Booth's algorithm is commonly used for multiplication of signed numbers because it is the most powerful algorithm to perform multiplication of signed numbers. This multiplier frequently performs addition of A or S with P and performs arithmetic right shift in P register. The algorithm of booth multiplier as follows:
    x—multiplicand and y—multiplier
    Step 1: fill the A with x bits and $y + 1$ zeros, similarly S with $-x$ and $y + 1$ zeros also p with $x + y + 1$ zeros
    Step 2: check the last two bit of p
        If it is 01 perform $P + A$.
        If it is 10 perform $P + S$.
        If it is 00,11 do no operations.
    Step 3: Shift the values obtained to one bit right.
    Step 4: Repeat step 2 and three y times.
    Step 5: Eliminate the MSB of P, and the resultant is obtained.

Wallace Multiplier:
    In Wallace multiplier the delay is always equivalent to logarithm of operand size. The operation of wallace multiplication follows:
    Step 1: The bits are grouped for the product.
    Step 2: Using carry save adders the product matrix is reduced to a two rows
    Step 3: fast carry propagate adder will add the remaining two rows and the final product is obtained.
    The Wallace tree is a family of column compression multipliers (Fig. 2).

Dadda Multiplier:
    The Dadda multiplier with carry look ahead reduces the number of rows in the multiplication process. The counter numbers are less in this method. This reduction in counter placement will provide a reduction in partial product. The figure shows the flow of the Wallace multiplier (Fig. 3).

Fig. 2 Wallace multiplier
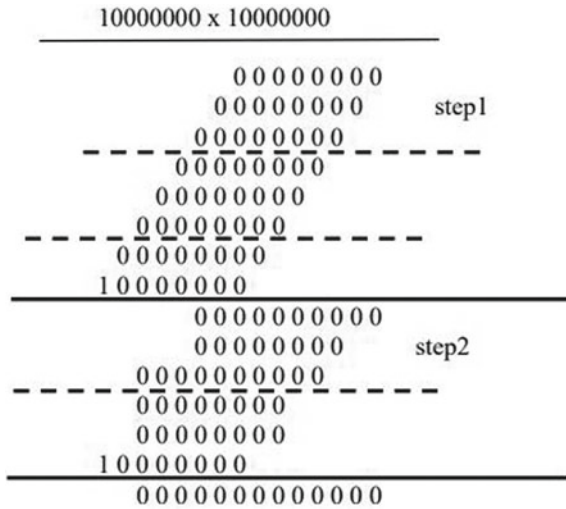
Fig. 3 Operation 8 × 8 bits dadda multiplier
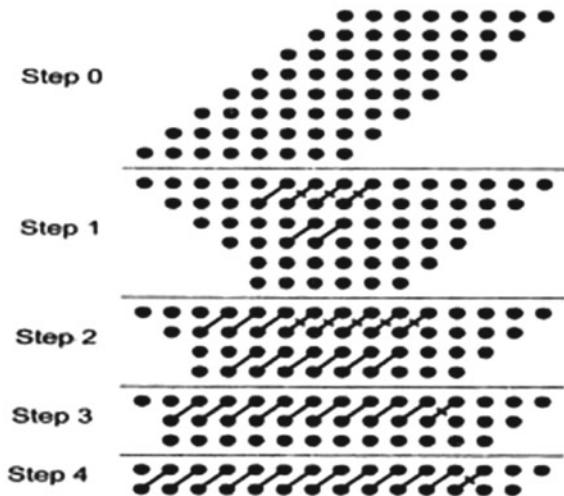
After analyzing the performance parameters like speed, area and power consumption of Booth, Wallace, and Dadda multiplier it can be concluded that Dadda multiplier is best in considering area, Booth multiplier provides better power optimization and Wallace tree multiplier provides better delay optimization. When the number of interconnections between the building blocks increases there will be increase in the intermediate stages of the multiplier. This paper shows that the Dadda multiplier has a 76.5% increase in interconnections as it has more number of inter-stages.

2.2 Vedic Multiplier

Sri Lakshmi et al. [1] proposed "Design and Implementation of High-Speed Vedic Multiplier Using Hybrid Full Adder". A novel multiplier architecture is introduced which includes the algorithms like Array, Modified Booth, Dadda multipliers etc. A multiplier based on Vedic mathematics is discussed in this paper. Vedic mathematics is an ancient method of sutras for mathematical calculations. The special purpose of Vedic maths is it reduces the complicated calculations by calculating the partial products in prior before the beginning of actual operations of multiplication so that the optimization in time and delay can be achieved. A novel hybrid vedic multiplier approach has been discussed. In this paper a Vedic Multiplier with CSA is used to increase the speed. The half adders and full adders in existing Vedic multipliers are replaced by hybrid adder and compressor adders and comparison has been performed. Compressors are the logic circuits that can add 3 or more bits simultaneously. Hybrid full adder is also a kind of full adder with a lesser number of gates so that the area is reduced. Therefore, by observing all the above results, we can conclude that the hybrid adder is more power and delay efficient. Though the compressor has more power and area efficiency compared with normal conventional full adder design, hybrid adder is giving even better results compared with compressors. So, hybrid adders give better performance among these three adders circuits (Fig. 4).

Compared with conventional adder the hybrid adder has a delay of 7.04%, power of 11.12% and area has 1.46% respectively. In comparison of hybrid adder with
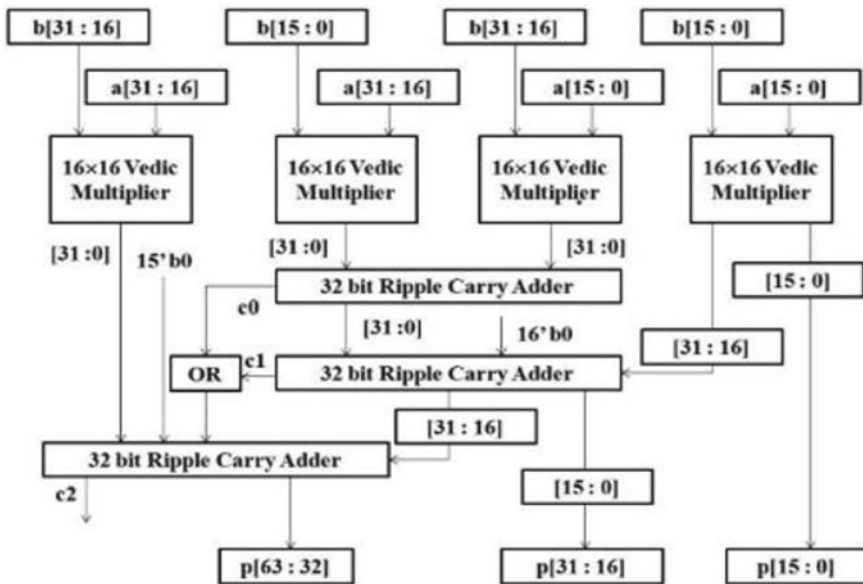
Fig. 4 Vedic multiplier

compressors the delay, power and PDP got reduced by 2.02%, 1.43% and 3.74% respectively. Finally conclude that hybrid adder with vedic is gives better efficiency compared with conventional full adder and compressors.

2.3   *Wallace Tree Multiplier with Kogge Stone Adder*

Sundhar et al. [5] proposed the "performance analysis of wallace tree-multiplier with Kogge stone Adder using 15–4 compressor" (Fig. 5).

Wallace multiplier using full adder and half adders to bring down the partial product from three to two rows and then two partial products are added using the final adder. In this paper an approximate 16×16 bit Wallace multiplier using 15–4 compressor architecture is designed and if we compare the performance of this multiplier with Kogge stone adder, it is using the same architecture of multiplier using parallel adder. This architecture infers that kogge stone adder is faster when
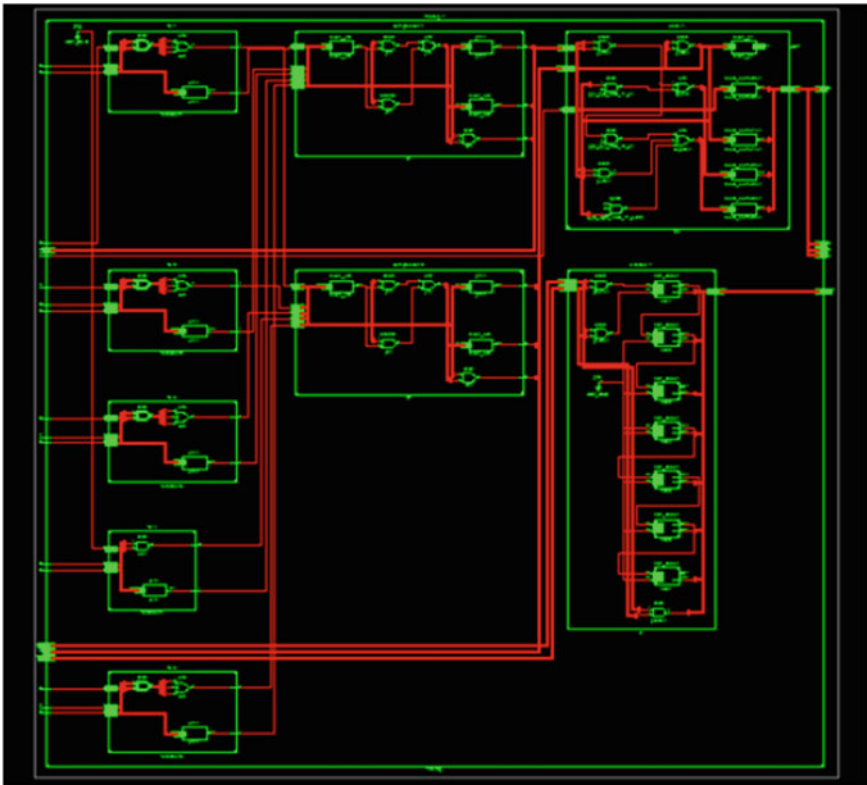
Fig. 5   Overall architecture of 16×16 Bit wallace tree

we are compared to the multiplier with parallel adder since Kogge stone adder is basically a parallel prefix adder. This adder has the fastest addition based on design time. Hence this multiplier occupies about 15% of total coverage area and dissipated power and delay are also highly reduced.

2.4 Array Multiplier

Devi et al. [6] proposed "low power optimized Array multiplier with reduced area". In this paper a 32-bit design of power and reduced area optimized unsigned array multiplier is proposed. This multiplier is a 32bit array multiplier which makes use of Carry Select Adders (CSAs) to execute partial product additions. This adder's area is expanded but delay is reduced. Ripple Carry Adders (RCAs) are the adders that have a closely packed area but they are the slowest type of adders (Fig. 6).

On the other hand, carry look ahead adder (CLA) is the high-speed adder but they are the worst from the area point of field of vision. Hence carry select Adder (CSA) have been considered as the best comparing with RCA and CLA because they offer a good counterbalance between compact areas of RCAs and short delay of CLAs. The array multiplier with CSA uses modified booth algorithm with greatest power
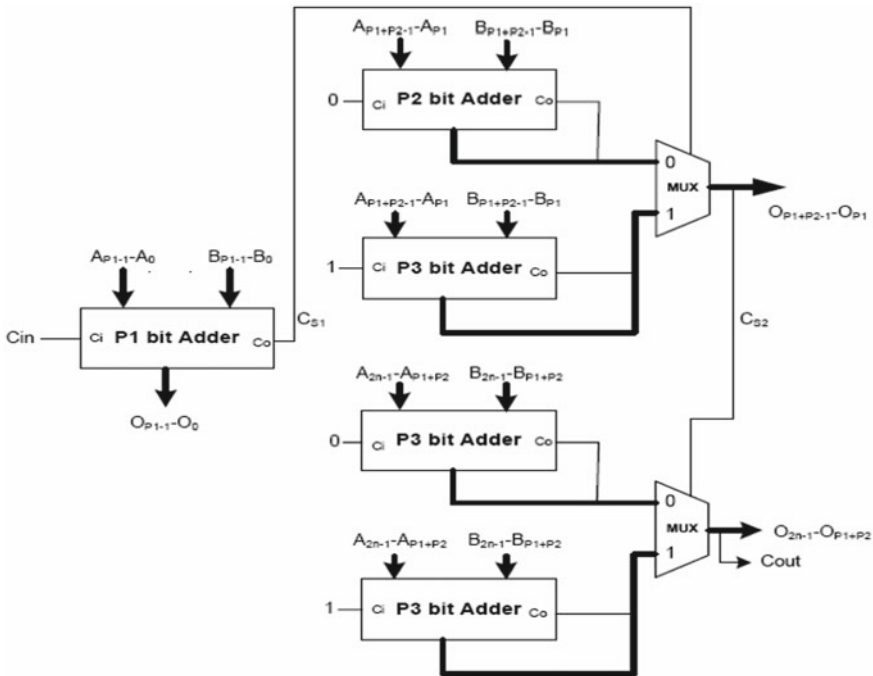
Fig. 6 Array multiplier with carry select adder

compression technique to minimize the power dissipation upto 22% and reduction in area is also achieved.

2.5 Array Multiplier with OTFC

Aruna et al. [2] proposed "A low power and high speed Array multiplier using on-the-fly conversion". In this multiplier design with single precision truncation, the product of 2 N-bits produces 2 N but partial products, excluding these 2 N bit partial products, are going to be divided into 2 N-(n/2) bits and N/2 bits. As a result finally, 2 N bits are created by adding the above bits using ripple carry adder (Fig. 7).

The array multiplier outlined in this paper is designed and enforced with no truncation or addition approach; instead it is executed using a typical array multiplier scheme. The proposed array multiplier produces the high order bit (MSB) of the final product. The multiplier design leverages on the fly conversion converter that is implemented as the tail end of the multiplier. This is to achieve the expedited carry propagation in the last leg of multiplication. The OTF conversion logic is achieved to
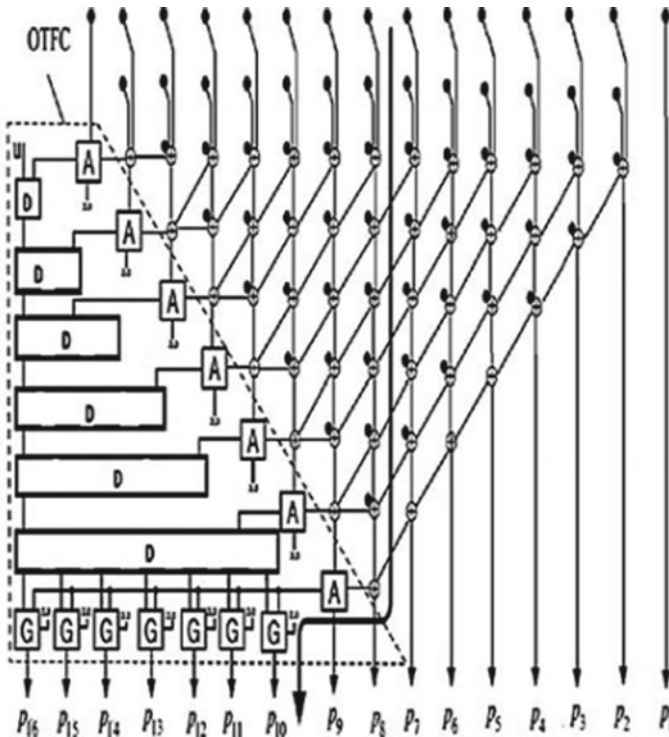
Fig. 7 Array multiplier with OTFC

Table 1 Comparing existing multiplier

S. No.	Multiplier	Delay (ns)	Power (mW)	Area
1	Booth multiplier	3.75	3.24	214 LUTs
2	Dadda + CLA multiplier	2.56	5.65	878 µm^2
3	Carry select multiplier	46.23	155.4	4034 LUTs
4	Variable carry skip multiplier	40.6	153.7	3970 LUTs

Table 2 Comparing recent multipliers

S. No.	Multiplier	Delay (ns)	Power (mW)	Area
1	Vedic + Full adder multiplier	13.64	173.2	3653 LUTs
2	Vedic + Hybrid adder multiplier	12.12	170.71	3396 µm^2
3	Array multiplier	3.14	9.89	5349 LUTs
4	Radix-4 S-P multiplier	21.28	4.35	87 LUTs
5	Wallace multiplier	3.125	0.42	1570 µm^2

provide constant delay irrespective of input bits processed. Hence the array multiplier with OTF conversion has a minimum array, smaller area and less energy.

3 Comparison on Existing Multipliers

See Table 1.

4 Comparison on Recent Multipliers

See Table 2.

5 Resultant Efficiency Chart

6 Conclusion

The proposed multiplier is divided into two sub circuits which operate at different critical paths. The design of CMOS Multiplier that will be highly adaptable to AI, ML and DSP kind of applications. The Multiplier is designed for bit width of 32 and 64, expecting an improvement of 3 × and 3.56 × respectively. On comparing the

standard parallel multiplier, this design gives 3.64 × improvement in area and delay. Upcoming works on application will take complete feature of two speed optimization.

References

1. Sri Lakshmi Y, Vigneswaran T (2020) Design and implementation of high-speed vedic multiplier using hybrid full adder. Int J Adv Sci Technol 29(3):10663–10669
2. Aruna S, Venkatesh S (2019) A low power and high speed array multiplier using on fly conversion. Int J Recent Technol Eng 7(5S4)
3. Moss DJM, Boland D, Leong PHW (2019) A two-speed, radix-4, serial-parallel multiplier. IEEE Trans VLSI Syst
4. Vasudeva G, Cyril Prasanna Raj P (2015) Study of 8 bits fast multipliers for low power application. Int J Soft Comput Eng 5(1). ISSN: 2231–2307
5. Sundhar A, Tharshini SD, Priyanka G, Ragul S Saranya C (2019) Performance analysis of wallace tree multiplier with Kogge Stone Adder using 15–4 compressor. In: 2019 International conference on communication and signal processing (ICCSP), Chennai, India, 2019, pp 0903–0907
6. Devi P, Singh G, Singh B (2011) Low power optimized array multiplier with reduced area. In: Mantri A, Nandi S, Kumar G., Kumar S. (eds) High performance architecture and grid computing. HPAGC 2011. Communications in Computer and Information Science, vol 169

Automated Boneage Analysis Using Machine Learning

J. K. Krithika, Yayha Norashikin, K. Porkumaran, and S. Prabakar

Abstract Bone age assessment is done to analyse the skeletal maturity according to their chronological age. This is done by radiograph method considering the left hand or wrist. Bone age assessment is done by comparing the chronological age to assess the endocrine disorders and pediatric syndromes. Earlier the manual method was used, where the radiologists compare the radiograph image with the atlas and estimate the age of the bone. In this study, the analysis and classification of the x-ray image of the left hand is experimented to determine the bone age. Here the bone age analysis method involves the segmentation of the image, feature extraction and classification using support vector machine(SVM). The results obtained is future used to assess the skeletal abnormalities.

Keywords Support vector machine · Feature extraction · Hand X-ray image · Region of interest

1 Introduction

Bone is connective tissue. There are 300 soft bones which are formed during the birth, then the cartilage grows and slowly replaced by hard bone. During the adolescence period there are 206 bones formed. These bones have their own blood vessels and made up of living cells. The longitudinal growth in long bones is due to process

J. K. Krithika (✉)
Dr.N.G.P Institute of Technology, Kalapatti Road, Coimbatore, India

Y. Norashikin
UniversitiTeknologi Petronas, Seri Iskandar, Perak, Malaysia
e-mail: norashikin_yahya@utp.edu.my

K. Porkumaran
Sri Sairam Engineering College, Chennai, India

S. Prabakar
Sona College of Technology, Salem, India
e-mail: srisornaprabu@gmail.com

of endochrondral ossification and width growth is by the development of skeletal tissue directly from the fibrous membrane [1]. Bone age assessment is done by comparing the chronological age, which is used for diagnosis of endocrine disorders and pediatric syndrome. Osteoporosis is a skeletal disease, which is caused due to low mineral density in the bone. This is diagnosed by measuring the bone mineral density (BMD) [2]. The Radiograph images are used as they are quite safe, the exposed radiation is less. The use of age assessment in social development is to identify the details about the refugee, this also used in criminal cases and also for forensic reports [3].

2 Problem Statement

Bone age helps to evaluate growth but the children does not mature at the same time, which makes it difficult to evaluate the growth. Usually the bone age assessment depends on radiologist experience, it is carried out by visual evaluation of the skeletal development using GP atlas. In manual approach, bone age assessment varies from one scientist to another. It is also time consuming and may lead to errors.

In this research method, it will focus on analysis of the left hand x-ray image. This is done by taking the ROI part and Extracting the image. This research will result in development of automated bone age prediction algorithm using machine learning technique.

3 Objectives

The assessments that will be carried out are about the bone age analysis from x-ray image radiograph, to achieve few goals:

1. To analyze the bone from the x-ray image at various level of age.
2. To develop and evaluate bone age prediction algorithm by using machine learning technique.

4 Literature Review

Osteoporosis is a skeletal disease due to low bone density; this is diagnosed by measuring the bone mineral density. Bone age analysis was also used for evaluating the growth and diagnosis of pediatric syndrome and endocrine diseases.

There were different methods for determining the bone age. CT was used but it was not preferred as its radiation is high. MRI is being developed, but requires more research. Ultrasonography is used but it is not accurate as radiographic methods [3]. In ultrasound method 2 transducers are used, one produce ultrasonic waves and other

one is the receiver. Radiograph method where considered as the predominant method for bone age analysis. A study was conducted by using the mammography x-ray, which has the high image quality. The method used was dual energy x-ray absorptiometry (DEXA) and sonography. The DEXA is more accurate but it is expensive and radiation emitted is more. It is used in spine and hip. Sonography method is radiation free and its cost is low. It is usually used to assess the bone mineral density of peripheral bones (heel, knee cap) [5].

Studies were not only done in hand bone but few studies done the bone age analysis by dental maturity, it was used for forensic purposes, but not used in diagnosis of endocrine diseases. The other method is by iliac bone, it was determined based on the degree of maturation of the iliac crest apophysis, but this was a drawback one as the ossification is not uniform in iliac crest apophysis. The other method is by femoral head, which is the alternative way by assessing the depth of the epiphysial cartilage of femoral head which is continuously being ossified. Ossification is complete and most of the cartilage is replaced by bone remaining cartilage, which is called as hyaline articular cartilage. The bone age analysis was first done manually in tooth, later in bones such as skull, foot bones, knees, spinal cord, pelvis, rib, femur, carpals and epiphyses of the ulna and radius. The main characteristics of the bone age assessment are fusion of bones. The problems in manual approach is time consuming, it varies from scientist to scientist, it may also lead to error and during surgery over estimation or under estimation of bone length may lead to deformity. But in automated bone age analysis, according to Dinesh M.S the radiograph must have preprocessing, segmentation, image enhancement, rotation of bone pixels to proper angel, measurement analysis and decision analysis [6].

An attempt using the SVM classification method has been introduced by Markus Harmsen et al. [4]. The semi-automatic BA analysis developed by combining the support vector machine with cross correlation to a prototype image for each class. The evaluation of BAA is presented by comparing the nominal and real- valued SVM with k nearest neighbor classification. According to this automated BAA, there are some step that have present: (1) Extract 14 epiphyseal regions from the radiographs; (2) retain image features using the image retrieval in medical application framework; (3) Use these features to build a classifier model (training phase); (4) Evaluate performance on cross validation (testing phase); (5) Classify unknown hand image (application phase). These all steps have been implement by using C++ and SQLite database version 2.8.17. Therefore, based on the 1097 hand radiographs of 30 diagnostic classes in range 0–19 years, the nominal and real-value SVM are resulting 91.57% and 96.16% accuracy respectively. However, in all the study done, can observe that the SVM classification method have the higher accuracy and state that the simple classification for the BA analysis. However, some improvement of feature extraction is required as to achieve the BA assessment more efficient and can classify to the age level categories.
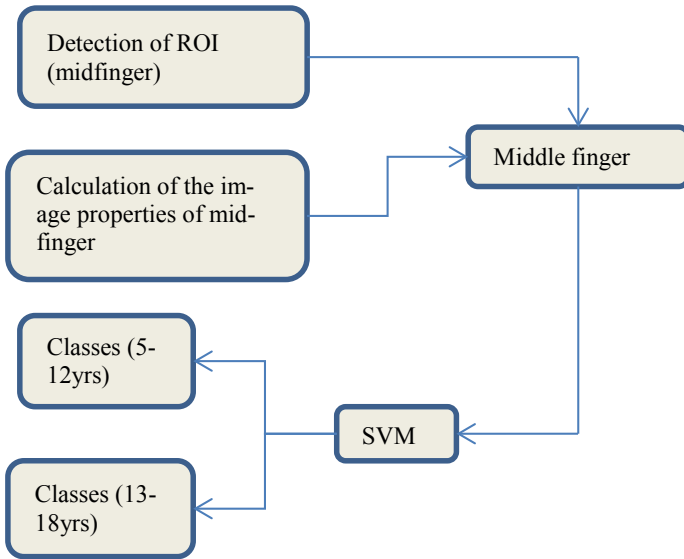
Fig. 1 Simplified performance model of automatic bone age analysis

5 Proposed Work

The detection and feature extraction of the hand bone x-ray of different image is the significant work done in the bone age analysis. The input images are taken from the website, and feature is extracted and the trained data is feeded into the classifier (Fig. 1).

5.1 Input Image

The image dataset of the left hand is taken from the website kaggle.com. The test dataset and trained datasets are downloaded, about 10,000 and above datasets for each dataset. The dataset of the images is obtained with the CSV, which contains the information regarding the person's age, sex that are used for training.

5.2 ROI Extraction

The first stage used for segmentation. The middle finger is labeled using the toolbox in mat lab for all 120 images. Middle finger is considered because of prominent

growth of bones. After labeling of the image the data is exported into the workspace in tabular form.

5.3 Feature Extraction

Feature extraction is the next stage. From the ROI image the mean, standard deviation, maximum value, energy, variance is determined. Normalization was also considered and the output was compared between before normalization and after normalization.

5.4 Classification

The feature extraction will undergo the classification method. All the classifier was used to find the better classifier for classification. From the work it was found that, linear and quadratic svm has better classification accuracy.

5.5 Result of Training the Dataset for 120 Images

See Figs. 2, 3, 4, 5, 6 and 7; Table 1.
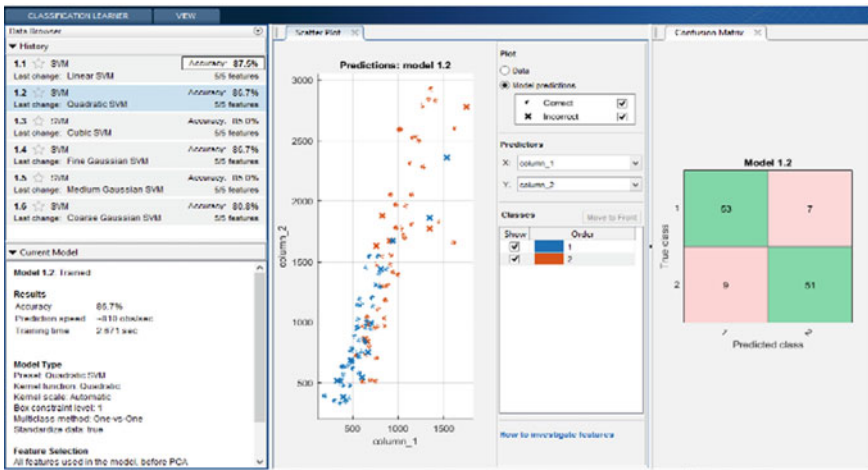
Fig. 2 Quadratic SVM after normalization
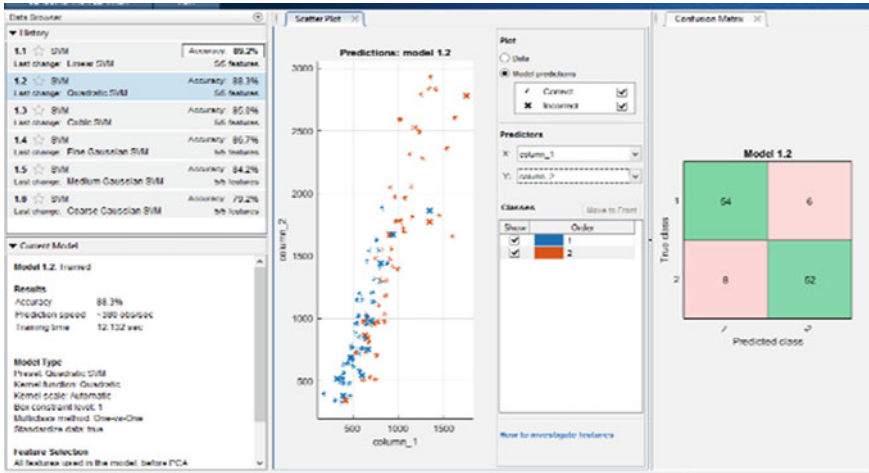
Fig. 3 Quadratic SVM before normalization
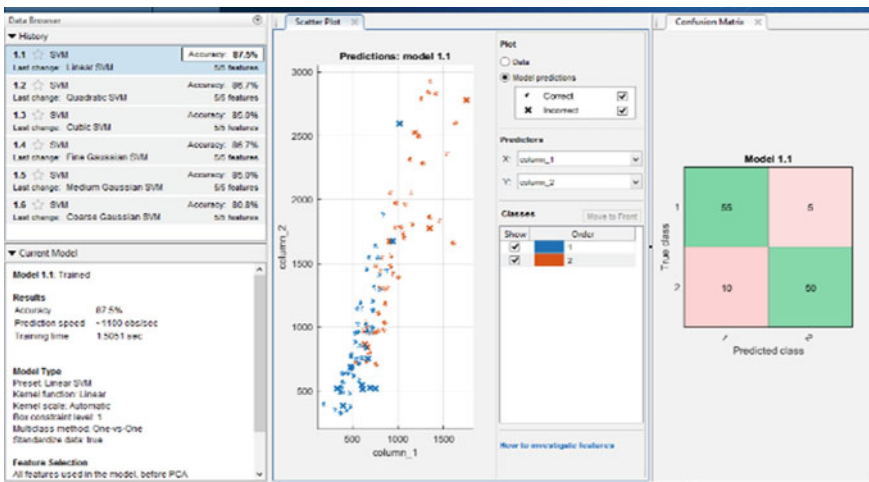
Fig. 4 Linear SVM after normalization

From the 120 images it is found that before normalization, the accuracy is 89.2%, which is greater when compared to after normalization. Because during normalization the raw data of the image may be missed. For less number of images, the parameters used for feature extraction was enough and from this it is proven that linear classifier gives the good result.
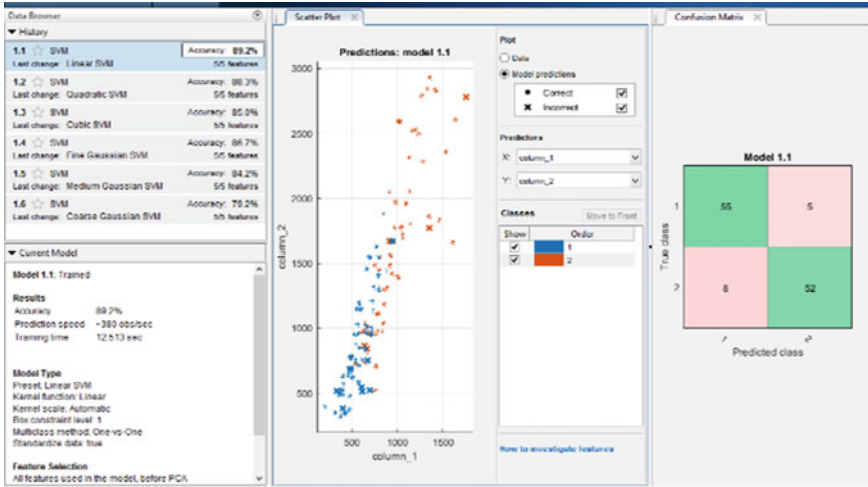
Fig. 5 Linear SVM before normalization
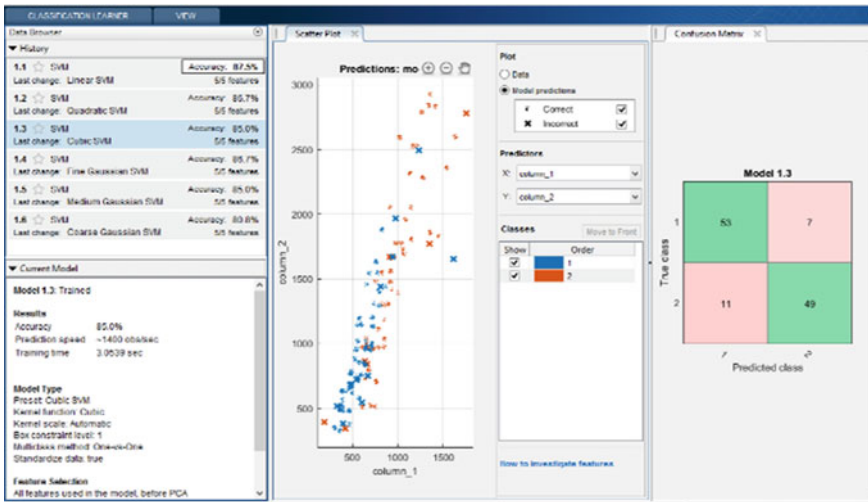
Fig. 6 Cubic SVM after normalization

5.6 Result of Training the Dataset for 336 Images

See Figs. 8, 9, 10, 11, 12 and 13; Table 2.

For more no of images, the accuracy after normalization and before normalization does not vary much, but comparing to 120 images, the accuracy obtained in this is less. Because for higher no of images the parameters used such as mean, standard
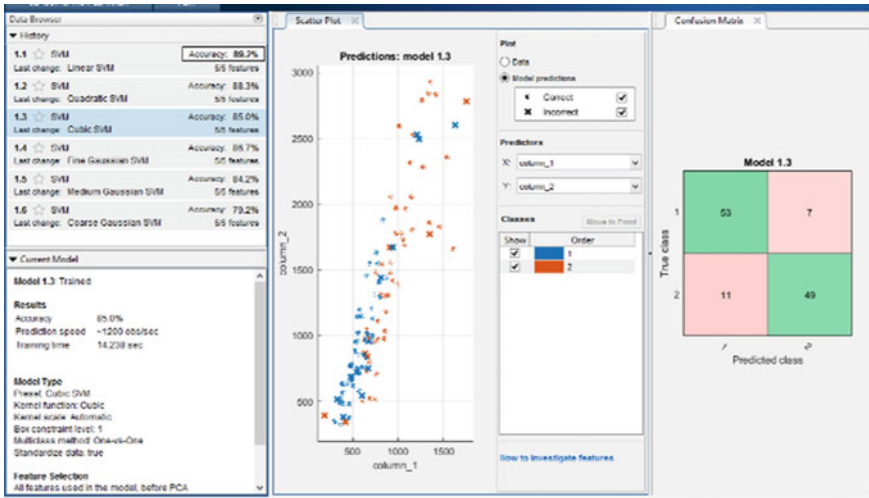
Fig. 7 Cubic SVM before normalization

Table 1 Accuracy obtained for 120 images

Classifier name	Accuracy % (after norm)	Accuracy % (before norm)
Linear SVM	87.5	89.2
Quadratic SVM	86.5	88.3
Cubic SVM	85	85
Fine Gaussian SVM	86.7	86.7
Medium Gaussian SVM	85	84.2

deviation, energy, minimum value, maximum value, variance for feature extraction is not enough. The accuracy always dependent on the feature extracted.

6 Classifier Result

Once the extracted features are fed into classifier, it is trained and the result is generated as a code in the workspace. From the generated code, the Validation accuracy and validation prediction of the images are obtained. For testing the obtained program, more than 10 images are labeled again and feature is extracted from it and fed into the code generated for the result Fig. 14.
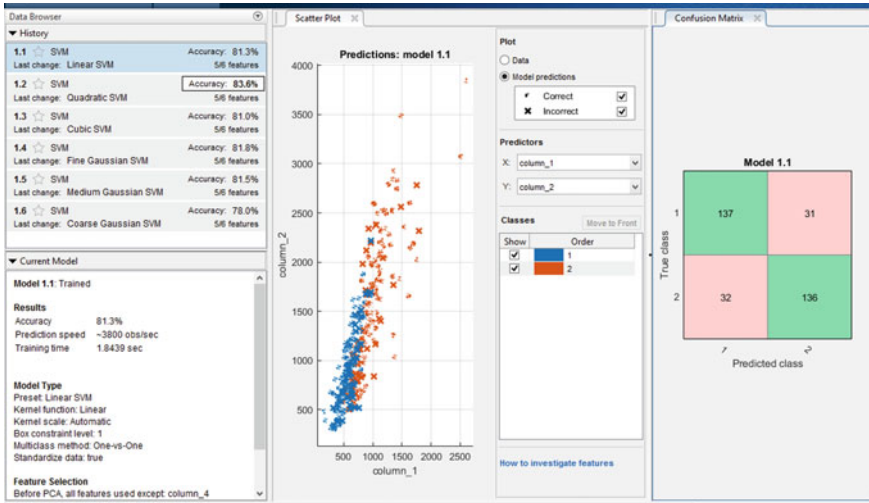
Fig. 8 Linear SVM before normalization
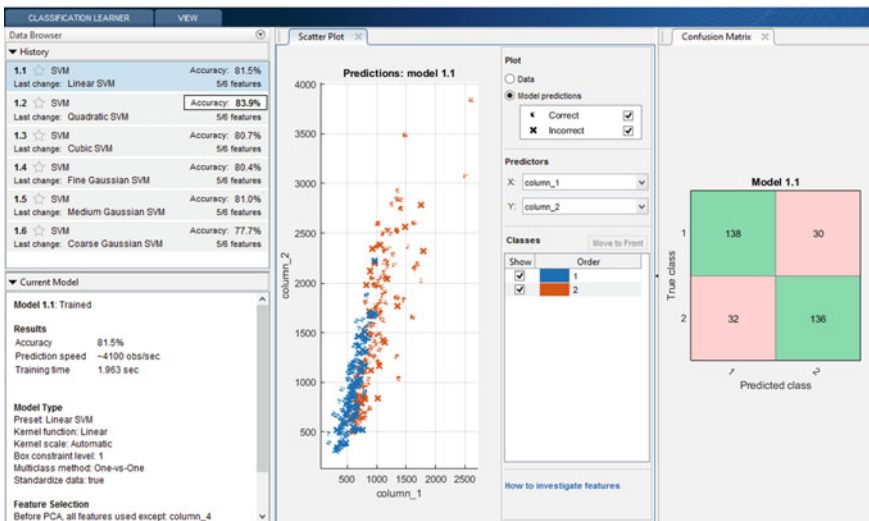
Fig. 9 Linear SVM after normalization

From the figure the result is obtained. The accuracy for classification is 89.2% as the code generated is by the linear classifier at accuracy 89.2%. 8 images for each class was fed., Totally 16 images were fed into the generated code and the validation accuracy is 0.8750 and for class 1, only 1 image was predicted wrong and for class 2, 1 images where predicted wrong.
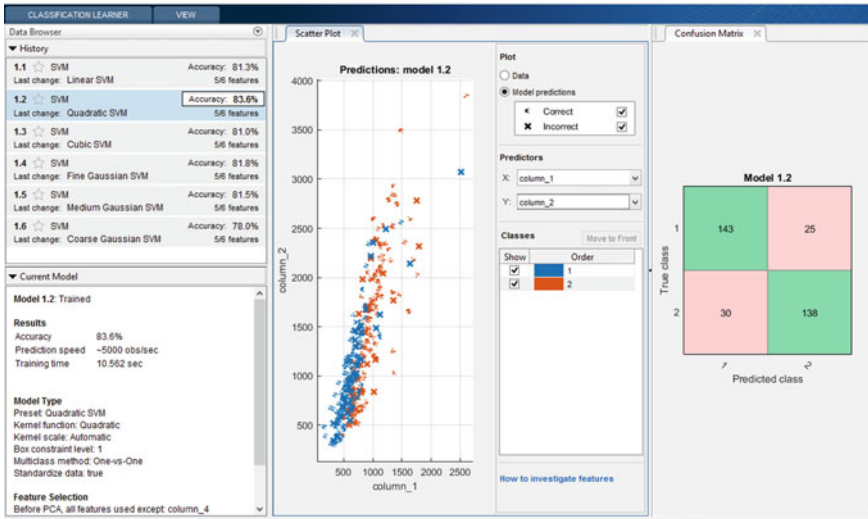
Fig. 10 Quadratic SVM before normalization
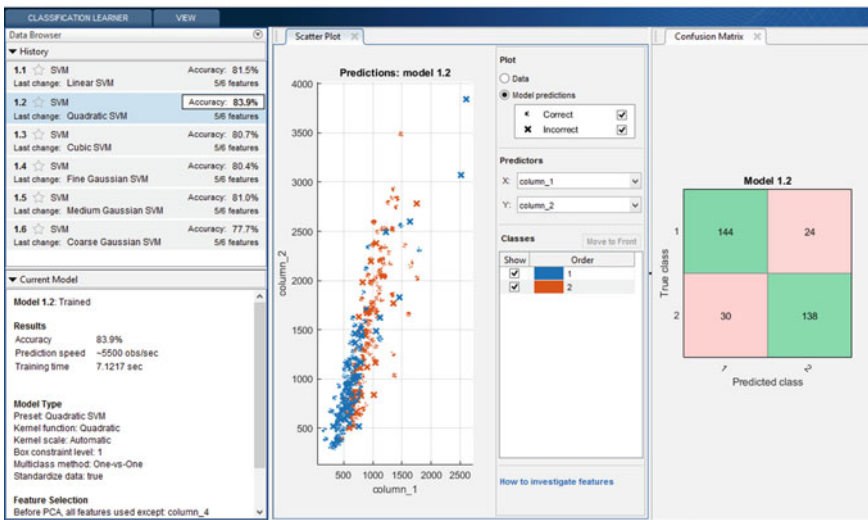
Fig. 11 Quadratic SVM after normalization

7 Conclusions

In conclusion, the bone age reflects the skeletal maturity when differing significantly from chronological age. There were a lot of method to automate the analysis of the bone age that will be time consume for the radiologist in analysis the radiographs of
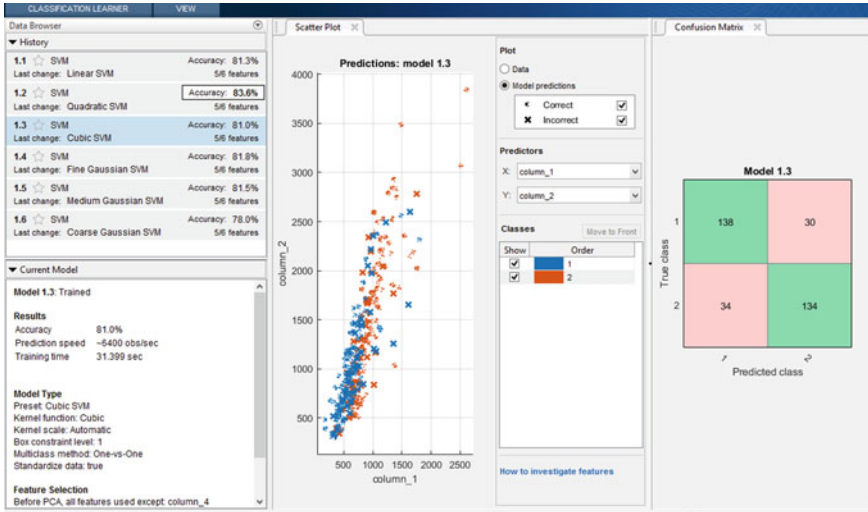
Fig. 12 Cubic SVM before normalization
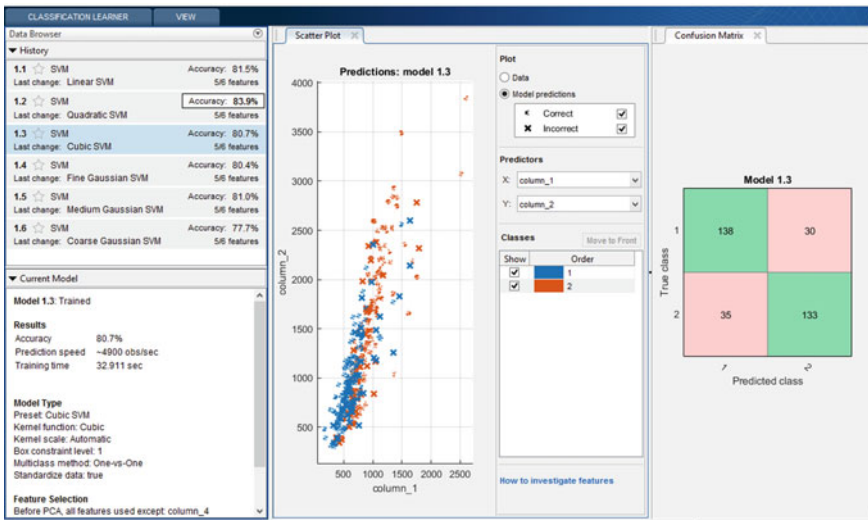
Fig. 13 Cubic SVM after normalization

bone age but in this analysis, the reliable feature extraction is the most important key point for analyze the development of the bone age. However, in the implementation of the image processing for feature extraction as inspected ROI image have been done as the proposed methodology. The method contains the detection of ROI using

Table 2 Accuracy obtained for 336 images

Classifier name	Accuracy % (after norm)	Accuracy % (before norm)
Linear SVM	81.5	81.3
Quadratic SVM	83.5	83.6
Cubic SVM	80.7	81
Fine Gaussian SVM	80.4	81.8
Medium Gaussian SVM	81.0	81.5
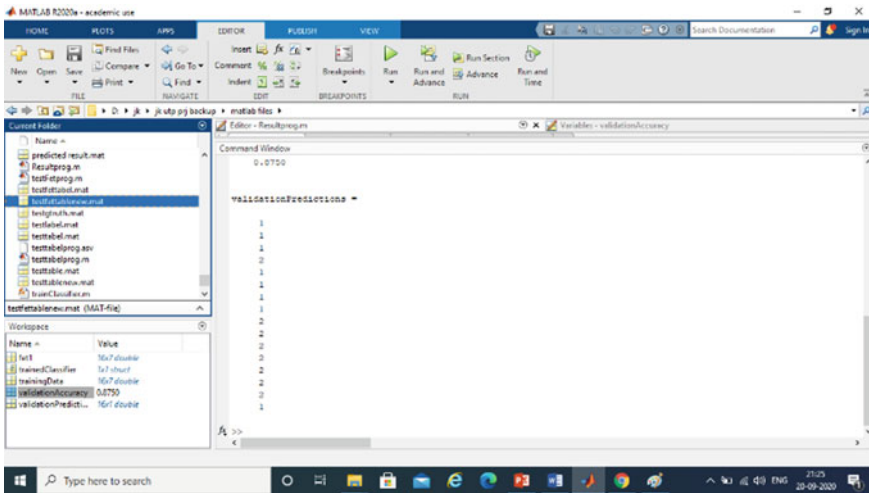
Fig. 14 Classified result of the given test images

parameters like mean, variance, energy and then the images are trained in the SVM classifier.

References

1. Lee H, Tajmir S, Lee J, Zissen M, Yeshiwas BA, Alkasab TK, Choy G (2017) Fully automated deep learning system for bone age assessment. Springer, Aug 2017
2. Vicente G, Osman R, Hand bone age. Springer
3. Mughal AM, Hassan N, Ahmed A (2014) Bone age assessment methods. PMC, Jan–Feb 2014
4. Harmsen M, Fischer B, Schramm H, Seidl T, Deserno TM (2013) Support vector machine classification based on correlation prototypes applied to bone age assessment. IEEE J Biomed Health Inform 17(1):190–197
5. Huo Z, Jyang CY, Purdum M, Bone Mineral density assessment mammography system. United States patent patent no.: us 7.965,813 b2 hu0 et al
6. Bakthula R, Agarwal S, Automated human bone age assessment using image processing methods. Int J Comput

Multi-Class SVM Prediction Model for Lung Cancer Diagnosis

D. Lakshmi, J. Sivakumar, and S. Ramani

Abstract Detection of lung abnormalities is essential in the field of medical diagnostics in order to reduce mortality. Early and correct diagnosis of these lung abnormalities leads to timely and appropriate treatment thereby reducing the mortality. This area of research always faces a challenging task to differentiate the cancerous tissues from non-cancerous tissues using CT images. This paper presents the SVM prediction model for the characterization of lung tissues namely fibrosis, suspicious of TB and carcinoma. This model is designed with three set of features formed with Gray-Level Co-Occurrence Matrix (GLCM) and Gray-Level Run- Length Matrices (GLRLM) and is evaluated using True Positive Rate, False Negative Rate, Positive Predictive Value, False Discovery Rate. The classifier model shows trustworthy results with classification accuracy ranging between 93.9% and 98.6% respectively.

1 Introduction

In India, Non-Communicable diseases lead to mortality rate of 9% and 11% due to cancer and chronic respiratory issues [1]. General clinical practice includes various non-invasive techniques in the management and diagnosis of lung abnormalities. These techniques offer great potential in the disease prediction and treatment. Developing countries conduct the lung cancer screening with a higher risk of diagnostic procedures even for benign tissues and extreme treatment for over diagnosed cancers [2]. Recently, integrated back scattering technique in endobronchial ultrasonography has been used in the lung cancer diagnosis [3]. Bayesian Network (BN) predicts better than SVM for two-year survival in non-small cell lung cancer patients with missing variables. The logistic regression model to the BN has an Area Under Curve

D. Lakshmi (✉) · J. Sivakumar
Department of Electronics and Communication Engineering, St. Joseph's College of Engineering, OMR, Chennai 600119, India

S. Ramani
Department of Computer Science and Engineering, Panimalar Engineering College, Poonamallee, Chennai 600123, India

(AUC) of 0.71, 0.75 and 0.76 for the Ghent, Toronto and Leuven datasets respectively [4]. Selection of textural features is another major challenge in the characterization of lung tissues. Interobserver variability in texture analysis including lesion morphology, image acquisition parameters, segmentation parameters, the number of texture features and the category must be dealt that result in satisfactory outcome in a clinical trial [5–9]. The role of non-linear classifiers in medical imaging diagnostics is significant in the non-invasive characterization of lung abnormalities. Multiple kernels based SVM has nowadays become effective in the characterization and staging of lung diseases from CT images.

2 Literature Survey

Cancer is a multifaceted disease and many automated cancer detection systems are proposed in literature. A predominant research work is based on classification of later stages of lung tumors commonly termed as malignant. Early detection of cancer also known as prognosis is not straight- forward as information about cancer biomarkers are not known clearly. Detection of cancer from CT images involves feature extraction and classification using machine learning algorithms such as ANN, Decision Tree, Linear Discriminant Analysis, SVM, Nearest Neighbor classifier and Bayes Classifier. Recent literature reports that SVM classifier outperforms other conventional classifier systems in lung cancer detection. SVM performs a two-class classification by transforming the input data to higher dimensional feature space through a linear or a non-linear mapping function. It identifies a decision boundary known as hyperplane that maximizes the separation between the input variables in feature space. Keshani et al. [10] proposed a SVM based lung nodule detection system from CT images. The classifier was trained on 2D stochastic and 3D anatomical features to detect the nodule areas. They reported an overall detection accuracy of 89% and false discovery rate of 7.3% per scan. Thomas et al. [11] proposed a data fusion technique to combine the genomic and proteomic data for breast cancer detection. They proposed a weighted average Least Squares SVM (LS-SVM) classifier for cancer classification. They report classification performance of various conventional classifiers based on ROC analysis. Shen et al. [12] proposed a bidirectional chain coding scheme and used SVM classifier to segment the lung nodules from CT images. They reported a segmentation accuracy of 92.6% with the proposed scheme. Huang et al. [13] compared the classification performance of standalone SVM and SVM ensemble classifier in breast cancer analysis. They reported that SVM ensemble classifier based on RBF kernel function outperformed other classifi-ers used in the analysis. Froz et al. [14] proposed a combination of textural features based on artificial crawlers and used SVM classifier for differentiating lung nodules from non-nodules of CT images. They reported a mean accuracy of 94.3% and AUC value of ROC curve of 0.922 for RBF kernel SVM classifier. A study done by Wang et al. [15] compared several classical machine learning methods like Random Forest(RF), Support Vector Machine(SVM), Adaptive Boosting (Ada-Boost), Back Propagation Artificial Neural Network(BP-ANN)

with Convolution Neural Network(CNN) architectures for the classification of lung nodules from metastasis of Non-Small Cell Lung Cancer(NSCLC). The effect of textural features over diagnostic features in the discrimination of these lung nodules is compared and the results are very attractive. Orozco et al. [16] presented RBF kernel based SVM classifier utilizing wavelet features for the clinical data taken from LIDC and ELCAP data base and obtained an accuracy ranging between 81 and 100. Makaju et al. [17] proposed SVM model to classify the lung nodules into benign or malignant with a maximum accuracy of 86.6%. Hassen et al. [18] experimented stepwise Forward Selection and Principal Component Analysis (PCA) based SVM for the classification of lung diseases. Hybridizing PCA, FCM and SVM improves the classification accuracy. Ajin and Mredhula [19] analyzed the behavior of combining Radial Basis Function and Polynomial Function in the classification of different Interstitial Lung Disease (ILD) patterns. Combining these two functions referred as Hybrid kernel based SVM is able to differentiate these ILD patterns with accuracy better than the ANN, K- Nearest Neighbor and Deep CNN. Optimizing the SVM kernel function may greatly reduce the prediction time and [20] obtains the classification accuracy of 99.17% for 70 number of nodules. A retrospective study done by Dennie et al. [21] evaluated 55 patients differentiating primary lung cancer with granulotamous nodules and results in an accuracy of 64.8%. Mohana et al. [22] proposed a method based on Nearest Neighbour for the detection of lung cancer. GLCM features are extracted and using these features the cells are classified as benign and malignant. Nearest Neighbour classifier achieved an accuracy of 98.76%. Senthil and Shubha [23] implemented a method based on Back Propagation Network for lung cancer detection. Seven features are selected using Principal Component Analysis (PCA). The parameters and weights for BPN are optimized using Ant Lion Optimization (ALO). Their experimental results shows an accuracy of 87.1%, 93.2% and 99.1% in Stage I, Stage II and Stage III respectively.

3 Proposed System

3.1 Subject Selection and Dataset

Our Cohort includes 17 subjects of multi-slice chest Low-Dose CT images. The specification of CT images includes tube voltage of 120kV, tube current ranging between 120 mA and 220 mA and slice thickness between 0.3 mm and 0.7 mm and is stored in Digital Imaging and Communications in Medicine (DICOM) format. Dataset comprises of individual slices of each patient expressed as 512 x 512 matrix with a 8- bit gray-level stored in Bit-Map format. The number of slices representing healthy person is 123, fibrosis is 103, carcinoma is 185 and the suspicious of tuberculosis case is 50. The other cases include fibrosis and the suspicious of tuberculosis cases. This work uses histopathological test as the ground truth for categorizing lung tissues. This paper proposes SVM prediction model for the differentiation of various

lung tissues namely suspicious of TB, fibrosis and carcinoma. The work aims at the implementation of highly accurate SVM prediction model with very less prediction time.

3.2 Methodology

As shown in Fig. 1 our classification procedure begins with unsupervised segmentation of lung tissues from 440 number of low dose CT images comprehensively. GLCM involves extracting first order features specifically contrast, correlation, homogeneity and energy. These features are computed along the four directions given by angle θ resulting in 16 number of features. Entropy is also calculated for the tissues inside the lung parenchyma and outside the lung parenchyma defining model with 18 features.

Similarly, 11 GLRLM features are computed across 4 four directions defining model with 44 features [24]. The features extracted are subsequently trained and validated with the different SVM kernels namely linear, quadratic, cubic and gaussian for characterization of the lung cancer with respect to other abnormalities as shown
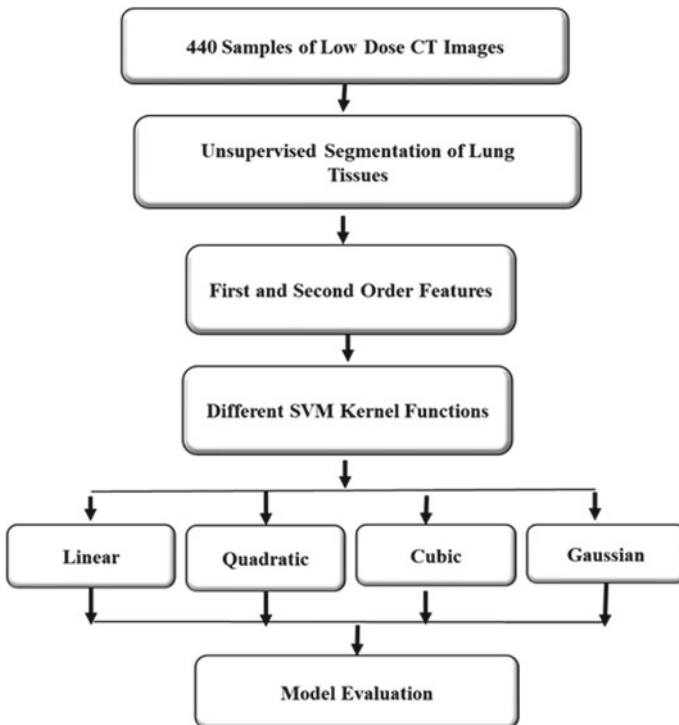
Fig. 1 Methodology

in Fig. 1. The strengths of our proposed work are 1. Processes the low-dose CT image comprehensively. 2. Prevents human intervention during diagnosis. 3. Prior knowledge of ROI is not required.

4 Experimental Results and Discussion

Prediction of SVM model is evaluated in terms of accuracy and prediction time with respect to kernels and type of feature set. Many researchers use SVM for binary classification problem and shown satisfactory performance. But, the use of binary SVM is restricted for multi-class problems. Our three class SVM classifier model performance with the three feature sets namely GLCM, GLRLM, both GLCM and GLRLM are evaluated by the Confusion matrix as in Tables 1, 2 and 3 respectively. It is well observed that the feature set type and the size contribute the classifier performance to a greater extent. Classifier Model with 62 features reduces misclassification of carcinoma and other cases significantly.

4.1 Performance of SVM Classifier Model Based on Features

Our proposed system implements three class SVM prediction model trained with GLCM based features, GLRLM based features and both GLCM and GLRLM based

Table 1 Confusion matrix for model with 18 features

Actual/predicted	Healthy	Other cases	Carcinoma
Healthy	113	–	–
Other cases	–	127	15
Carcinoma	–	7	178

Table 2 Confusion matrix for model with 44 features

Actual/predicted	Healthy	Other cases	Carcinoma
Healthy	113	–	–
Other cases	–	135	7
Carcinoma	–	14	171

Table 3 Confusion matrix for model with 62 features

Actual/predicted	Healthy	Other cases	Carcinoma
Healthy	113	–	–
Other cases	–	137	5
Carcinoma	–	5	180

features. Table 4 shows the classifier performance parameters namely true positive rate, positive predictive value, false negative rate and false discovery rate respectively. Table 5 presents the accuracy and prediction time of SVM classifier with gray-level co-occurrence matrix parameters and gray level run length matrix with respect to different kernels like linear, quadratic, cubic and Gaussian kernels respectively. Classifier accuracy is influenced by the selection of the different feature set numbering 18, 44 and 62 respectively. Moreover, the maximum accuracy is obtained for quadratic and cubic kernels in both the glcm and glrlm based feature set. The model categorizes the three cases with an accuracy of 95, 95.2 and 97.7% respectively.

4.2 Prediction Time for Different Kernels

Our experimental results clearly demonstrate that the proper selection of the information data from the training dataset and exploiting that information in the prediction function optimization reduces the prediction time. Figure 2 shows Prediction time of SVM with different kernel function gives interesting results that model with glcm features consumes maximum time of 7.65s compared to 1.05s for model with glrlm features. Figure 3 gives the Quadratic, Cubic and Gaussian kernel functions with GLRLM based features able to predict with a less timing of 1.11s, 1.069s and 1.246s compared with GLCM based features of 1.269s, 1.49s and 1.167s respectively. Combining both GLCM and GLRLM features shown in Fig. 4 does not yield significant change on prediction time. Finally, SVM prediction model with Quadratic and Cubic Kernels proves that GLRLM features is efficient in the characterization of the lung tissues namely fibrosis, suspicious of TB and carcinoma resulting in an accuracy of 98.6% each as shown in Fig. 5.

4.3 Multi-class SVM Classifier Model Prediction Accuracy

The classifier models are usually evaluated by their prediction accuracy. Our three class SVM classifier model is assessed by its prediction accuracy and is also compared with all the binary class SVM state of the art as shown in Table 6. Many researchers implemented Binary class SVM classifier model with GLCM and GLRLM features yielding less accuracy. Our proposed three-class SVM Classifier model shows significantly high value of prediction accuracy.

5 Conclusion

We developed three class SVM prediction model with 5 fold cross-validation for the differentiation of lung cancer from other lung tissues like suspicious of TB and

Table 4 SVM classifier performance parameters with three feature set

Performance Parameters	True positive rate			Positive predictive value			False negative rate			False discovery rate		
	Healthy Cases (in %)	Other Cases (in%)	Carcinoma Cases (in %)	Healthy Cases (in%)	Other Cases (in %)	Carcinoma Cases (in %)	Healthy Cases (in %)	Other Cases (in %)	Carcinoma Cases (in %)	Healthy Cases (in %)	Other Cases (in %)	Carcinoma Cases (in %)
Model with 18 features	100	89	96	100	95	92	–	11	4	–	5	8
Model with 44 features	100	95	92	100	91	96	–	5	8	–	9	4
Model with 62 features	100	96	97	100	96	97	–	4	3	–	4	3

Table 5 SVM prediction response for GLCM based features with respect to kernel function

Kernel function	GLCM		GLRLM		Both GLCM and GLRLM	
	Accuracy (in %)	Prediction time (insec)	Accuracy (in %)	Prediction Time (in sec)	Accurac y (in %)	Prediction time (in sec)
Linear	95	7.614	95.9	1.0518	98.2	7.2094
Quadratic	96.4	1.2694	98.6	1.1142	97.7	1.363
Cubic	96.4	1.492	98.6	1.0694	97.7	1.1665
Gaussia n	94.8	1.167	93.9	1.2463	94.3	1.313

Fig. 2 Prediction time of SVM with glcm features for different kernel functions
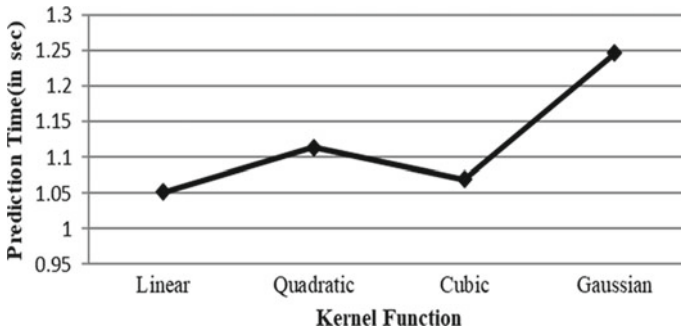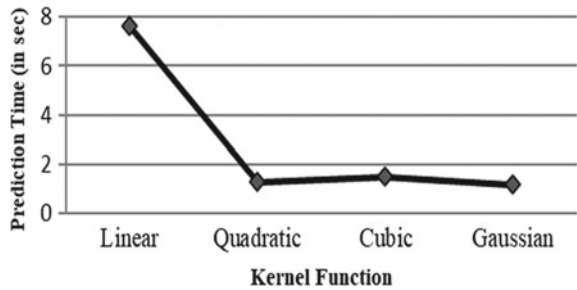
Fig. 3 Prediction time of SVM with glrlm features for different kernel functions

Fig. 4 Prediction time of SVM with different kernel functions for both glcm and glrlm features
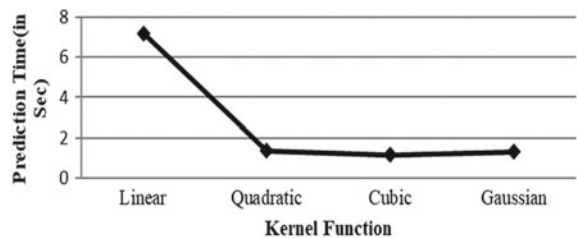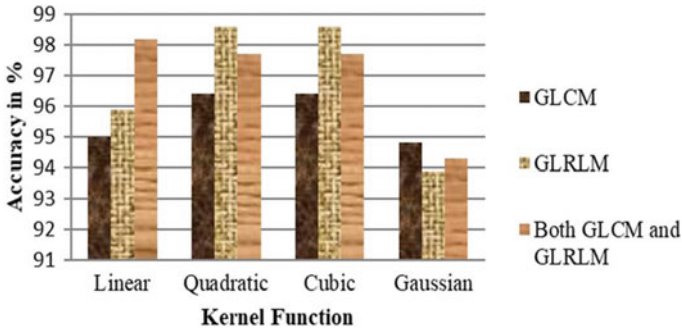
Fig. 5 Variation in accuracy of SVM Prediction model with three datasets

Table 6 Comparison of performance of SVM classifiers for different features

Author	Features	Images	Classifier/class	Accuracy (in %)
Hongkai Wang al et (2017)	GLCM based features	PET/CT	SVM/binary	69.23 ± 8.57
Hiram Madero Orozco et al. (2015)	Wavelet features	CT	SVM-RBF/binary	81–100
Suren Makaju et al. (2018)	Area, perimeter, centroid, diameter, eccentricity, mean intensity	CT	SVM/binary	86.6
Donia Ben Hassen et al. (2016)	Circularity, skewness, kurtosis, entropy, correlation, homogeneity, lacunarity, localisation	Chest Radiographs	Hybrid SVM/binary Fuzzy	74.19
Ajin M and Mredhula L (2017)	Linear ternary co-ocurrence pattern	CT	Hybrid kernel based SVM/ binary (RBF and Polynomial function	90.52
Tao Zhou et al. (2016)	Shape features, intensity features, texture features	CT	SVM/binary	99.17
Carole Dennie et al. (2015)	Correlation, sum of squares, sum variance, sum entropy, difference Variance	Contrast Enhanced- CT	SVM/binary	60

Table 6 (continued)

Author	Features	Images	Classifier/class	Accuracy (in %)
Our proposed work	GLCM based features	CT	SVM/three class – linear	95.0
			– Quadratic	96.4
			– Cubic	96.4
	GLRLM based features		– Gaussian	94.8
			SVM/three class – Linear	95.9
			– Quadratic	**98.6**
			– Cubic	**98.6**
			– Gaussian	93.9
	Both GLCM features and GLRLM based features		SVM/three class – Linear	**98.2**
			– Quadratic	97.7
			– Cubic	97.7
			– Gaussian	94.3

fibrosis. It is concluded that quadratic and cubic kernels for GLRLM based features yield a maximum accuracy of 98.6% with prediction time of 1.23sec. Combining these features and predicting with linear kernel increases the accuracy to 98.2%. The main advantage of our Classifier model requires no prior knowledge of ROI for segmentation, where the system process the low-dose CT comprehensively. The work may be enhanced for the detection of metastasis of the lung abnormalities by including dataset contributing images of different abnormalities.

References

1. http://www.who.int/nmh/countries/ind_en.pdf?ua=1
2. Veronesi G, Novellis P, Voulaz E, Alloisio M (2016) Early detection and early treatment of lung cancer: risks and benefits. J Thorac Dis 8(9):E1060–E1062
3. Ito F, Kawasaki M, Ohno Y et al (2016) Noninvasive tissue characterization of lung tumors using integrated backscatter intravascular ultrasound. An Ex Vivo Comp Study Pathol Diagnos Chest 149(5):1276–1284
4. Hatt M, Majdoub M, Vallières M et al (2015) F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor, volume in a multi-cancer site patient cohort. J Nucln Med 56:38–44
5. Jayasurya K, Fung G, Yu S, Dehing-Oberije C, Ruysscher D, Hope D (2010) A Comparison of Bayesian network and support vector machine models for two-year survival prediction in lung cancer patients treated with radiotherapy. Med Phys 37(4)
6. Fried DV, Tucker SL, Zhou S, Liao Z, Mawlawi O, Ibbot G, Court LE (2014) Prognostic value and reproducibility of pretreatment CT texture features in stage III non–small cell lung cancer. Int J Radiat Oncol Biol Phys 90:834–842

7. Doumou G, Siddique M, Tsoumpas C, Goh V, Cook GJ (2015) The precision of textural analysis in 18 F-FDG-PET scans of oesophageal cancer. EurRadiol 25:2805–2812

8. Tixier F, Hatt M, Le Rest CC, Le Pogam A, Corcos L, Visvikis D (2012) Reproducibility of tumor up take Heterogeneity characterization through textural feature analysis in 18 F- FDG PET. J Nucl Med 53:693–700

9. Nyflot MJ, Yang F, Byrd D, Bowen SR, Sandison GA, Kinahan PE (2015) TU-AB- BRA-04: quantitative radiomics:Sensitivity of PET textural features to image acquisition and reconstruction parameters implies the need for standards". Med Phys 42:3587

10. Keshani M, Azimifar Z, Tajeripour F, Boostani R (2013) Lung nodule segmentation and recognition using SVM classifier and active contour modeling. A complete intelligent system, Comput Biology Med 43:287–300

11. Thomas M, De Brabanter K, Suykens JAK, De Moor B (2014) Predicting breast cancer using an expression values weighted clinical classifier. BMC Bioinformatics 15. https://doi.org/10.1186/s12859 014- 0411–1

12. Shen S, Bui AAT, Cong J, Hsu W (2015) An automated lung segmentation approach using bidirectional chain codes to improve nodule detection accuracy. Comput Biology Med 57:139–4. https://doi.org/10.1016/j.compbiomed.2014.12.008

13. Huang MW, Chen CW, Lin WC, Ke SW, Tsai CF (2017) SVM and SVM ensembles in breast cancer Prediction. PLoS ONE 2017;12:e0161501. https://doi.org/10.1371/journal.pone.0161501

14. Froz BR, de Carvalho Filho AO, Silva AC, de Paiva AC, Nunes RA, Gattass M (2017) Lung nodule classificatio using artificial crawlers, directional texture and support vector machine. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2016.10.039

15. Wang H, Zhou Z, Li Y, Chen Z, Lu P, Ang E, Wanyu Liu W (2017) Comparison of machine Learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from F-FDG PET/CT images. EJNMMI Res 7:11.https://doi.org/10.1186/s13550-017-0260-9

16. Orozco HM, Villegas OOV, Sánchez VGC, Domínguezz HDJO. Alfaro (2015) Automated system for lung nodules classification based on wavelet feature descriptor and support vector machine, BioMedical Eng OnLine 14:9.https://doi.org/10.1186/s12938-015-0003-y

17. Makaju S, Prasad PWC, Alsadoon A, Singh AK, Elchouemi A (2018) Lung cancer detection using CT scan images. Procedia Comput Sci 125:107–114

18. Hassen DB, Zakour SB, Taleb H (2016) A hybrid fuzzy-SVM classifier for automated lung diseases diagnosis. Polish J Med Phys Eng 22(4):97–103.https://doi.org/10.1515/pjmpe-2016-0017

19. Ajin M, Mredhula L (2017) Diagnosis of interstitial lung disease by pattern classification. In: 7th International conference on advanc in computing & communications, ICACC -2017, 22–24 August 2017, Cochin, India, Procedia Computer Science vol 115, pp 195–208

20. Zhou T, Lu H, Zhang J, Shi H (2016) Pulmonary nodule detection model based on SVM and CT image feature-level fusion with rough sets. Hindawi Publishing Corporation BioMed Research International, vol 2016, Articl ID 8052436, 13 p. https://doi.org/10.1155/2016/8052436

21. Dennie C, Thornhill R, Carolina V-V, Souza A, Gupta A, Maziak D, Bayanati H (2016) Role of quantitative computed tomography texture analysis in the differentiation of primary lung cancer and granulomatous nodules. Quant Imaging Med Surg 6(1):6–15

22. Mohana RM, Devi RDH, Bai A (2019) Lung cancer detection using nearest neighbour classifier. Int J Recent Technol Eng (IJRTE) 8(2S11). ISSN: 2277-3878

23. Senthil S, Shubha BA (2019) Improving the performance of lung cancer detection at earlier stage and prediction of reoccurrence using the neural networks and ant lion optimizer. Int J Recent Technol Eng (IJRTE) 8(2). ISSN: 2277–3878

24. Lakshmi D, Santosham R, Ranganathan H (2014) Automated texture based characterization of fibrosis and carcinoma using low-dose lung CT images. Int J Imaging Syst and Technol 24(1):39-44. Online ISSN: 1098-1098.https://doi.org/10.1002/ima.22077

Survey on Fire Safety Robot & Implementation of Android Application

K. Sakthisudhan, S. Rubika, S. Sadhasivam, V. Suguna, and Sri Muruga Rajan

Abstract Detection and extinguishing of fire are dangerous job that puts the life of fire fighters in danger. Using robots to do these things will reduce the loss of lives. The roles of current systems are limited in certain ways so it can be improvised to a quicker and effective method. This paper offers an overview of a fire fighting robot with toxic gas calculation. Several gas sensors are used to determine the volume of poisonous gases such a carbon monoxide, carbon dioxide and ammonia. The fire is detected using a flame sensor and the information is processed by the processor and the procedure to extinguish it is initiated. Data and informatics is the essential factors for effective firefighting operation.

Keywords Fire-fighting · Android application · Carbon dioxide · Carbon monoxide and ammonia

1 Introduction

Fire battle is a dangerous job, but firefighters are still putting their lives on line against fire. The mortality rate of firefighters in service is high. The temperature steadily increases when a fire threat arises. The harmful gases such as carbon dioxide and carbon monoxide appear to impact humans. Fire fighters can also be replaced by robots that can do the same thing safely. This article reviews and describes problems and criteria for robot-based firefighting and presents prospects of productive fighting in the future. The programmed robot locates the fire and automatically extinguishes it. If the fire is not detected, the mobile application may monitor it manually. With mobile device, the robot can be run easily. It tests the safe concentrations of poisonous gases such as carbon dioxide, carbon monoxide and ammonia in the burn. This helps us to consider the strength of the fire and the wellbeing of victims. A mobile network for autonomous firefighting can track and manage a particular area for the occurrence of fire, and locate the exact place of the fire, extinguish the fire and

K. Sakthisudhan (✉) · S. Rubika · S. Sadhasivam · V. Suguna · S. M. Rajan
Dr.N.G.P. Institute of Technology, Coimbatore 638401, India

return safely after completion of the extinction. The obstacle monitoring is carried out using A / D conversion in the microcontroller to convert analogue output from the ultrasonic sensor to digital, and it is compared with the pre-specified threshold value for the AFFMP to decide whether there are any obstacles too close [1]. The Snake Fighter Anna Konda can drive the fire away from outer obstacles apart from a flat floor, and can use the nozzle that is fixed to the front with a hydraulic medium in joint action to extinguish the fire. It is a combined use of water to power, cool and extinguish hydraulic joints [2]. Robots will reach the area of the enemy and submit information via wireless camera for night vision. The fire sensor senses the fire inside the warfare and sends the fire extinguisher information, so the water pump can be triggered whenever possible. An ultrasonic sensor is used to detect all obstacles and to detect the distance between the robot and the obstacle. Without causing damage, ultrasonic sensors calculate the distance. When the robot identifies obstacles and responds. The ultrasonic sensor is incorporated such that the entire world is laid out [3]. The conventional robot firefighting IOT based firefighting robot can be replaced. The fire alarm from the cloud is issued to an android application. Then, the automatic receiver in the robot provides a live video of the fire position. The recipient also allows the consumer and fire victims to connect. This helps the people trapped in the position where fire has occurred to get guidance on evacuation routes. By calculating the amount of carbon monoxide within the fire area, the possible health effects can be calculated roughly [4]. A water tank and a wireless communication operated pump are used for pumping water. The robot body moves by the microcontroller's performance. The robot is operated through a Bluetooth app. Remote position IR flame sensor is used. Facilities provide wireless networking. Exactness and performance are maximized after fire detection with a minimum amount of time [5]. The robot is wirelessly operated and remotely controlled. The fire is sensed by a sensor. The robot is loaded with a water tank and all motors are operated by the Microcontroller. The fire and gas sensors are sent to the microcontroller when the fire threshold is over 450C or the gas threshold is over 300 parts per million. The message is received via Bluetooth technology on a mobile device [6]. On its own axis the robot rotates to scan for fire at various angles in the building. If the sensor LM35 detects heat, the robot stops and moves the fire by means of a centrifugal pump that casts water, if the temperature increases. In darker areas, the fire flame is best sensed. It is a robot that prevents fire because it can immediately detect and extinguish [7]. Nodes where fire incidents happen sometimes are located in an indoor environment. A sensor and microcontroller node is included. The information shall be transmitted to the coordinator unit if the node detects fire. The co-coordinating unit is the main unit linking all nodes. The coordinators unit performs two actions, one is to send fire alarm message to the department and the other to send details to a mobile robot consisting of an indoor map. An algorithm for the planning of a star path designs a route from the current position to the fire location, which is called the global path. Mobile robots planner [8] transforms the route information into separate movement instructions. Using its onboard camera the robot will conduct monitoring. The operator of the robot will direct it remotely around the entire site. The extinction of Water and CO_2 is conducted. Nebulous sprays are being used to avoid burns in the

robot. In harmful circumstances it becomes more helpful by bi-directional contact with victims. A Tmega 2560 microcontroller [9] is used to monitor it primly. The robot is generally automatic and can be changed to manual mode by remote control. By using the Bluetooth module, smart phone is connected to the device. The robot holds the fire away safely. The processing unit is Arduino microcontroller. Proteus is used for robot research and simulation. Robot is operated from a remote site based on an RF communication system, to ensure human protection. It is quick and easy to implement the open-loop control technique. Improved output is the closed-loop power. However, due to changes in the parameters [10], closed loop framework can be unstable. For the intelligent burning robot, the search-and-find algorithm is used. The robot began moving in the direction of the fire, when it was moved on. It recognized a room without a fire and eventually discovered the room with a fire and turned to the fire. It went back to where it began until the fire was extinguished. The robot effectively fuses the fire through this algorithm [11]. To monitor the robot and to acquire sensor data, the assembly language software is used. Using the wireless RF controller, the monitoring computer and a remote supervised computer to operate the mobile robot. An adaptive fusion algorithm is used to detect fires [12]. The robot moving vertically is useful in high-level firefighting applications. In a reasonable time, this robot can climb up a house. The handrail is used as a step on a balcony. It can go up from 2.50 [m] to 3.10 [m] [13] to the building with its floor inteval.

2 Materials and Method

The purpose of the work proposed is to develop an Android-operated fire-fighting robot to replace the conventional fire-fighting robot. The IR sensor data is read and the robot moves forward or the robot spins and moves forward if no collision occurs. The signal is sent to the microprocessor when the fire is detected by the fire sensor. The data is interpreted by the microprocessor and then instructed to pump water in the water pump. The pump works with the motor driver to extinguish the fuel. The quantity of harmful gases is calculated during this process and the data is forwarded to the customer. Multiple gas sensors can be used to detect gases such as carbon dioxide, carbon monoxide and ammonia. A fire sensor is used in five channels, as it can simultaneously sense fire in 120° and the fire can be detected more easily. The potential health impacts of the victims can be measured approximately according to the amount of gases and medical assistance is available in advance. A servo motor is an electric device that can precisely push or rotate an object. In compliance with the directions or inputs, the motor driver lets the motor shift. The control / processor listens to the low voltage and regulates an actual engine that needs high input voltages. A 3~6 V power supply will run the Submersible Pump engine. With a very low current consumption of 220 mA, it can take up to 120 L per hour. It is used to put the fire out. A flame detector module of 5 channels is used for detecting the wider area of flame (>120°). Ammonia (NH_3) has a strongly pungent odor, colorless gas or compressed
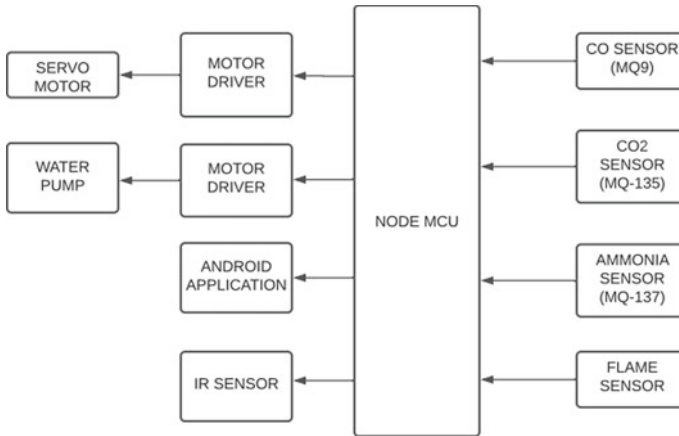
Fig. 1 Firefighting robot functional diagram

liquid. The detector of ammonia is based on an electrical theory. It disseminates ambient air into the liquid electrolyte in the sensor via a membrane (Fig. 1).

A CO_2 sensor or carbon dioxide sensor is a carbon dioxide measuring unit. Infrared gas sensors (NDIR) and chemical gas sensors are the most general concepts for CO_2 sensors. Carbon monoxide is an odorless, colorless and tasteless gas which cannot be detected in the absence of an effective detector. MQ-7 gas sensor sensitive material is SnO_2, which is less clean air conduction. The IR sensor is an electronic system that transmits light to sensing an object of the surroundings, and it is used to detect CO at low temperatures (heated to 1.5 V). An IR-sensor can measure and detect the heat of an object. When there is not an object, there is no signal from the infrared receiver; the infrared receiver receives signals from an object which blocks and reflects infrared light. Node MCU is an open source firmware that provides open source prototyping board designs. The term MCU Node (micro-controller unit) combines "node" and "MCU". The word "Node MCU" does not mean the related development kits but strictly speaking, the firmware. It is cost-effective and can be integrated in support of WIFI. The consumption of energy is poor.

3   Conclusion

This proposed work is aimed at using robots to detect and extinguish fires that are minimally harmful to humans. It also tests the amount of poisonous gases including carbon dioxide, carbohydrate and ammonia using gas sensors, as well as controlling fire. The gas sensor data helps to predict approximately the potential health effects of the injured, to ensure that medical assistance is prepared in advance.

4 Future Work

Water is used to extinct the fire in the proposed work, but the extinct material may be modified by fire type. The water capacity in the water tank can be increased to meet requirements. To find their safe levels, other gas sensors can be installed.

References

1. Khoon TN, Sebastian P, Saman ABS (2012) Autonomous fire fighting mobile platform. In: International symposium on robotics and intelligent sensors 2012 (IRIS 2012). Elseiver
2. Liljeback P, Stavdahl O, Beitnes A (2006) Snake fighter—development of a water hydraulic fire fighting snake robot. In: 9th International conference on control, automation, robotics and vision. ICARCV '06, pp 1–6
3. Kolambe K, Pote R, Jadhav A, Chennur V (2018) Spy robot with fire detection and water sprinkling. In: 2nd International conference on electronics, communication and aerospace technology (ICECA 2018) IEEE conference record # 42487; IEEE Xplore. ISBN:978-1-5386-0965-1
4. Kanwar M, Agilandeeswari L, IOT based fire fighting robot. IEEE
5. Zaman HU, Khan TA, Falgunee SR, Rashid GMS, Talukder FH (2018) Autonomous firefighting robot with optional bluetooth control. In: 2018 4th International conference on computing communication and automation (ICCCA). IEEE
6. Bose JSC, Mehrez M, Badawy AS, Ghribi W, Bangali H, Basha A, Development and designing of fire fighter robotics using cyber security. IEEE
7. Rakib T, Rashid Sarkar MA, Design and fabrication of an autonomous firefighting robot with multi sensor fire detection using PID controller. In: 5th International conference on informatics, electronics and vision. (ICIEV) –IEEE
8. Anantha Raj P, Srivani M (2018) Internet of robotic things based autonomous fire fighting mobile robot. In; 2018 IEEE international conference on computational intelligence and computing research (ICCIC)
9. Mittal S, Rana MK, Bhardwaj M, Mataray M, The fire fighting robot. In: IEEE international conference on advances in computing, communication control and networking
10. Fahim SR, Das SK (2018) High-Performance open-loop tracking control of a small scale firefighting robot. In: IEEE international conference on advancement in electrical and electronic engineering 22–24 Nov 2018
11. Kim JH, Keller B, Lattimer BY (2013) Sensor fusion based seek-and-find fire algorithm for intelligent firefighting robot. In: 2013 IEEE/ASME international conference on advanced intelligent mechatronics (AIM) Wollongong, Australia, 9–12 July 2013
12. Su KL (2006) Automatic fire detection system using adaptive fusion algorithm for firefighting robot. In: 2006 IEEE international conference on systems, Man, and cybernetics Oct 8–11, 2006.Taipei, Taiwan
13. Amano H, Osuka K, Tarn TJ (2001) Development of vertically moving robot with gripping handrails for firefighting. In: Proceedings of the 2001 IEEE/RSI international conference on intelligent robots and systems Mad, Hawaii, USA, Oct. 29 -Nov 03, 2001

Detection of Emergency Vehicles Using Radio Frequency Identification (RFID)

Usha Mittal and Priyanka Chawla

Abstract Road traffic is a major problem in major cities. With the increasing number of vehicles, the resolution of the problem is very crucial. Problems related to traffic not limited to road congestion, but also difficult for the passage of emergency vehicles. In order to build an intelligent traffic signal system, a complete traffic management system must be built that deals with all traffic-related problems. The aim of this article is to propose an efficient system for detecting and prioritising emergency vehicles on the junction. In this paper, emergency vehicles are identified and prioritized using Radiofrequency Identification (RFID). RFID tags serve to identify vehicles in a unique way. In this article, ultra-high-frequency RFID readers are used with a frequency of 865–928 MHz and cover a maximum distance of 9 m.

Keywords Emergency vehicle detection · RFID · Traffic lights · Priority · RF receiver · RF transmitter

1 Introduction

Urbanization, industrialization and population growth have resulted in a considerable increase in vehicular traffic. There is no strong traffic framework on the planet today, an approach to transmit effective traffic frameworks is through the unique control of traffic lights depending on the size of the traffic Furthermore, there are no priority services for priority vehicles such as ambulances, police vans or fire trucks. As a result, certain services other than regular services must be provided to priority vehicles [1].

A traffic signal that runs on timers is the most common type of traffic control. You are supposed to work somehow at a particular time of day, regardless of the traffic. We are not getting rid of unnecessary congestion.

When you are stuck in a traffic jam, you can address the problem of emergency vehicles with the means to clear the lane in which emergency vehicles travel. This

U. Mittal · P. Chawla (✉)
Lovely Professional University Phagwara, Phagwara, India

may be done as an emergency vehicle approaches with the nearest traffic light. During an emergency, the RFID tag acts as a transducer and transmits signals to the lighting system [2].

RFID is a technology that allows the automatic identification of an individual, packet or element with radio signals [3]. It relies on RFID tags to do so. These are small transponders that provide identity information when requested over a short length. The majority of RFID labels have at least two parts. One is a built-in circuit that stores certain information, modules and demodulates the signal for radio frequency and other special functions. The second antenna serves the purpose of receiving and transmitting the signal. There are primarily two types of RFID tags: active RFID tags that contain batteries and passive RFID tags that do not contain batteries [4].

1.1 Different Ranges of RFID

To transmit, RFID labels and readers must be adjusted to a similar frequency. There are various frequencies which may be used by an RFID gadget. Generally speaking, the more widely recognized are.

- Low recurrence, or LF, (125–134 kHz)
- High recurrence, or HF, (13.56 MHz)
- Ultra-high recurrence, or UHF, (433, and 860–960 MHz).

On the various frequencies, the radio waves act in a contrasting manner, so it is essential to choose the adjustment recurrence for the application.

Low-frequency labels, for example, have a frequency and may penetrate delicate metallic surfaces even more appropriately. LF RFID frameworks are likewise ideal for perusing high water content articles, for example, natural product or beverages, however the read go is just centimeters or centimeters. Common place LF RFID applications incorporate creature labeling and get to control.

High-frequency labels work very well for metal items and can be used to handle goods containing medium to high water. HF RFID frames usually operate in inches, but can have a greater reading range of about three feet (1 m). Following library books, following the patient stream, and travel tickets are the average HF RFID applications.

Generally, UHF frequencies provide essentially a better understanding of the extent and can move information quickly (e.g., read many more labels every second) than lower or high recurrence frequencies (inches to 50 + ft. as per the RFID framework arrange- ment). Since, in all cases, the frequency of UHF radio waves is shorter, their signal would be lowered (or weakened) and would not be communicated via metal or water. Other basic uses of UHF RFID integrate the electronic assortment of costs and the control of discontinuation of access because of the more stretched scope of understanding.

2　Literature Survey

Because of wasteful traffic frameworks, fuel time and cash are incredibly lost, it some of the time prompts lost life. So a great deal of work is being done to deal with those issues. In 2009, the IEEE issued a smart traffic light framework that uses RFID to filter the vehicle through the vehicle's RFID label and collect its EPC data [1]. The traffic volume, vehicle need and street and time are determined by sort of vehicle. A guideline calculation that suggests how the traffic signals are worked is given in the Decision Making Section (DMS).

In 2009 [5], the developer proposed a technique for the use of the system's sensors as well as the innovation implemented. Accordingly the exchanging of traffic signal builds traffic and street limit and may prevent traffic jams [6].

In 2014 [7], a web-based traffic management system that could be managed remotely and monitored simultaneously at different crossings was built using new web server technologies. The primary node is the Central Traffic Management Unit (CTMU), which uses Internet technology to remotely track and control individual nodes.

The author used the RFID label in 2015 [8] and consists of three parts: automatic monitoring of signals, clearance of vehicles and detection of stolen vehicles.

In [9], the utilization of RFID exercises controls is talked about, specifically those recognized in the image arrangements and in impedance techniques, to keep up a vital good ways from issues which generally emerge from standard movement control sys- tems. This multi-vehicle, multilateral and multi-track convergence framework is over- seen by RFID. It provides a beneficial time organization plan, in which, a component schedule proves to be consistently for the segment of each movement segment. The current use of the framework emulates the fundamental choice of a functioning cop. The amount and monitoring of vehicles in each segment is the property where estimates and decisions are completed. The disadvantage of this examination is that the techniques used for the emergency unit and the control frame of the moving panels are not attempted.

In 2013, the author [4] proposed a framework for RFID and GPS-enabled crisis vehicles. The focal errand of this work is to lessen the time the crisis vehicle arriving at the recuperating office by normally clearing up the course through which the salvage vehicle continues before the moving signs are reached. This should be possible when a crisis vehicle is at a specific division of its action convergence, turning the action signal in transit toward the green. The use of RFID perceives the hand brake and non-emergency cases, in which unnecessary movement blockage is seen. Correspondence is delivered through the phone and GPS between the crisis vehicle and moving signs. When development converges, the instrument is completely mechanical and does not require human mediation. The restriction of the structure is that it needs all the information about the start, the end of the tour. This that not work if the salvage vehicle needs to follow an alternate course or if the beginning stage is unquestionably not comprehended ahead of time.

Ongoing advances have been utilized to recognize a crisis vehicle nearness and to figure the real traffic dense[10], for example, infrared (IR) and worldwide situating framework (GPS). [11]. So as to recognize the nearness and check vehicles, RFID labels were utilized to distinguish the nearness of crisis vehicles [12]. The most widely recognized techniques for traffic recognition incorporate video, microwave, radar identification, ultrasonic location and so on [9, 13].

For crisis vehicle detection [2, 14] and vehicle tracking [5], RFID innovation has been recommended. So as to control traffic signals [15], RFID and GSM were additionally proposed. Such projects, be that as it may, depend on specifically taking care of various issues.

3 Proposed Methodology

Every vehicle has passive RFID labels with explicit Electronic Product Code (EPC) RFID label number. No outer force source should be connected. The intricacies, such as the vehicle number, the vehicle type and the owner's data are also stored on each RFID tag. The Ultra-High Frequency (UHF) recurrence band recognizes labels in the RFID radio wire. The peruser module peruses this information and it is then sent for additional preparing to the worker. Table 1 shows the details of emergency vehicles taken for experiment. The proposed system includes 2 main units that are linked together, namely: emergency vehicle detection and vehicle priority based on the type of vehicle.

Table 1 Details of emergency vehicles

EPC vehicle number	Vehicle type	Owner's	Details priority
1011 XXX	Ambulance	XXX	1
1012 XXX	Firebrigade	XXX	2
1013 XXX	Ambulance	XXX	1
1015 XXX	Ambulance	XXX	1
1016 XXX	Police	XXX	3
1017 XXX	Firebrigade	XXX	2
1018 XXX	Ambulance	XXX	1
1019 XXX	Police	XXX	3
1020 XXX	Police	XXX	3

3.1 Detection of Emergengy Vehicle

This system is made up of an RFID tag in emergency vehicles installed. During an emergency, it is used as a tracker. This means that the driver activates the RFID tag, which is then detected by RFID readers located a few meters away at the junction, whenever there is an emergency. These readers then continuously transmit the signals to the intersection where the traffic lights are controlled. The reader starts detecting signals and the conditions return to normal as soon as the emergency vehicle is tagged into the junction.

Access control is used to detect IDs entering or leaving the RFID reader area. The signals are passed to the junction unit following identification by RFID. The entire detection unit is shown in the Fig. 1.
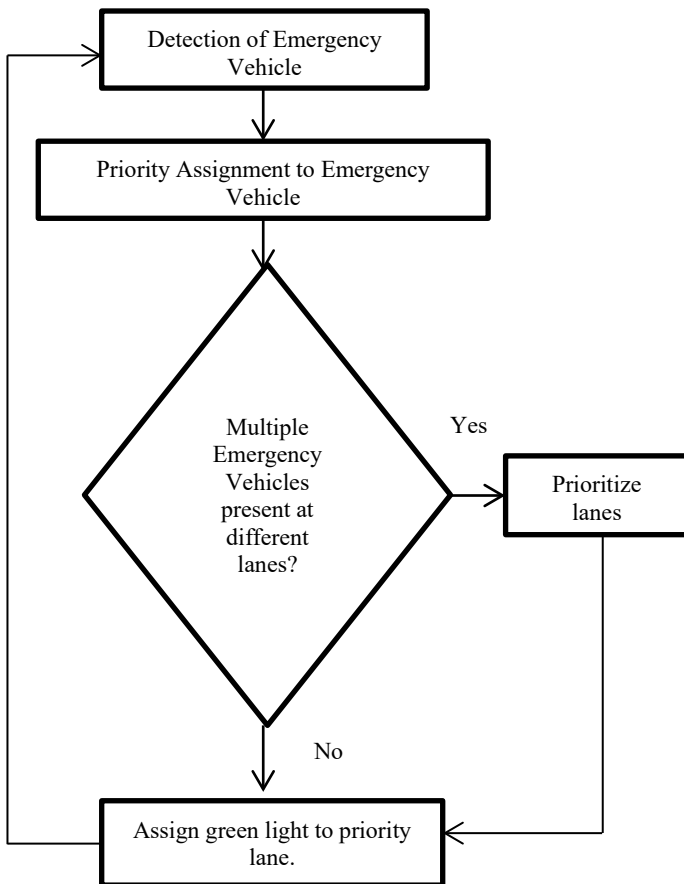
Fig. 1 Flow chart of proposed methodology

1. Signals sent by the RFID reader are collected by the transceiver. After recognizing the received signals, system identifies the type of emergency vehicle and assigns priority to it. For example, if more than one different or same emergency vehicle are present at different lanes, then according to their type, a numeric value say 1 for ambulance, 2 fire truck and 3 for police van.
2. System then prioritize the lanes according.

4 Results and Discussions

This paper has a great impact on the traffic problems faced in urban areas. Normally the traffic is controlled by a predetermined signal light controlled system. The working of the modules is based on the RFID tag information. In this paper, ultra high frequency RFID tags are used. The only requirement of this paper to be implemented is that the vehicles should be provided with a unique RFID tag. Proposed method is implemented and simulated in Python and the results are shows in the Table 2. Traffic light switching starts from north direction. Emergency vehicle present column represents whether any emergency vehicle is present or not. Third column represents

Table 2 Results of proposed method

Traffic light switching direction	Emergency Vehicle Present?	Number of emergency vehicles present	Direction and details of vehicles from where vehicles coming	Signal to be provided
North	Yes	2	From west, Firebrigade (EPC: 1017) From North, Police van (EPC: 1020)	Signal to be provided on West
West	Yes	1	From North, Police van (EPC: 1020)	Signal to be provided on North
North	No	NA	NA	Signal to be provided on South
South	No	NA	NA	Signal to be provided on East
East	Yes	2	From North, Police van (EPC: 1016) From South, Ambulance (EPC: 1018)	Signal to be provided on East

number of vehicles present. Fourth column shows type and direction of the emergency vehicle with RFID information. Last columns shows the decision of traffic light controller i.e. in which direction next signal to be provided.

Any RFID label that exists on a path for over 20 min will distinguish an anomaly. This might be an accident of the vehicle or a mishap by street. The vehi- cle with this EPC number will be tried for issues utilizing the phone number of the proprietor and, where pertinent, will give prompt help.

5 Conclusion

This paper presents a methodology for the arranging of traffic crisis vehicles. The pro- posed crisis location and organizing framework depends on radio recurrence ID. With the crisis vehicle location module it will guarantee that the crisis vehicles will arrive at the goal as quickly as time permits. Contrasted with recently executed plans, the program is both financially savvy and simple to actualize. The cost level of the gad- get is exceptionally valuable. Each RFID tag is gathered progressing by the system. That vehicle may utilize the RFID labels to spare significant data, for example, vehicle number, proprietor's information, vehicle model and number of encroachments by that vehicle. The accompanying can likewise be utilized to screen crooks, checking vehi- cles taken and e-ringing vehicles, assessing the most recent details for different sorts of street vehicles, identifying bottlenecks, analyzing territories for customary blockages and deciding the causes.

References

1. Bansal CS, Chandra A (2009) An intelligent traffic control system using RFID. IEEE Potentials 28(3):40–43. https://doi.org/10.1109/mpot.2009.932094
2. Sharma S et al (2013) Traffic light priority control for emergency vehicle using RFID. Int J Innov Eng Technol 2(2):363–366
3. Hua LC (2010) Automatic vehicle identification system based on RFID. Anti-Counterfeiting security and identification in communication (ASID) (2010), pp 281–284
4. Chellani N, Tahilyani C (2013) Traffic congestion detection and control using RFID technology. Int J Eng Res Technol
5. Chavan SS (walke) et al, Design of intelligent traffic light controller using embedded system. In: Proceedings of the second international conference on emerging trends in engineering and technology, ICETET-09
6. Bharadwaj R et al (2013) Efficient dynamic traffic control system using wireless sensor networks". In: Proceedings of the IEEE international conference on recent trends in information technology (ICRTIT). pp. 668–673
7. Pallavi P, Hussain MA (2014) Intelligent traffic control system using embedded web technology. Int J Comput Sci Inf Technol Secur 4(2)
8. Sundar R, Hebbar S, Golla V (2015) Implementing intelligent traffic control system for congestion control, ambulance clearance, and stolen vehicle detection. IEEE Sens J 15(2):1109–1113. https://doi.org/10.1109/jsen.2014.2360288

9. Balamurugan A et al (2015) Automated emergency system in ambulance to control traffic signals using IoT. Int J Eng Comput Sci 11533–11539
10. Farheena S, Chandak BM (2014) An approach towards traffic management system using density calculation and emergency vehicle alert. IOSR J Comput Sci 4:24–27
11. Fan XH, Zhang YL (2009) A Design of bi-verification vehicle access intelligent control system based on RFID. In: Proceedings of the ninth international conference on electronic measurement & instruments (ICEMI'2009)
12. Zhou ZG et al (2008) Secure design of RFID tags in the new type license plates automatic identification system. In: Proceedings of international symposium on computer sciences and computational technology, TBD, pp 694–697
13. Andronicus F, Maheswaran PG, Scholar (2015) Intelligent ambulance detection system. Int J Sci Eng Res 1462–1466
14. Mithun NC, Rashid NU, Mahbubur Rahman SM (2012) Detection and classification of vehicles from video using multiple time-spatial images. IEEE Trans Intell Trans Syst 13(3):1215–1225. https://doi.org/10.1109/tits.2012.2186128
15. Rifai I, Alj M, Yassine (2014) An energy-efficient WSN-based traffic safety system. 1–6. https://doi.org/10.1109/IACS.2014.6841980

Big Data-Enabled Solutions for Covid-19

Dirisala Saikrishna, Bavisetti Sai Dhanush, Madhusudhan Rao,
Priyanka Chawla, and Usha Mittal

Abstract In the time of the pandemic like CORONA, Covid-19, everyone is ftghting against this deadly virus. Besides, governments are looking for a barrier that stops spread of virus until the vaccine is made. In modern era, technology plays an important role. This paper brings the way by using a powerful technology called Big data. Big data know for handling a large amount of data and provide powerful insights into the data. Big data integrated with Artificial Intelligence is a powerful tool to ftght against this pandemic. Many countries like Taiwan, China with the use of Big Data stop this pandemic up to some extent. But the collection of data itself comes up with the big challenge of PRIVACY AND SECURITY. In the recent times, the world has seen the effect of data leaking whether by Facebook or by Google. Many European countries due to this big challenge will not be able to use this technology.

Keywords Big data · Artificial intelligence · Privacy · Security · Data collection

1 Introduction

1.1 Bigdata

The term describes it all Big data refers to data which is in large volume and different formats or variety. Generally, Big data is described using the following characteristics also known as 3 V's of Big Data. i.e., Volume, Variety, Velocity [1].

Volume—In Today's world every day Over 2.5 quintillion bytes of data produced in a single day i.e., approximate.1.7 MB of data in a single second. Big data helps to manage this huge amount of data with the help of its various frameworks and tools.

Velocity—It refers to the rate at which data is generated and by that time only acts. on that data. It is challenging but big data provides the various solution for this alsoby

D. Saikrishna · B. S. Dhanush · M. Rao · P. Chawla (✉) · U. Mittal
Lovely Professional University, Phagwara, Punjab, India

279

using Various frameworks like Apache Spark, Flink, Apache Storm we can perform the data analytics as soon as data enters the system.

Variety—It refers to different types of data that are produced. It is been Categorized into 3 formats i.e., structured, unstructured, semi-structured.

Structured—It consists of a traditional relational table. Data is represented in a tabular manner.

Unstructured—The data which is not in a particular format is referred to as. unstructured data. Audio, video is the type of unstructured data.

Semi-structured—The data which has format, but it not structured is called. structured data. JSON, CSV formats are the common example of this type of data.

2 How Big Data Help in Fighting Against Covid

Data has a crucial role in knowing the hidden facts. By collecting Data of individuals big data helps to provide the insights. Data which helps fight against this pandemic is:

2.1 Medical Data

Data provided by hospitals will play an eminent role. Through recent research, it is found that a person with heart disease, lung disease or a diabetic person will have more affected by coronavirus by the prior knowledge of this type of person [2]. Govt. can target this type of person and provide them more instructions during this coronavirus. Also, By the research, it is found that coronavirus has more effect on a certain age group of people, i.e., children and old age people.

2.2 Mobile Data Location

Usage of location data is quite very useful [3]. Because as we have a medical history of a patient, we can use the location of a person easily to provide sufficient aid to that person. Also, by this Govt. also able to see people are following social distancing or not.

2.3   Travel History

By getting this data, it is easy to identify which person has travel history (most favorably out of the country) [4]. So that quarantine period is assigned to that person. And a person is tested. By this government's priority to handle these people first.

3   Privacy Preserving in the Healthcare System

As mentioned in abstract, the biggest challenge in data collection is privacy. Though this type of data is sensitive data as it completely discloses the identity of an individual. By medical data we know the disease that person is suffering. By location data, there is no privacy of person. There will be a close eye on the govt. or hacker on a person. Countries like Korea, Taiwan, and China collect that data but still the privacy the fundamental rights of a person. How person will be protected is still a blank. Democratic countries need a better solution to concur this challenge. One solution for that is an individual can voluntarily share its data and authorities will assure him that its identity will be secured [5]. And after using your data is destroyed. But the problem here is how can one know its identity is not being compromised and its data is being destroyed or not. In recent years, many companies said that they will destroy the data. But they did not. Also, if there is an agreement and a person has full trust in authorities. There is another big challenge for participation.

3.1   Big Data Approach to Handling These Challenges

Before stepping forward let us know what the difference between Privacy and security is.

Privacy refers to control over how personal information is being collected or used. It is the power of an individual to stop information about themselves disclose to the public or using it for personal use.

Security refers to defend the information of an individual by using technology, process and by using training data generation is categorized into two ways i.e., Active Data Generation, Passive Data Generation. In Active data Generation data owner will give its data to the third party while in passive data generation data is generated by data owner while online activity and data owner do not know its data is being collected by the third party. Both risks are minimized by access restriction or by falsifying the data.

Access Restriction—In Access restriction, if data is collected by active data generation user cannot give its data if it discloses its sensitive information. If it is passive data generation data is restricted by using an anti- tracking extension or adblocker and encryption tools.

Falsifying Data—Sometimes it is not possible to neutralize access to the data. For that time data owners can distort the data before data gotten by a third party. Once the data is distorted it is not easy to reveal the true information. Socket Puppet is one of the falsifying tools that do its work by deception. Multiple Socket puppet the data about an individual is regarded as having a place with multiple people by the data collector.

4 Big Data Privacy When Data is Being Stored

Storing information is not hard in big data due to advancements in technology. But it is high-risk work. Because once a big data database is being compro- mised the privacy of an individual is disclosed. Moreover, Big data follows distributed approach data is placed in several datasets are several locations so privacy is a priority for this. The conventional security mechanism is divided into four categories Application- level security, Database level Security, file- level security, and media level security [6]. 3 V's of big data depicts its storage mechanism to be scalable.

4.1 Privacy-Preserving Techniques in Big Data

Privacy is being a major issue not only now but also, in the past. So, some techniques were used earlier.

De-Identification—It is a traditional technique for privacy- preserving, for persevering individual privacy. first, we need to generalize the data then suppres- sion we must give the data for data mining, following up de- Identification comes with the concepts of K-anonymity, l-diversity, t- closeness to increase the privacy of information [7]. Migrating de-identification in bigdata is a great deal for privacy preserving. Thus, big data also offer many methods of re-identification so using de-identification is not a great idea for privacy protection. For a description of K-anonymity, l- diversity, t-closeness let us first discuss some important terms.

Privacy-Preserving Aggregation—It is a homographic encryp- tion technique used for data collection. This works as follows given a homo- graphic public key source can convert their data into cipher texts. Different sources can use the same public key then the whole result is aggregated and recovered by using a specific private key. But this technique is purpose spe- cific. We can only use this technique for storing and collecting data. For data exploitation and mining, this technique has no use. So, it does not help in big data analytics.

5   Recent Techniques in Privacy-Preserving

Differential Privacy Differential privacy is one of the recent technologies which is used widely. This technique helps the analyst to provide useful information without revealing the identity of the user. It generally works by adding some distraction in data. This distraction is large enough to de- identified someone's privacy and small enough to perform analytics to data at the same time. So, both ways it is a useful technique. As we have seen in deidentification by hiding some data it pretends like data is de- identified and privacy is obtained but it is not following example support above statement. In the mid-'90 s when the Commonwealth of Massachusetts Group Insurance Commission (GIC) released the anonymous dataset by hiding the name, age of the people. One Ph.D. Student by taking publicly available voters dataset matches with the GIC dataset easily re- identified people and it destroys the privacy. While DP does not work that way there is an intermediate software present in between database and analyst. Only the information needed by the analyst is provided by the database through the software. This intermediate software is also called a privacy guard. It works as follows:

Step 1: Analyst makes the query to the database through the intermediate privacy guard.
Step 2: Privacy guard evaluated the query and the previously evaluated queries and evaluates privacy risk.
Step 3: Then the privacy guard will get the desired information from the database by adding some distraction in it.

The level of distraction is proportional to the evaluated privacy risk. If the privacy risk is high the distortion is high and vice versa.
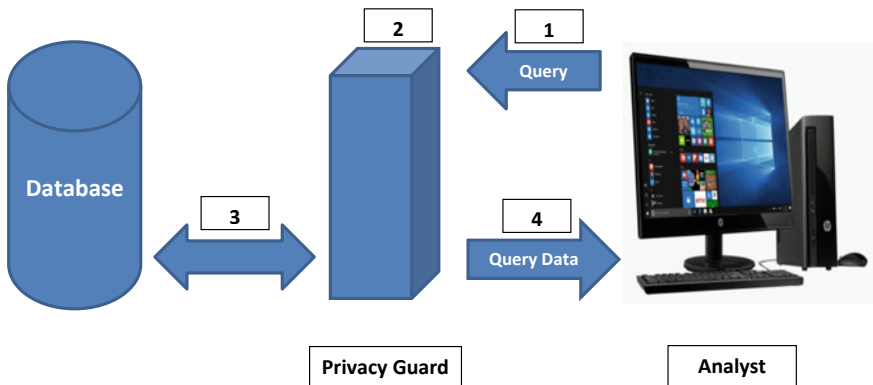
As Discussed earlier many algorithmic approaches of adding noise to the data but the fact is chances of reidentification is still in a method. Meanwhile, this algorithm also adds noise to the data, or we can say data to the noise, so it is quite hard. For the hackers to find the original data as the same way of finding a needle in the haystack

concept or we can say finding salt in sugar. This will surely create a more cost of computation for making a haystack (noisy data) to hide the needle. In Fig bellow, the white circles are noisy data while the black circles are the original data. The service provider adds a dummy as noisy data as original transactional data collected by the data provider. Likewise, a unique code is assigned both dummies as well as original data. This code is maintained by the service provider so that during the extraction of data dummy data and original data got to separate [8].

• Original data
○ Duplicate data
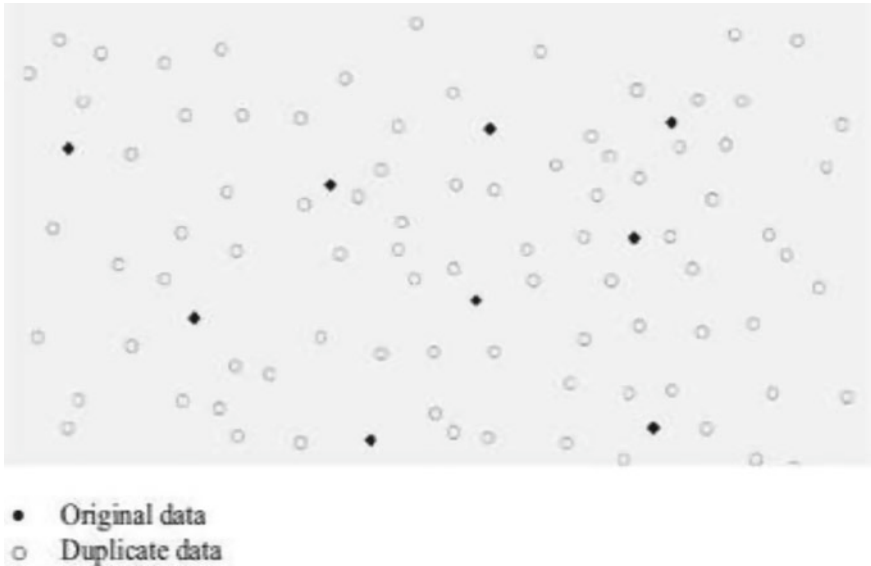
6   Privacy-Preserving in the Healthcare System

The way in which medical records are being digitized gives a new face to the health-care system. Thus, doingthis medical industry experiencing a heavy load of data. By heavy lead of data then big data comes into the play. Due to the real-time availability of records and more accurate test result helps doctors to take more constructive deci-sions in treatment. The real-time monitoring of embedded sensors attached to the patient gives the warning whenever there is an anomaly. Healthcare digitization with integrated analytics is one of the next big waves in healthcare information technology with electronic health records (EHR) being a crucial building block for this vision. By this healthcare organization came to result that with the EHR incentive programs the results are more accurate and thus eventually lead to improved healthcare of patients. But with great power comes great responsibility. Like that the ever- changing risks and vulnerabilities, security exploitations are expected to grow in the coming years.

Big data offers a comprehensive survey of different tools and techniques used in pervasive healthcare in a disease-specific manner [9]. It provides a quick result by finding major diseases and disorders to give a quick treatment. The adoption of big data in healthcare significantly increases the security risks and patient privacy risks [10]. As we have seen traditional techniques were not directly applied to the large and inherently diverse datasets. In today's world, we have seen there is a demand for cloud-stored.

6.1 Real-Time Security Analytics

Today's requirement in real-time analytics is predicting threat sources in real- time is of utmost need in the healthcare industry. The most common attack in the healthcare industry is distributed denial of service (DOS) attack to steal important information and in real-time analytics, the delay is not accepted [11].

6.2 Data Quality

Health data is usually located from different sources, and with different databases, this leads to anomalies in the data lot of missing values are there. These things should be handled because dirty data leads to incomplete analysis [12].

6.3 Data Sharing and Privacy

Health data contains personal information. Their privacy of an individual is always at risk. Healthcare information would be masked by de-identification techniques and later that data is provided to the data analytics based on legal data sharing agreement. But due to this, the quality of data I decreased that is not usable for analytics. So, to take care of privacy, there is always a balance between privacy and providing meaningful data [13].

7 Conclusion

In this paper, we discussed how big data is helpful in Fighting against COVID- 19 all we need to take care of challenges that come with big data. We have seen different techniques to secure the privacy of an individual. Big data leads to knowledge that provide better analytics. Yet a small amount of data is being used for analytics due to privacy issues. Privacy techniques with its advantages and disadvantages are

described in this paper. Concepts of differential privacy and other comprehensive techniques are being used. MapReduce framework is used for anonymizing techniques and it easily is scaled up using adding more mappers and reducers. As for future more effective solutions need to be presented. In the field of healthcare as well more scalable technique that can handle the true power of big data is needed. IoT is a future integrating big data with IoT is gamechanger in every field. Forms of data itself whether it is semi-structured or unstructured come with many challenges. As such, there is a lot more to do in privacy-preserving in big data. For a time being to fight against COVID-19 discussed techniques will improve big data analytics.

References

1. Oracle big data for the enterprise (2012) [online]. http://www.oracle.com/caen/technoloqies/biq-doto
2. Big data versus COVID-19: opportunities and privacy challenges, [online] https://www.bruegel.org/2020/03/bigdata-versuscovid-19-opportunitiesandprivacychallenges/?utm_content=bufferfdf0a&utmmedium=social&utm_source=facebook.com&utm_campaign=buffer
3. Mohammadian E, Noferesti M, Jalili R (2014) FAST: fast anonymization of big data streams. In: ACM proceedings of the 2014 international conference on big data science and computing, article 1
4. Big data at the speed of business, [online]. http://www-01.ibm.com/software/data/bigdata/2012
5. Mehmood A, Natgunanathan I, Xiang Y, Hua G, Guo S (2016) Protection of big data privacy. In: IEEE translations and content mining are permittedfor academic research
6. Lu R, Zhu H, Liu X, Liu JK, Shao J (2014) Toward efficient and privacy-preserving computing in big data era. IEEE Netw 28:46–50
7. Li N et al (2007) t-Closeness: privacy beyond kanonymity and L-diversity. In: Data engineering (ICDE) IEEE 23rd international conference
8. Abadi DJ, Carney D, Cetintemel U, Cherniack M,Convey C, Lee S, Stone- braker M, Tatbul N, Zdonik SB (2003) Aurora: a new model and architecturefor data stream manag ement. VLDB J 12(2):120–39
9. Jain P, Pathak N, Tapashetti P, Umesh AS (2013) Privacy preserving processing of data decision tree based on sample selection and singular value decomposition. In: 39th international conference on information assurance and security (lAS)
10. Wang C, Wang Q, Ren K, Lou W (2010) Privacypreserving public auditing for data storage security in cloud computing. In: Proceedings of IEEE international conference on infocom, pp 1–9
11. Sedayao J, Bhardwaj R (2014) Making big data, privacy, and anonymization work together in the enterprise: experiences and issues. Big data congress
12. Fong S, Wong R, Vasilakos AV (2016) Accelerated PSO swarm search feature selection for data stream mining big data. IEEE Trans Serv Comput 9(1)
13. Big data privacy: a technological perspective and review Priyank Jain*, Manasi Gyanchandani and Nilay Khare

Smart Traffic Monitoring System

V. Radhika, V. Madhan Babu, and S. Jayaprakash

Abstract In the modern era, traffic congestion is one of the major issues in smart cities. The existing traffic management is unable to handle the growing number vehicles. The main aim of this work is to automate the traffic signal by monitoring the entire system and allowing vehicles to pass through lanes based on the traffic density. The existing method consists of a signal with fixed timing and also human intervention for managing traffic. In this proposed work, a system is designed to program the traffic signal timing and reassign the green signal glowing time based on traffic density. The lane with higher traffic density will have maximum timing to pass through and the lane with less traffic density will have minimum timing to pass through. A system that gives highest priority to emergency vehicles is also developed.

Keywords Smart city · Traffic management · Traffic density

1 Introduction

Automation in traffic signals can be an effective transportation management system utilized in smart cities. It uses the advancement in Internet of Things (IoT), sensor technology and data analytic techniques to manage the traffic effectively and enhance the traffic management system. The four major components in the traffic management system are data collection, data transmission system, data analysis system and finally the collected information is sent to the end user. The aim of the intelligent traffic management is to create interactive transport system, manage the traffic and provide the real time information about the traffic to the common users [1]. IoT can change the transport system effectively and integrates the embedded sensors, actuators and other gadgets to collect and provide the information about the real time traffic activity. Traffic signals are generally used to ensure the safe and systematic movement of vehicles. These signals ensures the pedestrians or the vehicles to cross the intersection

V. Radhika (✉) · V. Madhan Babu · S. Jayaprakash
Department of Electronics and Instrumentation Engineering, Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India
e-mail: radhika.senthil@srec.ac.in

roads without any conflicts. But if a particular traffic signal fails it leads to increase in accidents, waiting time, increase in greenhouse emissions also. If a traffic signal in all the lanes have uniform timing it leads to more crowdie in a particular lane than the other lanes. If the timing is allotted to lanes based on the density of vehicles and also on any other specific criteria traffic jam can be avoided. So traffic signal timer programmed based on the traffic density can reduce the traffic congestion problems.

Traffic light system consists of traffic lights and traffic light control system to control the traffic signals. In the existing system the controllers used are based on electro mechanical mechanism. These controllers are based on open loop mechanism without any feedback mechanism and sensing network. So they lack in intelligence to adapt to the change in traffic density and control the traffic signal effectively. Traffic signal timer that works according to the traffic density by utilizing the image processing technique is also proposed in earlier works. In those methods high resolution cameras are used to sense the changing traffic density around the traffic signal and programs the traffic signal timing accordingly. In these methods images are captured on regular intervals, then the images are processed and given to the controller. Controllers on receiving the input, compares it with predefined thresholds and sends the required signal to traffic signal timer. But these methods need a high resolution camera and images also to be captured continuously for predefined time. So an alternative method to automate traffic signal that reduces the traffic jam effectively and minimize the air pollution is needed for current busy scenario.

This paper proposes an intelligent traffic management utilizing IoT to handle the different traffic situations. This proposed work reduces the flaws in the existing system traffic system more effectively. This system utilizes the ultrasonic sensors to detect the number of vehicles and the RFID tag placed in the emergency vehicles like ambulance, VIP vehicles and fire brigades to detect and prioritize the emergency vehicles. So this proposed system detect traffic congestion, provide predictable information about the traffic condition can reduce the problems like traffic congestion, accidents and parking issues.

In this paper Sect. 2 explains the related works, Sect. 3 elaborates the proposed work, the results and discussion is explained in Sects. 4 and 5 gives the conclusion and future scope.

2   Related works

This section presents an overview on the techniques and approaches used in the related works considered for monitoring the traffic system.

For working with the traffic congestion issues, predictive type or real time knowledge about the traffic is necessary. Traditional methods uses sensors like pneumatic tube and automatic counter to obtain real time data on traffic signal data's. These techniques fail to detect the vehicle movement in multiple lanes. Video/image processing techniques utilizing the number plate recognizing system to find the traffic density is proposed. This technique fails due to non-uniformity in the number plates and visibility problems in poor weather condition. Wireless Sensor Networks (WSN) and

Vehicle to Vehicle network is utilized to measure the traffic density is also proposed. Also various related works using fuzzy logic and neural networks learning algorithms are also utilized for controlling the traffic light is found in literature.

Combinational Interaction Testing (CIT) and Combinational interaction testing with constraint approaches supports the prioritized based test generation for traffic signal generation. Optimization algorithms were proposed for scheduling traffic light conditions. Simulated annealing applied to CIT technique is proposed for computing the test case generation for traffic signal conditions. Traffic analysis toolbox volume IV: guidelines for applying corsim micro simulation modeling software and Direct signal timing optimization: strategy development and results works uses Genetic Algorithm (GA) for optimizing the programming cycle for traffic lights [2–4]. In Intelligent Testing of Traffic Light Program approach a PFGM algorithm is proposed to generate the prioritize test suite for a given a traffic feature model. In this work an expert system with reduced set of scenarios with similar fault detection capacity is developed with minimum cost. Developed expert system numerically quantify and qualify the traffic cycle program for different traffic conditions. In the Isolation niches particle swarm optimization applied to traffic lights controlling work, particle swarm optimization algorithm is used for scheduling traffic lights. In this approach one way road with two intersections was proposed. A multimodal traffic responsive strategy using Particle Swarm Optimization (PSO) algorithm is proposed in this approach multi-objective PSO was used. This work uses predictive control model and simulations were carried out with private, and public vehicle model were carried out on a virtual urban road network.

3   Proposed Work

Existing traffic signal management system consists of traffic signal that are programmed such that the yellow, green and red signals glow for fixed interval of time. The system proposed here identify the traffic density on different lanes and adjust the traffic signal timing. But traffic jams happening in today scenario signs that traffic signal must be programmed based on the density of road. So the proposed system aims to reprogram the traffic signal based on the traffic density. This traffic monitoring system has two modules such as hardware and software module. The hardware module has ultrasonic sensor, RFID sensor, RF Module, LEDs, Arduino UNO microcontroller and relay module [5, 6]. The software module uses Arduino microcontroller programmed with Ardiuno IDE.

The primary element of this system is the ultrasonic sensor and RFID sensor. The ultrasonic sensor is used to detect the density of vehicles in a lane and feed the response to the centralized controller. Centralized controller make decisions to regulate the traffic signal based on the data given by the ultrasonic sensor about the traffic density. LEDs are used as traffic signal in this work. RFID Tag is fixed to special vehicles like Ambulance, fire brigades and VIP vehicles. RFID is a technique that automatically identifies the object through radio frequency waves. Various types

of RFID tags are available but majorly used tags are active and passive RFID tags in this traffic management system. Passive tags don't contain internal power source but the life time of it is more than active tag. Active tags have internal battery as power source and range of coverage is more than passive tags. Passive type tag is used in this work. RFID tag has an integrated circuit for processing the data, modulating and demodulating the RF signal and an antenna for transmitting and receiving the data. The RFID reader module is fixed to the particular traffic signal and RFID tag is placed to the emergency or special vehicles.

Arduino is an ATMega based microcontroller is used as a centralized traffic controller and receives the ultrasonic sensor data and manages the timing of switching of traffic light between red, green and yellow. The ultrasonic sensors fixed at four sides of road in a four way lane that detects the vehicle density in a particular lane. Ultrasonic sensor is generally used as a distance measuring sensor with a transmitter and a receiver [7, 8]. Ultrasonic sensors are placed at regular interval of distance across the lane. First ultrasonic sensor placed in the lane sends the ultrasonic wave that hits the target surface and it reflects the wave back reaches the sensor and the second sensor placed after 10 m of distance sends the ultrasonic wave get reflected back to the sensor. The microcontroller counts the number of vehicle in a lane based on the distance and timing for reflection of waves between sensors. RF module is used for the communication between Arduino and the sensor. Figure 1 shows the ultrasonic sensor placed in four lanes.
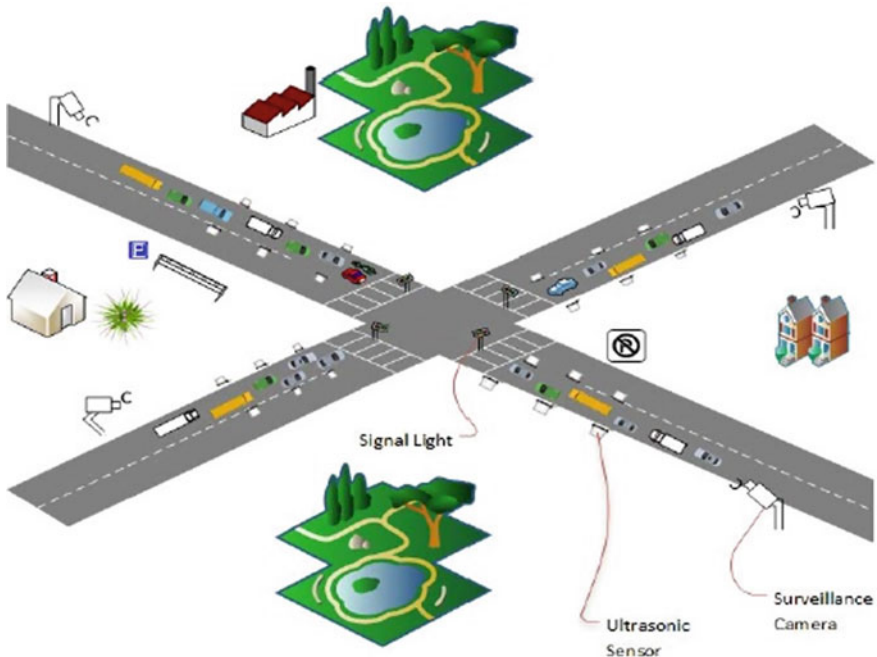
Fig. 1 Smart traffic management system with ultrasonic sensors
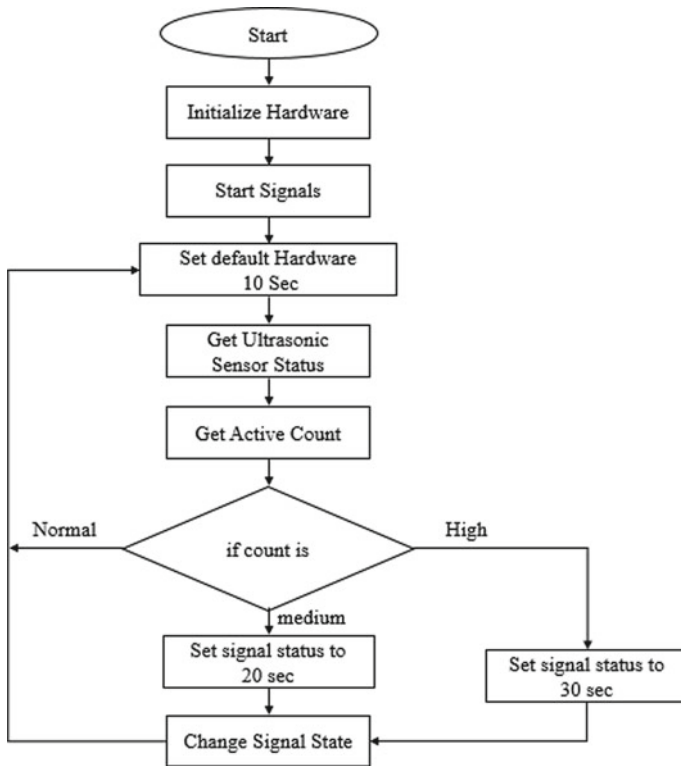
Fig. 2 Flow diagram for detecting vehicle density

The flow diagram for the count of the vehicles with ultrasonic sensor and microcontroller is shown in Fig. 2.

Data from ultrasonic sensor is processed by the Arduino microcontroller and programs the LEDs in traffic signal according to the density of vehicles. The maximum time for tracking 5 min and minimum time for tracking the density is 1 min. Based on the tracking data received from the ultrasonic sensor the microcontroller gives the priority to the lane with large vehicle density else the priority is programmed based on Round- Robin method.

This system can consider the density of traffic of road by utilizing the ultrasonic sensor and can also consider the priority of different types of vehicles [9, 10]. For this to happen a RFID tag is fixed to the special vehicles or emergency vehicles. RFID reader reads the RFID tag fixed to the special vehicles. If a special vehicle enters the road with heavy traffic RFID reader gets the signal from RFID tag and sends the signal to the microcontroller. Green signal is enabled in the particular lane till the emergency or special vehicle passes away and traffic get sequenced. The use of RFID tag differentiates the emergency and special vehicles from the other
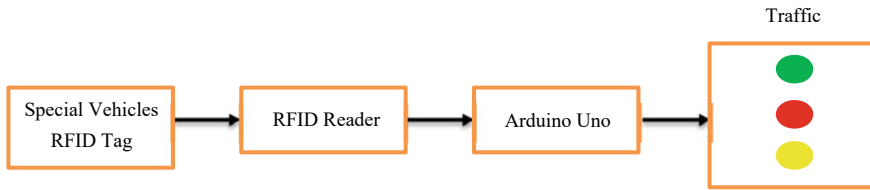
Fig. 3 Block diagram for detecting emergency vehicle

vehicles and avoids traffic congestion. The block diagram for the proposed work for identifying the special vehicles is shown in Fig. 3.

The Arduino UNO is used as the microcontroller communicates with RFID reader and the traffic signal. It establishes the communication between different modules and generates the control signal for controlling the traffic light and different peripherals. So the Microcontroller using RFID tag can alter the time of glowing of traffic lights upon the arrival of the emergency signals and reduce the waiting time of those vehicles.

4 Results and Discussion

Proposed work insights for controlling and managing the traffic by dynamically changing the timing of traffic light. This proposed work reduces the congestion, accident, waiting time and pollution. Results presented below shows the effectiveness of proposed smart traffic monitoring system. The ultrasonic sensors along with RF modules are placed at a regular intervals of distance in the road. The system is placed near road as a stand-alone device. Microcontroller detects the number of vehicles whenever the vehicle passes between the ultrasonic sensors in a particular lane. The microcontroller control the timing of glowing of traffic light based on the vehicle count and reduce the congestion at road intersections. Delays of traffic light are classified as LOW, MEDIUM and HIGH range. These ranges can be predefined based on the vehicle count. Emergency vehicles are equipped with RFID tags and these RFID tags are read by RFID reader placed at the traffic junction. If the emergency vehicles are detected by the RFID reader it is given as an interrupt to the microcontroller. Microcontroller on receiving the interrupt changes the glowing conditions of LEDs in the traffic signal. It makes the green LED in the particular lane to glow till the emergency crosses the junction and after that system returns to normal working based on the density of vehicles in each lane. The proposed system works effectively in managing traffic at junctions of road and also gives priority based on the type of vehicle. A prototype has been made to evaluate and show the arrangement of our entire system.

Fig. 4 Placement of sensors

The Arduino UNO used as a microcontroller and programmed with C language. Arduino is connected to different ultrasonic sensors and RFID reader. RF module used for serial communication between microcontroller and sensors.

Placement of sensors is shown in Fig. 4 and traffic light setup is shown in Fig. 5.

Figure 6 shows the green signal for lane with more vehicles, Fig. 7 shows the red signal for lane with less vehicles and Fig. 8 shows the green signal for an emergency vehicle in a lane.

Fig. 5 Traffic light setup

Fig. 6 Green signal for vehicles

Fig. 7 Red signal for vehicles

5   Conclusion and Future Scope

In this work the traffic is managed based on the priority and the density of road in different lanes of junction. The vehicle density is calculated as a first step in this work and control the signal timing accordingly. The time of glowing for green signal is reassigned. The lane with maximum density of traffic will receive maximum timing for the vehicles to pass through. The lane with less traffic density will receive

Fig. 8 Green signal for emergency vehicle

minimum timing for the vehicles to pass through. Ultrasonic sensors and microcontroller plays a key role in this work to automate the traffic management system. This work also gives preference to the emergency vehicles such as ambulance, fire vehicle, and police cars by allowing traffic signal in the particular lane to glow to green till the vehicle crosses the signal. RFID tag fixed to emergency vehicle and the RFID reader at the traffic junction allows the emergency or special vehicle to pass through with the help of microcontroller.

This work can be combined to a single system for detecting emergency vehicle and density of vehicle in a particular lane without using separate system.

References

1. Sanchez J, Gal M, Rubio E (2008) Applying a traffic lights ́evolutionary optimization technique to a real case: 'Las Ramblas' area in Santa Cruz de Tenerife. IEEE Trans Evol Comput 12(2)
2. Ferrer J, García-Nieto J, Alba E, Chicano F (2016) Intelligent testing of traffic light programs: validation in smart mobility scenarios. Hindawi Publishing Corporation, vol 3871046, no 1
3. Harrison C, Eckman B and R. Hamilton (2010) Foundations for smarter cities. IBM J Res Dev 54(1)
4. Mendonca M, Cowan D (2010) Decision-making coordination and efficient reasoning techniques for feature-based configuration. Sci Comput Program 75:311
5. Sivasankar BB (2016) IoT based traffic monitoring using raspberry Pi. Intl J Res Eng Sci Technol 1:2454
6. Perumalla BK, Sunil BM (2015) An intelligent traffic and vehicle monitoring system using internet of things architecture. Int J Sci Res 6:391

7. Parkhi1 AA, Peshattiwar AA, Pande KG (2016) Intelligent traffic light system using vehicle density. Int J Electr Electron Eng 8(172)
8. Lanke N, Koul S (2013) Smart traffic management system. Int J Comput Appl 75(93)
9. Tyagi V, Kalyanaraman S, Krishnapuram R (2012) Vehicular traffic density state estimation based on cumulative road acoustics. IEEE Trans Intell Transp Syst 23:253
10. Shruthi KR, Vinodha K (2012) Priority based traffic lights controller using wireless sensor networks. Int J Electron Signals Syst 1:2231

Harvesting Electrical Energy from Body Heat to Power Bio Medical Wearables

M. Preethika, Mohamad Radzi Bin Ahmad, K. Porkumaran, S. Prabakar, and R. Sundar

Abstract Energy harvesting has become a key factor in Biomedical applications. Research on energy scavenging from body motion has been investigated to evaluate the feasibility of powering a wearable or implantable systems. Energy from walking has been previously extracted using generators placed on shoes, backpacks, and knee braces while producing power levels ranging from milliwatts to watts. The research presented in this paper examines the available power from body locations. The finger, wrist, and palm were the chosen target localizations. There are many untapped renewable energy resources such as wireless charging, thermal energy, and vibration energy that can be used as an alternative to the conventional sources such as solar, wind, and hydro. As we are moving towards Industry 4.0, new advancements in technology has been introduced. Therefore, with a great advancement comes a greater need for powering up devices by thermal technique. Implantable biomedical equipment such as pacemaker and battery using biomedical wearables such as the blood pressure monitor consumes low power and provides greater consistency. In order to provide continuous supply to devices, a new approach of extracting power from the heat of the human body is implemented thus avoiding unnecessary invasive surgeries for implant.

Keywords Energy scavenging · Body motion · Power levels · Thermal technique · Implantable biomedical equipment · Body heat · Wearables

M. Preethika (✉) · R. Sundar
Dr.N.G.P. Institute of Technology, Dr.N.G.P. Nagar, Coimbatore, Tamil Nadu, India

M. R. B. Ahmad
Universiti Tecknologi Petronas, Bandar Seri Iskandar, Tronoh, Perak, Malaysia

K. Porkumaran
Sri Sairam Engineering College, Chennai, India
e-mail: porkumaran@gmail.com

S. Prabakar
Sona College of Technology, Salem, India
e-mail: srisornaprabu@gmail.com

1 Introduction

A fast-growing class of such devices is wearable, where sensor nodes are tightly coupled with the human body, low power consumption is crucial in wearable systems due to the light weight and size constraints for batteries, which severely limits the energy that can be stored in the device. Wearables are; however, very tightly constrained in terms of the size and weight and must also be compatable with the body. Therefore, the possibilities for EH systems are more restricted than for other applications. Energy can be harvested from various environmental sources [1] including light using photovoltaics [2], movement of the wearer [3], from radio frequency energy (RF), or from temperature differences using thermoelectric generators (TEG). The human body in contrast is a constant heat source. Typically, a temperature difference exists between body core and the environment. Because of the voltages produced by thermal harvesting are typically too low to power wearable electronics, a conversion stage (DC-DC) with high conversion efficiency needs to be included into the wearable system. A complete system analysis from the body heat to the application is required to maximize both the output power and wearability. Thermal energy harvesting is an another method of reducing the dependability on non-renewable sources. There are various energy harvesting available, among them are electromagnetic, electrostatic, solar, and thermal methods. However, these mechanisms only produce a very limited power. To meet the power necessity for most biomedical devices, we use a boost converter to obtain the required output in order to satisfy the biomedical wearable's power requirement. The chosen target locations for power generation here are the finger, palm, and wrist.

1.1 Problem Statement

Thermal energy from human body is rarely used to generate energy. The existing thermal energy harvester is minimally used for Biomedical applications. Hence, there is a need to develop an alternative charging method using thermal energy harvesting to prolong the life of the battery.

1.2 Objectives

The objectives of this harvesting techniques are:

i. To explore the concept of thermal energy harvesting method for specific type of Biomedical wearable devices.
ii. To model the thermal energy harvesting method and charge the Biomedical device.

iii. To boost the minimial output using a boost converter for required power to charge the biomedical device.

iv. A study of thermal energy harvesting method along the Electrical and Electronical Engineering to produce a successful Energy Harvesting device.

2 Literature Review

2.1 Energy Generation from Human Body

Research on energy scavenging from body motion has been investigated to evaluate the feasibility of powering a wearable or implantable systems. Energy from walking has been previously extracted using generators placed on shoes, backpacks, and knee braces while producing power levels ranging from milliwatts to watts. Thermal energy harvesting is an another method of reducing the dependability on non-renewable sources. The energy harvested from heat uses Thermoelectric Generators (TEG), which is based on the principle of Seebeck effect [4]. According to [5], it is a certain type of thermoelectric effect in which the temperature difference is converted to electrical energy which can be represented by this equation:

$$E_{emf} = S_e \Delta T$$

or

$$V = \propto (T_H - T_C) \tag{1}$$

- Se / $\propto$: Seebeck coefficient
- ΔT: Temperature difference
- T_H: Hot side temperature
- T_C: Cold side temperature

A TEG comprises of two thermoelectric semiconductors n-type and p-type. These semi-conductors are series-connected at the top and bottom of the junctions and are subjected to two temperature difference which are hot and cold. The representation is shown in Fig. 1.

Because of the different temperature at the top and bottom of the semiconductors, the electrons and holes mobilize towards the opposite side, which gives rise to the electric field between the cold and hot sides. As a result, an open circuit voltage is produced at the terminals. To generate an output power, a load resister RL is connected across the bottom of the cold side, i.e., the output terminal. Output power, PL can be calculated from the voltage and current across the load resistor of the TEG. According to [6], the electric power that is generated from the TEG is directly proportional to the temperature difference between the surfaces of the TEG. Therefore, the internal resistor, Rin = RL, a maximum power, Pmax is produced and can be expressed in the equation as follows:

Fig. 1 Semi-conductor thermocouple seebeck effect
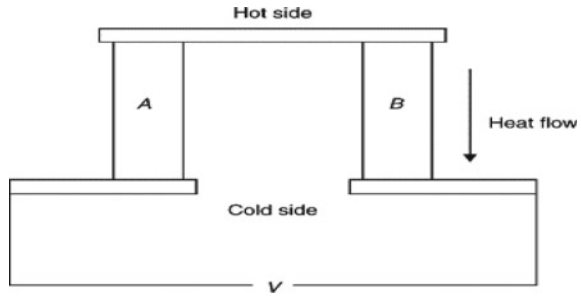
$$P_{L\,\text{max}} = \frac{\propto^2 \Delta T^2}{4Rin} \tag{2}$$

However, in [1], it is also known that the open-cicuit output voltage, Voc of a TEG module is also directly proportional to the temperature difference between the two surfaces. The equation for that can be expressed below:

$$V_{OC} = \propto m\Delta T$$

Thermal energy harvesting is in DC. Thus, only a DC-DC converter is needed. This is used to scale the voltage from the input to the output load. To develop an efficient DC-DC voltage conversion, it is recommended to use an inductor-based converter which stores the energy using an external inductor and then releasing it to the load. However, since the voltage and current needs to be stepped up, usually a boost convertor is introduced as a step-up transformer to overcome the drawbacks of MOSFET switching; however, it is unable to sustain very low voltages (under 50 mV), which is the case with the thermoelectric generation.

2.2 Human-Based Energy Harvesters

According to [7], the typical self-powered wearable device technology's power requirement is in the range of 0.01–10 mW and is dependent on its applications. It is found that the power levels from the thermal energy, generated by the human body is greater than the power generated by the photovoltaic cells (Fig. 2).

Figure shows the heat transfer of a thermal generator for a TEG on the human skin.

T = Temperature of the body core, skin, radiator, ambient air.
R = Thermal resistances.
W = Heat flow.

It is found statistically from the research that at the usual location of a watch, wrist, thermal resistance at an ambient temperature of 22.7 °C is 440 cm^2 K\W, while under the watchstrap, which is the radial artery, it decreases to 120–150 cm^{2}K\W. Thus,

Fig. 2 Thermal circuit equivalent of a TEG
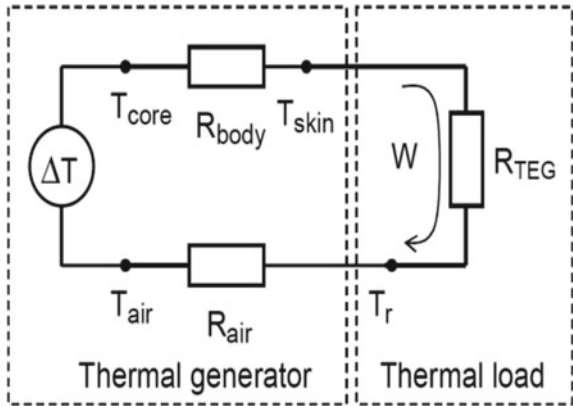
Fig. 3 (**a**) Off-the-shelf thermopile
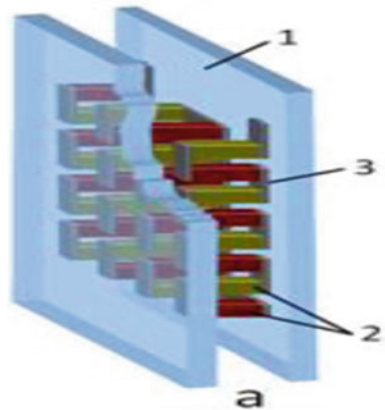
the thermal resistance in a normally clothed person changes a little or not at all in the trunk if the ambient temperature decreases as the average body's thermal resistance is 140 cm^{2}K\W.

Figures shows the miniaturized thermopile attached to the skin [2] (Fig. 3).

From Figure (a):

The miniaturized thermopile attached to the skin is shown in the figure provided with 1. Ceramic plates which is a metal that acts as an outer covering for the thermopile holding the 2. Thermocouple legs which are interconnected along the 3. Metal interconnects altogether attached to the ceramic plate (Fig. 4).

From Figure (b):

4. The human skin is kept in contact with the 5. hot plate which is placed parallel to the 6. cold plate from where the temperature difference ΔT is noted between the two plates. The 7. holding, encapsulating, and anti-shock elements are provided for holding, covering, and providing resistance to the shock, respectively. Primarily the

Fig. 4 (**b**) Simulated TEG
on the human body
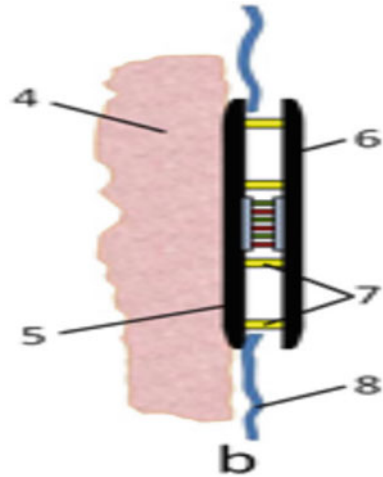
Fig. 5 (**c**) Simulated TEG
on the human body
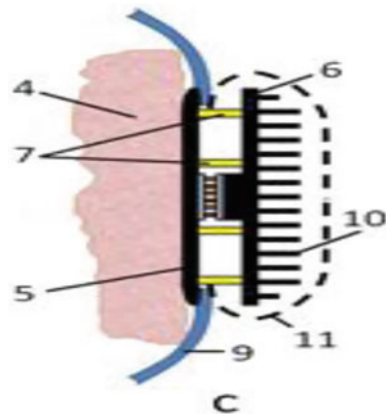
temperature difference is recorded only with the placement of a 8. piece of clothing
between the skin and the two plates (Fig. 5).

From Figure (c):

Secondly, along with the arrangements of figure (b) 9, watchstrap is kept in
contact with the 5. hot plate instead of 8. a piece of clothing, and the tempera-
ture difference ΔT is recorded. 10. Pins of radiator are the heat exchangers used to
transfer the thermal energy between [2] the human skin and two plates. Finally, the
11- Shock protection grid which is an integral component of electrical safety from
which the normal voltage will be identified and will provide the information needed
to determine the shock prevention method.

3 Proposed Work

Primarily, after understanding the concept behind the thermal energy harvesting, I have decided to use a method for heat generation from the human body and to construct a mathematical model for the thermal energy harvesting.

Secondly, based on the constructed mathematical model, a simulation was done using MATLAB Simulink and that was to test whether it is feasible or not for the required output, in order to charge the battery of Biomedical wearables. In order to obtain the output in volts, I used a boost converter to boost the output obtained from the thermal energy harvester and is then fed to the battery of the biomedical wearables.

3.1 Mathematical Equation

Thermal Energy Harvesting:

$$V_{oc} = \alpha \times \Delta T \tag{1}$$

where Voc is the output voltage, α is the Seebeck coefficient, and ΔT is the difference in temperature between the hot and cold surfaces.

$$P = \eta \tag{2}$$

where P is power, η is the thermal efficiency, and Q is the heat generated.

From another equation,

$$P = \frac{V_{OC^2}}{4R_e} \tag{3}$$

where R_{el} is the electric resistance

Substituting Eqs. (2) in (3) we get,

$$Q = \frac{V_{OC^2}}{4\eta \mathrm{Re}l} \tag{4}$$

Substituting Eqs. (1) in (4) we get,

$$Q = \frac{\alpha \times (\Delta T)^2}{4\eta R} \tag{5}$$

Thus Eq. (5) is the equation we simulate in MATLAB Simulink.
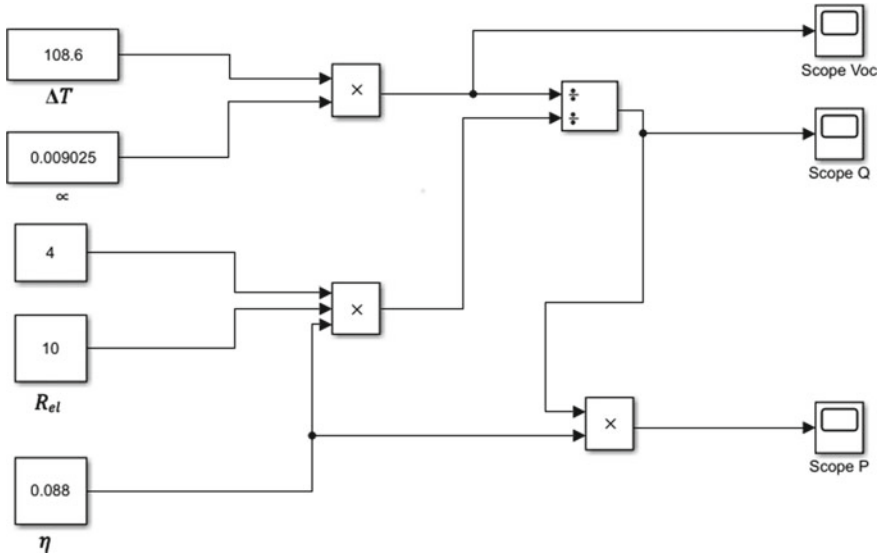
Fig. 6 MATLAB Simulink Block Diagram

3.2 Simulink Block

Thermal Energy Harvesting (Fig. 6).

3.3 Thermal Energy Harvesting

The output voltage of thermal energy harvesting is shown in figure based on the simulation done using MATLAB Simulink.
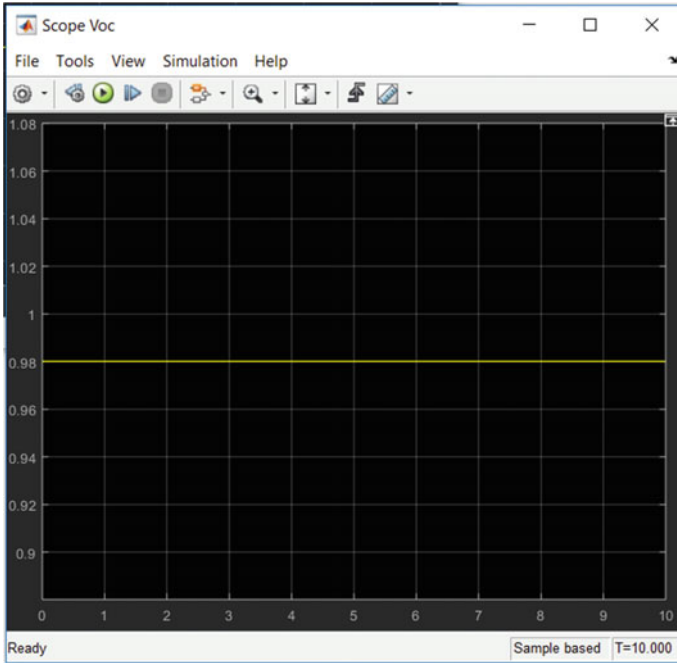
According to the theoritical calculatuons of Eq. (1) of the output voltage, V_{OC}, it is found that $V_{OC} = 0.98$ V. However, the experimental value based on the simulation done shows that the output voltage, $V_{OC} = 0.976$ V, which is close to the theoritical value. Figure representing the output power, P of the thermal energy harvester.
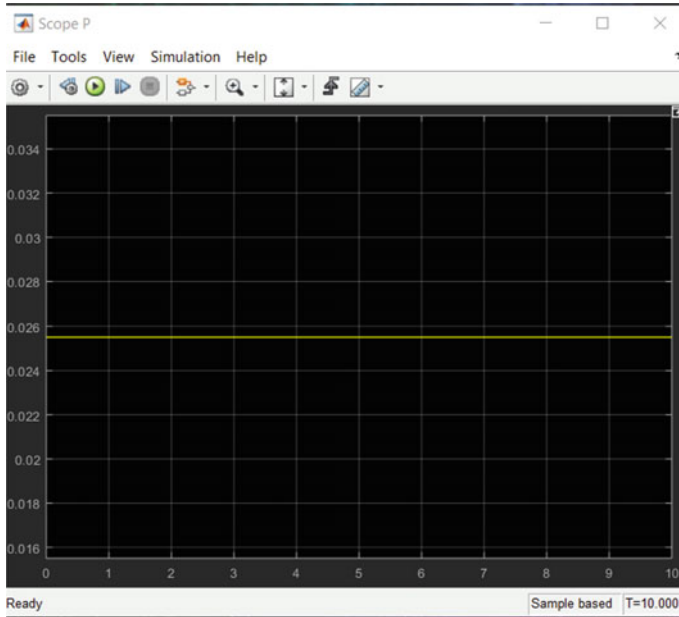
According to the theoritical calculations of Eq. (2) of the output power, P, it is found that P = 22.56 mW. However, the experimental value based on the simulation done shows that the output voltage, P = 24.4 mW, which is much similar to the theoritical value.

3.4 Thermocouple

The thermoelectric generator used as the thermal energy harvester is an off-the-shelf 20 mm × 20 mm thermocouple with a 10-cent coin as scale. The datasheet of the thermocouple given by the supplier, where the parameters for the hot side temperature is 200 °C and cold side temperature is 30 °C.

Thermocouple Output

The thermocouple is tested to see how much voltage it can produce by using a voltmeter. As this will be used to power up a wearable device, multiple points on the hand on two separate surfaces is tested to determine which will produce more voltage. In addition to the obtained output. the boost converter circuit is attached to increase the output voltage to power the biomedical wearable devices. Table 1 states the maximum voltages obtained from the test.

Table 1 Output voltage results from multiple points in hand on two surfaces

Plastic lid output in V			Counter top output in V		
Part of hand	Thermal output in V	Boosted output in V	Part of hand	Thermal output in V	Boosted output in V
Finger	10.8 mV	4.95 V	Finger	17.8 mV	8.0 V
Wrist	17.88 mV	8.4 V	Wrist	22.5 mV	9.94 V
Palm	20.79 mV	9.7 V	Palm	22.45 mV	9.93 V

3.5 Boost Converter Multisim

Considering the below boost converter circuit as a sample for the generation of output. The values from the TEG are fed to the boost converter circuit for output appraisal and then, it is fed to the biomedical wearables (Fig. 7).

4 Problem Solution

Based on the results in Table 1, it is found that when the thermocouple is on a counter-top, the voltage produced is higher than the plastic lid. However, it is difficult for the thermocouple to maintain its stability and temperature. Thus, the voltage fluctuates and does not give a stable reading.
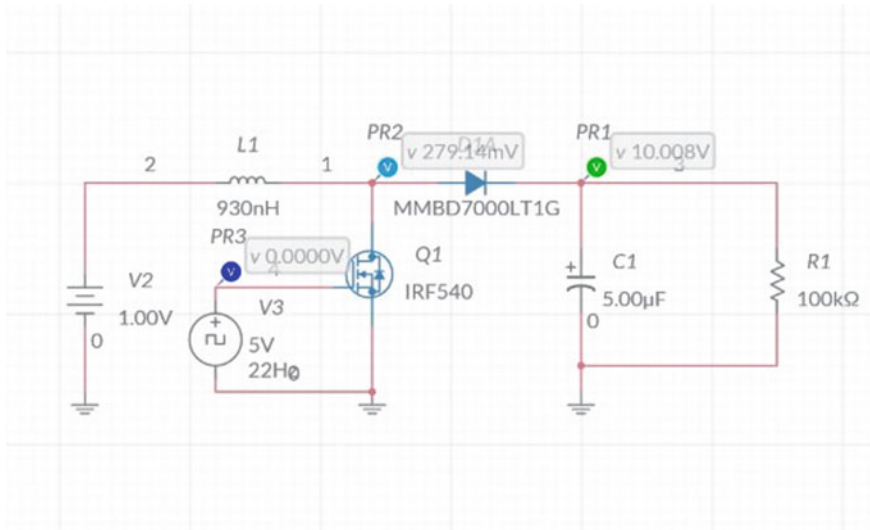
Fig. 7 Boost converter multisim

4.1 Boost Converter Output

Provided with the output of the finger, palm, and wrist readings by Multisim, the given input is boosted to obtain the required output in order to charge the biomedical wearables. The output will be generated by using the boost converter Multisim and fed to biomedical wearable devices such as the blood pressure monitor, pulsoximeter, heart rate sensor, temperature sensor, and pulse sensor.

5 Conclusions

In conclusion, it is found that the thermal energy harvesting simulation from MATLAB Simulink produced an output voltage Voc of 0.98 V, which is very much similar to the theoretical value 0.95 V. However, when conducting an experiment at 3 points on the hand; fingers, wrist, and palm from the purchased TEG, it is found that the output voltage produced a significantly lower power due to external factors such as thermal equilibrium and fluctuating atmospheric temperature. The maximum output voltage obtained from the fingers, wrist, and palm are 13.6 mV, 22.4 mV, and 17.2 mV, respectively. Therefore, the boost converter is used to increase the voltage output from the TEG. Thus, the appraisal values are 4.95 V, 8.4 V, and 9.7 V, which is in volts. Thus, the obtained output is very much supportive and effective to charge a Biomedical Wearable's Battery.

References

1. Udalagama CJ (2010) Electrical energy generation from body heat. IEEE ICSET
2. Arabee SBM (2018) A study of wireless charging and thermal energy harvesting methods for IOT applications
3. Thielen M, Sigrist L, Magno M, Hierold C, Benini L (2017) Human body heat for powering wearable devices: from thermal energy to application. Energy Convers Manage 131:44–54
4. Alhawari M, Mohammad B, Saleh H, Ismail M (2013) A survey of thermal energy harvesting techniques and interface circuitry. IEEE, pp 381–384
5. Ghosh A, Meenakshi Khalid S, Harigovindan VP (2015) Performance analysis of wireless body area network with thermal energy harvesting. In: Proceedings of 2015 global conference on communication technologies (GCCT 2015), pp 916–920
6. Ahiska R, Mamur H (2013) Design and Implementation of a new portable thermoelectric generator for low geothermal temperatures. IET Renew Power Gener
7. Leonov V, Su J, Vullers RJM (2010) Calculated performance characteristics of micromachined thermopiles in wearable devices. EDA Publishing/DTIP

The Significant Relationship Between E-Governance System and Government Operation Excellence

Almahdy Alhaj Saleh, Ahmed Rashid Alkhuwaylidee, Murugan Thangiah, and Arif Razzaq

Abstract Each government is looking for to provide the best services to establish efficiency and quality of performance. This goal could be accomplished by improving the service performance of entire sectors in society. The government of Syria has realized the importance of moving in the direction of information technology. Therefore, E-governance initiatives were launched in Syria as a part of overall country information technology in 20 s century. Each government sector has since upgraded the performance by having its websites and e-services application. However, there are gaps and loose connections exist among the sectors, which has accordingly tarnished the image of Syrian E-governance. This has led to significant questions about the requirement of modification and enhancement of such service. Hence, the purpose of this research is to investigate and explore the factors that drive the E-governance implementation and affect government performance as well as the government-citizen relationship in Syria.

Keywords E-governance · Government of Syria · Government-Citizen

A. A. Saleh (✉) · M. Thangiah
Postgraduate Centre Limkokwing University of Creative Technology Selangor, Selangor, Malaysia

M. Thangiah
e-mail: murugan.thangiah@limkokwing.edu.my

A. R. Alkhuwaylidee
Ahmed Rashid Alkhuwaylidee Computer Engineer Technology Mazaya University College Nassirya, Nassirya, Iraq

A. Razzaq
Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia

1 Introduction

In recent years, new technologies and concepts have therefore altered government contact with relation aspect such as organizations and citizens [1–3] by launching original service styles, similar to e-learning, e-banking, e-health, e-commerce, e-business, e-voting and others, the electronic government has been popularized with the term E-governance.

This refers to the services supplied by governmental to citizens via providing new information and communication technology technologies that may allow creating provision faster and easier. It had been first presented at the '90 s, it is regarded as an innovation introduced by using high-tech development from information technology increase [4].

Most scholars have divided E-governance into three main categories which are: Government to Government (G2G), Government to Citizens (G2C) and Government to Business (G2B). According to Fang (2002) and a few scholars have invented a fourth category which is Government to Employee (G2E) [5, 6], however, none of the scholars has explored the interaction between the opposite categories beyond the government. Category as they consider them to be the e-commerce a part of the E-governance. For instance, Business to Citizen (B2C), Business to Employee (B2E) and Employee to Citizen (E2C), don't seem to be yet identified within the E-governance system.

Moving closer to E-governance would provide several benefits to a republic's kind of services, collectively with finances procedure and control. Certainly, it would enhance the processing of the government in various components for the improvement of Government Operation Excellence (GOE). A range of these benefits is decreasing expenses and growing activity thanks to the discount of printed paper and acquisition of documents and area for storage [7–9].

Reducing the workload on employees is some other advantage to the work environment due to the fact employees are concerned with the minimal quantity of citizens and this can create an increased and serene environment. E-governance will advantage country areas together with large cities since citizens can practice for services lengthy as they're a part of the republic and in some instances and services, citizens may additionally practice for and request a service whilst they're overseas. Further advantages to the environment, knowledge, etc., should even be counted for the improvement of the country [10]. Additionally, E-governance has some challenges and downsides, and these should not be forgotten or ignored. Just to name a few of these difficulties are illustrated underneath cost, unreachability, hyper-surveillance, loss of transparency and accountability [11, 12].

2　The Proposed Method

The present research will use a quantitative approach to research given that it will examine the relationship between several independent variables as well as the acceptance of E-governance system by users in Syria. The dependent variables of the study are the Government Operation Excellence (GOE), while the independent variables are Citizens-centricity, facilitating conditions, performance expectancy, effort expectancy, flexibility system. And the mediation variable are behavioral intention.

Given the quantitative leanings of the research, the study will thereby focus on existing statistical data, facts and figures. The accompanying data collection tools have been identified as questionnaires. Questionnaires were particularly effective in collecting quantitative data, particularly if researchers are aware of how the variables are to be tested [13]. In addition to this, questionnaires are also effective when it comes to collecting data from a wide sample of respondents, dispersed across a number of locations. Using a questionnaire also allows researchers to ensure that respondents are presented with a uniform set of questions, this will later result in uniform sets of data which are not challenging as far as analysis is concerned.

Questionnaires, if designed effectively are unlikely to be time-consuming, therefore allowing data to be collected in a relatively short space of time [13–15]. Given the fact that the research sample is geographically dispersed across Syria, the researcher relied on a questionnaire approach to investigate.

3　Research Method

According to [16], the method is a set of principles that in any particular situation has to be reducing to a method uniquely suited to a given situation. The research methodology for this research, it presents a systematic and organized procedure that enables the investigation of the relationship between the different variables attached, namely E-governance, behavioral intention to use and government operation excellence.

In achieving the objectives, the research methodology deals with procedures involved in this research. Questionnaire targeting to survey the adoption information technology for promoting E-governance system among citizens in Republic of Syria are including in the instrument.

Series of tests such as Structural Equation Model (SEM), and reliability analysis were conducted to confirm that the instruments were valid (Reliability and Convergent Validity). Also, the research methodology that was followed and adopted by this research was detailed out. Includes all the procedures, process, and guidelines that according to research literature is suitable for this kind of the investigation.

The main aim of this research is to study and investigate the mediation effect of behavioral acceptance on the relationship between E-governance and government operation excellence in Syria.

3.1 Research Framework

The review of the information system relevant literature indicates that the acceptance of information systems, especially E-governance systems are influenced by three factors, namely: (1) individual factors (2) system factors (3) environmental factors [17–22].

Concerning theoretical contributions, the principal theories that are used in this study are the Implications of DeLone and McLean model, Technology Acceptance Model (TAM), the Task-Technology Fit (TTF) and Unified Theory of Acceptance and Usage Technology (UTAUT). In this research we combining thus theories into a new framework in order to achieve the optimum GOE for the Syrian citizens.

Hence, the evaluation of GOE will be supported by the equal common criterion as is stipulated within the SERVQUAL approach. It will be classified into assurance, tangibles, reliability, responsiveness and empathy [23].

The focus of this study is to examine the effect of E-governance on GOE.

H1: There is a significant relationship between an E-governance system and Government Operation Excellence.

Conceptualization of Factors.

Information and communication technology review of the literature has shown that a number of studies have been conducted to investigate users' acceptance of E-governance systems [24–26].

The majority of previous studies have focused on employees and citizens environments within economically developed nations. While developing countries have much to gain from utilising information technology, they usually receive little to no attention in regards to research [24–30]. To do so, they need to improve upon three aspects: the implementation environment, individual factor, and system factor. In that respect, the current study will aim to develop an understanding of the various factors which influence the acceptance of E-governance system by Syrian citizens.

Behavioral intention.

Behavioral intention to use: Citizens get motivated to use a certain system or application if it was used or marketed by social groups, friends or family or any matter that touches the person's behavior or feeling [31]. A number of authors [31–34] have attested to the fact that BI has been employed and utilized as a dependent variable in the examination of E-government acceptance.

Compeau and Higgins [34] Argues that intention is the determinant of the theoretical foundation in information system research in the sense that individual behavior dictates whether or not to use or not use the technology. Also, BI refers to specific prompts that cause individuals to act in certain ways or perform certain behaviors [35]. This study adding three related determinants of behavioral intent to use E-governance. They are as follows: system flexibility, citizens-centricity, and facilitating conditions. They are added to reflect the individual factor because they have a significant influence on the intention to use of E-governance.

H2: Behavioral Intention to Use moderates the relationship between an E-governance system and Government Operation Excellence.

Performance Expectancy (PE).

A number of authors including [36, 37] outline PE as an individual's expectations of improvement, when using a specific system which could enhance their work performance. A host of the literature suggests that performance expectancy is a key variable in predicting user intention in information and communication technology contexts [32, 37–40].

Adapting this construct (PE) to E-governance system suggests that Syrian citizens think that using E-governance system will enhance their learning and consequently will be reflected on their efficiency. Within an E-governance context, there are many studies which suggest that PE directly influences user intent to use an E-governance system [21, 41, 42].

H2: Performance expectancy will have a positive influence on behavioral intentions to use E-governance services.

Effort Expectancy (EE).

Effort expectancy, as described by both [36, 37], is the level of ease felt by individuals when using information systems. Several studies have demonstrated that effort expectancy has a powerful effect on behavioral intention. That is to say, the intent to use information system technology [36, 38, 43]. In that sense, then, effort expectancy theory suggests that individuals, in this case, will usually accept a system if they feel that said the system is easy to use. This study will aim to determine the degree to which independent factors such as effort expectancy correlate with the intent to use E-governance system.

H3: Effort expectancy will have a positive influence on behavioral intentions to use E-governance services.

Citizens-Centricity (CC).

A citizen centric involves the availability of services from the top user's point of view instead of the attitude of the government department [44, 45]. E-governance project should have a citizen centric perception. However, citizen-centric service delivery is additionally an elaborate issue with many perspectives that require to be considered at the very beginning of a change project. His provision of citizen centric service has been identified by some researchers as a critical success factor [46].

The citizen-centric delivery E-governance services determinants in developing countries would allow for better understanding of the citizens' needs, desires, requirements and priorities that must be taken into consideration by governments to ensure the success of services [47].

H4: Citizens-Centricity will have a positive influence on behavioral intentions to use E-governance services.

System Flexibility (SF).

System flexibility within the context of E-governance refers to the level of freedom that citizens' feel E-governance permits them, in terms of time and place [48].

[49] Study concluded that flexibility greatly impacts behavioral intent. Moreover, [50] explored the differing variables which influenced user satisfaction in regards to online courses. [50] Concluded that, in most cases, perceived flexibility is vital regarding user satisfaction. System Flexibility, which is the ability to implement different processes and apply different facilities to achieve the same goals. It consists of items such as: product volume flexibility, people flexibility, etc.; [51].

E-governance literature has also provided us with many studies that indicate the predictive ability of flexibility of E- governance system in determining behavioral intention [21]. In this research, Syrian citizens will adopt the E-governance system if they perceive that they can use the system at any time from any place.

H5: System Flexibility will have a positive influence on behavioral intentions to use E-governance services.

Facilitating Conditions (FC).

Facilitating conditions, as defined by Venkatesh (2003) are the factors which lead users to believe that the technical and organizational infrastructure of Syria governance can support E-governance application system. In that respect, facilitating conditions can result in the rejection of information technology systems due to lack of faith in the respective organizations and their systems themselves [37]. [39] Also stated that facilitating conditions resources, such as money and time would motivate a person to use a particular system. Several studies in the information technology context have indicated that the facilitating conditions construct has a significant direct influence on user behavior [52–54].

Adapting this construct to E-governance system suggests that Syrian citizens perceive that technology infrastructure, and governance infrastructure will enable them to use the E-governance system without any problem, and to interact with other users [55].

H6: Facilitating conditions will have a positive influence on behavioral intentions to use E-governance services.

SERVQUAL.

Jong and Wang [56] Refined their previous service quality evaluation model [57] and created the SERVQUAL. In short, the authors suggest that a practical service quality perception stems from the chasm between consumer expectations and therefore the performance of the service that's actually delivered. Thus, finishing up new studies, with new data collections and analysis, they were able to improve their scale, reducing the initial ten dimensions to seven: (1) Tangibles; (2) Reliability; (3) Responsiveness; (4) Communication, Credibility, Safety, Competence; 5) Courtesy; 6) Knowing the customer; and (7) Access.

During a second stage, which entailed another improvement of the study, SERVQUAL was reduced to 5 dimensions (Table 1). According to [57], in spite of the kind of service, consumers basically use an equivalent criterion to assess quality. Service quality could be a general opinion the client forms regarding its delivery, which is constituted by a series of successful or unsuccessful experiences.

It is difficult to live the standard of service operations because they need the characteristic intangibility. Geared toward solving this problem, [57] developed a

Table 1 SERVQUAL dimensions

dimensions	Definition
Tangibles	The physical appearance of equipment and staff involved in the provision of services
Reliability	The ability to deliver the offered service in the scheduled date and time, independently of any problems that may arise
Responsiveness	The ability to help every client in an unhesitating way and to effectively and positively promote the service
Assurance	The ability to inspire trust, security and technical quality on the part of the staff when delivering a service
Empathy	The available and unhesitating assistance provided to the client, individualized and focused on his main interests
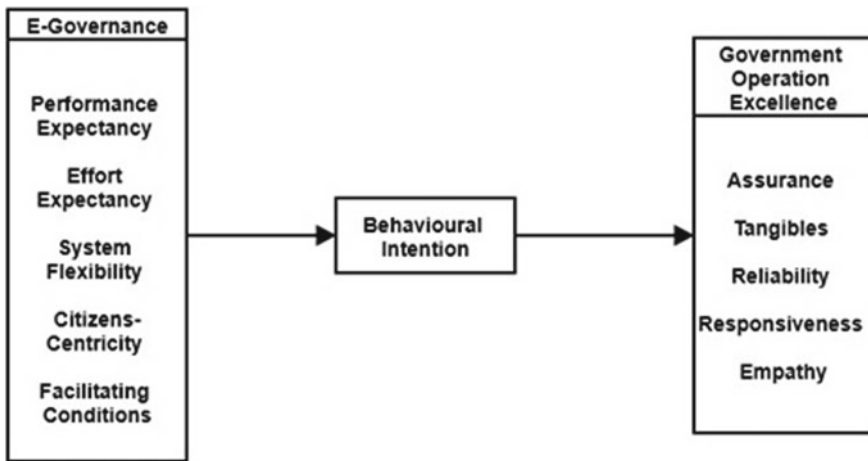
Fig. 1 Research framework

strategy within which there's a comparison between several orders of expectations and perceptions of service quality by the patron (Fig. 1).

4 Results and Discussion

The measurement framework in this study was made up of 42 items to measure 11 constructs, namely: Citizens-Centricity (CC), Facilitating Conditions (FC) System Flexibility (SF), Effort Expectancy (EE), and Performance Expectancy (PE). Empathy (EMP), Responsiveness (RES), Reliability (REL), Assurance (ASSU), Tangibles (TAN), and Behavioral Intentions (BI).

Table 2 Table type styles

Fit index	Modified model	Recommended values	Source
Df	1013		
CMIN ($\chi 2$)	1238.809		
p-value	0.000	> 0.05	[58, 59]
$\chi 2/df$	1.223	≤ 5.00	[58]
GFI	0.881	≥ 0.90	[59]
AGFI	0.867	≥ 0.80	[60]
CFI	0.976	≥ 0.90	[58, 61]
TLI	0.974	≥ 0.90	[62, 63]
IFI	0.976	≥ 0.90	[62, 63]
RMSEA	0.020	≤ 0.10	[64]

4.1 Reliability and Convergent Validity

Once the uni-dimensionality of the constructs was achieved, each of the constructs was assessed for their reliability and validity. Reliability is assessed using Cronbach's alpha, construct reliability (CR) and average variance extracted (AVE), while for validity using construct, including convergent and discriminant Table 2. Represents the result of Cronbach alpha and convergent validity for the modified measurement model with 42 reminder items.

As shown in Table 2, the remaining indicators have high factor loadings ranging from 0.863 to 0.915 indicating that these indicators have preserved the meaning of the factors.

Table 2 also shows that the AVE, which reflects the overall amount of variance in the indicators accounted for by the latent construct, was above the cut-off 0.5 for all constructs as suggested by [65], ranged from 0.531 to 0.684.

The composite reliability values, which depict the degree to which the construct indicators indicate the latent construct, exceeded the recommended value of 0.6 for all constructs as recommended by [66], ranging from 0.797 to 0.915.

The Cronbach's Alpha values, which describes the degree to which a measure is error-free, range from 0.795 to 0.914 which were above the threshold of 0.7 as suggested by [58]. Therefore, the achieved Cronbach's Alpha for all constructs was considered as sufficiently error-free.

4.2 Goodness of Fit Indices

Based on the current GOF results, the chi-square is substantial at the 0.010 level. The normed fit index of lowest difference chi-square can be removed if the sample size for the research is more than 200 [82, 83]. After correcting for degrees of freedom

Table 3 Type styles

Construct	Internal reliability cronbach alpha	Convergent validity	
		Average variance extracted (AVE)[a]	Composite reliability (CR)[b]
CC	0.795	0.567	0.797
EMP	0.843	0.531	0.819
RES	0.856	0.599	0.857
FC	0.854	0.576	0.844
REL	0.896	0.684	0.896
TAN	0.858	0.632	0.872
ASSU	0.809	0.597	0.815
BI	0.862	0.557	0.863
SF	0.862	0.611	0.863
EE	0.914	0.641	0.915
PE	0.892	0.577	0.891

in based on the number of variables, the adjusted GFI (AGFI) was 0.881, which was greater than the recommended cut-off value of according to [84, 85]. The model predicted 86 percent of the variances and covariance in the survey results, according to the report. The model showed a satisfactory fit of collected data on the CFI, TLI, and IFI indices with values more than the cutoff value of 0.9 (i.e., 0.976, 0.973, and 0.975, respectively) [86, 87]. In addition, the root-mean square error of approximation (RMSEA) was 0.020, and that was lower than the suggested threshold of 0.1 [88, 89, 90]. Furthermore, the fact that the Comparative CMIN/df (1.224) has been less than 5 revealed the good fit of the model [89, 90]. See Table 3.

5 Conclusion

The study has distributed 600 questionnaires to Syrian citizens. From the total, only 423 questionnaires were returned, with only 393 questionnaires were usable (65.5%). 16 respondents returned the questionnaire with missing data of more than 30% for each questionnaire. According to [67–70] the questionnaires that have missing data of more than 25% should be omitted. Additionally, 14 of the respondents returned the empty questionnaire. Therefore, only 385 cases have been used for the study analysis.

A total of six hypotheses has been tested in achieving the objectives of the research. From the total, five direct relationships have been found to be statistically significant to users' intention to use framework of E-governance system, since the study indicates that the first significant relationship is between Performance Expectancy and Behavioral Intention (0.000). The second significant relationship is between Effort

Expectancy and Behavioral Intention (0.000). The third significant relationship is between Citizens-Centricity and Behavioral Intention (0.026). The fourth significant relationship is between System Flexibility and Behavioral Intention (0.002). The fifth significant relationship is between Facilitating Conditions and Behavioral Intention (0.000).

When taking into consideration the indirect relationship, two relationships were proposed to statistically affect between E-governance systems (EGOV) and Government Operation Excellence (GOE). The coefficient parameters estimate of the structural model are examined to determine whether E-governance (EGOV) as independent variables have positive indirect effects on Government Operation Excellence (GOE). Through Behavioral Intention (BI).

The first important indirect relationship was proposed between E-governance (EGOV) and Government Operation Excellence (GOE) with a total effect of (0.119). The second important indirect relationship was proposed between E-governance (EGOV) and Government Operation Excellence (GOE) with Satisfaction Services with a total effect of (0.096). The third important indirect relationship was proposed between E-governance (EGOV) and Government Operation Excellence (GOE) through Satisfaction Services with a total effect of (0.184). The fourth important indirect relationship was proposed between E-governance (EGOV) and Government Operation Excellence (GOE) with a total effect of (0.119). The fifth important indirect relationship was proposed between E-governance (EGOV) and Government Operation Excellence (GOE) with Behavioral Intention (BI) with a total effect of (0.088). The sixth important indirect relationship was proposed between E-governance (EGOV) and Government Operation Excellence (GOE) through Behavioral Intention (BI) with a total effect of (0.031).

References

1. Lee M-C (2010) Explaining and predicting users' continuance intention toward e-learning: an extension of the expectation-confirmation model. Comput Educ 54:506–516
2. Rokhman A (2011) E-Governance adoption in developing countries; the case of Indonesia. J Emerg Trends Comput Inform Sci 2:228–236
3. Gil-Garcia JR, Zhang J, Puron-Cid G (2016) Conceptualizing smartness in government: an integrative and multi-dimensional view. Gov Inf Q 33:524–534
4. Grönlund Å, Horan TA (2005) Introducing E-Gov: history, definitions, and issues. Commun Assoc Inf Syst 15:39
5. Bertot JC, Jaeger PT, Mcclure CR (2008) Citizen-Centered e-governance services: benefits, costs, and research needs. In: Proceedings of the 2008 international conference on digital government research, 2008. Digital government society of North America, pp 137–142
6. Carter L, Belanger F (2004) The influence of perceived characteristics of innovating on e-governance adoption. Electron J E-Governance 2:11–20
7. Tolbert CJ, Mossberger K (2006) The effects of e-government on trust and confidence in government. Public Adm Rev 66:354–369
8. Helbig N, Gil-García JR, Ferro E (2009) Understanding the complexity of electronic government: implications from the digital divide literature. Gov Inf Q 26:89–97

 9. Sharma SK, Govindaluri SM, Sharma RK (2013) A two-staged regression-neural network model for understanding and predicting the quality determinants of e-governance services in Oman. Int Found Res Dev (Ifrd) 57
10. Huang Z, Bwoma PO (2003) An overview of critical issues of e-governance. Issues Inf Syst 4:164–170
11. Bertot JC, Jaeger PT, Grimes JM (2010) Using Icts to create a culture of transparency: e-governance and social media as openness and anti-corruption tools for societies. Gov Inf Q 27:264–271
12. Alzahrani AI (2011) Web-Based e-governance services acceptance for G2c: a structural equation modelling approach
13. Sekaran U (2007) Metodelogi Penelitian Untuk Bisnis (Terjemahan). Edisi Empat, Jakarta, Salemba Empat
14. Wozencroft A, Hardin R (2011) Examining counselor satisfaction in an outdoor therapeutic recreation camp. Entertainment Manage
15. Alkhuwaylidee A, Almahdy AS (2019) Syrian e-governance framework toward government excellence service. In: Paper presented at the technologies for developing information systems Tris-2019
16. Checkland P (1981) Systems theory. Syst Pract
17. Zhang JQ, Craciun G, Shin D (2010) When does electronic word-of-mouth matter? a study of consumer product reviews. J Bus Res 63:1336–1341
18. Saleh AA, Alyaseen IFT (2021) Successful factors determining the significant relationship between e-governance system and government operational excellence. Bull Electr Eng Inf 10(6):3460–3470
19. Algharibi A, Keung LC, Niukyun S, Zhao L, Tyler E, Taweel A, Hobbs R (2012) Adapting the unified theory of acceptance and use of technology (Utaut) as a tool for assessing future technology adoption of e-trial software tools. In: Paper presented at the proceedings of the 10th international conference on information communication technologies in health
20. Alzahrani ME, Goodwin RD (2012) Towards a utaut-based model for the study of e-governance citizen acceptance in Saudi Arabia. In: Proceedings of world academy of science, engineering and technology (No. 64). World academy of science, engineering and technology
21. Alkhuwaylidee AR (2019) Extended unified theory acceptance and use technology (utaut) for e-learning. J Comput Theor Nanosci 16(3):845–852
22. Cody-Allen E, Kishore R (2006) An extension of the utaut model with e-quality, trust, and satisfaction constructs. In: Paper presented at the proceedings of the 2006 ACM Sigmis CPR conference on computer personnel research: forty four years of computer personnel research: achievements, challenges & the future
23. Nanayakkara C, Widdett D (2005) A model of user acceptance of e-learning technology: a case study of a polytechnic in New Zealand. In: The fourth international conference on information system technology and applications. New Zealand, pp 180–189
24. Huai J (2011) Quality evaluation of e-governance public service. In: 2011 International conference on management and service science, 2011. IEEE, pp 1–4
25. Chau PYK, Hu PJH (2002) Investigating healthcare professionals' decisions to accept telemedicine technology: an empirical test of competing theories. Inf Manage 39(4):297–311
26. Al-Khatib H, Lee H, Suh C, Weerakkody V (2019) E-Governance systems success and user acceptance in developing countries: the role of perceived support quality
27. Prieto-González L, Matsuo MT, Colomo-Palacios R (2014) Electronic business and marketing
28. Susanto TD, Diani MM, Hafidz I (2017) User acceptance of e-governance citizen report system (a case study of city113 app). Procedia Comput Sci 124:560–568
29. Taherdoost H (2018) Development of an adoption model to assess user acceptance of e-service technology: e-service technology acceptance model. Behav Inf Technol 37:173–197
30. Maharjan M (2019) Citizens' acceptance of e-governance services in Nepal: applying utaut model. In: Advanced engineering and ict–convergence 2019 (ICAEIC-2019), p 122
31. Abbad MM et al (2009) Looking under the bonnet: factors affecting student adoption of e-learning systems in Jordan. In: The international review of research in open and distance learning, vol 10, no 2

32. Hung S-Y, Chang C-M, Kuo S-R (2013) User acceptance of mobile e-governance services: an empirical study. Gov Inf Q 30:33–44
33. Alkhuwaylidee A, Almahdy AS (2019) Syrian e-government framework toward government excellence service. Технологии Разработки Информационных Систем Трис-2019
34. Compeau DR, Higgins CA (1995) Computer self-efficacy: development of a measure and initial test. MIS Q 189–211
35. Hsia J-W, Tseng A-H (2008) An enhanced technology acceptance model for e-learning systems in high-tech companies in Taiwan: analyzed by structural equation modeling. In: 2008 International conference on cyberworlds. IEEE, pp 39–44
36. Hasan H, Ditsa G (1999) The impact of culture on the adoption of it: an interpretive study. J Glob Inf Manage (JGIM) 7(1):5–15
37. Barua M (2012) E-Governance adoption in government organization of India. Int J Manag Public Sector Inf Commun Technol 3(1):1
38. Venkatesh, V., Morris, M. G., Davis, G. B. & Davis, F. D. 2003. User Acceptance Of Information Technology: Toward A Unified View. Mis Quarterly, 425–478
39. Davis F, Bagozzi R, Warshaw P (1992) Extrinsic and intrinsic motivation to use computer in the workplace. J Appl Soc Psychol 22(14):1111–1132
40. Naor, O., & Geri, N. (2008). Easy as e-mail? Probing the slow adoption of an online assignment submission system. In: Proceedings of the chais conference on instructional technologies research 2008, vol 5, pp 94–101
41. Taylor S, Todd P (1995) Understanding information technology usage: a test of competing models. Inf Syst Res 6(2):144–176
42. Venkatesh V, Davis FD (2000) A theoretical extension of the technology acceptance model: four longitudinal field studies. Manage Sci 46:186–204
43. Goussal DM, Lezcano MSU (2003) Synchronous distance learning and virtual classrooms: a case study on student expectations and conditioning factors. Australas J Educ Technol 19
44. Sahin I, Shelley M (2008) Considering students' perceptions: the distance education student satisfaction model. Educ Technol Soc 11(3):216–223
45. Marchewka J, Liu C, Kostiwa K (2007) An application of the utaut model for understanding student perceptions using course management software. Commun IIMA 7(2):93–104
46. Karim MA (2003) Technology and improved service delivery: learning points from the Malaysian experience. Int Rev Adm Sci 69:191–204
47. Themistocleous M, Irani Z, Love PE (2005) Developing e-governance integrated infrastructures: a case study. In: Proceedings of the 38th annual Hawaii international conference on system sciences. IEEE, pp 228–228
48. King S, Cotterill S (2007) Transformational government? The role of information technology in delivering citizen-centric local public services. Local Gov Stud 33:333–354
49. Hermans CM, Haytko DL, Mott-Stenerson B (2009) Student satisfaction in web-enhanced learning environments. J Instr Pedagogies 1
50. Davis FD, Bagozzi RP, Warshaw PR (1989) User acceptance of computer technology: a comparison of two theoretical models. Manage Sci 35:982–1003
51. Elsheikh Y, Azzeh M (2014) What facilitates the delivery of citizen-centric egovernment services in developing countries: model development and validation through structural equation modeling. Int J Comput Sci Inf Technol 6(1):77
52. Hsia JW, Tseng AH (2008) An enhanced technology acceptance model for e-learning systems in high-tech companies in Taiwan: analyzed by structural equation modeling. In: International conference on cyberworlds, pp 39-44
53. Almahamid SM (2013) E-Governance system acceptance and organizational agility: theoretical framework and research agendas. Int J Inf Bus Manage 5:4
54. Selim HM (2005) Critical success factors for e-learning acceptance: confirmatory factor models. Comput Educ 49(2):396–413
55. Folorunso O, Ogunseye OS, Sharma SK (2006) An exploratory study of the critical factors affecting the acceptability of e-learning in nigerian universities. Inf Manage Comput Secur 14(5):496–505

56. Jong D, Wang TS (2009) Student acceptance of web based learning system. In: 2009 International symposium on web information system and application (Wisa'09), pp 533–536
57. Tarcan E, Varol ES, Kantarcı K, Fırlar T (2012) A study on kazakh academicians' information technology acceptance. Bilig 62:205–230
58. Li, H. And Suomi, R., 2009. A Proposed Scale For Measuring E-Service Quality. International Journal Of U-And E-Service, Science And Technology, 2(1), Pp.1–10
59. Nunnally JC, Bernstein IH (1994) Psychometric Theory. Mcgraw-Hill, New York
60. Bagozzi, R.P. And Yi, Y. (1988) 'On The Evaluation Of Structural Equation Model', Journal Of Academy Of Marketing Science, Vol. 16, No.1, Pp.74–94
61. Hoyle RH (1995) Structural equation modeling: concepts, issues, and applications. Sage Publications
62. Chau PYK, Hu PJH (2001) Information technology acceptance by individual professional: a model comparison approach. Dec Sci 32(4):699–719
63. Byrne BM (2013) Structural equation modeling with lisrel, prelis, and simplis: basic concepts, applications, and programming: psychology press
64. Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL (2006) Multivariate data analysis, 6th edn. Pearson Prentice Hall, United State Of Amreica
65. Parasuraman A, Zeithaml VA, Berry LL (1988) Servqual: a multiple-item scale for measuring consumer perc. J Retail 64:12
66. Parasuraman A, Zeithaml VA, Berry LL (1985) A conceptual model of service quality and its implications for future research. J Mark 49:41–50
67. Ho R (2006) Handbook of univariate and multivariate data analysis and interpretation with Spss. United States of American: Chapman & Hall/Crc, Taylor & Francis Group
68. Schumacker RE, Lomax RG (2010) A beginner's guide to. Structural equation modeling, 3rd edn. Taylor & Francis Group, New York
69. Razzaq A, Mohammed AA (2020) Cloud Erp in Malaysia: benefits, challenges, and opportunities. Int J 9.5
70. Mohan NR et al (2014) Using cloud–based web application for universityes: a case of university of technology in Iraq. Adv Comput 4.1:15–17
71. Saleh AA, Alkhuwaylidee AR, Thangiah M (2021) Successful factors determining the user behaviour in information system. J Phys Conf Ser 1963(1):012161

Trajectory Tracking Control of Industrial Robot Manipulator Using VS Control System with STHWS Method

S. Nandhakumar ⓘ**, B. Shreeram** ⓘ**, R. Thirumalai** ⓘ**, K. Sakthisudhan** ⓘ**, and P. Sivaprakash** ⓘ

Abstract In this paper, an experimental investigation has been made to evaluate the effectiveness of proposed integration of single term haar wavelet series method with a variable structure control system. The proposed method is applied to non linear system to obtain solution for trajectory tracking problem by considering dynamics of the robot as singular system time varying model. The aim is control the variation or error occurs when the robot is working. In order to minimize the error, integration of STHWS method with VS control has been designed and implemented in two dof industrial robot. The experiment has been conducted with various time intervals and results are analyzed. The obtained results reveals that the proposed integration of STHWS method with VS control system works well and robust with singular system time varying model for control of positions.

Keywords Trajectory tracking · Manipulator dynamics · Single term haar wavelet series (STHWS) · Position control · Singular and non-singular systems · Time varying and invariant · Variable structure control (VS)

1 Introduction

In today's competitive world, it is very clear that the product manufactured should satisfy the needs of the customer. This scenario increases demand of quality in manufacturing and it tends to improve the infrastructure to automation and robotics. This demand increases the considerable research in robot dynamics and control. The correct robot manipulator dynamic model plays a vital in control mechanisms. The formulation of the dynamic model involves various internal and external factors. As

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022
R. Ibrahim et al. (eds.), *International Conference on Artificial Intelligence for Smart Community*, Lecture Notes in Electrical Engineering 758,
https://doi.org/10.1007/978-981-16-2183-3_31

it operates, certain physical parameters like load and speed controls these forces [1–4]. In order to achieve the proper control the exact dynamic model is very important. In this proposed research, the dynamic model is formulated based on the Euler–Lagrange principle [5]. To attain the required response and performance laws and strategies are to be determined within the dynamic model.

Many literatures are available and the performance of robots has been studied using differential equations. These differential equations are easy to solve and provides a simple analytical solutions without non linearity's [6–8]. For representing any robot configurations where constraints causes rise in contact forces, differential equations with singular systems are more suitable ([9] McClamroch 1986). Robots that are applied for advanced manufacturing jobs, the complexity involved in building a mathematical model is a direct indication. Hence, it is necessary to study the behavior of proposed model robot arm and its performance for designing robustness arm for industrial use. Moreover, it is essential to study various control techniques to analysis problem and issues pertaining to control robot.

Due to various control issues and limitations of traditional controllers, the VS control is highly desirable which has numerous interesting and significant properties such as exact knowledge of the manipulator's model is not needed and under certain conditions, the system performance is insensitive to bounded external disturbances. These properties are unique and cannot be easily obtained by other controllers which makes VS control a perfect model for manipulator controls [10–14].

The objective of proposed work is to investigate the effectiveness of the proposed integration of STHWS method with VS control by analyzing trajectories of each joint of 2 DOF industrial pick and place robot. The analyses involves the tracing the positions and velocities of each joint for find out the experimental results follows the desired trajectories which is obtained by using cubic polynomial method with zero speed at end point. During continuous operations, the trajectories has been traced and plotted. Based on the task, the control system calculates the torque if a trajectory goes beyond limits, the mathematical block programmed in the controller calculates the new torque and controls the variations in position by varying input torque. The time taken for completing one cycle is 48 s. The results are plotted and compared with the reported values. From the results it has been observed that STHWS to VS control non singular system time model and supports in attaining preferred trajectories within the accuracy range expected.

2 Singular System Time-Varying Case Robot Arm Model

The dynamics of a robot arm can be represented in terms of system of non linear equations by Nandhakumar et al. [15]

$$T = P(Q)\ddot{Q} + R(Q, \dot{Q}) + S(Q) \tag{1}$$

where:

$P(Q)$ Is the Coupled inertia matrix,

$R(Q, \dot{Q})$ Is the matrix of coriolis and centrifugal force

$S(Q)$ Is the Gravity matrix,

T Is the Input torques applied at various joints.

For a manipulator with 2 DOF, under the assumption of lumped equivalent masses and massless- links, the dynamic equations is indicated below

$$T_1 = K_{11}K_{12}q_1 + K_{122}(q_2)^2 + K_{112}(q_1q_2) + K_1$$
$$T_2 = K_{21}K_{22}q_2 + K_{211}(q_1)^2 + K_2 \qquad (2)$$

where

$K_{11} : (W_1 + W_2) d_1{}^2 + W_2 d_2{}^2 + 2W_2 d_1 d_2 \cos(q_2)$

$K_{12} : W_2 d_2{}^2 + W_2 d_1 d_2 \cos(q_2)$

$K_{21} : W_{12}$

$K_{22} : W_2 d_2{}^2$

$K_{112} : -2W_2 d_1 d_2 \sin(q_2)$

$K_{122} : -W_2 d_1 d_2 \sin(q_2)$

$K_{211} : W_{122}$

$K_1 : [(W_1 + W_2)d_1 \sin(q_1) + W_2 d_2 \sin(q_1 + q_2)]g$

$K_2 : [W_2 d_2 \sin(q_1 + q_2)]g$

A linear time-varying singular system discussed by Nandhakumar et al. [15], which is in the following form

$$M(t)\dot{x}(t) = C(t)x(t) + D(t)u(t) \qquad (3)$$

with initial condition $x(0) = x_0$.

where $x(t)$ is a n-state vector and

$u(t)$ Is the p-input control vector and

$M(t)$ Is a $n \times n$ singular matrix,

$C(t)$ and $D(t)$ are $n \times n$ and $n \times p$ matrices respectively.

The said algorithm has been reported to be used for solving position and velocity errors in robot arm during operation [16–18].

3 Implementation on an Industrial Robot

The experiment is conducted using 2 dof industrial robot manipulator. The configuration and hardware details are discussed by Nandhakumar et al. [5] as shown in the Fig. 1. In the experiment has carried out in four stages which include two rotations and two linear movements. Initially the arm is in home position i.e. in the top position. So it has to move to some degree in rotation (2 radians) followed by linear

Fig. 1 Industrial Robot
Manipulator

down movement (0–10 cm) to pick the component. After picking up the object, it has
to move upward from (10–5 cm) and reaching the third position, it rotate and place
the object. During operations, each joint trajectory is traced. Whenever new position
is out of control, the proposed model evaluates the desired torque and compares it
with the existing one. If it goes out of control, the proposed integration estimates the
torque based on the varying positions, and compares with existing torque. Finally,
the exact trajectories have been achieved and results of velocity errors and positions
errors have been plotted and shown in Figs. 2, 3, 4, 5, 6, 7, 8 and 9. From these
figures, it has been observed that the proposed integration controls the errors and
provides the projected range of accuracy.

4   Results and Discussion

Figures 2, 4, 6 and 8 shows the positions and velocity profiles of Joint 1 and 2
respectively arising out of the proposed algorithm. From these profiles, it has been
observed that trajectories very closely follow exact trajectories which are obtained by
cubic polynomial method. The position errors and velocity errors are as shown in the
Figs. 3, 5, 7 and 9. The errors obtained are large than the numerical simulations [15].
The obtained level of accuracy is comparatively good with steady errors. A detailed
comparison of the results with [5] shows that the position and velocity error for joint
1 and joint 2 is less. From Fig. 3, at joint 1—Position error—4.0×10^{-3} rad, Velocity
error—3×10^{-2} and at joint 2—Position error 5.0×10^{-2} rad and Joint velocity
error 4×10^{-2} rad/s which is approximately 20% less than the model reported by
Nandhakumar el al 2013.
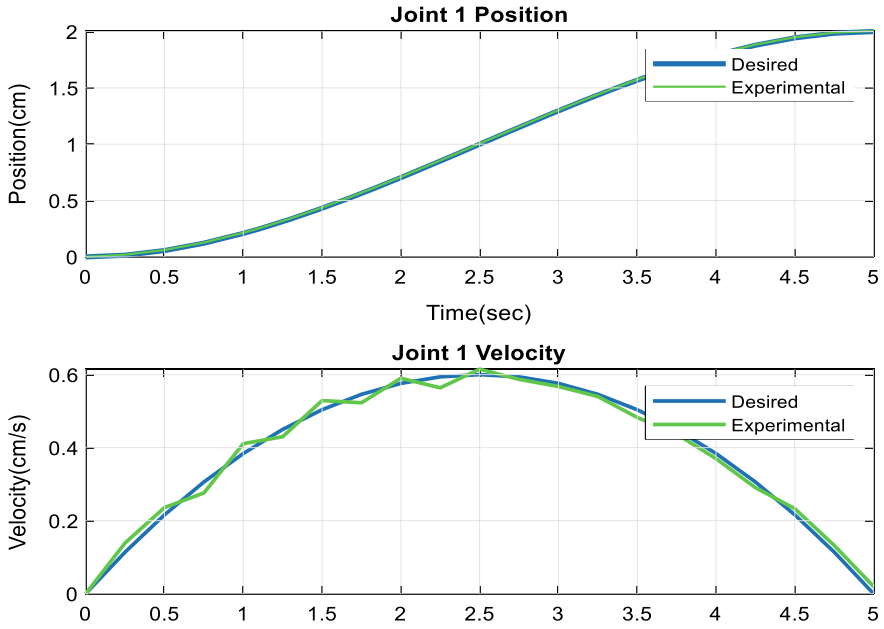
Fig. 2 Joint 1 position and velocity
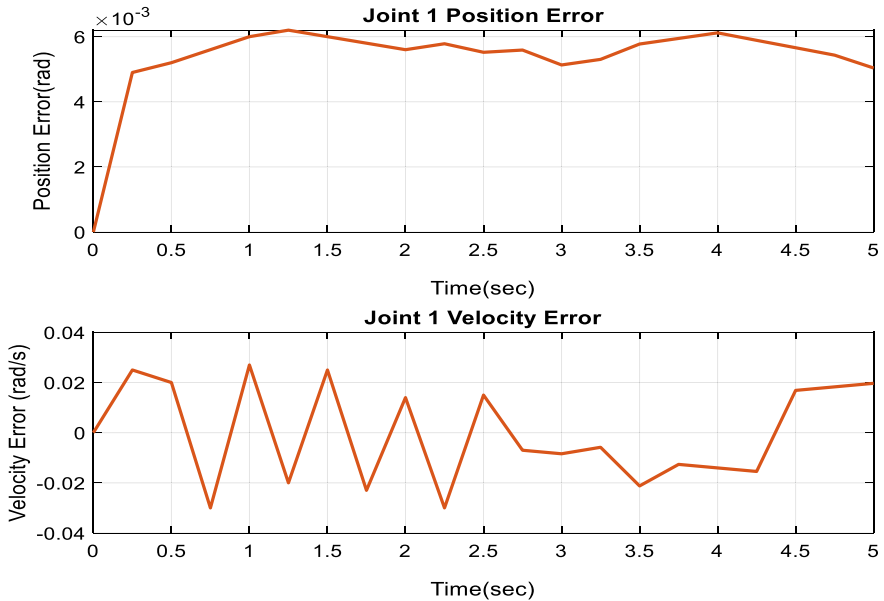
Fig. 3 Joint 1 position and velocity errors
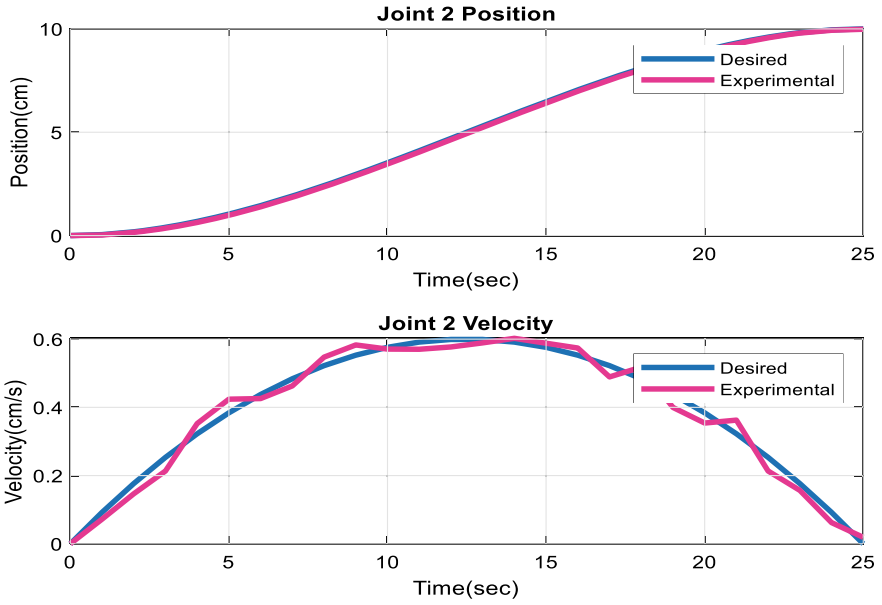
Fig. 4 Joint 2 position and velocity
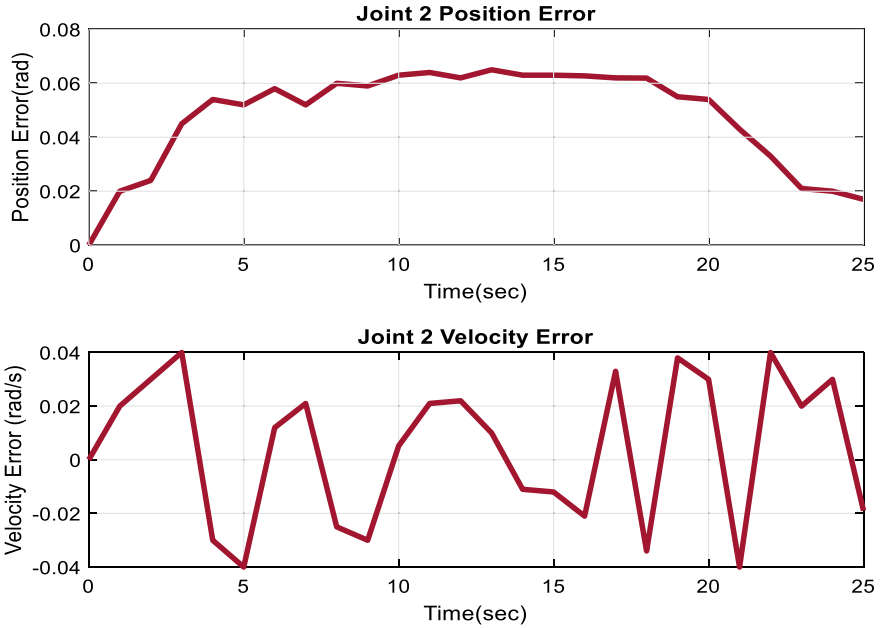
Fig. 5 Joint 2 position and velocity errors
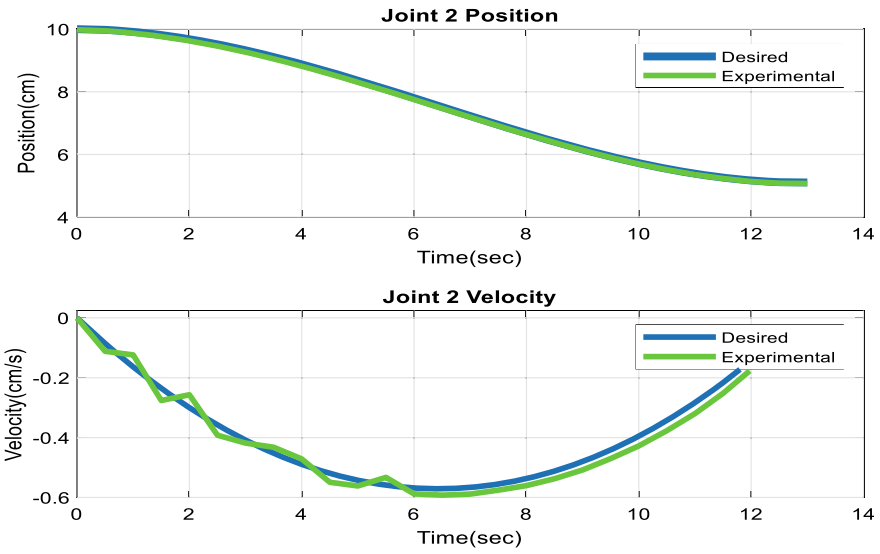
Fig. 6　Joint 2 position and velocity
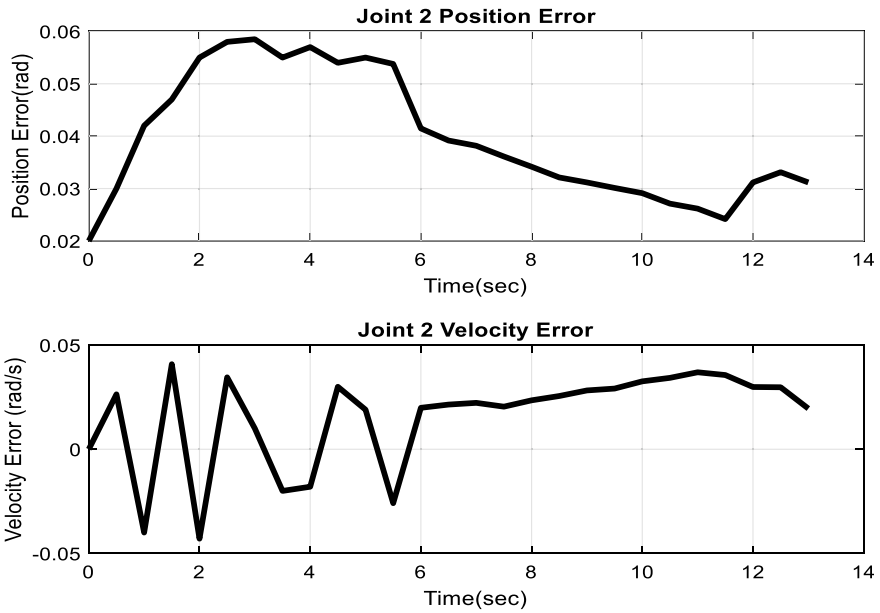
Fig. 7　Joint 2 position and velocity error
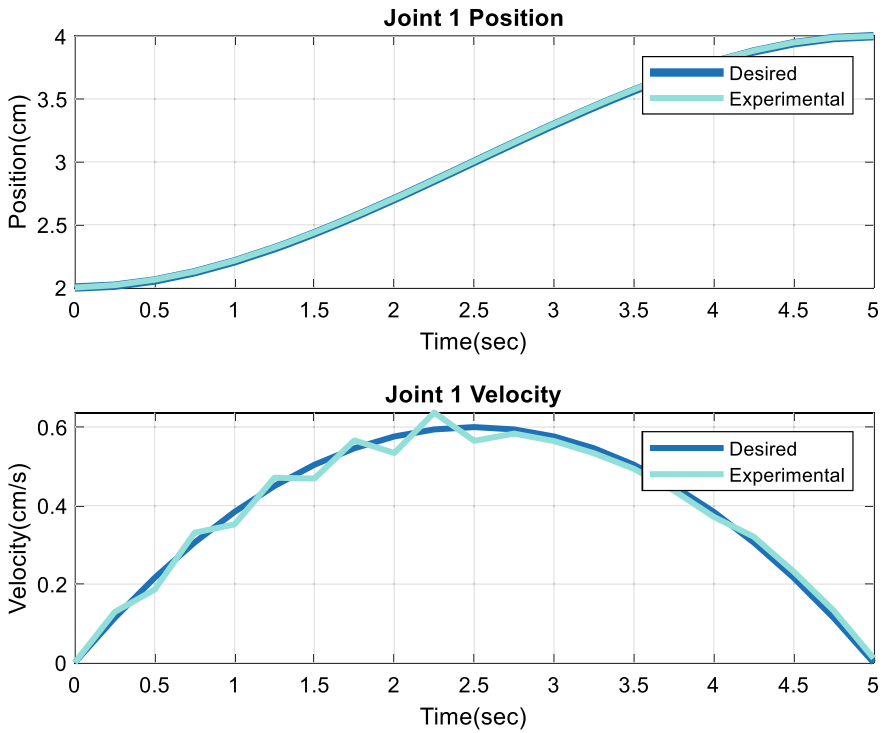
Fig. 8 Joint 1 position and velocity

5 Conclusions

In this paper, an experimental investigation is made to analyze the effectiveness proposed integration of STHWS with VS Control method. An appropriate law for controlling the manipulator has been designed for integration and implemented in two link robot arm. An experiment was carried out by taking into consideration time varying model as singular system time. The objective was to exhibit the ability of proposed integration of STHWS method with VS control system for practical use with diverse arm dynamics. The results are compared with previous studies in the yields good results and the performance of the manipulator good and fall within estimated range of accuracy. The healthiness of proposed integration with VS control system also justified from the results.
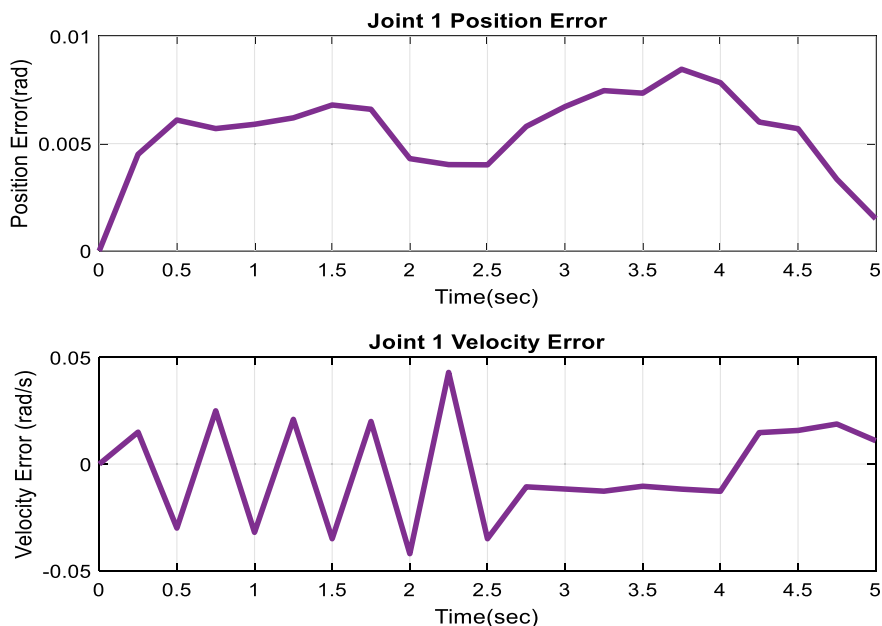
Fig. 9 Joint 1 position and velocity errors

References

1. Sekar S, Prabakaran K (2011) Numerical Solution for linear and nonlinear singular systems using single term haar wavelet series, Int J Commun Comput Inf Sci 140:204–212
2. Sekar S, Prabakaran K, Paramanathen E (2012) Single-term Haar wavelet series technique for time varying linear and non-linear singular systems. In: International conference on pattern recognition, informatics and medical engineering (PRIME-2012), pp 364–369. https://doi.org/10.1109/ICPRIME.2012.6208373
3. Nandhakumar S, Selladurai V, Sekar S (2009) Numerical investigation of an industrial robot arm control problem using haar wavelet series. Am J Eng Appl Sci 2(4):584–589
4. Nandhakumar S, Selladurai V (2011) Haar wavelet approach to second order robot arm control problem. Eur J Sci Res 56(4):573–583
5. Nandhakumar S, Muthukumaran V, Soorya Prakash K, Shunmughanaathan VK (2013) Position control of industrial robotic manipulator using variable structure control system with single term haar wavelet series method. J Vib Control 21(12):2465–2483
6. Ata AA, Johar H (2004) Dynamic simulation of task constrained of a rigid-flexible manipulator. Int J Adv Robot Syst 1:61–67
7. Park BS, Yoo SJ (2018) A low-complexity tracker design for uncertain nonholonomic wheeled mobile robots with time-varying input delay at nonlinear dynamic level. Nonlinear Dyn 89(3):1705–1717
8. Yamawaki T, Yashima M (2007) Effect of gravity on manipulation performance of a robotic arm. In: IEEE international conference on robotics and automation, Rome, pp 4407–4413
9. Liu W, Li F, Wan Y et al (2020) Design of manipulator based on two-dimensional force feedback. Intel Serv Robot. https://doi.org/10.1007/s11370-020-00321-y
10. Castri C, Messina A (2007) Exact modeling for control of flexible manipulators. J Vib Control 18(10):1526–1551

11. Cheng-Yuan C, Hsu K-C, Chiang K-H, Huang G-E (2008) Modified fuzzy variable structure control method to the crane system with control dead zone problem. J Vib Control 14(7):953–969
12. Hannan MA, Hussin I, Ker PJ, Hoque MM, Hossain Lipu MS, Hussain A, Blaabjerg F (2018) Advanced control strategies of VSC Based HVDC transmission system: issues and potential recommendations. IEEE Access 6:78352–78369
13. Pai MC (2015) Sliding mode control of vibration in uncertain time-delay systems. J Vib Control 16(14):2131–2145
14. Quynh NX, Nan WY, Yen VT (2020) Design of a robust adaptive sliding mode control using recurrent fuzzy wavelet functional link neural networks for industrial robot manipulator with dead zone. Intel Serv Robot 13:219–233. https://doi.org/10.1007/s11370-019-00300-y
15. Nandhakumar S, Selladurai V, Muthukumaran V, Sekar S (2010) Haar wavelet approach to time varing model robot arm control problem. Int J Acad Res 2(4):66–74
16. Bujurke NM, Salimath CS, Shiralashetti (2007) Numerical solution of stiff systems from nonlinear dynamics using single-term Haar wavelet series. Int J Nonlinear Dyn 51(4):595–605
17. Hsiao CH (2004) Haar wavelet approach to linear stiff systems. Math Comput Simul 64(5):561–567
18. Hsiao CH (2005) Numerical solution of stiff differential equations via Haar wavelets. Int J Comput Math 82(9):1117–1123

Medical Fitness Device for Driving Assessment

K. Chamundeswari, R. Saranya, M. Smitha, S. Prabakar, and K. Porkumaran

Abstract This project is to check Medical Fitness for accidental Drivers and paralysed patients health condition using pressure and heart beat sensor with RTC unit. This can be helpful to analyze the accident person's health improvement condition or periodic monitoring of that person. In hospitals, accidental patients' body conditions (Leg and Hand) must be monitored constantly, which is typically done by doctors or other paramedical staff. And when the conditions go abnormal then we sense those values by Displaying in LCD and buzzer sound. The suitable parameters which should be monitored are first determined and also the respective sensors are connected to the patient's body and periodically checked by a PIC microcontroller. And these parameters have fixed threshold levels which are monitored and also the parameters are transmitted to a LCD for continues monitoring purpose.

Keywords Accidental drivers · Analyze · Parameters · PIC microcontroller

1 Introduction

An accident is knock or miss event that at times has convenient or more undesirable consequences, other times being inconsequential. The term implies that such an incident might not be preventable since, their leads to circumstances go unrecognized and unaddressed. Accident may lead to injuries which can affect the nervous system. This may also lead to paralysis, trauma, stroke etc. Paralysis may be a loss of muscle function partly of our body which may be temporary or permanent. This

K. Chamundeswari (✉) · R. Saranya · M. Smitha
Dr.N.G.P Institute of Technology, Coimbatore, Tamil Nadu, India
e-mail: chamundeswaribme@ngpit.ac.in

S. Prabakar
Sona College of Technology, Salem, India
e-mail: srisornaprabu@gmail.com

K. Porkumaran
Sri Sairam Engineering College, Chennai, India
e-mail: porkumaran@gmail.com

could flow from to attack within the nerves of the peripheral systema nervosum by the body's system which may eventually result in temporary paralysis of arms and legs [Guillain –Barre syndrome]. The patients with such disabilities don't seem to be ready to perform their everyday actions like feeding, toilette usage, movement through space and driving. After recovery, to assess their driving ability this project are going to be helpful by measuring the parameters like pressure with which they apply the break and therefore the time with which they react and heart beat. Since, till now only the software and scans like MRI, CT, X-RAY are present to assess the anatomical functions and therefore the overall fitness. Occupational therapists, as well as certified driver rehabilitation specialist are required to create important recommendation concerning the fitness to the drive of their clients on daily. To help in formulating these recommendations they're going for scanning technologies. Hence the goal of our project is to assess the fitness without scanning. This project will have high accuracy than the prevailing ones. This device are going to be helpful for simple observation of patients exact health condition and improvement. it's easy to control and handle. It may be of great use thanks to it's compactness and price effectiveness.

2 Related Works

A wireless and reliable health monitoring and warning system was proposed by Khemapech [1] in that the vital signals has been recorded using the wearable device and further analyzed for providing real time warnings. The heart rate and blood pressure and pulse rate are the measured as the vital signals for depicting the patient's cardiovascular symptoms by comparing ranges with the standard values given by the doctors. One of the key drawback they faced is the, range between wearable device and the software of the system is the not in the limited communication aspects. Omer Al-Salihi [2] described about an android mobile application which was successfully built and consequently the data transmission using internet (WiFi) to both the application and to the remote server by was done perfectly,. In addition to the current, the next objective of this was achieved by examining the vital parameters remotely and predicts the consequences. Although the results were far considerable there still exist a problem in both hardware and software interface. Appelboom et al. [3] proposed a project by implementing smart wearable sensors into the practice patient care could improve the relationship between doctor and patient which increased the independence and involvement of patients regarding to the healthcare aspects and remote monitoring provides the most effective way of monitoring which brings a drastic change in the healthcare organization.

Kumara et al. [4] designed a device to assist these patients fingers of the hand that play a major role in this system. The system ensures the patient to precise his needs by converting the finger bending into auditory speech. If the identical action is repeated thrice, then there is an efficiency of the system to send a text message to the concerned caretaker with the assistance of GSM module. In future more flex sensors can be accustomated to perform multiple tasks and more health Parameter may be included. Aminian and Naji [5] proposed a system which implements all

sufficient aspects of monitoring to measure the vital parameter for the multiple users or the patient. In this work a coordinator node is assigned which is attached to the body of the patient which in turn it collects all the vital information's by the wireless sensors and it transmits to the base station for future analysis. The placed sensors on the body of the patient act as a wireless body sensor network (WBSN) and are capable of sensing the vital parameters. In case any abnormalities in the measured parameters the system indicates an alarm to the patient and message also has been passed to the doctor's end by an Email or SMS.

While comparing with the pervious findings there is a large amount of reduction in the energy consumption and also there is a speed in the network lifetime which in turn improves the communication coverage to extent the comfortness and independence for ensuring the quality of life, these are considered as a most advantageous of the proposed system. Deepasri et al. [15] proposed a monitoring system, for indicating the patients health aspects. In this work the most efficient biosensors are used to observe or record the physiological parameters such as blood pressure, pulse rate, heart rate and measured data are continuously sent to the nurse through by using GSM. Future it been processed in Microcontroller (MSP430). Roshdy et al. [14] proposed a design for the paralyzed hand people a easy portable device with light weight and low cost. Based on all these consideration we proposed a device for monitoring the parameters after recovery of the paralyzed patient in the aspects of driving.

3 Methodology

The project uses Pressure and heart beat sensor for transferring data to a microcontroller unit fitted in doctors monitoring area. The pressure sensor value and reaction time take by patient is displayed through LCD screen. Respective sensors are connected to the patient's body whenever medical test is carried out and periodically values are checked by a PIC Microcontroller. All these parameters have fixed threshold levels which are monitored. And the parameters are transmitted to a LCD for continues monitoring purpose. Depending upon the value treatment maybe carryout further.

The project uses Pressure sensor for transferring data to a microcontroller unit fitted in doctors monitoring area. The pressure sensor value and latent period take by patient is displayed through LCD screen. Depending upon the worth treatment maybe carryout further.

4 System Implementation

The proposed system makes use of two sensors namely flexi force or pressure sensor and heart beat sensor to measure the appropriate parameters (Fig. 1).
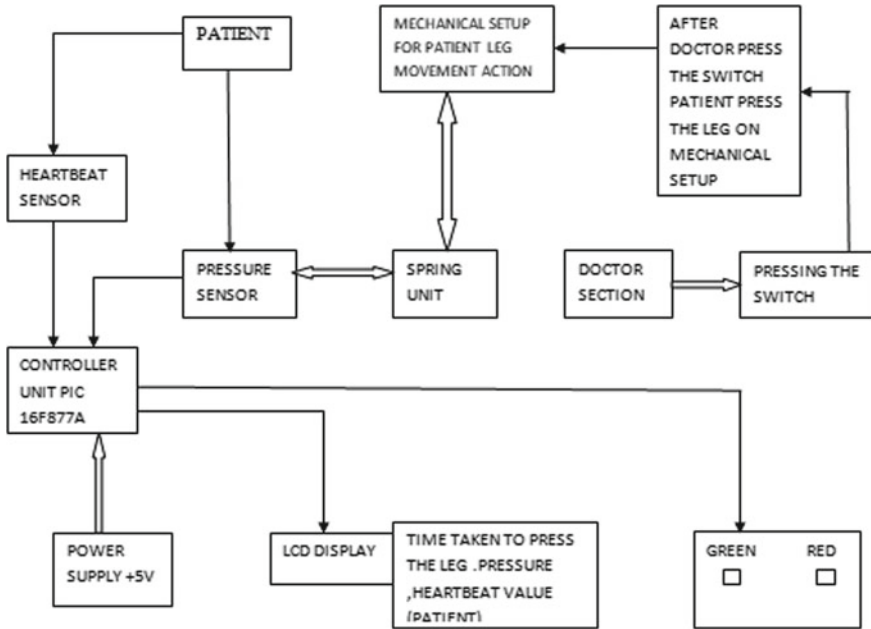
Fig. 1 Block diagram

4.1 PIC Micro Controller

In this system we use the PIC 16F877A microcontroller for the conversion of the parameters. The pin of the this microcontroller is shown in the Fig. 2.

4.2 Pressure Sensor

FSR are used to continuously monitoring the pressure of the patients especially the pressure applied to the peddle of accelerator and brake at the time of driving. The Fig. 3 shows the FSR used in the proposed work.

4.3 Heartbeat Sensor

The heart sensor is placed on the patient's body to continuously measure the heart rate of the patient while driving. This is value is send to the physicians at the other end. This is a digital value.
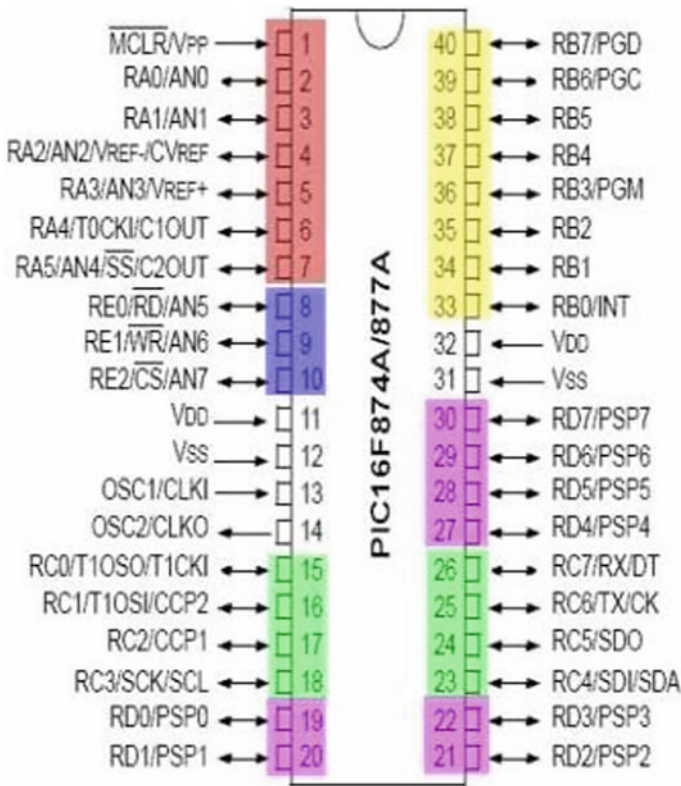
Fig. 2 Pin diagram of 16F877A

Fig. 3 Flexi force sensor

4.4 LCD Display

LCD (Liquid Crystal Display) is used to display the values of the pressure and the heart rate. Here in this work we use a 16 × 2 LCD display is very basic module and is very commonly used in various devices and circuits. The LCD display used in our work is shown in the Fig. 4.
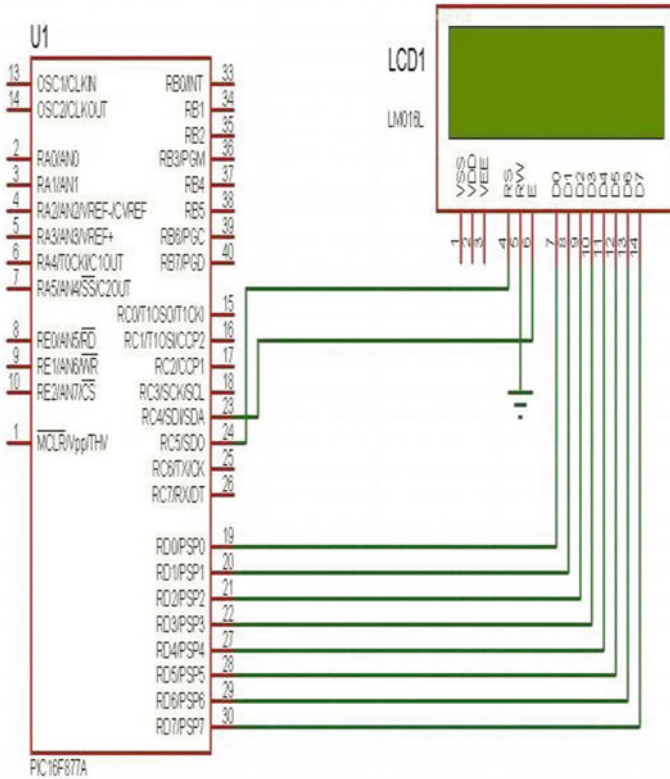
Fig. 4 LCD display

4.5 Power Supply Description

The power supply we use in this work is ac voltage, typically 220V rms, is connected to a transformer, which steps that ac voltage down to the level of the desired dc output.

4.6 Software Required

- Embedded C.
- PIC Boot loader.
- CCS C Complier.

5   Results and Discussion

The working modules of the proposed system are shown in Fig. 5 and Fig. 6 which are very helpful for the paralytic patients and accidental people.

The system continuously monitors heart beat rate and blood pressure of patient.

The system also has the feature to check the physical fitnesss of the patient for driving with a high accuracy. Patients exact health condition can be easily observed. Due to its compactness and low cost this product can be used easily in all healthcare centers, even in home. Also it is easy to operate and handle.

Fig. 5 Working module of the proposed work
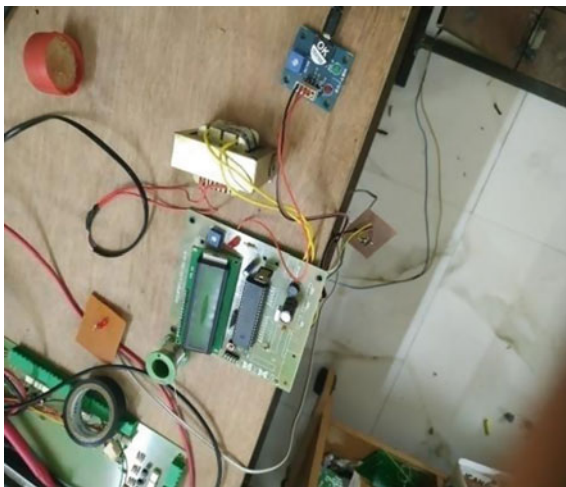
Fig. 6 Working module of the proposed work determining the heart rate and BP
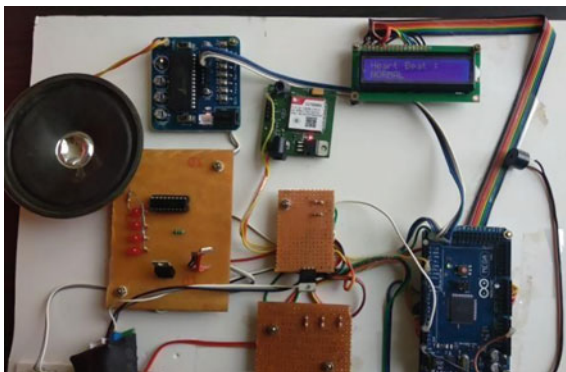
6 Conclusion

This project presents the design and implementation of accidental health monitoring system by using flexi force sensor module. The proposed system will be a helping hand for accidental and paralytic patients. It is concluded that Programmable Interface Controller (PIC) has been the low cost implementation used for recording and transmitting the bio-medical signals and very useful to the patients. This system was developed to minimize the device's size and allow for daily life usage.

References

1. Khemapech I (2015) A real time health monitoring and warning system. TRON Forum
2. Omer RMD, Al-Salihi NK (2017) Healthmate: smart wearable system for health monitoring (SWSHM). In: 2017 IEEE Students conference on electrical, electronics and computer science
3. Appelboom G, Camacho E, Abraham ME, Bruce SS (2014) Smart wearable body sensors for patient self—assessment and monitoring. Archives of Public Health
4. Kumara KR, Kadam A, Rane N, Gouda A, Sensor based wearable system to assist paralytic patient with continuous health monitoring. Int J Future Revolution Comput Sci Commun Eng
5. Aaminian M, Naji HR, A hospital healthcare monitoring system using wireless sensor networks. J Health Med Inf
6. Nelson A, Shyam Kumar P (2013) Wearable multi sensor gesture recognition for paralysis patients. In: 2013 IEEE, University of San Francisco
7. VasanthRao Y, Swamy Bhukya (2019) Gait Assistive exoskeleton device for semi paralyzed stroke survivors. Int J Eng Adv Technol (IJEAT) 8(5c)
8. Agrawal SK, Sangwan V (2007) Assessment of motion of a swing leg and gait rehabilitation with a gravity balancing exoskeleton. IEEE Trans Neural Syst Rehabil Eng 15(3)
9. Yorozu A, Moriguchi T (2011) Improved leg tracking with data association considering gait phase during turning motion in walk tests. In: 2015 IEEE International conference on advanced intelligent mechatronics [AIM] July 7 2011
10. Milovanovic I (2008) Radial basis function (RBF) networks for improved gait analysis. In: 2008 IEEE 9 th Symposium on neural network applications in electrical engineering, NEUREL
11. Sayegh F, Fadhli F, Karam F, BoAbbas M, Mahmeed F, Korbane JA (2016) A wearable rehabilitation device for paralysis. In: IEEE 2016, American University of the Middle East Egaila, Kuwait
12. Hasegawa Y, Mikami Y, Watanabe K, Firouzimehr Z (2008) Warble handling support system for paralyzed patients. In: 2008 IEEE/RSJ International conference on intelligent robots and systems acropolis convention center nice, France, Sept 22–26, 2008
13. Hasan HN (2019) A wearable rehabilitation system to assist partially hand paralysed patients in repetitive exercises. Biomedical Engineering Department, College of Engineering Quality Assurance , University of ThiQar, First International Scientific Conference Al-Ayen University IOP. J Phys Conf Series 1279 (2019) 012040
14. Roshdy A, Al Kork S, Said S (2019) A wearable exoskeleton rehabilitation device for paralysis—a comprehensive study. Adv Sci Technol Eng Syst J 4(1):17–26
15. Deepasri. T Gokulpriya. M Arun kumar. G (2017) Automated paralysis patient health care monitoring system. Dep Electron Commun Eng South Asian J Eng Technol 3(2):85–92

Application of Artificial Intelligence for Reservoir Storage Prediction: A Case Study

Abdus Samad Azad, Pandian M. Vasant, José A. Gámez Vintaned, and Junzo Watada

Abstract There are many relevant and interesting contributions using Artificial Intelligence (AI) based techniques, with different purposes. It has been used as an effective way for estimating the forecasted data of reservoir daily storage value. The efficiency of various AI methods was explored in this article and later the best method is selected for reservoir storage level prediction. In estimating reservoir storage levels several regression algorithms and artificial neural network (ANN) approaches have been evaluated. There is better agreement between the ANN model compared to regression algorithms. The findings were demonstrated by significant correlation coefficient (R^2) rate among the expected and calculated training outcome variables up to 0.91 and the highest validity outcome of Root Mean Square Error (RMSE) was 5.1989. Consequently, this method is therefore adequate for robustness and generalizability abilities and is ideal for forecasts.

Keywords Hydropower · Reservoir · Artificial intelligence technique · Reservoir inflow · Forecast · Artificial neural network · Machine Learning

Nomenclature

ANN	Artificial neural network
AI	Artificial intelligence
PNN	Back propagation neural network
ELM	Extreme learning machine

A. S. Azad (✉) · P. M. Vasant
Department of Fundamental and Applied Sciences, Universiti Teknologi PETRONAS, 31750 Tronoh, Perak, Malaysia

J. A. Gámez Vintaned
Department of Geosciences, Faculty of Sciences and Information Technology, Universiti Teknologi PETRONAS, 31750 Tronoh, Perak, Malaysia

J. Watada
IPS Research Center, Waseda University, Tokyo, Japan

LM   Levenberg marquardt
ML   Machine learning
MSE   Mean square error
R   Correlation coefficient
RMSE   Root mean square error

1 Introduction

A reservoir is a physical (artificial and natural) system used to preserve water for the storage, management and regulation of the flow of water [1]. Reservoir level controlling parameters are inflows, water storage, water release, evaporation, and infiltration from reservoir. These parameters are unpredictable and must be considered in the operation of water supplies. The higher results can be achieved if the dam reservoir level is calculated correctly [2]. Unfortunately, if a reservoir is full on the maximum storage level due to excessive inflow, the water level rises and the discharge rate over the drainage line rise. This event will lead to downstream flooding [3]. Therefore, reservoir water level forecasting is important for flood caution, flood management, river water potential assessment, hydroelectric power generation, domestic management and for river water potential assessment [4]. Also, the significance of water level forecasting would give the operator the advantage of optimum water level drawing and a sustainable hydropower generation strategy [5].

There have been several methods used in the forecasting of hydrological activities in recent years. Previously, the reservoir water level forecasting software used a traditional linear mathematical relationship approach focused on operator knowledge, mathematical curves and instructions [6]. Overestimation of features in data and wide missed variables, however, causes low performance and undermines computational models because of difficulty and shortage of explosion to the data. With regard to these problems, the use of Machine Learning (ML) techniques has been adopted, provided that nonlinear process modelling and forecasting effects are better than conventional models, such as moving average processes [7–9]. Machine Learning (ML) focuses on allowing systems through a provided data to forecast, clustered, absorb classification models, or render choices. The main importance of this modelling is the software's ability to visualize scenarios of data without the above-mentioned understanding from variables influencing the projection variables [10, 11]. Studies have continued to implement this ML to forecast a range of modeling methods and variables to increase efficiency and model's durability prediction description [12]. Regression Algorithms is considered an alternative approach to ANN for implementing nonlinear prediction. It has been increasingly used in hydrological prediction over the years [13]. In graphical or mathematical equations, predictive methods are typically provided using computed historical data and statistical analysis [14]. This research hence emphasizes on predicting improvements to the amount of storage water through ML algorithms.

Key objective in this research to come up with reservoir operant by reliable storage level predicting method and to investigate the simulation methodology of Regression Algorithms likely linear regression (LR) models, decision trees (DTs), Gaussian process regression (GPR) models, support vector machines (SVMs), as well as Artificial Neural Network (ANN) and ensembles of tree models the models were trained. Secondly, the efficacy of the best method will be used in studying the water level parameter pattern. Lastly, the selected method will be used for industrial operations and development and increases the reliability of forecasting for water level storage. It will help to establish better river operating strategies, leading to better conditions for water users, including agricultural operations, manufacturing practices, and hydropower production. As a consequence, the exact prediction for the level of stored water would have a positive effect on improved planning for all relevant industrial and commercial operations. In a summary, the purpose over this analysis is to predict volume of stored reservoir water thus it is an essential consideration for decision-making or the reservoir operator to manage the planning of water supplies.

2 Materials and Methods

2.1 Study Area and Dataset

Mangla Dam is located on the Jhelum River in the Mirpur District in Pakistan. Coordinates are latitude: 33.142083 °N and longitude: 73.645015 °E with Height 147 m (482 ft) and Length 3140 m (10,302 ft). The total capacity 9.12 km^3 (7,390,000 acre ft) with catchment area 33,334 km^2. Turbine numbers are 10 MW × 100 MW, installed capacity 1150 MW (15% overload) 1500 MW (max. planned). Figure 1 illustrate the Dam location.

Data for this study were 10 years Meteorological Historical data points, collected for average monthly temperature, rainfall, wind speed, humidity, sun hours and UV index from January 2009 to October 2020. Hydrological average monthly data of inflow, irrigation demand and reservoir storage for 2014–2020 was collected. The average monthly data was used for the study purpose (Fig. 2a, b).

2.2 Methodology

In this study, the dataset was split into training and test dataset, to avoid over-fitting of the method. Using the dataset of the training section for various statistical models and Neural Network Regression (NNR).

Fig. 1 Location of Mangla Dam

Regression Learner App (RLA)

The RLA is a feasible framework available with MATLAB online and latest versions. RLA models were developed for all the training, using influential hydrological (inflow, irrigation demand) and meteorological variables (temperature, rainfall, wind speed, humidity, sun hours, uv index), variables as independent variables, and the reservoir storage water level of these sites as dependent variables.

Artificial Neural Network (ANN)

Three-layer ANN, consisting of one input layer, one hidden layer, and one output layer, is widely preferred in practical engineering implementations because it can approximate virtually any kind of complex relationships to excellent accuracy between the inputs and target outputs. The sketch map of a standard three-layer ANN is shown in Fig. 3.

Evaluation Criteria

The performance of the Regression Models and the ANN was evaluated using the coefficient of determination (R^2), mean–squared error (MSE) mean absolute error (MAE) and Root Mean Square Error (RMSE). The value of R^2 closer to 1 shows good correlation between the data sets. Following equations are used to measure the errors:
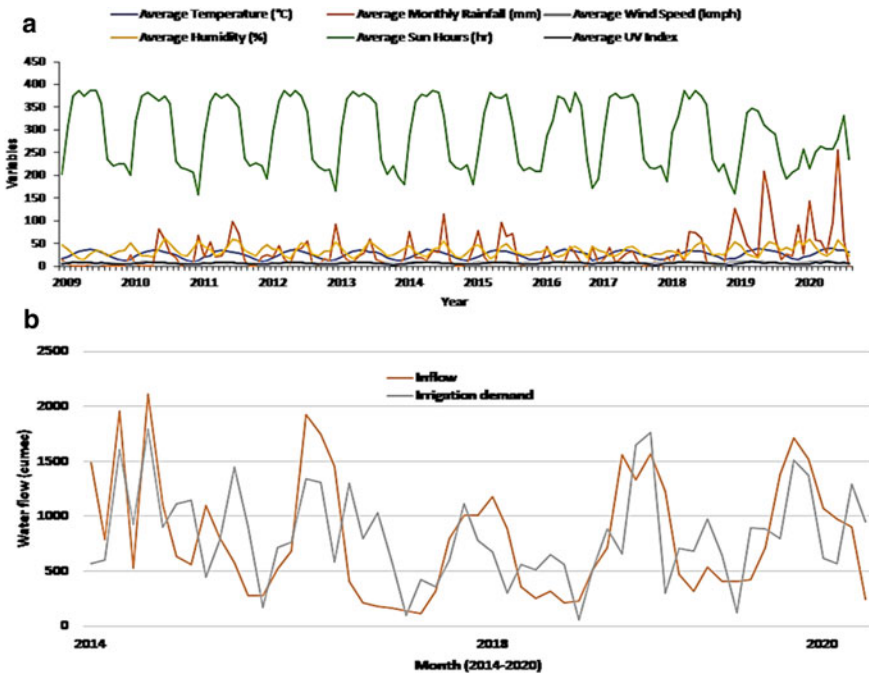
Fig. 2 **a** Meteorological data. **b** Hydrological data
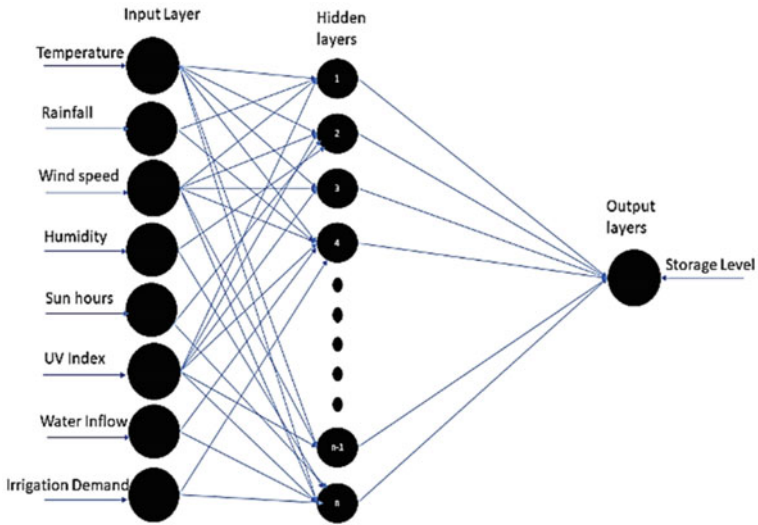
Fig. 3 Structure of the ANN

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_{pred,i} - y_{exp,i}\right)^2}{\sum_{i=1}^{n}\left(y_{pred,i} - y_m\right)^2} \tag{1}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_{pred,i} - y_{exp,i}\right)^2 \tag{2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_{pred,i} - y_{exp,i}\right| \tag{3}$$

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{\left(y_{pred,i} - y_{exp,i}\right)^2}{n}} \tag{4}$$

where, respectively, ypred and yexp are the predicted and experimental values. In the data, N denotes the total number of values. The mean answer value is denoted by ym.

3 Result and Discussions

3.1 RLA and ANN Modeling

The RLA enables system to instantly train several standard regression techniques simultaneously time analyze their findings and thus choose the most appropriate one. The RMSE, quadratic regression (R-square) parameter, mean absolute error (MAE), mean squared error (MSE), predicted speed, and time of training of Regression models and Artificial Neural Network are shown in Table 1. The table shows that ANN had better accuracy in terms of errors evaluation than regression linear algorithms. Therefore, ANN is selected for further study purpose.

The efficacy of the ANN model at various training protocols is listed in Table 2 here. The most efficient learning strategies for ANN training was back propagation (BP) algorithm. This algorithm can be quickly applied and incorporated in functional forecasting systems as given in early research works. It is possible to split back propagation into two phases: A process of feed-forward and a backward step. In current study feed forward BP algorithm had been used based on its minimum value of MSE. The input information was delivered to the input layer, the hidden layer and the output layer in order to the feed-forward stage, to receive the output information. ANN was trained and tested further to reach the near-optimal relationship between the input data set and the output data set.

It must be remembered, though, that the participation of several neurons in the hidden layer may induce the over-adaptation, resulting in the lack of network generalizability, and that a limited number of hidden neurons may also fall short of the

Table 1 Performance values of regression algorithms

Prediction model		Results					
		RMSE	R^2	MSE	MAE	Prediction Speed (obs/s)	Training Time (s)
Linear Regression Models	Linear	7.34	0.8	53.89	5.71	~800	1.835
	Interaction Linear	169.39	9.72	28,691	65.876	~1100	0.4768
	Robust Linear	7.3077	0.82	53.402	5.6857	~2100	0.57047
	Stepwise linear	7.7387	0.80	59.887	6.2356	~2200	6.8793
Regression Trees	Fine Tree	15.216	0.21	231.53	11.273	~3400	1.137
	Medium Tree	14.753	0.26	217.65	11.439	~6200	0.1182
	Coarse Tree	17.135	-0.00	293.62	14.898	~6000	0.1274
Ensembles of Trees	Boosted Trees	18.269	0.14	333.75	16.436	~1100	1.5164
	Bagged Trees	12.015	0.51	144.35	10.185	~1400	1.0856
Support Vector Machines	Linear SVM	8.0991	0.78	65.595	6.6522	~2700	0.61451
	Quadratic SVM	8.6011	0.75	73.98	7.019	~5500	0.17838
	Cubic SVM	12.47	0.47	155.5	8.74	~4900	0.13694
	Fine Gaussian SVM	15.918	0.14	253.38	13.677	~5100	0.17306
	Medium Gaussian SVM	8.0131	0.78	64.209	6.4718	~5100	0.13812
	Coarse Gaussian SVM	11.814	0.52	139.58	9.5742	~5300	0.13056
Gaussian Process Regression Models	Squared Exponential	8.5007	0.75	72.262	6.8081	~3000	1.1217
	Matern 5/2 GPR	7.8666	0.79	61.883	6.3482	~4100	0.25802
	Exponential GPR	7.5524	0.81	57.038	6.085	~4100	0.2701
	Rational Quadratic GPR	7.7866	0.79	60.632	8.0132	~4100	0.33501
ANN		5.1989	0.91	27.0289	9.1293	~4900	5.6785

Table 2 Trial-and-error results for development of reservoir efficiency ANN model

Algorithm	Function	Transfer function in hidden layer	Transfer function in output layer	Correlation coefficient
Conjugate gradient backpropagation with Polak–Ribiere updates	traincgp	Tansig satlin purelin	Purelin	0.7217
Levenberg Marquardt backpropagation	trainlm	Tansig satlin purelin	Purelin	0.912
Gradient descent with momentum and cascade-forward backpropagation	traingdx	Tansig satlin poslin	Purelin	0.164
elman backpropagation	trainrp	Tansig satlin poslin	Purelin	0.427
Feed-forward backpropagation	trainbfg	Tansig satlin poslin	Purelin	0.419
Layer recurrent	trainscg	Tansig satlin poslin	Purelin	0.539
Bayesian Regularization	trainbfg	Tansig satlin purelin	Purelin	0.812

data and that the network will not then remember. Therefore, if the structure fails to meet the experimental outcomes, the water treatment plant production value becomes disorganized and the plant management challenging. The trial and error approach were used to identify the most effective network model, the ideal number of cached layers and the maximum possible number of neurons in the cached layer. There were two maximum number of hidden layers. The addition of more hidden layers increases the system's complexity and thus simulator time and expense. Authors have shown that a framework with a hidden layer will estimate any continuous function if adequate freedom is given.

The required design generated both training and testing data with the minimum of errors. Table 3 tests the output of each network model by evaluating the MSE for each test performed to find the required architecture. Backpropagation network and 14–20 neurons have been shown to achieve minimum MSE levels, so that all models are limited to 20 neurons and evaluated.

Simulated data was analyzed for the testing and validating of model networks with only one low, one medium, one in high-value storage datasets. The findings for the predictive process modelling in Table 3 show the number of neurons in hidden layer. Several experiments for each group have been performed to achieve the required

Table 3 Number of neurons in hidden layer for modeling of prediction

Number of neurons	MSE train	MSE test	R^2 train	R^2 test
1	1132.40885e−0	1131.82874e−0	3.86626e−1	3.39020e−1
2	1130.51469e−0	1110.32990e−0	3.22761e−1	4.94977e−1
3	1221.4858e−0	825.41410e−0	2.89077e−1	3.77285e−1
4	719.83942e−0	871.38243e−0	6.81527e−1	6.18165e−1
5	585.61699e−0	714.74720e−0	7.27652e−1	6.59739e−1
6	663.92940e−0	858.15404e−0	7.00667e−1	5.64750e−1
7	608.52763e−0	910.74590e−0	7.17946e−1	6.12950e−1
8	625.97087e−0	845.29416e−0	7.22004e−1	6.03408e−1
9	512.81392e−0	439.05576e−0	7.71174e−1	7.44953e−1
10	568.93684e−0	410.62168e−0	7.36236e−1	8.17789e−1
11	589.04647e−0	617.56938e−0	7.30761e−1	7.43482e−1
12	581.47511e−0	691.45059e−0	7.55153e−1	6.71959e−1
13	447.71431e−0	823.96620e−0	8.05743e−1	5.85416e−1
14	403.09447e−0	651.54030e−0	8.33410e−1	7.23248e−1
15	636.77776e−0	561.76090e−0	7.23871e−1	7.49778e−1
16	635.03722e−0	576.82133e−0	6.97208e−1	7.53079e−1
17	541.03630e−0	976.30986e−0	7.49574e−1	6.13714e−1
18	488.40561e−0	784.90172e−0	7.85160e−1	6.83428e−1
19	409.98544e−0	703.54441e−0	8.29223e−1	6.46347e−1
20	187.03860e−1	740.59603e−1	9.22101e−1	6.57288e−1

network layout, before the appropriate learning rate, the number of hidden layers and the number of neurons per hidden layer have been achieved.

If a model ANNs can precisely define the highest standards in the Correlation coefficient (R2) and minimum standards in the Root Mean Square of Error (RMSE) it can simply be the accurate and superior ANN model. The number is noticeable in Fig. 4 that the result directions for training (R-value = 0.9221), validation (R value = 0.82371) and testing (R value = 0.65729) are aimed very strongly. These values may equate to a total R value of = 0.86872. In this scenario, the network reaction has so far been optimistic, and simulation can be used to reach new inputs.

The Regression (R) plot highlights the assessment between the target that is believed to be the optimal output and the ANN output that is known to be the actual output, which indicates that the ANN output is exactly the same as the target output. Notably, the Regression (R) value indicates the significance of outputs and priorities. In fact, since R can exceed 1, this means an accurate linear relevance between goals and outputs. The validation and the performance checked display values of (R) are both greater than 0.90. It shows high performance of the model.
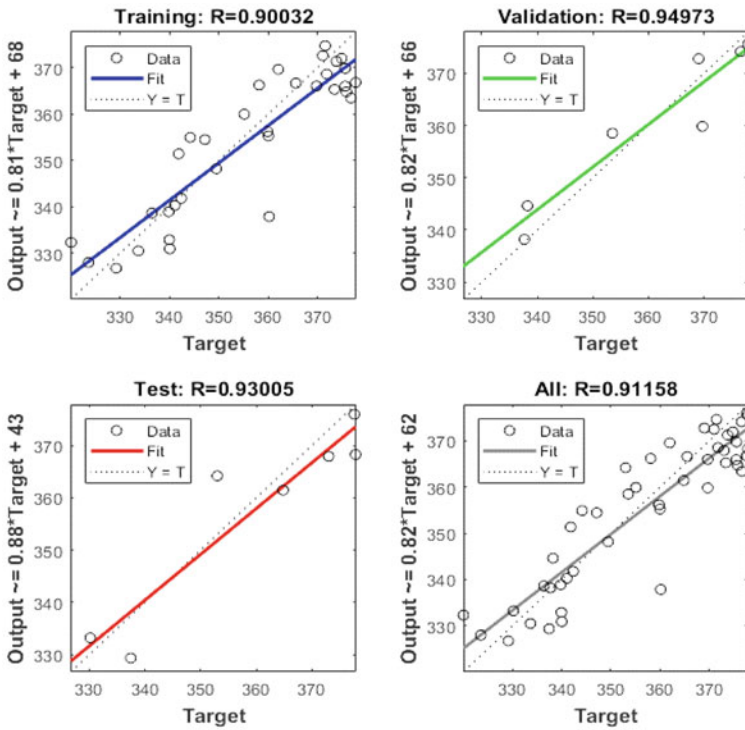
Fig. 4 Regression plot by ANN

3.2 Prediction of Storage Data

The trained model was used for the prediction of storage data for water level of January–November 2020. Figure 5 shows that the data was well predicted using the trained ANN model. The prediction accuracy was 92% and RMSE value 5.1989. The lower value of error shows that ANN is a reliable ML tool for prediction of reservoir storage data provided with suitable meteorological and hydrological data.

4 Conclusion

The networks used to train and forecast water storage levels using the Matlab software in this study. The learning algorithm, such as Levenberg–Marquardt, was qualified in training to train the model. The results were shown by a high value of the correlation coefficient (R^2) between the expected and observed training variables of up to 0.91 and the highest validity efficiency value of the RMSE was 5.1989. Consequently, it
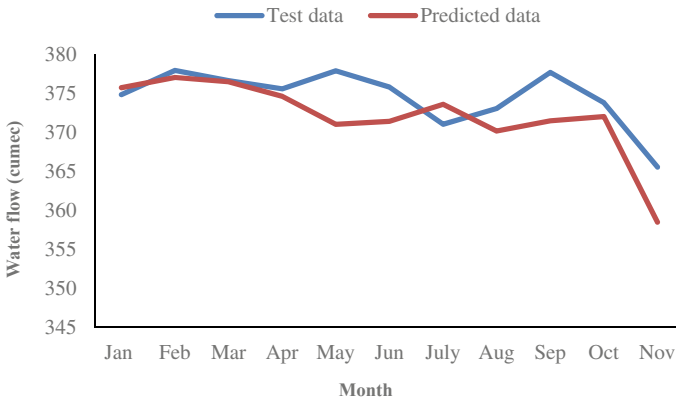
Fig. 5 Prediction by ANN

is concluded that for accuracy and generalization capability, this developed model has been satisfactory and is suitable for prediction purpose.

References

1. Hussain W, Ishak W, Ku-mahamud KR, Norwawi N (2014) Neural network application in reservoir water level forecasting and release decision, pp 54–63 (2014)
2. Üneş F, Demirci M, Taşar B, Kaya YZ, Varçin H (2019) Estimating dam reservoir level fluctuations using data-driven techniques. Pol J Environ Stud 28:3451–3462
3. Ashaary NA, Hussain W, Ishak W, Ku-mahamud KR (2015) Forecasting the change of reservoir water level stage using neural network, pp 103–7 (2015)
4. Dogan E, Isik S, Toluk T, Sandalci M (2007) Daily streamflow forecasting using artificial neural networks. In: International congress on river basin management, pp 449–59
5. Choi C, Kim J, Han H, Han D, Kim HS (2020) Development of water level prediction models using machine learning in wetlands: a case study of Upo wetland in South Korea. Water (Switzerland) 12
6. Valizadeh N, El-Shafie A, Mirzaei M, Galavi H, Mukhlisin M, Jaafar O (2014) Accuracy enhancement for forecasting water levels of reservoirs and river streams using a multiple-input-pattern fuzzification approach. Sci World J
7. Grenney WJ (1995) Neural networks for river flow prediction. J Comput Civ Eng 9:293–293
8. Saadon A, Abdullah J, Muhammad NS, Ariffin J (2020) Development of riverbank erosion rate predictor for natural channels using NARX-QR Factorization model: a case study of Sg. Bernam, Selangor, Malaysia. Neural Comput Appl 32:14839–49
9. Yang T, Asanjan AA, Welles E, Gao X, Sorooshian S, Liu X (2017) Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. Water Resour Res 53:2786–2812

10. Najah A, El-Shafie A, Karim OA, Jaafar O (2011) Integrated versus isolated scenario for prediction dissolved oxygen at progression of water quality monitoring stations. Hydrol Earth Syst Sci 15:2693–2708
11. Hipni A, El-shafie A, Najah A, Karim OA, Hussain A, Mukhlisin M (2013) Daily forecasting of dam water levels: comparing a support vector machine (SVM) model with adaptive neuro fuzzy inference system (ANFIS). Water Resour Manag 27:3803–3823
12. Zantalis F, Koulouras G, Karabetsos S, Kandris D (2019) A review of machine learning and IoT in smart transportation. Future Internet 11:1–23
13. Wang Q, Wang S (2020) Machine learning-based water level prediction in Lake erie. Water (Switzerland) 12:1–14
14. Zhao T, Minsker B, Salas F, Maidment D, Diev V, Spoelstra J et al (2018) Statistical and hybrid methods implemented in a web application for predicting reservoir inflows during flood events. J Am Water Resour Assoc 54:69–89

Fabrication of Parallel Mechanism Actuated by Pneumatic Artificial Muscle for Rehabilitation Therapy

S. Krishnan, A. M. A. Rani, Laxamanan G. Kurappa, and Sivajothi Paramasivam

Abstract Ankle–Foot injuries are amongst the most common injuries of the lower limb and almost 25,000 people experience ankle injuries each day due to vigorous activities. Traditionally, ankle injuries are rehabilitated via physiotherapy using simple equipment like elastic bands and rollers, requiring intensive efforts of therapists and patients. Currently, Stewart platform rehabilitation devices are actuated by various methods including double acting pneumatic cylinder, hydraulic, electric motor and shape memory alloy. The limitation of using these actuation methods is that it provides lower range of motions and requires higher maintenances. A less known type is Pneumatic Artificial Muscles (PAMs). PAMs are operated by pneumatic pressure and are contractible naturally upon inflation. The objective of this study is to fabricate Stewart platform using Pneumatic Artificial Muscle, to analyze the range of motion produce by various diameter of the PAMs. The diameters of PAM used were 8, 10 and 12 mm. It was observed that Stewart Platform actuated by PAM of diameter of 12 mm produced maximum platform angle of 31.73°, whereas PAM with the diameter of 10 and 8 mm produced maximum platform angle of 28.62° and 25.31°, respectively.

Keywords Pneumatic Artificial Muscle · Stewart Platform · Range of motion (ROM)

S. Krishnan (✉)
Manipal International University (MIU), Nilai, Malaysia
e-mail: krishnan.subramaniam@miu.edu.my

A. M. A. Rani
Universiti Teknologi PETRONAS (UTP), Tronoh, Malaysia

L. G. Kurappa
University Nilai, Nilai, Malaysia

S. Paramasivam
UOW Malaysia KDU University College, Glenmarie Campus, Shah Alam, Malaysia

1   Introduction

1.1   Background of Foot and Ankle

Ankle joint is one of the significant joints in human body which helps maintain body balance during ambulation. In a human skeleton, human ankle joint is a very complex bony structure. Ankle sprain is a common injury and the number of injuries is rising day by day due to vigorous activities carried out by people and this can happen to athletes and non-athletes, children and adults. Beside vigorous activities, it can also happen when you simply step on an uneven surface, or step down at an angle. Ligament play important role in human ankle where it holds the ankle bones and joint in position. For all the abnormal movements-especially twisting, turning, and rolling of the foot, the ligament provides protection to the ankle joint. Since ligament is an elastic structure it has the ability go back to their normal positions after it stretches within its limits. Sprain occurs when ligament is forced to stretch beyond its limit. Actual tearing of the elastic fibers caused by the severe sprain. Apart from ankle sprain, stroke is one of the most typical impairment which causes the patient difficult to lift their foot due to weakness in the dorsiflexion muscles. Figure 1 shows anatomy of human ankle.

Physical therapies method which rehabilitates patient one-to-one is known as traditional rehabilitation. Obtaining desired results from this traditional rehabilitation method solely depends on the therapist skills. This traditional method affects the therapists in a negative way since the rehabilitation process contains repetitive and intensive exercises and it takes longer period of time to cure. Hence, there will be difficulties in delivering the same rehabilitation process by the therapist. Therefore, from the viewpoint of patients, therapists and rehabilitation process by studies, the
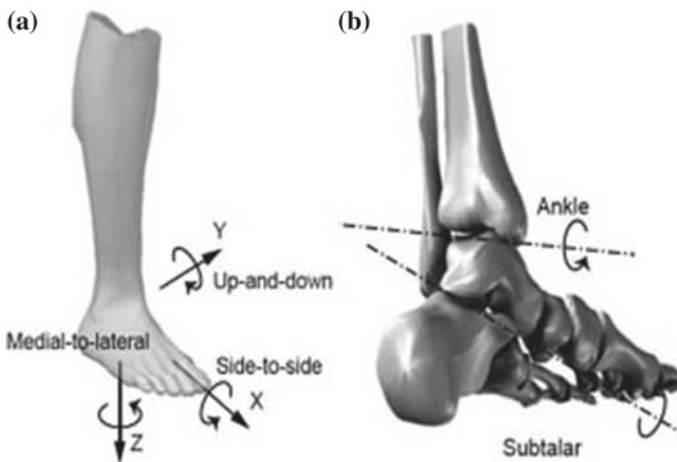
Fig. 1   a Rotation and **b** joints of the ankle [1]
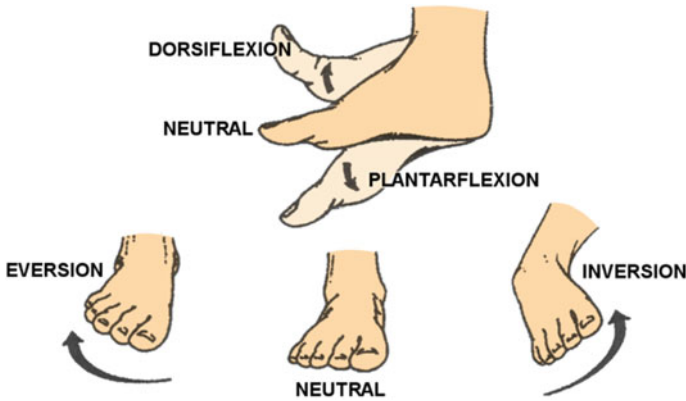
Fig. 2 Movement of human ankle [2]

importance and requirements of robotic-based rehabilitation has been emphasized. One of the advantages of using ankle rehabilitation robots is that it provides a precise measurement of data to the therapists to records and observes every details of the rehabilitation process and makes changes of the diagnosis, prognosis, and customization of conceivable treatment. Patient's recovery can be accelerated by delivering more accurate, repetitive and intensive therapies through this rehabilitation process using ankle rehabilitation robots. Flexibility of the ankle joints can be increased by using these rehabilitation robots and allow patient's ankle undergo all four movement: dorsiflexion, plantar-flexion, inversion and eversion. Figure 2 depicts movement of human ankle.

2　Literature Review

2.1　Pneumatic Artificial Muscles

The braided pneumatics actuators were invented by McKibben to help the movement of polio patients 1950. They are also called a McKibben Muscle or Pneumatic Muscle which is actuated by compressed air. Pneumatic muscle was used as the original applications such as, door opening arrangement and on industrial robot [3]. The displacement movement performed by pneumatic muscle is shown in Fig. 3.

The main element of PAM is thin membrane which makes it lightweight and it is an advantage during replacement of a defective muscle. There are different types of braided sleeves discovered in the past literature such as latex and silk pneumatic Muscle [5–7]. Another advantage of PAM in their inherent complaint behavior is when force is exerted on the PAM, which gives in without increasing the force in the actuation.

Fig. 3 The displacement movement carried out by pneumatic muscle [4]
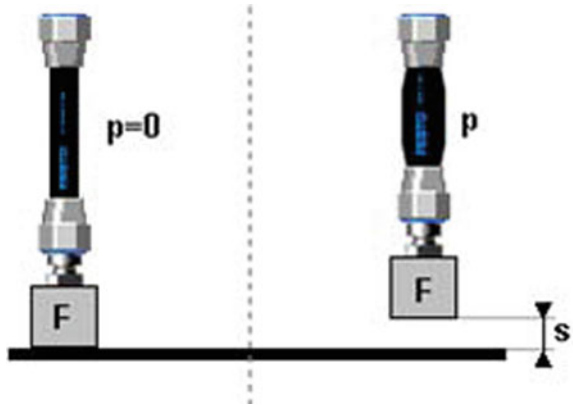
2.2 Application in Robotics

Wickramatunge and Leephakreeda [8] reported that the low assemble weight and high power to ratio of PAMs is most desirable to be considered for use in mobile robotics. In certain situation, the impotency of the PAMs is higher than the conventional pneumatic cylinder. The development of a model gives effective description to understand the mechanical behavior of the PAMs. The stiffness parameter of the muscle denotes as K and it is considered a function of the actuated air pressure, P and total length, L_S. The elastic force is denoted by $F_{elastic}$ and the expression given in Eq. 1.

$$F_{elastic} = K(P, L_s)L_s \tag{1}$$

In this study, the stiffness parameter K is taken as a second order polynominal function of L_S and P and it is expressed in Eq. 2. However, a universal approximation method with fuzzy logic or neural network can be used as alternative choice for accuracy viewpoint [9].

$$K = a_o P^2 + a_1 P L_s + a_2 L_s^2 + a_3 \tag{2}$$

where a_o, a_1, a_2 are constant cerfficient, which can obtained from experiment data. Typical values of coefficient in K for operating on construction mode under low pressure:

$$a_0 = -2.977, a_1 = -5.890, a_2 = 3.203 \quad \text{and} \quad a_3 = 104.824$$
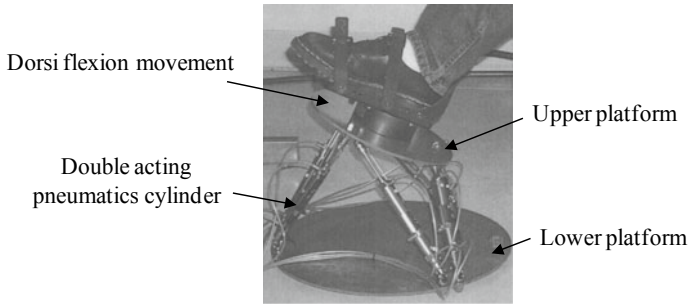
Fig. 4 The "Rutgers Angle": the system view [10]

2.3 Parallel Mechanism

Recently there has been great interest in some research center around the world to propose and develop automated system for ankle rehabilitation. Much of the work focuses on the limitation of existing commercial rehabilitation devices, i.e. obtaining combinations of basic ankle movement. Figure 4 shows the Rutgers ankle platform. Rutgers ankle platform is a Stewart platform that generates displacement to the ankle–foot of the patients during rehabilitation process by actuating six pneumatic cylinders. The Stewart platform was designed to control the displacement of the cylinder in six degrees of freedom movement of the foot–ankle.

The electro pneumatic controller is used to actuate the Stewart platform. The electro pneumatic controller modulates the air pressure in the platform. The Rutgers actuation is based on pneumatic system, and this has disadvantages in terms of portability due to the heaviness of the device and sounds that is unsuitable for medical rehabilitation environment.

3 Methodology

The schematic of Stewart platform is shown in Fig. 5 consisting of upper platform, lower platform and Delrin panel. Basically, one end of the Pneumatic Artificial Muscle is connected to upper platform and the other end is connected to pneumatic fitting. When the Pneumatic Artificial Muscle is pressurized, the muscle will expand and shorten leading to the displacement of the upper platform. The upper platform will be deformed to the downward position which creates dorsi/plantar flexion movement. The gyroscope is fixed in between the upper platform to measure the top angle and force sensitive resistor (FSR-402) is fixed on top of the upper platform to measure the force applied on the platform. Figure 5 shows the position of upper platform before and after it is pressurized in 3D modeling.

(a) 0°   **(b) 31.73°**
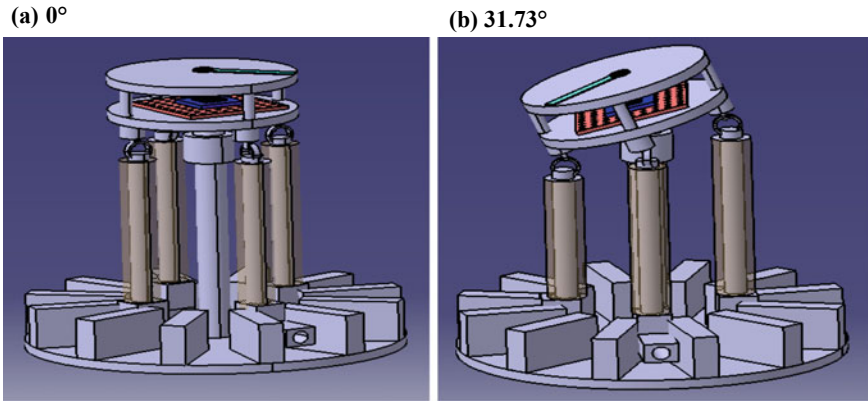
Fig. 5 Stewart platform actuated by PAM **a** 0° and **b** 31.73°

The Pneumatic Artificial Muscle used in this experiment is enclosed with nylon braided sleeves. The PAM is pressurized from 1 bar up to 5 bars and diameter of PAM used is 12 mm. The gyroscope is fixed in between the upper platform to measure the top angle and force sensitive resistor (FSR-402) is fixed on top of the upper platform to measure the force applied on the platform. The developed Stewart platform is depicted in Fig. 6.
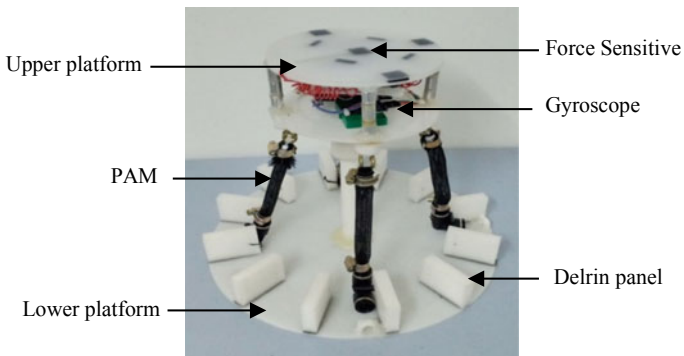
Fig. 6 Developed Stewart platform actuated by PAMs

Fig. 7 Force produced by various diameters of PAM
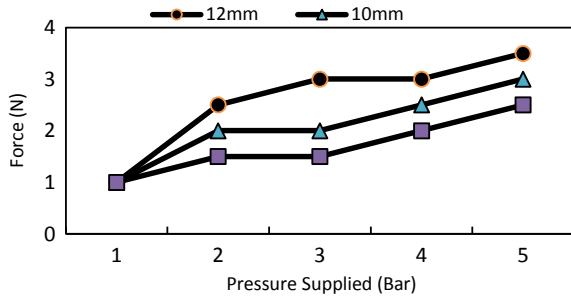
4 Results and Discussion

4.1 Force Produced by PAMs

Preliminary experiment was conducted with different diameters of 8, 10 and 12 mm to identify the force produced by the pneumatic artificial Muscle. Figure 7 shows the comparison force produced by different diameters of Pneumatic Artificial Muscle. It is observed that the PAM with the diameter of 12 mm produced maximum pulling force of 3.5 N when PAM is pressurized maximum to 5 bars. The PAM with the diameter of 10 mm has a pulling force of 3.0 N and 8 mm have a pulling force of 2.5 N when PAM is pressurized to 5 bars. According to PAM with the diameter of 12 mm the difference between maximum force and minimum force is 2.5 N followed by PAM with the diameter of 10 mm and 8 mm which is 2 N and 1.5 N respectively.

4.2 Range of Motion Produced by the PAMs

Figure 8 presents the comparative study of Range of Motion produced by Pneumatic Artificial Muscle with different diameters of 8, 10 and 12 mm. It was observed that Pneumatic Artificial Muscle with the diameter of 8 mm produced minimum range

Fig. 8 Range of motion produced by various diameter PAMs
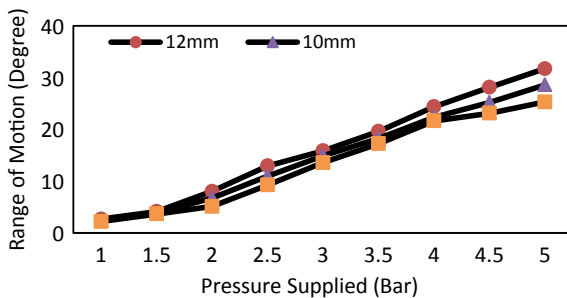
of motion of 2.21° and the maximum range of motion of 25.31°. The difference between minimum and maximum range of motion of the platform is 23.1°. PAM with the diameter of 10 mm obtained minimum range of motion of 2.43° and maximum range of motion of 25.31°. The difference between maximum and minimum range of motion is 22.88°. There is a slight increase in range of motion when the diameter increases. The maximum and minimum range of motion obtained for 12 mm PAM is 31.73° and it was 24.8° ROM actuated by SMA [11, 12].

5 Conclusion

In this study, it is established that the use of Stewart platform actuated by Pneumatic Artificial Muscle improves the range of motion (ROM) for patient's ankle–foot rehabilitation in biomedical applications. As the PAM is pressurized, the muscle gets expand and shorten in length pulling the upper platform downwards. Hence, dorsiflexion and plantar flexion movement is achieved. Pneumatic Artificial Muscle with the diameter of 12 mm was fabricated in the Stewart platform. The Stewart platform actuated by Ø 12 mm diameter of PAM produced maximum platform angle and force of 31.73° and 3.5 N, respectively. The developed Stewart platform actuated by Pneumatic Artificial Muscle produced range of motion up to 31.73° which is acceptable to promote dorsi/plantar flexion.

References

1. Krishnan S, Nagarajan T, Rani A, Ambaraj W, Ramiah R (2016) Rehabilitation for foot/ankle-continuous passive motion (CPM) using shape memory alloy (SMA) actuated stewart platform. ARPN J Eng Appl Sci 11:22
2. Brockett CL, Chapman GJ (2016) Biomechanics of the ankle. Orthopaedics Trauma 30:232–238
3. Krishna S, Nagarajan T, Rani A (2011) Review of current development of pneumatic artificial muscle. J Appl Sci 11:1749–1755
4. Deaconescu T (2007) Studies regarding the performance of pneumatic Muscles. In: Proceedings of international conference on economic engineering and manufacturing systems, pp 271–275
5. Krishnan S, Nagarajan T, Rani AMA, Rao T (2012) Silk pneumatic artificial muscle (SPAM) construction for bio-medical engineering application. In: 2012 IEEE business engineering and industrial applications colloquium (BEIAC), pp 302–306
6. Nagarajan T, Krishnan S, Amirtham V, Abdul-Raniand AM, Rao T (2013) Experimental investigation-natural fiber braided sleeve for pneumatic artificial muscles actuation. Asian J Sci Res 6:596
7. Laksanacharoen S (2004) Artificial muscle construction using natural rubber latex in Thailand. In: The 3rd Thailand and material science and technology conference, pp 10–11
8. Wickramatunge K, Leephakpreeda T (2009) Empirical modeling of pneumatic artificial muscle. In: International multi conference of engineers and computer scientists, pp 1726–1730
9. Vanderborght B, Van Damme M, Van Ham R, Verrelst B, Lefeber D (2006) Pleaded pneumatic artificial muscles for robotic application. In: Proceedings of the IEEE international conference on robotics and automation, pp 4324–4328

10. Krishnan S, Krishna A, Nordin A, Amirtham C, Rani A, Rao T (2020) Design and fabrication of a parallel mechanism for foot/ankle rehabilitation therapy. In: Advancement in emerging technologies and engineering applications. Springer, pp 133–141
11. Nagarajan T, Krishnan S, Rani A, Amirtham V, Rao T, Ambaraj W et al. (2015) A novel approach of using SMA wires actuated stewart platform for continuous passive motion (CPM) of ankle rehabilitation therapy. In: Advanced materials research, pp 494–498
12. Abdul-Rani AM, Krishnan S, Nagarajan T, Rao T, Ramiah R, Ambaraj W (2017) Early mobilization using heat treated shape memory alloy (SMA) for rehabilitation phases. In: Key engineering materials, pp 112–116

A Review of Artificial Intelligence-Based Techniques to Estimate Atmospheric Parameters Influencing the Performance of Concentrating Photovoltaic/Thermal Systems

F. Masood [ORCID], **P. Nallagownden** [ORCID], **I. Elamvazuthi** [ORCID], **J. Akhter** [ORCID], and **M. A. Alam** [ORCID]

Abstract Concentrating photovoltaic/thermal (CPV/T) technology is regarded as the most auspicious part of renewable energy capable of reducing reliance on fossil fuels due to its superior performance and hybrid output nature. CPV/T technology aims to reduce the cost of the renewable systems by replacing the costly solar cell material with relatively cheap optical devices that concentrate the light collected from the sun to a small solar PV cell and simultaneously generating useful heat energy for process heat applications. However, the electrical and thermal performances of systems utilizing the methodology mentioned above get strongly affected by atmospheric parameters like solar radiation, ambient temperature, and the solar spectrum. In recent years, due to the advantages tendered by Artificial Intelligence tools to solve ambiguous and non-linear problems, many authors have used intelligent system-based techniques for the prediction of the above-mentioned atmospheric parameters. This paper presents a review of artificial intelligence-based techniques, including Artificial Neural Network, Genetic Algorithm, and their composite models for the estimation of atmospheric parameters that significantly influence the working of hybrid concentrating PV/thermal systems. The review demonstrates the feasibility and accuracy of artificial intelligence-based tools for precise solar insolation and ambient air temperature prediction.

F. Masood (✉) · P. Nallagownden · I. Elamvazuthi
Department of Electrical and Electronics Engineering, Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Malaysia

M. A. Alam
Department of Mechanical Engineering, Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Malaysia

J. Akhter
Department of Mechanical Engineering, University of Engineering and Technology Taxila, Rawalpindi 47080, Pakistan

F. Masood
Department of Electrical Engineering, University of Engineering and Technology Taxila, Rawalpindi 47080, Pakistan

1 Introduction

Owing to the rapid population growth and the fantastic advancement of the industrial sector, energy consumption is rising globally at a much faster pace. Renewable energy resources have emerged as a promising alternative to conventional energy generation systems due to their abundance and free availability. They have gained significant importance in overcoming the energy crisis [1]. Among all renewable energy resources, solar energy has particularly drawn enormous consideration in various engineering and industrial applications. Solar energy can be harnessed either by employing solar photovoltaic panels, which directly convert sunlight into electricity, or by utilizing solar thermal collectors, which extract heat energy from the solar radiation and transfer it to a working fluid, usually air or water, for further use in thermal applications [2, 3]. The combination of photovoltaic panels, solar thermal collectors, and reflective or refractive optical concentrators, also collectively called concentrating photovoltaic thermal collectors, has attracted the attention of researchers and developers in the field of solar energy due to their hybrid nature and tremendous performance as compared to their constituent elements operating alone [4]. The optics attached to the PV module also called the CPV system, is used to concentrate sunlight on PV cells to generate the same amount of power with the relatively lesser solar cell material, thus replacing costly solar cells with cheap optical concentrators. In contrast, the purpose of solar thermal collector appended to the back of PV cells, called photovoltaic/thermal system, is to remove heat from solar cells for enhancing their efficiency on the one hand and producing thermal energy on the other hand [5]. A CPV/T collector generates an output of hybrid nature consisting of both electrical and thermal forms.

The CPV/T systems are usually classified on the basis of the concentration ratio of the incident sun rays over the receiver [6]. The concentration ratio demonstrates the extent to which sunlight is concentrated on the target. The concentration is sometimes expressed in the form of 'suns,' which indicates the level up to which the solar power is aggregated. Based on concentration ratio (CR), the CPV/T systems are grouped into low ($1 < CR < 10$), medium ($10 < CR < 100$) and high concentration ($CR > 100$) photovoltaic/thermal systems [7].

In a CPV/T system, the prediction of electrical and thermal attributes is mandatory for the detailed assessment of operational constraints and the economic viability of these systems to promote their market proliferation [8]. However, the modeling of CPV/T systems is entirely dissimilar and more complicated than CPV and PVT systems operating alone in real operating conditions due to diverse factors like their dependence on spectral distribution, difficulty in measuring cell temperatures, and solar flux distribution on the receiver surface [9].

Artificial intelligence-based techniques are incredibly beneficial in resolving multifaceted problems and modeling complex systems [10]. For example, Artificial neural networks (ANN) have been extensively used in the field of photovoltaics to estimate and forecast solar radiation and for modeling and simulation of both standalone and grid-connected photovoltaic systems [10, 11]. Due to the superior performance shown by ANNs for modeling of non-linear systems together with the complicacy associated with the optical, electrical, and thermal modeling of CPV/T systems, many researchers have applied ANNs for modeling, simulation, characterization, and performance estimation of both low concentration and high concentration photovoltaic/thermal systems [10, 12]. The utilization of AI tools offers an alternate solution to the problems that are still complicated from a fundamental physical perspective due to the complexity of multi-physics phenomena associated with the operation of these systems [13]. Such methods no longer require particularized information regarding the materials needed for the fabrication of a given system. Besides, advanced knowledge of optics, semiconductor physics, and thermal sciences is no longer required as compared with most of the modern modeling techniques available in the literature [14]. However, ANNs also exhibit some disadvantages as compared to conventional analytical methods. For example, the inter-relationship between desired input and output variables is sometimes unknown, and the historical data required for modeling purposes must have high accuracy [15].

In this paper, a review of AI-based techniques to address the diverse issues related to the estimation of atmospheric parameters affecting CPV/T systems is presented. The primary aspiration is to prepare a summary of the published research papers about the applications of AI tools like Artificial Neural Network (ANN), Genetic Algorithms (GA), and their derivatives for the estimation of solar radiation as well as ambient air temperature.

2 Application of ANNs for the Estimation of Solar Radiation

The HCPV/T systems usually employ refractive optical elements to concentrate the sunlight over solar cells. Such systems respond only to direct normal irradiance abbreviated as DNI. Hence, this segment of sunlight is compulsory for the performance prediction and characterization of HCPV/T structures. On the other hand, the performance of LCPV/T systems depends on both direct and diffuse parts of solar radiation due to their wide acceptance angles. The prediction of direct and diffuse components of solar radiation is difficult when using analytical methods. Therefore, many authors have employed ANNs for their prediction.

Lopez et al. [16] proposed an artificial neural network based on Bayesian function for modeling of direct normal irradiance. The authors demonstrated that clearance index and air mass are major parameters pertinent to the operation of the system.

Alam et al. [17] used an ANN model to estimate beam solar radiation using a parameter called reference clearance index. The root-mean-square error for the model fluctuates in the range of 1.66–2.78% for different states in India. In reference [18], the authors established two ANN models for the prediction of DNI indirectly from the clearance index. A close agreement between predicted and measured values was observed for the Indian region. An adaptive alpha model for forecasting the hourly global, diffuse, and direct irradiances was developed by Millet et al. [19]. However, the results were found to be inferior compared to a feed-forward neural network. Premalatha and Naveen [20] developed six models based on ANN using the training algorithm, TRAINLM, and a transfer function, log-sigmoidal, to predict global solar radiation. The authors used 32 combinations of different input parameters for prediction purposes. Renno et al. [21] developed a multi-layer perceptron model for the prediction of global radiation and DNI. Two ANN models were developed using different parameters like climatic, astronomic, and radiometric variables. For hourly DNI estimation, the root-mean-square error value was found to be 3.15%. Hameed et al. [22] proposed a low-cost methodology for predicting solar insolation, a so-called substitute to a pyranometer, based on ANN. The authors used the open-circuit voltage and short circuit current of a solar PV module and then used these values as input parameters to ANN to estimate the irradiance. The authors concluded that the proposed method could produce solar irradiance values with fair approximation. Table 1 summarizes the AI base techniques used for the prediction of both components of insolation, i.e., direct, and diffuse.

From the literature survey, it was observed that most authors employed a particular type of ANNs called feed-forward back-propagation artificial neural networks to estimate the constituents of solar radiation like direct and diffuse components and global irradiance. Most techniques were found to be dependent on atmospheric parameters like temperature and humidity. Astronomic and radiometric parameters like longitude, latitude, and zenith angle were also used as input variables. The value of root-mean-square error for the selected papers ranges from 0.20 to 24.5%. Few authors used a combination of ANN and GA for the prediction of solar insolation. Rodriguez et al. [23] used the ANN model and GA to predict global solar radiation for Spain. The predicted values had an accuracy of 99% and a root-mean-square error of 6.74%. In another study, Rodriguez et al. [24] used a combination of ANN and GA models to predict DNI. The improvement in estimation performance was found to be 35%, whereas the root-mean-square error was found to be 24.29%.

3 Application of ANNs for the Estimation of the Ambient Air Temperature

Another atmospheric parameter that strongly affects the performance of CPV/T systems is the ambient temperature. An ANN model was used to predict, twenty-four hours in advance, the average hourly value of ambient temperature for Saudi Arabia

Table 1 Summary of AI models for prediction of solar radiation

S. No	Authors	Type of AI tool	Inputs	Outputs	Error
1	Lopez et al	Bayesian Neural Network	clearance index, air mass	DNI	–
2	Alam et al	Feed-Forward Neural Network	Latitude, longitude, altitude, month number	Ref clearance index	RMSE = 2.7%, MBE = 0.07%
3	Mishra et al.	Feed-Forward Neural Network, Radial Basis Function Network	Latitude, longitude, altitude, month number, mean sunshine, rainfall, relative humidity	Clearance index	RMSE = 5.4% RMSE = 12%
4	Millet et al.	Feed-Forward Neural Network	Temperature, humidity, sunshine, advance hour insolation	Direct and diffuse radiation	$R^2 = 98.34\%$ $R^2 = 98.2\%$
5	Premalatha and Naveen	Feed-Forward Neural Network	Sunshine hours, extraterrestrial radiation, daily max, and min temperatures	Global solar radiation	RMSE = 3.97%
6	Renno et al.	Feed-Forward Neural Network	clearance index, declination angle, hour angle	DNI	RMSE = 3.15% $R^2 = 0.992\%$
7	Hameed et al.	Feed-Forward Neural Network	Open circuit voltage, short circuit current, Irradiance	DNI	RMSE = 2.5%

by Tasadduq et al. [25]. The percent difference between forecasted and measured temperatures was 3.17, 4.18, and 2.84 for three years. For hourly temperature prediction, Abdel-Al [26] developed an alternative machine learning model. The authors used measured hourly temperature values for a period of seven years as input to the model. The model was used to estimate hourly temperature for the following day and the next hour temperature having a mean absolute error of 2.01 °F and 1.05 °F respectively. In another study, Smith et al. [27] developed an ANN model for the prediction of ambient air temperature for the entire year by using the actual weather data set measured by an organization, namely the Georgia Automated Environmental Monitoring Network. The present magnitudes of ambient air temperature together with 24 h of advance values, solar insolation, relative humidity, wind speed, and rainfall were used as model inputs. The model was able to predict the values showing a mean absolute error of only 0.516 °C for a one-hour horizon. To predict the average

ambient air temperature in Denizli, Turkey, Dombayci, and Golcu [28] established an ANN model. The authors used three years of temperature data noted by the Turkish state metrological service to train the network. The average previous day temperature, the day, and the month were used as input to the model. The root-mean-square value was found to be 1.96 for the testing data set. To estimate temporal and spatial air temperature, Deligiorgi et al. [29] developed various ANN models. For one hour, two hours, and three hours advance temporal air temperature prediction, the authors trained different ANN models using Levenberg-Marquadrt back propagation algorithm. The proposed ANN model performance was found to decrease with rising forecasting lag. The mean absolute error was less than 1.4 °C in all cases. The radial basis function network and the multi-layer perceptron non-linear feed forward neural networks were compared for spatial forecasting. It was concluded that both models produced accurate results for the air temperature.

An ANN model was developed by Almonacid et al. [30] for the prediction of air temperature on an hourly basis. The daily maximum, minimum, and average air temperature, latitude, and longitude of a site in Spain were used as inputs to the model. The results were compared with those produced by conventional techniques. The root-mean-square error value lies from 0.53 to 1.98 °C. Cobaner et al. [31] suggested applying ANN together with Multiple linear regression models and adaptive neuro-fuzzy inference system (ANFIS) for the prediction of the largest, smallest, and mean temperatures for a whole month, taking into account the geographical coordinates and the month number for a given site in Turkey. The input variables include longitude, latitude and altitude, and the month number. When compared, the results obtained by the ANFIS model were found to be best, with a mean absolute error of 1.25% for the mean temperature. The value of R^2 was found to be 0.966 for the average temperature. The comparison of ANN models with ANFIS for predicting air temperatures on a monthly basis was conducted by Kisi and Shiri [32]. For training the network, data obtained from twenty different metrological stations located in Iran were utilized. The longitude, latitude, altitude of the site, and the time span were used as input parameters. The performance of the ANN model was found to be superior to the ANFIS model in the test duration.

Venkadesh et al. [33] used a GA-based approach to adjudge the duration and resolution of prior input data for an ANN model to predict air temperature. The GA classified contributive roles of different input variables for the estimation of air temperature for one to twelve hours prediction horizons. The authors concluded that ANN models established based on the GA approach produced more accurate results as compared to the usual ANN. The mean absolute error showed an improvement with the proposed technique.

4 Conclusions

The atmospheric parameters like solar irradiance, ambient air temperature, and solar spectrum strongly influence the performance of CPV/T systems. A comprehensive

review of artificial intelligence-based tools for predicting these atmospheric parameters was presented in this paper. From the review of published research articles, it was found that most authors developed ANN-based models for the prediction of solar radiation and ambient temperature. A few authors employed a combination of ANN and GA for a more accurate prediction of irradiance and temperature. In contrast, characterization of the solar spectrum for CPV/T systems using AI-based models is still unexplored. Moreover, AI-based models were found to be robust and accurate tools for the prediction of atmospheric parameters. Due to their wide acceptance angles, low concentration systems possess the ability to collect direct and diffuse components of solar radiation. In contrast, most research studies focused on the prediction of direct solar irradiance only, so more accurate hybrid models are required to be developed for the prediction of diffuse radiation and the solar spectrum, particularly for LCPV/T systems.

References

1. Baig H, Sarmah N, Heasman KC, Mallick TK (2013) Numerical modelling and experimental validation of a low concentrating photovoltaic system. Sol Energ Mater Sol Cells 201–219
2. Chandan DS, Kumar PS, Reddy KS, Pesala B (2020) Optical and electrical performance investigation of truncated 3X non-imaging LCPV/T systems. En Con Mgt 113056
3. Zhang G, Wei J, Wang Z, Xie H, Xi Y, Khalid M (2019) Investigation into effects of non-uniform irradiance and photovoltaic temperature on performances of photovoltaic/thermal systems coupled with truncated compound parabolic concentrators. Appl Enrg 245–56
4. Al-Nimr MA, Mugdadi B (2020) A hybrid absorption/thermo-electric cooling system driven by a concentrated photovoltaic/thermal unit. Sus En Tech Assmnts 100769
5. Kim SM, Kim JH, Kim JT (2019) Experimental study on the thermal and electrical characteristics of an air-based photovoltaic thermal collector. Energies 2661
6. Sharaf OZ, Orhan MF (2015) CPVT solar collector systems: part I—fundamentals, design considerations and current technologies. Ren Sus En Rev 1500–65
7. Sharaf OZ, Orhan MF (2015) CPVT solar collector systems: part II—implemented systems, performance assessment, and future directions. Ren Sus En Rev 1566–1633
8. Diwania S, Agrawal S, Siddiqui AS, Singh S (2019) Photovoltaic–thermal (PV/T) technology: a comprehensive review on applications and its advancement. Int J En Env Eng 33–54
9. Almonacid F, Fernandez EF, Mellit A, Kalogirou S (2017) Review of techniques based on ANNs for the electrical characterization of cpv technology. Ren Sus En Rev 938–953
10. Shahsavar A, Moayedi H, Al-Waeli A, Sopian K, Chelvanathan P (2020) Machine learning predictive models for optimal design of BIPV/T collectors. Int J En Rsh 5675–95
11. Qazi A, Fayaz H, Wadi A, Raj RG, Rahim NA, Khan WA (2015) The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review. J Cnr Prdn 1–12
12. Motahar S, Bagheri-Esfeh H (2020) Artificial neural network based assessment of grid-connected pv/t systems in heating dominated regions of Iran. Sus En Tch Assesments 100694

13. Bermejo JF, Fernandez JF, Polo FO, Marquez AC (2019) A review of the use of artificial neural network models for energy and reliability prediction. A study of the solar PV, hydraulic and wind energy sources. Appl Sci 1844

14. Almaktar M, Rahman HA, Hassan MY, Saeh I (2015) Artificial neural network-based photovoltaic module temperature estimation for tropical climate of Malaysia and its impact on photovoltaic system energy yield. Prog Phvltaics Res Appl 302–18

15. Almonacid F, Fernandez EF, Rodrigo P, Perez-Higueras PJ, Rus-Casas C (2013) Estimating the maximum power of a HCPV module using an ANN. Energies 165–172

16. López G, Batlles FJ, Tovar-Pescador J (2005) Selection of input parameters to model direct solar irradiance by using artificial neural networks. Energy 1675–1684

17. Alam S, Kaushik S, Garg SN (2006) Computation of beam solar radiation at normal incidence using artificial neural network. Ren Enrgy 1483–1491

18. Mishra A, Kaushika ND, Zhang G, Zhou J (2008) Artificial neural network model for the estimation of direct solar radiation in the Indian zone. Int J Sus En 95–103

19. Mellit A, Eleuch H, Benghanem M, Elaoun C, Pavan AM (2010) An adaptive model for predicting of global, direct and diffuse hourly solar irradiance. En Cnvrsion Mgt 771–782

20. Premalatha M, Naveen C (2018) Analysis of different combinations of meteorological parameters in predicting the horizontal global solar radiation with ANN approach: a case study. Ren Sus En Rev 248–258

21. Renno C, Petito F, Gatto A (2015) Artificial neural network models for predicting the solar radiation as input of a concentrating photovoltaic system. En Con Mgt 999–1012

22. Hameed WI, Sawadi BA, Al-Kamil SJ, Al-Radhi MS, Al-Yasir YIA, Saleh AL, Abd-Al-Hameed RA (2019) Prediction of solar irradiance based on ANN. Inventions 45

23. Linares-Rodriguez A, Ruiz-Arias JA, Pozo-Vazquez D, Tovar-Pescador J (2013) An artificial neural network ensemble model for estimating global solar radiation from Meteosat satellite images. Energy 636–45

24. Linares-Rodriguez A, Quesada-Ruiz S, Pozo-Vazquez D, Tovar-Pescador J (2015) An evolutionary artificial neural network ensemble model for estimating hourly direct normal irradiances from Meteosat imagery. Energy 264–73

25. Tasadduq I, Rehman S, Bubshait K (2002) Application of neural networks for the prediction of hourly mean surface temperatures in Saudi Arabia. Ren Energ 545–554

26. Abdel-Aal, Radwan E (2004) Hourly temperature forecasting using abductive networks. Eng App Artficial Intell 543–556

27. Smith BA, Hoogenboom G, McClendon RW (2009) Artificial neural networks for automated year-round temp prediction. Comp Electro Agric 52–61

28. Dombaycı ÖA, Gölcü M (2009) Daily means ambient temp prediction using artificial neural network method: a case study of Turkey. Ren Energ 1158–61

29. Deligiorgi D, Philippopoulos K, Kouroupetroglou G (2013) ANN based methodologies for the spatial and temporal estimation of air temp. In: International conference on pattern recognition applications and methods

30. Almonacid F, Pérez-Higueras P, Rodrigo P, Hontoria L (2013) Generation of ambient temperature hourly time series for some Spanish locations by ANNs. Ren Energ 285–91

31. Cobaner M, Citakoglu H, Kisi O, Haktanir T (2014) Estimation of mean monthly air temperatures in Turkey. Comp Electro Agric 71–79

32. Kisi O, Shiri J (2014) Prediction of long-term monthly air temperature using geographical inputs. Int J Climato 179–86

33. Venkadesh S, Hoogenboom G, Potter W, McClendon R (2013) A genetic algorithm to refine input data selection for air temperature prediction using ANNs. App Soft Comp 2253–60

On the Performance of MIMO-UVLC System over Turbulence-induced Fading Channels

Sajid Mumtaz, Azrina Abd Aziz, and Komal Masroor

Abstract Visible light communication in underwater environments is replacing the RF and the acoustic signals due to its high bandwidth, low latency, and small attenuations. One of the performance degrading factors in underwater channels is the turbulence which is caused by a continuous change in refractive index of water. Spatial diversity can overcome the optical scintillations as well as the pointing errors. This paper evaluates the ergodic capacity and the BER performance of MIMO as well as SSK-MIMO UVLC system in presence of underwater turbulence, modelled by Lognormal, Gamma, and Weibull distributions. The capacity analysis is based upon varying number of transceivers and the water filling algorithm. Impact of spatial correlation is also considered for the capacity analysis. The results find their validation via Monte Carlo simulations.

Keywords Underwater visible light communication (UVLC) · Multiple input multiple output (MIMO) · Optical turbulence · Ergodic capacity · Distribution models · Space shift keying (SSK)

1 Introduction

Optical signals are currently in target stream of researchers due to their bandwidth and power efficiency, least latency, and reliability compared to traditional RF and acoustic signals. However, the transmission of optical signals underwater faces many challenges, such as absorption, scattering, turbulence, noise and pointing errors, that need to be resolved. Absorption and scattering of optical signals have been widely studied and modelled. Moreover, the blue and green window of optical spectrum offers minimal absorption and scattering [1], making it to be an optimized option

S. Mumtaz (✉) · A. A. Aziz · K. Masroor
Department of Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia

K. Masroor
Department of Electronic and Telecommunications Engineering, NED University of Engineering and Technology, Karachi, Pakistan

for wireless communication, hence called underwater visible light communication (UVLC).

Underwater turbulence, caused by rapid variations in temperature, salinity, air-bubbles and pressure, severely degrades the optical signals [2–4] and it is still an open challenge. Various statistical distributions, such as Lognormal, Gamma, Weibull, Generalized Gamma, Gamma-Gamma (GG), and some mixture distributions have been developed to characterize the scintillations due to optical turbulence underwater. For weak turbulence-induced fading with scintillation index $\sigma_i^2 < 0.1$, Lognormal distribution is the most widely used because of the weighted sum approximation of its random variables (RVs) [5–8]. For very low fading due to temperature gradients with scintillation index $\sigma_i^2 < 0.001$, Gamma distribution fits well with the optical intensity fluctuations [9], and its mathematical simplicity provides tractable solutions. For salinity-induced turbulence fading with scintillation index $0.1 < \sigma_i^2 < 1$, Weibull distribution model is the most appropriate one [10].

The channel capacity for underwater VLC channels has been studied and analyzed. Zhang [11] studied the ergodic capacity and the BER performance of UVLC system with pointing errors (PEs) under different water types, and observed that water turbidity and long travelling distance of photons affect both the capacity and the BER performance. Li [12] an analytic and capacity model based upon the PEs variance for turbulence-induced UVLC links. Xu [13] analyzed the scintillation index of plane wave propagation in anisotropic water turbulence under moderate-to-strong turbulent conditions, and noticed the large receiving apertures and wavelengths to improve significantly the UVLC quality and capacity.

The scintillations due to the optical turbulence are reduced using the methods, like Adaptive optics [14], Rytov method [15], aperture averaging [16], and higher-mode flat topped lasers [17]. MIMO is a proven technology to overcome fading, the misalignments, and the possible blockage due to the obstructions in free space optical (FSO) communications. Like RF, the capacity of MIMO Optical Wireless Communication (OWC) systems is also investigated rigorously. The ergodic capacity and probability outage for MIMO-FSO links under Lognormal, Rayleigh and Gamma-Gamma channels under different power constraints and modulation techniques have been investigated and analyzed [18–20], and the benefits of MIMO are proved in optical domain as well.

However, the performance of UVLC systems under spatial diversity techniques received lack of attention. Jamali [5, 8] analyzed the performance of UVLC system under MIMO technique, and proved that the spatial diversity remarkably improved the system's performance in weak underwater turbulence regimes. Abdallah and Khalighi [21] modeled UVLC channel based upon angular MIMO scheme, and found that compared to conventional MIMO, it withstood against the misalignment conditions for the channel capacity. Li [22] analyzed the effect of spatial correlation on imaging MIMO, and verified the fact that the correlation degraded the UVLC system performance. Jiang [23] evaluated the performance of MIMO-UVLC systems in combination with OFDM technique, and proved its worth to cope with the underwater turbulence-induced fading.

Moreover, for increased spectrum efficiencies, MIMO space shift keying (SSK) technique has been widely applied to OWC systems; where instead of signal constellations, the data bits are transmitted via antenna indices providing reduced system complexities. Abaza [24] assessed the performance of optical SSK-MIMO under atmospheric turbulence and observed an increase in spectral efficiency of 3 bits/s/Hz. Normally the existence of correlated signals in MIMO always reduced the system's performance. But Priyadarshani and Jaiswal [25] reported that the usage of SSK-MIMO over correlated FSO channel provided a significant increased BER performance up to 13 dB.

This paper evaluates the ergodic capacity for MIMO UVLC systems under Lognormal, Gamma and Weibull distributed channels based upon their achieved SNR. These models characterize the optical scintillations for a particular range of underwater turbulence conditions. Sections 2 and 3 cast light over the employed system and the channel models, respectively. Section 4 explains about the ergodic capacity with MIMO technology with some light over the correlated cannels. For enhanced MIMO spectrum efficiency, an overview about Space Shift Keying (SSK) is provided in Sect. 5. Section 6 presents result and discussion, and Sect. 7 gives the conclusion and the future work.

2 System Model

Consider an UVLC link with M number of transmitters (LEDs/LDs) pointing towards N number of receivers (PDs). We assume that this link utilizes intensity modulation direct detection (IMDD) with OOK modulation. We assume the performance of the given system is affected by background radiations, thermal noise, shot noise, which are normally modelled as independent and identically distributed (iid) additive white Gaussian noise (AWGN). To obtain the uncorrelated optical signals at the receiver, we assume that the separation between the receivers is greater than the irradiance spatial coherence distance, and the receiver's aperture diameter is less than the spatial correlation width. If the scattering is not too high, then the correlated signals will be received at the receiver, and that results in the reduced capacity and low data rates. The received signal at the nth receiver aperture is

$$r(t) = \eta s R P_t H s(t) + n(t) \tag{1}$$

where $s \in \{0, 1\}$ denotes the transmitted signal, and P_t is the transmitted power. η represents the optical-to-electrical conversion coefficient, R is the receiver's responsivity, $n(t) \in \mathbb{R}^{M \times N}$ is the AWGN with zero mean, and $N_0/2$ noise power spectral density (PSD). Moreover, H is the MIMO channel matrix comprising of row and column vectors, and $s(t)$ is the transmitted intensity OOK modulated optical beam.

3   Channel Model

For the analysis, we considered Lognormal, Gamma and Weibull distribution models that can characterize the optical scintillations due to weak, temperature-as well as salinity-induced turbulence, respectively.

3.1   Lognormal Distribution Channel

The intensity fluctuations caused by underwater weak turbulence are characterized by the fading coefficient I, which can be modelled by the PDF of Lognormal distribution given by

$$f_I(I; \mu_x, \sigma_x^2) = \frac{1}{2I\sqrt{2\pi\sigma_x^2}} \exp\left(\frac{(\ln(I) - 2\mu_x)^2}{8\sigma_x^2}\right) \tag{2}$$

where $\mu_x = \ln(I_0)/2$ and $\sigma^2 = \left[\mathrm{E}\{I^2\}/\mathrm{E}^2\{I\}\right] - 1$ is the logscale parameter (first moment) and the scintillation index (second moment) of normal distributed log-amplitude factor respectively, with I_0 as the received irradiance in absence of the considered channel.

3.2   Gamma Distribution Channel

The PDF for Gamma distribution is

$$f_I(I; a, b) = \frac{1}{\Gamma(a)b^a} I^{(a-1)} \exp\left(-\frac{I}{b}\right) \tag{3}$$

with a and b as shape and scale parameters. $\Gamma(.)$ is the Gamma function which is related to the S.I by the relation $b = 1/a = \sigma_I^2$. For UVLC links with very low temperature-gradients, i.e., $\sigma_I^2 \leq 0.001$, this simple distribution can be used to characterize the fluctuations of received optical beam.

3.3   Weibull Distribution Channel

The PDF for Weibull distribution is given by
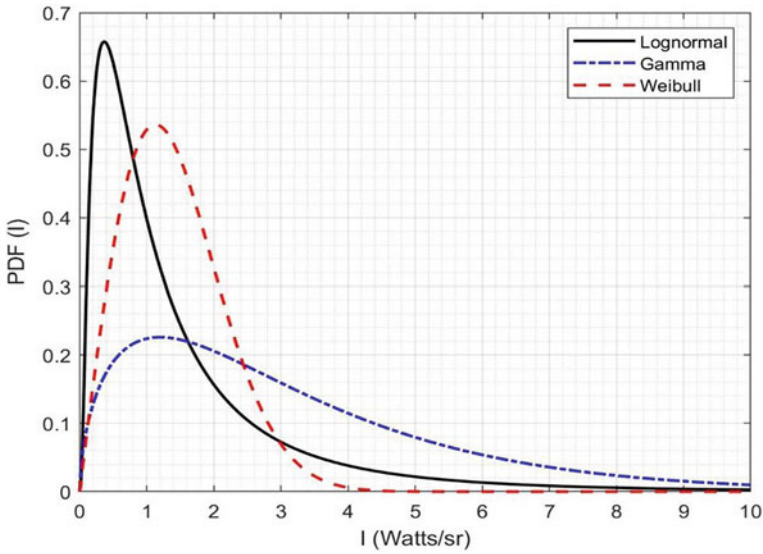
Fig. 1 Lognormal, Gamma and Weibull distributions with zero mean and unit variance for intensity fluctuations

$$f_I(I; a, b) = \frac{a}{b}\left(\frac{I}{b}\right)^{(a-1)} \exp\left(-\left(\frac{I}{b}\right)^a\right) \tag{4}$$

with a and b as shape and scale parameters, respectively. I, a and b are all positive valued. The parameter b is related to the S.I by the relation $\sigma_I^2 = b^2\{\Gamma[1 + (2/\lambda)] - \Gamma[1 + (1/\lambda)]\}$. It is noted that for $a = 1$, this distribution yields $f(I; 1, b) = (1/b)\exp(-(I/b))$ which is negative exponential distribution, and for $a = 2$, this distribution is converted to the Rayleigh distribution. The Weibull distribution fits well for salinity-induced turbulence fading in underwater channels.

The graph between the probability of irradiance because of the turbulence and the optical power per unit solid angle for the considered distribution models is shown in Fig. 1.

4 Ergodic Capacity

The utmost mutual information that can be attained by changing the PDF of transmit signal vector is called the capacity, which is an important parameter in characterizing the performance of a wireless communication system. The instantaneous channel capacity is defined as

$$C_{inst} = \mathrm{E}\{\log_2(1 + cI)\} \tag{5}$$

where c is a constant term, with $c = 1$ and $c = e/2\pi$ for heterodyne detection as well as intensity modulation and direct detection (IMDD) respectively [26].

We assume that the optical channel is memoryless, stationary, and perfect channel state information (CSI) is available at both transmitter and receiver of the system. In terms of MIMO channel, the capacity relation can be written as

$$C = \mathrm{E}\left\{\log\left[\det\left(I'_M + (1/M)H^H H\right)\right]\right\} \tag{6}$$

where I'_M denotes $M \times M$ identity matrix, and H is an $M \times N$ channel matrix with elements being I_{mn}. Since the instantaneous SNR is a random variable (RV) with a given PDF, so we assume the channel capacity as a RV. The average of this RV is the ergodic capacity, and can be written as

$$\overline{C} = \int_0^\infty B \log_2(1 + I) f_I(I) dI \tag{7}$$

where B denotes the transmission bandwidth. By substituting the values of $f_I(I)$ from Eqs. (2), (3), and (4) into Eq. (7), the ergodic capacity closed-form expressions can be obtained.

The expression for ergodic capacity under MIMO-UVLC Lognormal distribution channel is

$$\overline{C_{LN}} = \int_0^\infty B \log_2(1 + I) \frac{1}{2I\sqrt{2\pi\sigma_x^2}} \exp\left(\frac{(\ln(I) - 2\mu_x)^2}{8\sigma_x^2}\right) dI \tag{8}$$

The ergodic capacity for MIMO-UVLC Gamma distribution channel will be

$$\overline{C_{Gam}} = \int_0^\infty B \log_2(1 + I) \frac{1}{\Gamma(a)b^a} I^{(a-1)} \exp\left(-\frac{I}{b}\right) dI \tag{9}$$

Similarly, the ergodic capacity expression for MIMO-UVLC for Weibull distribution channel can be expressed as

$$\overline{C_{Wb}} = \int_0^\infty B \log_2(1 + I) \frac{a}{b}\left(\frac{I}{b}\right)^{(a-1)} \exp\left(-\left(\frac{I}{b}\right)^a\right) dI \tag{10}$$

5 Space Shift Keying MIMO

Underwater communication severely needs the wireless systems to be power as well as bandwidth efficient. Although MIMO technique can provide reduced signal fading [8], yet it could not provide high spectrum efficiencies. Opting higher order modulations can be a viable way out to attain spectrum efficient (SE) systems, but they result in higher peak to average power ratios (PAPR) which is another problem. There is an emerging method for achieving high SE in presence of PAPR is the spatial modulation (SM) MIMO [27] which utilizes both signal constellations as well as antenna indices at once. Some of the incoming bits are transmitted via signal space, while the remaining are transmitted through the transmitter index, thus increasing the SE. Being one antenna active at a time, power efficiencies are also achieved. SSK is a special case of SM which uses only the antenna indices for data transfer. SSK provides zero inter channel interference (ICI) and least detection complexities, making the system simpler compared to SM [28, 29]. These features make SSK-MIMO a good candidate for UVLC systems. This paper also analyzes the BER performance of optical SSK-MIMO UVLC system under Lognormal, Gamma and Weibull distribution models.

6 Results and Discussion

This section discusses the simulation-based results. As spatial diversity mitigates with the fading and increase the data rates, so with increasing number of transmitters and receivers, we increase the data rate, and obviously the capacity for the given channel. Figure 2a, b depict the channel capacity under different values of SNR with increasing number of transmitters and receivers. More receivers will receiver different SNR values that can be combined by maximal ratio combining (MRC) and equal gain combining (EGC) techniques, yielding net higher SNR, and offering huge
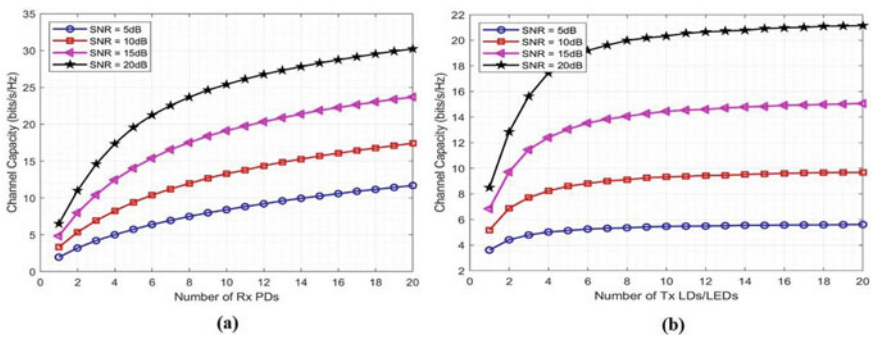
Fig. 2 **a** Channel capacity with varying number of receivers when number of transmitters is fixed, **b** Channel capacity with varying number of transmitters when number of receivers is fixed
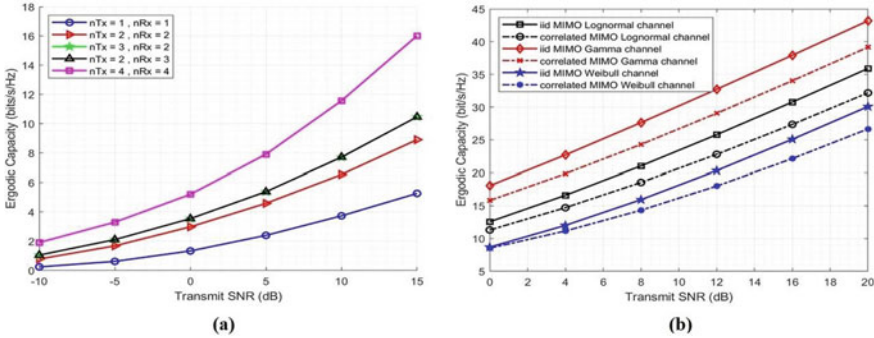
Fig. 3 a Channel capacity versus received SNR based on water filling algorithm, **b** comparison between iid and correlated MIMO UVLC channel under lognormal, Gamma, and Weibull distributions

channel capacity. For lower SNR values less than a specific threshold, an increment in the number of receivers does not help in providing the higher capacity and the data rates.

Figure 3a shows the channel capacity based on water-filling algorithm, where the optical receivers adapt themselves according to the channel conditions to attain the desired power points. When the channel state information (CSI) is not known at the transmitter, then the transmitter throws the power isotopically towards the receivers. Due to path losses, some of the power goes wasted, and the channel capacity is affected. However, in presence of CSI known to the transmitter, the transmitted power is then optimized, and it acts as a beamforming MIMO. For known CSI, with increase of transmitters and receivers, the channel capacity increases and thus offers high data rate transmission. Compared to SISO systems, MIMO provides an increased capacity of up to 11 dB, that can further be increased employing a greater number of transceivers, but at the cost of increased system complexities.

Much of the study is done for analyzing the capacity of MIMO channels when all the optical signals are independent and identically distributed (iid). In a scattering environment, the channel impulse response (CIR) has independent fingerprints helping the receivers to identify and differentiate among the signals. On the receiver side, this condition can be achieved by making separation between the receivers more than the channel correlation length. However, in some practical scenarios, the realization of this condition is not possible, and this makes the analysis of correlated channels more crucial and important. In less scattering environments, there exists the correlation among the signals, and hence the performance of the link is degraded. Figure 3b shows the ergodic capacity of Lognormal, Gamma and Weibull distribution channels with varying SNR values. The results show that Gamma channel yields higher capacity and data rates compared to Lognormal and Weibull channels. The results also show that in presence of correlation among signals, the capacity as well as the data rate under all channels reduce significantly.
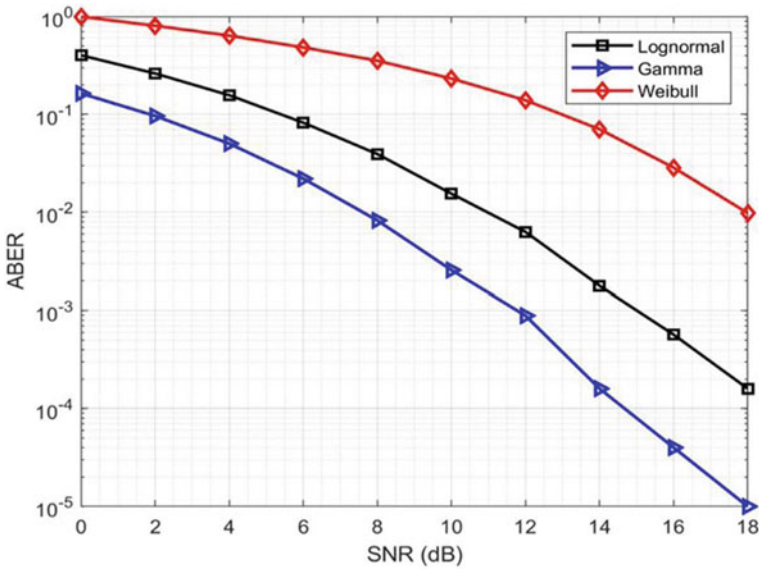
Fig. 4 BER versus SNR for optical SSK-MIMO UVLC system under Lognormal, Gamma and Weibull distributions

The BER performance of optical SSK MIMO system under Lognormal, Gamma, and Weibull distribution channels is manifested in Fig. 4. As SSK uses only the antenna indices for data transfer, so the system and detection complexity are reduced, and spectrum efficiency is achieved. BER versus SNR plot shows that OSSK-MIMO under Gamma distribution channel outperforms Lognormal and Weibull channels. However, its performance under Lognormal channel is much better than that of Weibull distribution.

7 Conclusion and Future Work

This paper evaluates the ergodic capacity and BER performance of MIMO underwater visible light communication (UVLC) under Lognormal, Gamma, and Weibull distributions which can characterize the optical scintillations due to weak, temperature-induced, and salinity-induced turbulence fading, respectively. More transceivers can transmit and receive more power, and that can be combined using MRC and EGC techniques to produce high SNR value, and hence that of the channel capacity. To optimize the transmit power, CSI is required at the transmitter, because an unknown CSI yields a reduced amount of channel capacity. Moreover, it is observed that in presence of correlation among optical signals at the receivers, the ergodic capacity for all channels reduces. The results also show that Gamma channel provides higher capacity and data rates; however, it can model only low temperature gradient

turbulence induced underwater fading. The results also show that the performance of SSK-MIMO under Gamma distribution channels is much better than those of Lognormal and Weibull channels.

The mixture distribution models, named Exponential and Lognormal, Exponential and Gamma, Exponential and Weibull, Exponential and Generalized Gamma (EGG), are also found to fit well with the experimental data for different underwater turbulence conditions. The performance and capacity analysis of MIMO as well as MIMO-OFDM UVLC systems under such channels with suitable coding and modulation techniques are still open research issues in such regimes.

References

1. Duntley SQ (1963) Light in the sea. JOSA 53:214–233
2. Kaushal H, Kaddoum G (2016) Underwater optical wireless communication. IEEE Access 4:1518–1547
3. Tang S, Zhang X, Dong Y (2013) Temporal statistics of irradiance in moving turbulent ocean. In: 2013 MTS/IEEE OCEANS-Bergen, pp 1–4
4. Andrews LC, Phillips RL (2005) Laser beam propagation through random media
5. Jamali MV, Nabavi P, Salehi JA (2018) MIMO underwater visible light communications: comprehensive channel study, performance analysis, and multiple-symbol detection. IEEE Trans Veh Technol 67:8223–8237
6. Huang A, Tao L, Wang C, Zhang L (2018) Error performance of underwater wireless optical communications with spatial diversity under turbulence channels. Appl Opt 57:7600–7608
7. Peppas KP, Boucouvalas AC, Ghassemloy Z (2017) Performance of underwater optical wireless communication with multi-pulse pulse-position modulation receivers and spatial diversity. IET Optoelectron 11:180–185
8. Jamali MV, Salehi JA, Akhoundi F (2016) Performance studies of underwater wireless optical communication systems with spatial diversity: MIMO scheme. IEEE Trans Commun 65:1176–1192
9. Oubei HM, Zedini E, ElAfandy RT, Kammoun A, Abdallah M, Ng TK et al (2017) Simple statistical channel model for weak temperature-induced turbulence in underwater wireless optical communication systems. Opt Lett 42:2455–2458
10. Oubei HM, Zedini E, ElAfandy RT, Kammoun A, Ng TK, Alouini M-S et al (2017) Efficient Weibull channel model for salinity induced turbulent underwater wireless optical communications. In: 2017 opto-electronics and communications conference (OECC) and photonics global conference (PGC), pp 1–2
11. Zhang H, Dong Y, Hui L (2015) On capacity of downlink underwater wireless optical MIMO systems with random sea surface. IEEE Commun Lett 19:2166–2169
12. Li Y, Zhang Y, Zhu Y (2019) Capacity of underwater wireless optical links with pointing errors. Opt Commun 446:16–22
13. Xu G, Lai J (2020) Average capacity analysis of the underwater optical plane wave over anisotropic moderate-to-strong oceanic turbulence channels with the Málaga fading model. Opt Exp 28:24056–24068
14. Holohan M, Dainty J (1997) Low-order adaptive optics: a possible use in underwater imaging? Opt Laser Technol 29:51–55
15. Ata Y, Baykal Y (2014) Scintillations of optical plane and spherical waves in underwater turbulence. JOSA A 31:1552–1556
16. Yi X, Li Z, Liu Z (2015) Underwater optical communication performance for laser beam propagation through weak oceanic turbulence. Appl Opt 54:1273–1278

17. Yousefi M, Golmohammady S, Mashal A, Kashani FD (2015) Analyzing the propagation behavior of scintillation index and bit error rate of a partially coherent flat-topped laser beam in oceanic turbulence. JOSA A 32:1982–1992

18. Chakraborty K (2005) Capacity of the MIMO optical fading channel. In: Proceedings of international symposium on information theory, ISIT 2005, pp 530–534

19. Luong DA, Thang TC, Pham AT (2013) Average capacity of MIMO/FSO systems with equal gain combining over log-normal channels. In: 2013 fifth international conference on ubiquitous and future networks (ICUFN), pp 306–309

20. Zhang J, Dai L, Han Y, Zhang Y, Wang Z (2015) On the ergodic capacity of MIMO free-space optical systems over turbulence channels. IEEE J Sel Areas Commun 33:1925–1934

21. Ghazy AS, Hranilovic S, Khalighi M-A (2019) Angular MIMO for underwater wireless optical communications: channel modelling and capacity. In: 2019 16th Canadian workshop on information theory (CWIT), pp 1–6

22. Li Y, Qiu H, Chen X, Fu J, Musa M, Li X (2019) Spatial correlation analysis of imaging MIMO for underwater visible light communication. Opt Commun 443:221–229

23. Jiang H, Qiu H, He N, Popoola W, Ahmad Z, Rajbhandari S (2020) Performance of spatial diversity DCO-OFDM in a weak turbulence underwater visible light communication channel. J Lightwave Technol 38:2271–2277

24. Abaza M, Mesleh R, Mansour A, Aggoune E-HM (2015) The performance of space shift keying for free-space optical communications over turbulent channels. In: Broadband access communication technologies IX, p 93870V

25. Priyadarshani R, Jaiswal A, Bhatnagar MR, Ghassemlooy Z, Zvanovec S (2018) Performance of space shift keying over a correlated gamma-gamma FSO-MISO channel. In: 2018 11th international symposium on communication systems, networks and digital signal processing (CSNDSP), pp 1–6

26. Ansari IS, Alouini M-S, Cheng J (2014) On the capacity of FSO links under lognormal and Rician-lognormal turbulences. In: 2014 IEEE 80th vehicular technology conference (VTC2014-Fall), pp 1–6

27. Jaiswal A, Bhatnagar MR, Jain VK (2016) BER analysis of optical space shift keying with gamma-gamma fading and pointing error. In: 2016 international conference on advanced technologies for communications (ATC), pp 369–374

28. Di Renzo M, Haas H, Ghrayeb A, Sugiura S, Hanzo L (2013) Spatial modulation for generalized MIMO: challenges, opportunities, and implementation. Proc IEEE 102:56–103

29. Jeganathan J, Ghrayeb A, Szczecinski L, Ceron A (2009) Space shift keying modulation for MIMO channels. IEEE Trans Wirel Commun 8:3692–3703

Fabrication of One Degree of Freedom Knee Exoskeleton for Nervous Disorder

J. Duwaraka, N. Sudharsana, S. Prabakar, K. Porkumaran, Iraivan Ilamvaluthi, and Ramani Kannan

Abstract Instrumented Cane is a wearable sensory system for patient's motion intention estimation in order to control the trajectory of knee exoskeleton. This Instrumented Cane will be a helpful assistive device in term of balance and motion for those who suffering neurological disorder, in specific Paralysis and spinal cord injury. This report describes the study on method utilizing Instrumented Cane as a attainable intermediate between the patient and exoskeleton to enhance in terms of movement in their day-to-day activities. The structure is designed with inertial measurement unit integrated with wireless communication module in order to assist real-time control of exoskeleton. This report aims the purpose of an Instrumented Cane that able to contribute to the assisted locomotion technology as well as upgrade the mobility of physically challenged people.

Keywords Knee · Wireless · One DOF · Paralysis · Serial communication · Instrumented cane and control trajectory

1 Introduction

Paralysis is a common term used to define the defective sensation due to following damage to the nervous system. In medical terms accurate level of injuries are used which is supportive in determining the parts of body that affected by paralysis and also loss of function. Generally, this disease can be within a temporary period or even worst

J. Duwaraka · N. Sudharsana (✉)
Biomedical Engineering, Dr. N. G. P Institute of Technology, Coimbatore, India

S. Prabakar
Sona College of Technology, Salem, India

K. Porkumaran
Sri Sairam Engineering College, Chennai, India
e-mail: porkumaran@gmail.com

I. Ilamvaluthi · R. Kannan
Electrical and Electronic Engineering, Universiti Teknologi Petronas, Seri Iskandar, Malaysia

to be permanent. In common cases, paralysis affects half side of human body, called as unilateral. Under classification of paralysis, there is a particular medical condition defined as Paraplegia [4]. Paraplegia is defect affecting the lower extremities of human body. This term describe the paralysis that affects both legs and partially the trunk, but not the arms. Contracting the study of Paraplegia, there is another form of paralysis that affects just one limb known as Monoplegia. Crucial Monoplegia is scientific term for paralysis affecting the single lower limb [5]. Patients with these disabilities have trouble in walking as well as reflecting their muscle. Conversely, rapid development of technology has formed a solution with a design mechanism of exoskeleton in conjunction of instrumented cane for the control.

Development in this field creates an enthusiasm towards patients to experience the transaction from undergoing rehabilitation therapy to a "real world" walking practice. Human motion intention depends on certain bioelectrical signal of neurology system. Because of variation on their bioelectrical difficulties in mobility. Research community developed several wearable gait, exoskeleton, instrumented cane that will compute and estimate the motion intention by disable. Nevertheless, there are several constrains on the design mechanism and followings are:

1. Modelling with variable electronic hardware modules are composite.
2. In data transmission power supply for wireless is ineffectual.
3. Immense design and additional weight restrain, uncomfortable cane.

Movement transparency is critical when wearing a robot for gait rehabilitation. In other words: When wearing the exoskeleton, its movement should be synchronized and consistent with a patient's natural movement. "If not, it exerts extra forces on the human joint "And this extra force causes patient discomfort and unnatural movements."

1.1 Limitation in Current Design

Human motion intention depends on certain bioelectrical signal of neurology system. Because of variation on their bioelectrical difficulties in mobility. Research community developed several wearable gait, exoskeleton, instrumented cane that will compute and estimate the motion intention by disable. Nevertheless, there are several constraints on the design mechanism and followings are:

- Modeling with variable electronic hardware modules is composite.
- In data transmission power supply for wireless is ineffectual.
- Immense design and additional weight restrain uncomfortable cane.

Movement transparency is critical when wearing a robot for gait rehabilitation. In other words: When wearing the exoskeleton, its movement should be synchronized and consistent with a patient's natural movement. "If not, it exerts extra forces on the human joint" And this extra force causes patient discomfort and unnatural movements.

1.2 Objectives

- The hardware connection is simple with wireless transmission of information.
- The design is light weight to control the knee exoskeleton.
- The battery can be replaced easily and used for long duration.

2 Scope of Proposed System

The expected result of this project is to develop an instrumented cane that able to communicate and control a lower limb knee exoskeleton together with control algorithm that able to notice human motion trajectory. This work is to contribute in the use of rehabilitation for remote assessments as well as for the disable with their day-to-day activities. Throughout this project, research will be performed on the required balance in motion, ability, upper body strength, and the use of arm strength with the instrumented cane. Besides, the scope of this project is also to concentrate on the wireless communication system to assist in real-time control of exoskeleton via a standard I2C serial communication protocol.

2.1 Proposed Work

IMU sensor was attached to the instrumented cane to detect the movement trajectory that is discharged. The output signal obtained from the sensor will be illuminated using a microcontroller that is compatible with the sensor and wireless communication module. The data will then be mediate to the microcontroller used at the exoskeleton in order to control the trajectory based on, the information is carried via Bluetooth communication sensor. Instrumented cane and exoskeleton enhance the compatibility using Bluetooth communication protocol as the best option. As the information is conveyed to the exoskeleton at real-time, it begins to initiate the motor to act as actuator in order to make motion. Fundamentally, humans target in motion triggers first, in sequence to command the robotic structure make motion. The push button will be assigned as a safety circumstance if there is any sudden detach or drop of cane from the arm, it will not send any trajectory control data to the Exoskeleton. This prevents the exoskeleton to make any false motion trajectory which could cause weakness. In related parts for Exoskeleton section, the Microcontroller, Controller, and Motor subsystem will be continuation from previous project.

Fig. 1 One degree of freedom knee exoskeleton
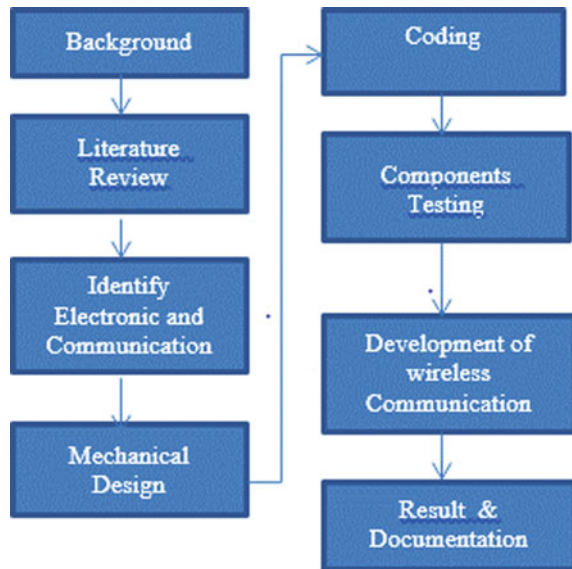
3   Methodology

After a thorough study on several research papers, there are several different types of instrumented cane with different design mechanism in order to control the trajectory of exoskeleton. Two major factors is considered in this project, first the sensor for angle reflective properties and second on the communication medium. Realtime assessments on the motion intension can be establish using rotational attributes computing sensing technology integrated with radio-communication technology in order to control the trajectory of exoskeleton. Wireless communication medium plays an important role as exoskeleton response based on acquired real-time data signal that receive from the cane (Fig. 1).

4   Hardware Implementation

The development of exoskeleton design comprise of instrumented cane which control the knee trajectory of one degree of freedom motion using the exoskeleton. The degree of freedom that the exoskeleton can perform is flexion and extension. The instrumented cane consists of IMU (inertia measurement unit) sensor which is unavoidable and provides best information about body motion. It detects the movement trajectory of push button. It consists of two major specialities, compatible with Arduino and low power consumption. Sensor consist of a single chip embedded with accelerometer, gyroscope and provide information with high accuracy.

The chip uses I2C (inter integrated circuit) protocol for serial communication to function according to various body condition. The weight of sensor is 5 g which is helpful in reducing the weight of the prototype. Arduino nano,interpret information from MPU6050 and sends through I2C protocol to Bluetooth HC05.The board is user friendly and smaller in size. The power requirement is 5 V. The Bluetooth module HC05 act as the transmitter and has the capacity to act as both master and slave according to the preference. In this work it performs as master and can be paired with any slave device automatically. The consumption of power is low and the average current fluctuation range of 25 mA.

The communication range of distance required is 1–3 m and it is applicable for HC05 module. The HC06 bluetooth module in exoskeleton executes as the receiver and performs as slave. It weighs 0.9 g and suitable for wireless communication. The exoskeleton consist of controller boards Arduino UNO and MEGA. UNO act as the intermediate and act as middle device and information from Bluetooth receiver and convert into digital signals and relay it to Arduino MEGA. The coded program in the mega rotates the servo/cylindrical motor. The controller circuit controls the torque and provides the flexion and extension of the exoskeleton.

The design and development of electromechanical system for instrumented cane was done by modifying the mechanical design of a regular cane and instrumented with respecting sensor and microcontroller boards. The main objective is to able to modify an off-the-shelf cane such that it can house position, orientation and force sensor together with wireless microcontrollers to interpret and send data in order to control the exoskeleton trajectory. Based on the design of instrumented cane as illustrates in Fig. 2, additional weight factor on the cane was took into consideration in order to achieve one of the targeted objective. Execution on this project believed to deliverable an instrumented cane that able to sense the motion trajectory either forward or backward by the user and sends the real-time data in order to initiate trajectory of exoskeleton. This instrumented cane will be developing strong spirit to experience "real world walking" especially for those who suffering in Cerebrovascular Accidents (CVA) and Spinal Cord Injury (SPI). Critical consideration on the design specification whereby the instrument cane should be wireless, portable, light in weight as well as meets clinical requirements so that this cane is use comfortably. In conclusion, this instrument cane technology target to assist disability patients as all of the objectives is achievable.

5 Conclusion

IMU sensor was attached to the instrumented cane to detect the movement trajectory that is discharged. The output signal obtained from the sensor will be illuminated using a microcontroller that is compatible with the sensor and wireless communication module. The data will then be mediate to the microcontroller used at the exoskeleton in order to control the trajectory based on, the information is carried via

Fig. 2 Block diagram of instrumentation system
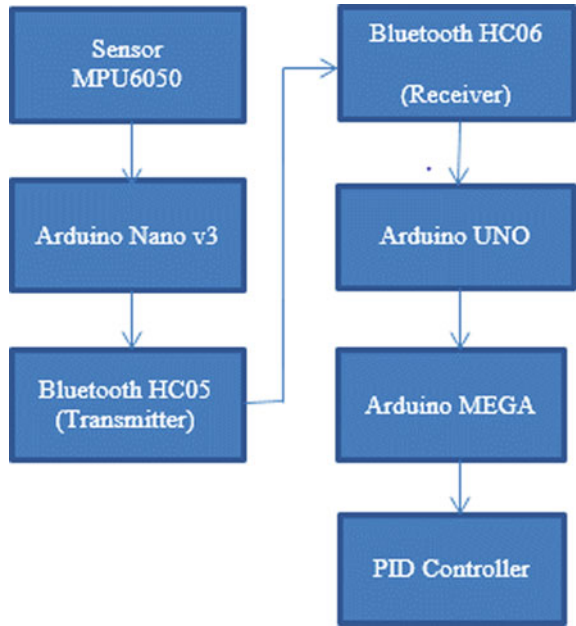
Bluetooth communication sensor. Instrumented cane and exoskeleton enhance the compatibility using Bluetooth communication protocol as the best option (Fig. 3).

Fig. 3 Fabricated of one degree of freedom knee exoskeleton

As the information is conveyed to the exoskeleton at real-time, it begins to initiate the motor to act as actuator in order to make motion. Fundamentally, humans target in motion triggers first, in sequence to command the robotic structure make motion. The push button will be assigned as a safety circumstance if there is any sudden detach or drop of cane from the arm, it will not send any trajectory control data to the Exoskeleton. This prevents the exoskeleton to make any false motion trajectory which could cause weakness. In related parts for Exoskeleton section, the Microcontroller, Controller, and Motor subsystem will be continuation from previous project. Consideration is made on the microcontroller that been used on exoskeleton, is the same microcontroller that will be used for instrumented cane. This is important feature, since to establish successful communication and suitable between two device in terms of control.

References

1. SCI stats and facts. Available http://danceforparalysis.org/sci-stats-and-facts/
2. Ibrahim A, Lee KY, Kanoo LL, Tan CH, Hamid MA, Hamedon NMA et al (2013) Epidemiology of spinal cord injury in hospital Kuala Lumpur. Spine 38:419–424
3. Selvathurai G (2016) Development of instrumented cane for control of exoskeleton of lower limb. Dissertation, 18662
4. Osman M, Tageldeen M (2015) Development of a wearable exoskeleton for arm rehabilitation. Dissertation, 17868
5. Huo W, Mohammed S, Amirat Y (2018) Impedance reduction control of a knee joint human-exoskeleton system. IEEE J 1063_6536
6. Smith BY (2015, Apr 9) What is paraplegia? Available http://www.news-medical.net/health/What-is-Paraplegia.aspx
7. Sadek A-R, Parmar NK, Sadek N-H, Jaiganesh S, Elkhodair S, Jaiganesh T Spontaneous upper limb monoplegia secondary to probable cerebral amyloid angiopathy. Int J Emerg Med 5:1–1
8. Chang WH, Kim Y-H Robot-assisted therapy in stroke rehabilitation. J Stroke 15:174–181
9. Hassan M, Kadone H, Suzuki K, Sankai Y (2013) Wearable gait measurement system with an instrumented cane for exoskeleton control. Sensors (Basel) 14:1705–1722
10. Claire P, Joyce F (2011) An instrumented cane devised for gait rehabilitation and research. J Phys Therapy Educ 25(1):36–41
11. Kyle BH (2017) Design and control of a lower limb exoskeleton emulator for accelerated development of gait exoskeleton. Proquest number 10283413
12. Mahmoodi SN, Hwan-Sik yoon KW, Haskew TA (2015) Design and control of power limb prosthesis. Proquest number 10239909
13. Zetteberg JH (2015) Design and development of two degree of freedom haptic device for neurorehabilitation. Proquest number 10020547. [13] Mussay SA (2016) Development and assessment of a control approach for lower limb exoskeleton for use in gait rehabilitation post stroke, May 2016. Proquest number 101551706
14. Gong C (2016) Development of robotic exoskeleton system for gait rehabilitation. Proquest number 10310874
15. Hao MA (2015) Mechatronic design and control of assistive knee brace for gait rehabilitation, September 2015. Proquest number 10297313
16. Luna CO (2016) Non linear control of a SVEN degree of freedom exoskeleton robot arm, March 21, 2016. Proquest number 10673642
17. Pehlivan AU (2016) Subject adaptive control paradigm for robotic rehabilitation, May 2016. Proquest number 10308368

18. Zhang J (2016) Towards systematic controller design in rehabilitation robot, August 2016. Proquest number 10308368
19. Bing C (2017) Design and implementation of a lower extremity exoskeleton for motion assistance to paraplegic patients
20. Jhawar V (2018) Design of knee exoskeletongait assistance, August 2018. Proquest number 10846724

Semi-Automatic Detection and Measurement of Fetal Parameters from Ultrasound Images and the Scope Automatic System Using LabVIEW

S. Prabakar, K. Porkumaran, J. Samson Isaac, R. Karthikeyan, G. Gopu, and Ramani Kannan

Abstract The idea of developing an automated fetal parameter detection and measurement system using image processing techniques available in LabVIEW is proposed in this research work. The six fetal parameters to be measured from the ultrasound images are crown rump length, biparietal diameter, femur length, head circumference, abdominal circumference and humerus length. We propose the semi-automatic and automatic process using the vision assistant tool in LabVIEW. The semi-automatic process is wherein the parameters are segregated in to six SUB VI's and are programmed into one single MAIN VI. This is then followed by manually dragging the region of interest ROI which could be either circle or line annotation and the values of the result is obtained. The automatic process is being performed only for four parameters namely AC, HC, HL and FL. The trimester fetal scan image is loaded and by applying certain threshold levels and performing convolution as well as morphological operations using different icons such as contour extract, the search direction and the contour selection are used to locate the images from left to right, top to bottom, maximum length, etc., for automation. Here, it selects the appropriate ROI and the values of the result are obtained. The main advantage of this work is that all the conventional methods of fetal anatomical parameter detections are improved with customized LabVIEW and the vision assistant tool. The conventional procedure

S. Prabakar (✉)
Sona College of Technology, Salem, India

K. Porkumaran
Sri Sairam Engineering College, Chennai, India
e-mail: porkumaran@gmail.com

J. Samson Isaac
Biomedical Engineering, Karunya University, Coimbatore, India

R. Karthikeyan
Dr NGP Institute of Technology, Coimbatore, India

G. Gopu
ECE, Sri Ramakrishna Engineering College, Coimbatore, India

R. Kannan
EEE, Universiti Teknologi Petronas, Seri Iskandar, Malaysia

consumes 20–30 min for a single subject scanning whereas in case of these proposed techniques the time taken for automatic detection is less than 5 min based on the system used and the manual intervention also be reduced in diagnostic procedures.

Keywords Fetal parameters · Crown rump length · Biparietal diameter · Femur length · Head circumference · Abdominal circumference · Humerus length

1 Introduction

Accurate fetal ultrasound measurements are one of the most important factors for high quality obstetrics health care. Common fetal ultrasound measurements include Biparietal diameter (BPD), Head circumference (HC), Abdominal circumference (AC), Crown rump length (CRL), Femur length (FL), Humerus length (HL). These measures are used to estimate the getational age(GA) of the fetus and are considered as vital parameters in an important diagnostic tool to identify the viability of the fetus.

In clinical settings, the manual measurement by specialists of BPD, HC, AC, CRL, FL, HL present the following issues such as the quality of measurements are user dependent, the exam can take more than 30 min and specialists can suffer from Repetitive Stress Injury (RSI) due to these lengthy exams. Therefore, the automatic system of these ultrasound measures has the potential of improving productivity and patient throughput, enhancing accuracy and consistency of measurements, and reducing the risk of RSI to specialists.

1.1 Existing Fetal Parameter Measurement

Automatic delineation and robust measurement of fetal anatomical structure in ultrasound images are a challenging task due to the complexity of the object appearance, noise, shadow and quantity of information to be processed. The previous solutions rely on explicit encoding of prior knowledge and formulate the problem as a perceptual grouping task solved through clustering or variational approaches.

These methods are known to be limited in the validity of the underlying assumptions and cannot capture complex structure appearances. Approaches have been used to distinguish between the appearance of the object of interest and background by training a discriminative constrained probabilistic boosting tree classifier. This system is able to perform effective segmentation of fetal abdomens.

A new class of nonlinear filter, the Combinational Non-Linear Mean Median filter (CNLMM) filter as well as image detection using Ada Boost, The MH boosting algorithm and probabilistic boosting tree were also proven efficient.

Table 1 Details of parameters for all trimesters

S. No	Trimester scans	Fetal parameters
1	First trimester	Crown Rump Length
2	Second trimester	BPD, HC, AC, HL and FL
3	Third trimester	BPD, HC, AC, FL an HL

The detailed study of literature in the area of fetal parameters measurement and medical image processing techniques and algorithms, especially the problem associated with the automatic segmentation and feature extraction for clear visualization issues.

The points inferred from results of literature review are:

- Loss of traits and application in outcomes of the measurement systems
- Fetal anatomy measurement requires manual intervention
- Long computing time for various parameter measurements.

This procedure is followed for all the six parameters and the parameters measured during each trimester are enlisted below (Table 1).

2 Proposed Research Work

The proposed system for the measurement of fetal parameters such as BPD, AC, HL, FL, HC and CRL is based on two processes namely semi-automatic and automatic process which is performed with the help of the Image processing criterion in the LabVIEW software known as VISION ASSISTANT.

The semi-automatic process is wherein the parameters are segregated into six SUB VI's and are programmed into one single MAIN VI. The trimester fetal scan image is loaded and is converted into 1000×1000 resolution. This is then followed by manually dragging the region of interest ROI which could be either circle or line annotation or the values of the result is obtained.

The automatic process is being dealt only for four parameters namely AC, HC, HL and FL. The trimester fetal scan image is loaded and by applying certain threshold levels and performing convolution as well as morphological operations using different icons such as contour extract, the search direction and the contour selection are used to locate the images from left to right, top to bottom, maximum length, etc. for automatic process. Here, it selects the appropriate ROI and the values of the result are obtained.

3 Semi-Automatic Fetal Parameter Measurement System

3.1 Process Involved in Semi-Automatic System

The functional blocks of semi-automatic fetal parameter measurement system is shown in Fig. 1. The six parameters are divided into six different SUB-VI's which runs only when that particular parameter is selected. All these SUB-VI's are together linked into a single MAIN-VI. The three trimesters were studied to know which parameter is measured during each trimester. The data is usually an ultrasound image which is in the PNG form and is read and displayed using the IMAQ tool.

The while loop is programmed and allows the user to wait for drawing the ROI and the invoke node is given to the input which stops the loop once the user uses the ROI. It is because the code is done only for the 'DRAW' event. Once them ROI is drawn, the ROI's coordinates (pixel indices) will be sent out through the 'COORDINATE' terminal as an array. If the ROI is not drawn, it will return an array without any elements. Once the ROI is drawn, the array with the coordinates of the ROI will be sent. This stops the loop, if the coordinates array is not empty, which means the ROI is drawn.

The property node is used for the processed input to read and write the properties of reference. This is used to set the properties based on the applications. The ROI is
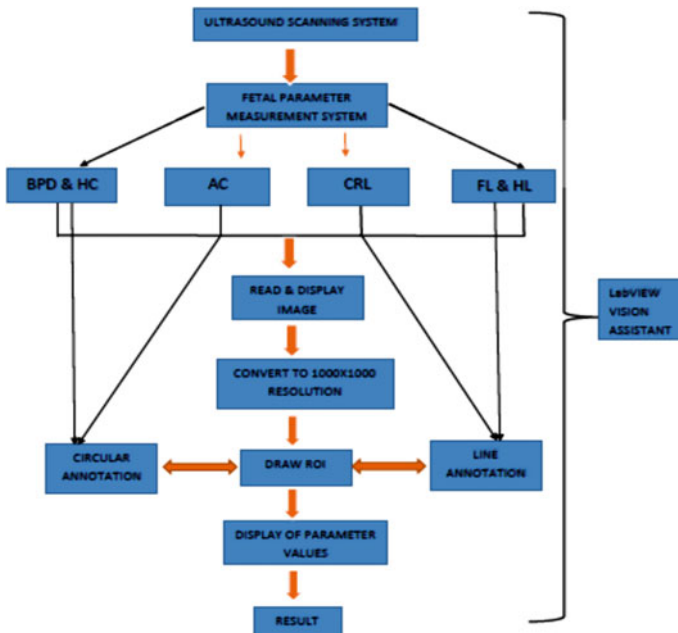
Fig. 1 Block diagram for semi-automatic system

converted into a mask. A mask is nothing but a group of pixels with the programmer specified value in it. Here the value is 255. The image is converted into an array with the array elements corresponding to the pixel's colour value/intensity. The output will be a 2-D array of rows and columns pixel colour values of the input image. Here, it is in the mask or other words, number of pixels in the ROI drawn by the user.

A function called unbundle by name is used. We use the element of global rectangle. The output from the global rectangle will be an array of 4 numeric elements namely; ROI's left, right, top and bottom coordinates respectively. Certain expressions like negation, addition, subtraction, division, etc. are used. The DBL converts a double-precision number to floating point number. The linear fit VI is used This VI uses the iterative general Least Square method and the Levenberg–Marquardt method to fit experimental data to a straight line of the general form described by the following equation:

$$f = ax + b \tag{1}$$

where x is the input sequence X, a is slope, and b is intercept. The expression node is used to calculate the expressions that contain a single variable such as acosh, etc. The calibration factor of 1000×1000 is done and multiplied, the sine of the expression node is then divided to get the result. This is then linked with the main VI.

4 Scope for Automatic Fetal Parameter Measurement System

The trimester fetal scan image is loaded and by applying certain threshold levels and performing convolution as well as morphological operations are used to locate the images from left to right, top to bottom, maximum length, etc., for automatic system. Here, it selects the appropriate ROI and the values of the result are obtained.

4.1 Process Involved in the Automatic System

The image is created using IMAQ CREATE and IMAQ READ and a specific memory is given to the image. This is then converted into an array to work with each pixel. This is enclosed in a nested for loop, in which the outer loop runs according to the number of rows in the pixels of the array and the inner loop runs for the number of columns. A threshold is set inside the loop to convert grayscale image to binary image.

Then the particles are removed from the image for proper segmentation of the appropriate region from the surrounding particles. If there are large numbers of particles it would be difficult to draw the region of interest ROI. The number of

particles or pixels which has been removed from the image is identified with the help of IMAQ PARTICLE ANALYSIS report. A filter of the nth order is used in which the X axis coordinate is -1 and Y axis coordinate is 5. For every value in the X axis, it is multiplied with 5 to define the organ of measurement as they are stronger in densities. This filtering operation is done only for giving intensity for the appropriate region.

Morphological operations such as dilation, erosion, opening and closing are performed in the vision assistant tool of LabVIEW. Dilation is a transformation that produces an image that is as the same shape as the original, but of different size. Dilation stretches or shrinks the original figure. It increases the valleys and enlarges the width of the maximum regions. Erosion is used to reduce the objects in the image and known that erosion reduces the peaks and enlarges the widths of minimum regions. Opening of an image is erosion followed by dilation with the same structuring element and closing operation is the reverse of opening operation.

The forthcoming operation is performed to select the ROI automatically. Hence, we use the IMAQ Extract Contour icon which extracts a single, best contour from an image. If the input image is calibrated, the contour will be learned with calibration.

5 Result of Proposed Semi-Automatic Fetal Measurement System

The main VI is the composition of all the six SUB-VI's. The user or the sonographer can select the type of trimester and perform the semi-automatic measurement. On selecting the appropriate trimester, the parameters that are measured only remains enabled while the other parameters are disabled. The main VI is displayed in Fig. 2.

The ultrasound scan image is loaded in the file path and the window shows two sets of images. The first image shows the original image and the second image shows the conversion of the original image to 1000×1000 resolution as shown in Fig. 3. This is set as the standard resolution for all parameters for greater throughput and accuracy.

In this semi-automatic process, the region of interest is drawn manually either by using circular annotation or line annotation. The region of interest ROI of BPD, HC and AC are drawn using circular annotation while the ROI for CRL, HL and FL are drawn using line annotation. This ROI depends on the calibration factor. The calibration factor for 1 pixel in the 1000×1000 resolution image is calculated as 0.016129 cm. The image is then converted back to array using Build array icon. This is followed by the important mode where the values are then converted from degrees to radians which are followed by other mathematical operations and the required value for the result is obtained. The output of the ROI drawn manually is shown in Fig. 4.

Fig. 2 Main VI for semi-automatic system
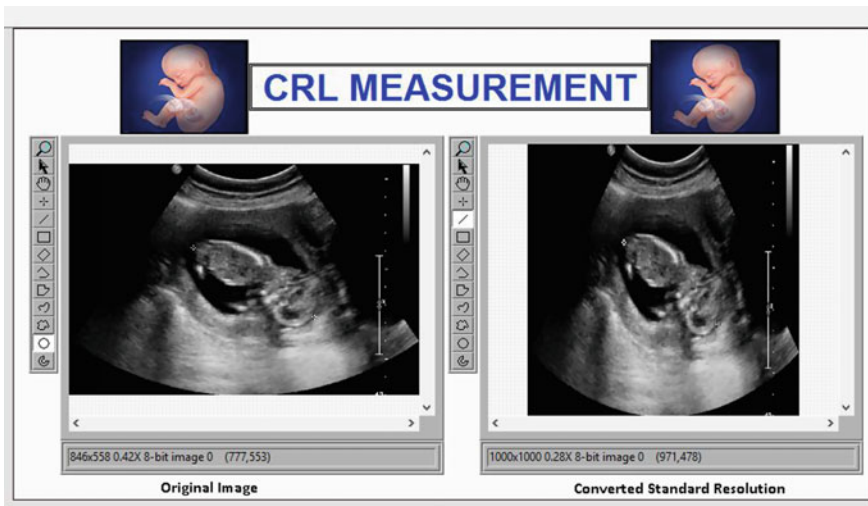
Fig. 3 Conversion of resolution to 1000 × 1000

6 Conclusion

Thus, the fetal parameters are measured manually and are similar to the process carried out by the sonographers and also LabVIEW being highly effective software with very minimal time duration of 5 min per parameter. The manual intervention drawbacks in this work is analyzed and hence to overcome these we proceed with

Fig. 4 Semi-automatic ROI
estimation and measurement
of CRL

the automatic measurement system. The characteristics of different technologies
have been reviewed. This proposed method is an efficient process of measuring the
fetal parameters by semi-automatic system and further automating the measurement
system. The work contains 50 different individuals 120 ultrasound images per person
that cover various weeks of the trimester scans. The subjective appearance of the
output images are found good and the performance of the proposed technique are
good since they marginally overcome the drawbacks and inferences of the existing
ultrasound fetal parameter measurement systems and algorithms.

References

1. Carneiro G, Georgescu B, Good S, Comaniciu D (2008) Detection and measurement of fetal
   anatomies from ultrasound images using a constrained probabilistic boosting tree. IEEE Trans
   Med Imaging 27(9):1342–1355
2. Priestly Shan B, Madheswaran M (2010) A generalized despeckling filter for enhancing fetal
   structures to aid automated obstetric pathologies. Int J Comput Theor Eng 2(3):1793–8201
3. Georgescu B, Good S, Comaniciu D (1983) Automatic fetal measurements in ultrasound using
   constrained probabilistic boosting tree. Siemens Corporate Research, USA
4. Shi Z, Fung KB (1994) A comparision of digital speckle filters. In: Proceeding IEEE on IGARSS,
   vol 4, pp 2129–2133
5. Schluter PJ, Pritchard G, Gill MA (2004) Ultrasonic fetal size measurements in Brisbane,
   Australia, vol 48(4). Department of Radiology, pp 480–486
6. Loughna P, Chitty L, Chudleigh TET (1994) Fetal size and dating: charts recommended for
   clinical obstetric practice. Academic Division of Obstetrics and Gynaecology, Nottingham
   University
7. Jardim Hussain S, Figueiredo M (2005) Segmentation of fetal ultrasound images. Ultrasound
   Med Biol 31(2):243–250
8. Tang S, Chen S-P (1969) A fast automatic recognition and location algorithm for fetal genital
   organs in ultrasound images

A Framework for Enhancing Network Lifetime in Internet of Things Environment Using Clustering Formation

Abdulrahman Aminu Ghali, Rohiza Ahmad, and Hitham Alhussian

Abstract Internet of Things (IoT) is one of the current technology that gains the highest acceptance from various industries and academia nowadays. The growth of this technology has streamlined the day-to-day activities of individuals. It can be stated that day-to-day activities in the industries without the aid of the IoT is becoming harder and harder due to the manual process. This study investigates the issues of energy consumption in the IoT environment due to denial of service (DoS) attack in the IoT environment. Besides, a solution that remedies the issues is proposed in this study. The contributions of this study are categorized into two aspects. Firstly, the study described the components of the IoT, which includes the sensor nodes. Secondly, the study emphasizes in improving the energy of the nodes to enhance the network lifetime in the IoT environment. To achieve the milestone the study was implemented using MATLAB, and thereby considering the enabling technology such as wireless sensor network (WSN) and the RFID in the perception layer of IoT. The results reveal that the network lifetime has improved significantly with about 6% improvements as compared with security-based modified LEACH (MS-LEACH).

Keywords IoT · Nodes · Network · WSN · Sensors · RFID and DoS

1 Introduction

Internet of things (IoT) is a collection of various objects (devices) that are interconnected to perform a specific task without human endeavor. The connected devices share data and information in order to achieve the intended goals in IoT application areas [1]. Besides, IoT has been implemented in many application areas such as oil and gas, transportation, health care, homes, education, and sports with the aim to transform the industries and individual's activities [2].

The widespread of using various technologies such as wireless sensor networks (WSN) and smart devices that are equipped with several sensors, such as radio

A. A. Ghali (✉) · R. Ahmad · H. Alhussian
Computer and Information Sciences Department, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Perak Darul Ridzuan, Malaysia

frequency identification (RFID) and near field communications (NFCs) have contributed immensely in developing the new concept of IoT technology [3].

To understand the concept of IoT technology, sensors are tiny devices that are connected wirelessly and coordinate to perform a specific task. Their small and their ability to support a wide range of applications has rendered the IoT to be one of the emerging wireless networking technology with extensive deployment across the globe [4].

A sensor device consists of batteries, a radio antenna, and a board to accommodate the aforementioned features. The sensor network consists of the ability to identify, process, communicate, in IoT environment [5]. Therefore, users can communicate in the environment using this network. Hence, it is easy to communicate with the physical phenomena with the aid of connected devices and the sensor nodes unlike the traditional ways. The wireless sensor network consists of tens to thousands of nodes. These nodes captured information, process, and forward the information collected to the cluster head (CH) and the CH will aggregate the data and sent to the base station (BS). Additionally, the nodes have specific features such as short duty cycle, power limitations, and battery life.

With all the advantages of IoT, on the other hand, it suffers a setback from DoS attack. Addressing the aforementioned problem and setting up the proper operation of WSN to increase the lifespan of the network and thereby minimize energy consumption is one of the main challenges. In the existing literatures varieties of improvements have been made to minimize the energy consumption in the WSN, and most of these approaches focusing only on the energy consumption improvement and left the DoS attacks un attended which allowed the attacks to be escalated in the IoT perception layer [6]. Therefore, solving the DoS attacks and placing the CH over the grid and to find how many CHs are there in the cluster will improve the efficiency of the network and thereby reduces the energy consumption. To present the method in order to solve the challenges the study proposed a framework using clustering formation to withstands with the challenges.

The study is organized as follows: Sect. 2 describes the application areas of the IoT. Section 3 consists of the methodology. Section 4 discusses the results. Section 5 concluded the paper.

2   Application Area of IoT

The following sections will highlight the application area of IoT and its challenges. These applications areas include transportation, health care, education, homes, sports, and automobile. Table 1 summarizes the security challenges in the IoT environment [6, 7].

Based on Table 1 the security challenges of the IoT application areas are summarized with their security challenges and proposed recommendations. Thus, the security challenges may caused the degradation of the network performance and communication denial between the sensor nodes and CHs. Hence, the study provides

Table 1 Summary of the security challenges in the IoT application area of IoT and their solutions

Application area of IoT	Security challenges	Proposed solutions
Transportation	DoS attack	Authentication mechanism
Health care	DoS and Zingbox	Confidentiality and authentication mechanism
Education	DoS attack	Authentication and availability mechanism
Homes	DoS and Replay attack	Integrity and confidentiality mechanism
Sports	DoS and Snipping attack	Communication security
Automobile	Man in the middle and DoS attack	Mutual authentication and authorization technique
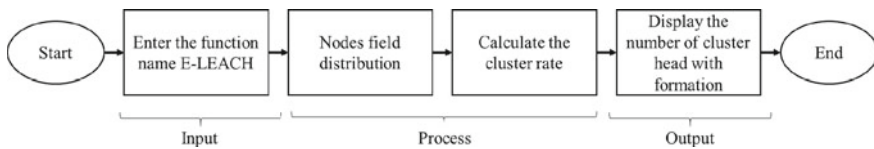
Fig. 1 Clustering formation arrangement

a framework that will improve the efficiency of the network lifetime in the IoT environment.

3 Methodology

In this section the study presents the method on how clustering formation is formed. Therefore, Fig. 1 displays the clustering formation method.

Initially the function name and number of the nodes are required to compute the clustering formation arrangement. Next, the nodes are distributed based on the number of scale design. After the nodes distribution, the CH is assigned. The calculation of the clustering rate is obtained to provide the adequate distribution between the cluster and the CH (For further details see Sect. 5). The output of the formation along with the CHs are displayed.

4 Results and Discussion

Based on the method presented, the result of the proposed method for enhancing the network lifetime in the IoT environment is shown in Table 2. The illustration of the figures and tables are displayed. Table 2 and Fig. 2 shows the nodes distributions.

Table 2 Nodes distributions

No	Items	Description parameter
1	Simulation area	350 × 500
2	No. of points	350
3	Radius meters	50 by 50
4	Field area	X by Y in meters

Fig. 2 Nodes distributions
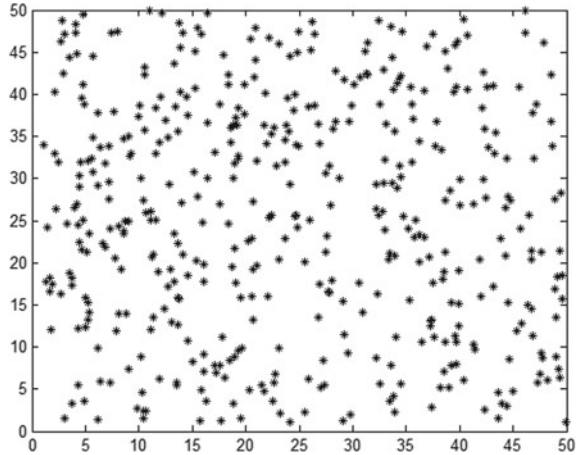
As shown in Fig. 2, the initial field distributions of the nodes are distributed based on 50 m × 50 m. The nodes are randomly distributed in the clusters, before the CHs are selected, the nodes are randomly installed into the scales waiting for the CHs selection. The CHs can be placed at the center or the edge of the field scale but it depends on the nodes selected to be the CH. The insertion of the CH in the center or the edge is convenient to assist the node to finds its transmission range to CH. Table 3 and Fig. 3 shows the advertisement phase.

In Fig. 3 the nodes with the red colors are the two CHs. The CH sent an advertising message to the BS for communication after a few rounds. The nodes usually transfer their energy to CH in the initial data transmission rounds. Thus, the data are aggregated and sent to the BS. Most importantly, the CHs uses their energy faster than non-CH for data transmission. Table 4 and Fig. 4 display clustering formation.

Table 3 Advertisement phase

No	Items	Description parameter
1	Simulation area	350 × 500
2	No. of points	350
3	Radius meters	50 by 50
4	Field area	X by Y in meters
5	Estimated no. of clusters	2

Fig. 3 Advertisement phase
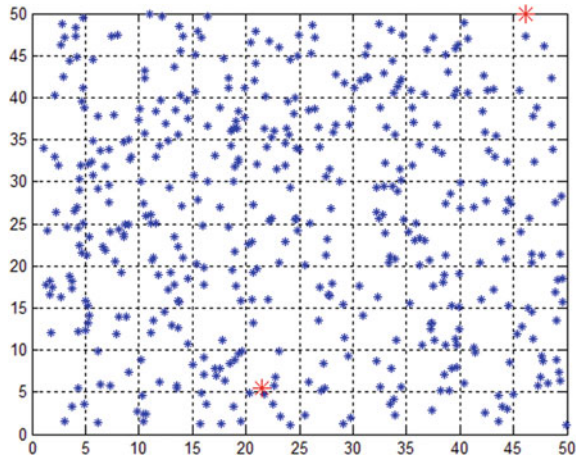
Table 4 Clustering formation

No	Items	Description parameter
1	Simulation area	350 × 500
2	No. of points	350
3	Radius meters	50 by 50
4	Field area	X by Y in meters
5	Estimated no. of clusters	2

Cluster Head 9

Cluster Head 3
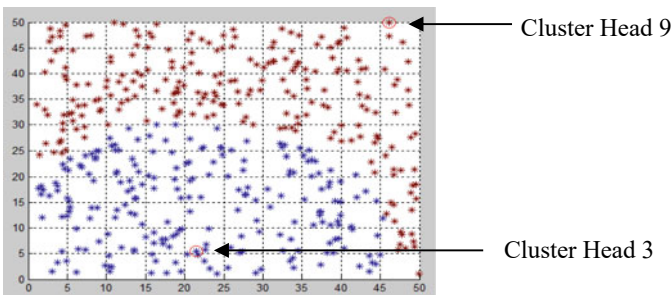
Fig. 4 Clustering formation arrangement

Figure 4 indicates the clustering formation for the nodes in the cluster. In this figure, the formation are divided into two which are brown and blue. The brown color indicates the clustering formation with a circle node representing its cluster group as the CH. Therefore, having the formation in the cluster will reduce the number of energy consumption from the nodes significantly. The more energy consumption is reduce the more network lifetime increased. Figure 5 indicates the energy consumption per nodes in the cluster.
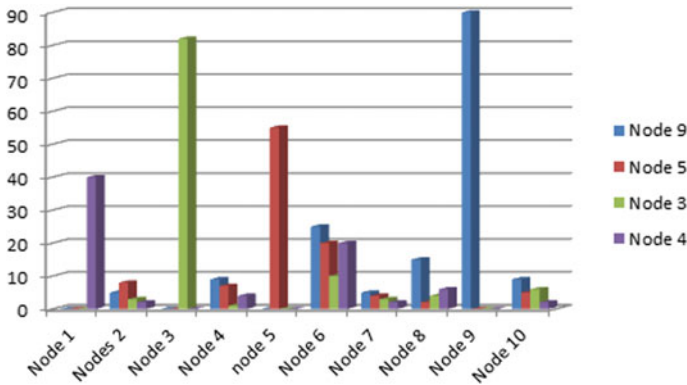
Fig. 5 Energy consumption per node in the cluster

From Fig. 5, it is evident that the energy dissipation by the nodes in the cluster is minimal. Among all the nodes, node nine (9), which served as the CH consumed the highest amount of energy since all the remaining nodes in the cluster transmit their data to the cluster head. The next highest node is (3), which is also the CH. This signifies that the other nodes transmit their data to the CHs and this allow them to have a rest in order to save their energy and by reserving their energy the network performance increase significantly.

5   Conclusion and Future Work

This study investigated the issues of energy consumption in the IoT environment and further provides solutions that remedy the challenges. Besides, the IoT security challenges specifically the DoS attacks can be addressed from the early stage of the IoT implementation. The contributions of this study are categorized into two aspects. Firstly, the study described the initial components of the IoT, which includes the sensor nodes. Secondly, the study provides a solution that improve the network lifetime in the IoT environment using a method clustering formation. The simulation results revealed that network lifetime is increase with 6% improvement as compared MS-LEACH.

References

1. Mahmoud R, Yousuf T, Aloul F, Zualkernan I (2015) Internet of things (IoT) security: current status, challenges and prospective measures. In: 2015 10th international conference for internet technology and secured transactions (ICITST). IEEE, pp 336–341
2. Asghari P, Rahmani AM, Javadi HHS (2019) Internet of Things applications: a systematic review. Comput Netw 148:241–261
3. Dehkordi SA, Farajzadeh K, Rezazadeh J, Farahbakhsh R, Sandrasegaran K, Dehkordi MA (2020) A survey on data aggregation techniques in IoT sensor networks. Wirel Netw 26(2):1243–1263
4. Ghali AA, Ahmad R, Alhussian H (2021) A framework for mitigating DDoS and DOS attacks in IoT environment using hybrid approach. Electronics 10(11): 1282
5. Xia Y, Li W, Clark W, Hart D, Zhuang Q, Zhang Z (2019) Entangled radiofrequency-photonic sensor network. arXiv preprint arXiv:1910.08825
6. Ghali AA, Ahmad R, Alhussian HSA (2020) Comparative analysis of DoS and DDoS attacks in Internet of Things environment. In: 9th computer science on-line conference. Springer, Cham, pp 183–194
7. Kouicem DE, Bouabdallah A, Lakhlef H (2018) Internet of things security: a top-down survey. Comput Netw 141:199–221

An Empirical Analysis on Big Analytics for e-Healthcare and Agriculture

N. Purandhar and S. Ayyasamy

Abstract There is a lot being said and done in the field of data analytics. Using large amounts of data for analytics has become one of the rising trends in the business world but, implementing this business intelligence into different sectors of government hasn't still progressed well. We have discussed two major applications of data analytics in government sectors where the government and eventually the citizens could benefit from all the available big data. The applications include (i) Agriculture, where the big data analytics could result into better crop planning, yield analysis, improved soil health and irrigation as well as reduce the support cost incurred. (ii) The section on data analytics in healthcare mainly points out the importance of predictive analytics in improving personalized healthcare and healthcare infrastructure as a whole. It also talks about how the government can unlock value through big data and machine learning to provide better health insurance than the existing ones and how data analytics is helping with fraud detection while providing the health insurances.

Keywords Big data · Data analytics · Recommendation engine · Predictive analytics

1 Introduction

Executive Vice President of Gartner, which is considered to be the world's important organization in Computer science and Technology Research and Advisory stated, "Information is the oil of the 21st century and analytics is the combustion engine" [1]. We need data analytics in e-governance in order to optimize performance in all the fields that come under the roof of governance. Data analytics uses historical data for providing intelligent solutions in order to make better decisions and strategies for the betterment of the society. Therefore, the field of data analytics is not just relevant but also very important for e-governance. The complete process of Data Analytics insinuates the data collection, extraction of information from that data, transformation

N. Purandhar (✉) · S. Ayyasamy
Department of CSE, Dr. N.G.P. Institute of Technology, Coimbatore, Tamil Nadu, India

of data, analyzing and interpretation of the data and reporting. Monetizing this data, i.e. creation of data to wealth is a challenge as well as an opportunity for the future. For different objectives different types of analytics is to be carried out. For which we need to classify data analytics.

The IBM organization in 2013 categorized analytics into three important ways like descriptive, predictive and prescriptive analytics. To which in 2015 diagnostic analytics was added.

A. **Descriptive Analytics**: Data mining techniques are used along with business intelligence to inquire "What has happened?" That means it analyses the historical data for insights and uses that data to determine the future [2]. Though it provides insights into historical data and trends, this technique is most time-intensive and produces the least value of all the types. Therefore, descriptive analytics alone is not used very often. But, combining it with any other form of analytics can be practical and productive.

B. **Diagnostic Analytics**: It is a form of sophisticated analytics which examines data or information to give solutions for the question "Why did it happen?". It uses historical data with other data to identify patterns and find dependencies [3]. Diagnostic analytics is used for root cause analysis of a problem.

C. **Predictive Analytics**: Statistical models are used to forecast "What could happen?" [2]. It uses the results of descriptive and diagnostic analytics to find out patterns, clusters and outliers to estimate future trends. Predictive analytics is widely used because of its numerous advantages but, it still depends largely on the data quality and therefore it requires continuous optimization.

D. **Prescriptive Analytics**: Methods for optimization and simulations are used to ponder "What should we do?" [2]. This is the action-oriented type and focuses on taking any endeavors possible to deal with the problem at hand. This advanced type of data analytics needs historical data coupled with information from external sources for applying statistical algorithms. Other than that, prescriptive analytics uses machine learning as well as business intelligence, which makes it expensive to implement and manage [4].

2 Big Data and e-Governance

On Feb 1st 2018, while delivering his speech on the Union Budget, the Finance Minister told the parliament that the government's think tank—Niti Aayog, would initiate a national program helping the government to focus on the development of cutting edge technologies in digital space namely Big Data, Artificial Intelligence (AI), Machine Learning, Internet of Things (IoT), 3D printing and others. Further accentuating to invest in research, training and skill development in robotics, AI, manufacturing, big data analysis and IoT, he announced to double the budget for Digital India Programme to 3073 crores in 2018–19 [5]. This shows that the government of India is keen on using the technological advancements in different fields for good.

2.1 Big Data in Agriculture

Farming is the cornerstone of India's wealth and this agriculture is responsible for 18% of India's gross domestic product (GDP) and facilitates employment to more than 50% of the countries' manpower but, because of rapid development of the industrial and service sectors the agricultural industry has declined to a great extent. For improving the yield of small farmers, the government of India had an agreement with IBM to develop an Artificial Intelligence based model for crop yield prediction [6]. When IBM will employ Artificial Intelligence for developing the technological model for improvement of agricultural output and productivity of various types of crops and soil species for the specified districts, NITI will use this data insights generated through these AI models to help farmers and other stakeholders [7]. Also, by providing soil health cards the government has empowered the farmers to better analyze the requirements of the water for irrigation and the crop to be sown.

To add value chain and increase profitability of the agricultural industry the government is keen on employing technology. Also, for creating opportunities in the field, the Ministry of Agriculture and Farmers "Welfare had a deal signed with the Indian Space Research organization (ISRO) to use satellites for geotagging the agricultural assets [8]. The big data acquired through such agricultural infrastructure will help develop vigorous responses for the issues pertaining to climate change, pest and disease outbreaks, and land degradation. Applying these "big data" approaches by the government for agriculture guarantees to search out new ways to build a food-secure society.

Big data is seen as a prospective technology for selection of appropriate agricultural inputs but, acquiring data is the biggest challenge. For any agribusiness increasing the profitability means increased yield and decreased support costs. Big data aids in enhanced yield analysis. Increased yield and decreased support costs play principal role to gaining profitability and enhanced customer experience for any small or large agribusiness. Big data analytics can enable improved analysis of crop yield and other crucial methods for a profitable and in-depth root-cause analysis resulting in actions for enhanced crop quality and reduced overall cost [9]. The data associated with turnover, utilization of available resources and the effectiveness of equipment, can be combined with further analysis for improved quality.

Predictive analytics may be used to foresee the demand for seeds, requirement of fertilizers and animal feed needs, for empowering the agribusiness supplier to take appropriate steps to match production to demand. Table 1 shows different fields of agriculture and how big data is used for their management.

Table 1 Big data approaches in agriculture

Focus area	Ref	Proposed techniques/solutions	Possible improvements
Soil health	[10]	Geocoded soil data is generated using soil genome sequencing and this data is characterized based on microbial diversity, further it is combined with AI to estimate the soil health combining the ground level and above ground features to describe soil microbial diversity	Soil health cards are provided by the government of India to all the farmers. They provide a comprehensive report of the soil health [11]. The data collected through the soil health cards could be modeled with a ML algorithm to identify steps to improve soil quality
	[12]	The soil data is compiled based on different attributes such as fertility, pH, Nitrogen, phosphorous, potassium contents and others. Chi-square Map Reduce method is used for feature selection. An Advanced Decision Tree classifier is suggested for pruning methods to build a tree	
	[13]	Information gain, Gain Ratio and Chi Square are calculated for feature selection of the soil contents using Map Reduce	
Crop planning	[14]	Parameters like year and crop type are used in the clustering algorithms along with special attributes such as time or crop type to identify which crop is to be sown	Farmers will be provided with soil health records, they can study the soil management techniques and accordingly, plan the upcoming crops [11]
Productive irrigation		The state of Andhra Pradesh is using the satellite data for predictive analytics to forecast water-stressed farms and based on the crop type use rain guns as a productive irrigation technique	This model could be implemented at various places in India where the crops are needed to be rain-fed. Also, innovative methods for irrigation are needed to be developed

(continued)

Table 1 (continued)

Focus area	Ref	Proposed techniques/solutions	Possible improvements
Reduced support cost		Many small and marginal farmers are utilizing M-trading platforms to sell their farm produces right away to businesses across the country through smart phones	To reduce cost of entire agriculture value-chain the farmers should be enabled with data centric methods for improving the efficiency of supply-chain and the marketplace of agrarian items
Yield analysis	[15] [16] [17]	Different data mining techniques like linear and multiple regression, clustering, decision tree and k-nearest neighbors have been used with the big datasets for finding the yield of the crop	With the soil health card, the farmers can know the nutrient contents of the soil better and hence, get an idea of which crops to invest in and what kind of fertilizers they need ultimately, increasing the crop yield [11]. IBM's AI based model for crop yield prediction is also in use by small farmers
Crop protection and weed control	[18]	Big data as well as machine learning approaches have been suggested for modeling and prediction of herbicide/pesticide resistance, detection of non-native species plants and flowers, decision support system for crop protection and using AI based systems for plant and flower control	There are no crop protection initiatives by the government available as yet

2.2 Big Data in Healthcare

A real-time alerting system for instant care through wearable devices is fast growing trend in the first world countries [19]. The world is rapidly moving from traditional medical practices to wearable medical devices but, India is still struggling for providing primary healthcare to majority of population. Despite of the government recommendations for standards of Electronic Medical Records (EMR), there are no concrete steps taken for centralizing EMRs. The Ministry of Health and Family Welfare suggests a standard for structured record fields to be inserted into the database viz., Name of the patient, DoB, Residential addresses, Doctor's name, Hospital name and addresses, and medical treatment reimbursement codes for easier management through a database system but, the lack of acceptance to field-code the data and use of EMRs by the medical practitioners and caregivers is a major barrier [20].

Moreover, the traditional databases are inadequate to handle the large chunks of data being produced every day. There are three dimensions to the role of government in Healthcare in India.

2.3 Using Big Data and Predictive Analytics to Improve Health Care by Selecting the Best Treatment Plans and Improving Diagnostics for Personalized Healthcare

The major shortcoming of the government is its lack of insight for providing development in personalized healthcare sector. The main focus needs to be on research and development models for epidemiology research clubbed with big data. Big data can help to identify problem areas where there is a need for improvement in personalized healthcare. It can also be useful for statistically analyzing the course of treatment employing the historical data. Whether it be biomedical sciences or pharmaceutical researches data is everywhere. One-way Predictive analytics can help hospitals is anticipate the number of incoming patients at different times during the day, week, month and year. This can solve the difficulty caused by over-staffing or under-staffing in hospitals because in the first scenario the cost of labor increases and in the latter the patient services might be hampered. The other area where the predictive analytics can help is by using data for identifying the cause of different diseases through available data.

2.4 Using Big Data and Predictive Analytics to Improve Health Care Infrastructure

The annual National Health Profile document; a publication of Ministry of Health and Family welfare has six indicators for measuring the fundamental health information. They are Demographic, Socio-Economic, Health Status, Health Finance, Health Infrastructure and Human Resources [21]. Furthermore, the measures for Health Infrastructure are categorized as educational infrastructure and service infrastructure. The educational infrastructure lists out number of medical colleges, students enrolled for diploma, undergraduate and postgraduate medical courses, as well as students enrolled for dental schools, AYUSH Institutes, nursing courses and paramedical courses. Health information indicators for Service infrastructure include details of allopathic hospitals, number of beds in a hospital, Ayurvedic and Homeopathy hospitals, sub centers, primary and community health centers, blood banks, mental healthcare hospitals and eye banks [21]. According to CRISIL"s opinion report, one subcentre for the population of 5000 in plain areas and for 3000 population in desert, tribal and hilly areas is a norm. But, there is still a gap between the planned structure and the present structure of the health and wellness sub-centers. Big data is central

to achieve this goal, by statistically measuring the population density, understanding the demographics and thus making a predictive analysis about where a sub-centre can be setup in a cost effective, efficient and impactful manner. The analysis will also need to focus on the density of public and private hospitals in urban, sub-urban and rural areas.

2.5 Using Big Data for Provisioning Health Insurances

Providing healthcare at affordable prices is one of the main objectives of the government in India. Under the initiative of Ayushman Bharat, the government of India is firm on contributing to better healthcare for all. The Pradhan Mantri Jan Arogya Yojana under Ayushman Bharat aims to cater to more than 50 crore beneficiaries. It covers not less than 10.74 crore underprivileged families from rural areas and workers' families from urban and suburban parts of the country as per 2011 Socio-Economic Caste Census (SECC) data [23]. It has a provision of health cover of rupees 5 lakhs for secondary and tertiary healthcare services. This scheme has a lot to offer to private hospitals, insurance companies as well as the main stakeholders—the beneficiaries. According to media reports of 13th Feb, 2019, there have been discrepancies in some of the procedures, and a few hospitals receiving a greater number of patients than the other. Therefore, the government has employed analytics methods to ensure Ayushman Bharat remains without any scams. The analytics is also necessary for detecting fraudulent behavior as the Aadhar card—the 12-digit unique identity number were not mandatory for availing the services and any identity cards were accepted in the initial stages. Utilizing the research evidence Recommendation engines can be built based on collaborative filtering for creating healthcare plans for patients with similar symptoms. These recommendation engines use best-fit classification algorithms to classify patients based on different parameters like, age, disease type, symptoms and seasonal changes among many. This would help to build a better prediction model for precision medicine. The most commonly used strategies for predictive analytics include Bayesian networks, Decision Tree learning, Markov and Monte Carlo Simulations [24]. In the paper [25], the authors suggested a combination of Principal Component Analysis and Neural Network for sentiment analysis. They had promising results when checked with different performance parameters. A feedback mechanism can be developed, where the patient or the caregivers can provide feedback on different parameters relating to the overall experience at the hospital. A mechanism can be implemented by applying sentiment analysis on the feedbacks, where (1) this analysis can be provided to the recommender systems for personalized healthcare plans, (2) the hospital can utilize the data to improve the infrastructure and (3) it can help to improve the efficiency of the nursing team by providing useful insights.

3   Conclusion

The applications discussed in this paper specify the potential of data analytics in their respective fields and the possible solutions using different methods of data analytics. It presents the sources from which the data is acquired for various applications, and how this data can help to improve the citizen services provided by the government. In Agriculture domain the government does the soil analysis of each farm and gives suggestions to the farmers for crop planning, soil health and yield analysis. But this is not enough for entire growth of the agrarian market and therefore, the paper discusses different big data analytics methods and machine learning algorithms for an inclusive agri-business setup. Healthcare is the sector where data analytics could do wonders with personalized medical plans, for infrastructure development and patient service improvement. Recommendation engines could be built for personalized medicine based on EMR's. But, first there is a need to develop and deploy centralized electronic health records systems. Once these systems are in place the data analysis could be done to increase efficiency and save lives.

References

1. Sondergaard P (2019) Information is the oil of the 21st century, and analytics is the combustion engine. Gartner Research, Datasciencecentral.com
2. Delen D, Demirkan H (2013) Data, information and analytics as services. Decis Supp Syst 55(1):359–363. Available https://doi.org/10.1016/j.dss.2012.05.044
3. Fleckenstein M, Fellows L (2018) Modern data strategy, 1st ed. Springer, Cham, Switzerland, p 133. ISBN 9783319689937
4. 4 types of data analytics to improve decision-making. Scnsoft.com, 2019. [Online]. Available https://www.scnsoft.com/blog/4-types-of-data-analytics
5. Ministry of Finance (2018) Budget speech 2018–19
6. Misal D (2019) How big data is the game changer for indian government in E-governance. Analytics India Magazine
7. D'Monte L (2019) To improve weather forecasting for farmers in India, IBM is relying on AI. https://www.livemint.com
8. Deoras S (2019) Top use cases where modi government used big data, AI for reform. Analytics India Magazine
9. (2016) Improving manufacturing performance with big data. Oracle Enterprise Architecture White Paper
10. Lesueur D, Burra D, Bui D, Nguyen M, Zhong D (2019) Geocoded soil data and AI to provide an estimate and indicator for soil health. CGIAR Platform for Big Data in Agriculture
11. Soil Health Card|National Portal of India. India.gov.in, 2015
12. Rajeswari S, Suthendran K (2019) C5.0: advanced decision tree (ADT) classification model for agricultural data analysis on cloud. Comput Electron Agric 156:530–539
13. Rajeswari S, Suthendran K (2018) Chi-square mapreduce model for agricultural data. J Cyber Sec Mob 7(1):13–24
14. Sekhar C, Kumar J, Kumar B, Sekhar C (2018) Effective use of big data analytics in crop planning to increase agriculture production in India. Int J Adv Sci Technol 113:31–40
15. Sellam V, Poovammal E (2016) Prediction of crop yield using regression analysis. Indian J Sci Technol 9(38)

16. Fan W, Chong C, Xiaoling G, Hua Y (2015) Prediction of crop yield using Big Data. In: 8th international symposium on computational intelligence and design
17. Rani S (2017) The impact of data analytics in crop management based on weather conditions. Int J Eng Technol Sci Res 4(5):299–308
18. Ip R, Ang L, Seng K, Broster J, Pratley J (2018) Big data and machine learning for crop protection. Comput Electron Agric 151:376–383
19. Firouzi F et al (2018) Internet-of-Things and big data for smarter healthcare: from device to architecture, applications and analytics. Futur Gener Comput Syst 78:583–586
20. Raghupathi W, Raghupathi V (2014) Big data analytics in healthcare: promise and potential. Health Inform Sci Syst 2(1)
21. Central Bureau of Health Intelligence (2018) National health profile—2018. Ministry of Health and Family Welfare, Government of India, New Delhi
22. CRISIL Opinion, Bharat A (2018) Improvement in quality of government infrastructure and leveraging private sector at right price to be crucial tasks for the scheme. CRISIL
23. Ministry of Family and Welfare, PIB Delhi (2019) Beneficiaries of Ayushman Bharat Yojana
24. Janke A, Overbeek D, Kocher K, Levy P (2016) Exploring the potential of predictive analytics and big data in emergency care. Ann Emerg Med 67(2):227–236
25. Vinodhini G, Chandrasekaran RM (2014) Sentiment classification using principal component analysis based neural network model. In: International conference on information communication and embedded systems (ICICES2014). https://doi.org/10.1109/icices.2014.7033961

Enhanced Approach in VANETs for Avoidance of Collision with Reinforcement Learning Strategy

A. Ganesh and S. Ayyasamy

Abstract   VANET (Vehicular Ad Hoc Network) is the most prominent technology that helps in the enhancement of performance and safety of the transportation system. The ad hoc network established between the moving vehicles on the road defines the VANET. VANET is distinguished by the mobility and self-organization of nodes, vehicles are considered as nodes here. V2V (Vehicle-to-Vehicle), V2I (Vehicle-to-Infrastructure) communications use VANET to warn vehicles nearby about emergencies (i.e., traffic jams, accidents), usually of the order of a few meters. Latency, secure transmission of messages, high-reliability criteria are the key challenges that one faces when implementing VANET. The primary goal of this paper is to predict the collision from the sensor readings and prevent the happening of collision and V2V communication using ESP-NOW protocol. ESP-NOW helps in the transmission of wireless messages between the vehicle nodes which are controlled by ESP32 microcontroller. Reinforcement Learning (RL) with deep policy gradient networks is chosen here. Policy gradient (PG) methods are widely used algorithms for RL. Using Deep Reinforcement Learning (DRL) on the ESP32, the car node maintains a safe distance from the nearest obstacle. ESP-NOW protocol helps in exchanging the information between vehicle nodes, by finding all the MAC addresses of vehicle nodes that are present in its range and communicates the information. The focus of the proposed system is on taking the appropriate action using the DRL algorithm to avoid the collision risk and enables V2V communication which helps to broadcast the vehicle's information to the other vehicles that are present in the V2V protocol range to maintain collision-free environment.

Keywords MANET · VANET · Human machine interface · Deep reinforcement learning · ESP32 microcontroller

A. Ganesh (✉) · S. Ayyasamy
Department of CSE, Dr. N.G.P. Institute of Technology, Coimbatore, Tamil Nadu, India

1 Introduction

VANET is essentially a subset of the Mobile Ad Hoc Network (MANET) [1–3] established to create a wireless communication system for vehicles. Vehicles now-a-days are equipped with intelligent and automated systems to work on real-time safety requirements. VANET's comprises of V2V, V2I, Vehicle-to-Grid (V2G) and Vehicle-to-Anything (V2X) modes of communication. Vehicles in VANET are integrated with OBU (On-Board-Unit), RSU (Road-side-Unit). OBU and RSU share information through the wireless channel. OBU deals with the data transmission from vehicle to vehicle, whereas RSU communicates the information from vehicles to road-side network. The OBU talks to RSU and with other OBU via DSRC (Dedicated-Short-Communication-Range). DSRC enables high speed, low latency V2V, and V2I communications, using the IEEE 802.11p and WAVE (Wireless Access in Vehicular Environment) standards. ITS (Intelligent Transportation system) is achieved by employing VANET. ITS facilitates creative traffic management capabilities through the use of various smart modes of communication and networks. Safe and secure message transmission is the main challenge VANET face to have a smooth and safe transportation system [4–6]. High vehicle mobility causes a wide range of problems, such as the disruption of the interaction between vehicles, in which vehicles fail to establish safe communication between them. It makes difficult to transmit a warning message to other devices in prior time before the deadline. Latency plays a key role in the communication link in sharing alert messages or various physical parameters. Minimum latency is the crucial requirement in providing the collision-free environment in VANETs. Different sensors are integrated into the vehicle to measure various physical parameters like acceleration, speed, and distance from the nearby vehicle. These parameters are displayed on the Human Machine Interface (HMI) device which is already connected with the vehicle [13]. Communication of alert/warning messages, certain physical parameters helps to avoid wrong actions of the driver which in turn prevents the occurrence of collisions.

The primary objective of VANET is to provide safety applications. In this paper, V2V communication is shown between the vehicle nodes which are equipped with ultrasonic sensor and LM393 speed sensor which gives two physical parameters i.e. distance and speed. The wireless communication is done by employing ESP-NOW protocol. Two-way communication is chosen here with the help of ESP-NOW protocol. Training of the DRL model on the ESP32 microcontroller is done here to avoid the nearest obstacle in the traveling path of a vehicle node [7, 8]. DRL involves developing a deep learning model that enables function comparison between input features and expected discounted reward values often referred to as Q values. Q-learning is a model-free reinforcement learning algorithm to practice policy by asking an agent what actions to take in what situations. The PG algorithm is used, which is one of the RL family. PG network typically uses a neural network. The network here consists of 1 input layer, it carries the distance to the nearest obstacle measured by the ultrasonic sensor, the output layer corresponds to 5 possible directions at which the car node can travel.

Numerous vehicle applications include efficient traffic control, active road safety, etc. The main goal is to eliminate car accidents and to provide a collision-free, driver-safe environment by sharing information on the danger of collisions and obstacles. To extend the driver's perception by allowing him/her to respond much more quickly is the fundamental idea. This is possible through the wireless communication of alert/warning messages. And also, the traffic jam situations are avoided by maximizing the movement of vehicles. Discovering the nearby vehicles and sharing the obstacle/hazard information at low latency is the best technique to maintain a collision-risk free environment.

2 Sensor Modules Used

2.1 Ultrasonic Sensor

An Ultrasonic sensor measures the distance between sensor and object. It has transmitter and receiver, in which transmitter sends an ultrasonic wave, if the wave comes in to contact with any object, the wave gets reflected back to the sensor receiver. It's measuring range is 2–400 cm with the accuracy of 3 mm. Distance is measured using the formula:

Distance (cm) = Speed * Time Speed = ultrasonic wave speed (**0.034 cm/μs**)
Time = time traveled by the wave to and from the object (consider half of the time, **time** (μs)/2)

2.2 LM393 Speed Sensor

Figure 1 is the IR (Infrared) based LM393 speed measuring sensor. This is used to measure the wheel rotation speed of a vehicle, the two vertical columns consist of IR

Fig. 1 LM393 Speed Measuring Sensor
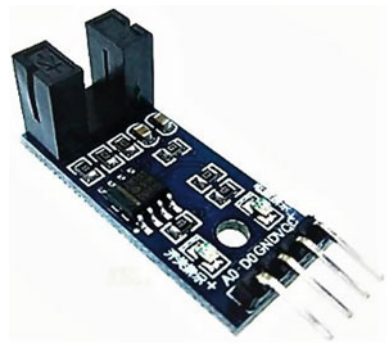
LED on one side and phototransistor on the other side. The speed is determined by wheel encoder which passes in between the two vertical columns. RPM (Revolutions per Minute) defines the completion of all the 20 slots passage between the two vertical column slots. This gives the RPM.

From RPM, the speed (cm/s) is measured by: (2/r * rpm)/60, where r (cm) = Radius of the wheel.

3 ESP-NOW Protocol

ESP-NOW protocol provides a fast, connectionless Wi-Fi communication involving short packet transmission. ESP-NOW is perfect for remote control systems, sensors, and other applications. There is a vendor-specific action frame in which the message is embodied and gets transmitted from one device to another device, the transmission is connectionless. For the information exchange between devices and to deliver a broadcast message, adding a peer with the broadcast address is done. Local device and peer device details are preserved in the low-level layer of ESP-NOW. Sending and receiving call back functions informs us about the success/failure of packet transmissions in broadcast/multicast communication.

4 Reinforcement Learning

RL consists of an agent, environment, action, reward, state [9, 10]. The agent performs the action in an environment which results in the state transition from s to s´. s, s´ is the current state and next state respectively. Interaction of the agent with the environment with certain actions leads to maximum reward. The combination of artificial neural networks and RL architecture enables defined agents to determine the best possible actions in a simulated environment to achieve their goals defines the DRL. In other words, it combines function approximation and target optimization, mapping state-action pairs to the predicted rewards.

Figure 2 shows the overview of the RL architecture for this application. The main task of the agent here is to perform an action, so that vehicle is prevented from the collision. The sensor readings are observed by the agent to get an idea about the environment state. After observations from the environment, the agent has to generate correct actions like moving forward/backward, left/right directions, or stopping the vehicle, this prevents the vehicle involving in a collision/accident. The agent learns a policy, which helps in the generation of correct actions from the observations. The repeated attempts of the agent to have the collision-free environment maximize the reward. Depending on the observations, actions, and rewards received, the training algorithm adjusts the policy of the agent. After training, the vehicle will be able to travel without involving in any collisions using tuned policy and sensor readings.
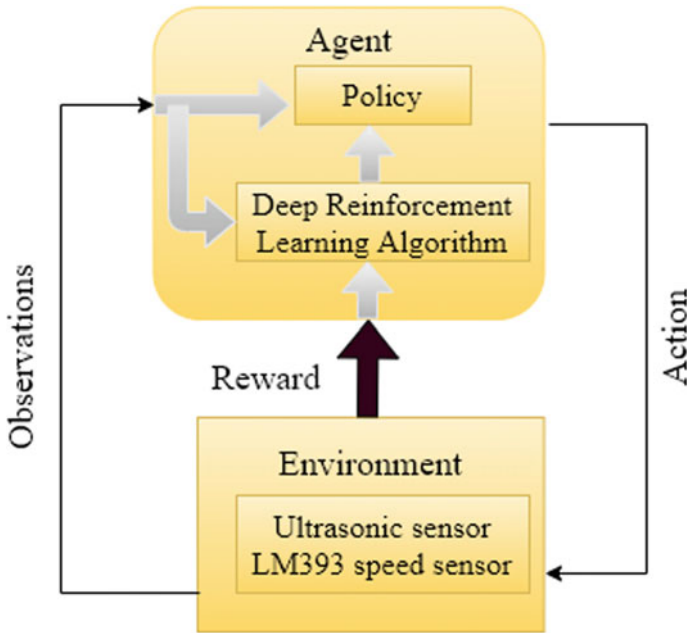
Fig. 2 Depiction of reinforcement learning

4.1   Q-Learning

Q in Q-learning indicates quality. Quality here defines how effectively an action helps in gaining some reward in the future. Mapping of state-action pairs to rewards is done by Q. For each state-action pair, a Q-value is present in Q-table. Q-learning employs Temporal Differences (TD) to determine the value of $Q^\pi(s, a)$. π is the policy, it maps the states to actions, where the actions guarantee the maximized reward. s and a represents the state and action respectively.

Figure 3 depicts the Q-learning algorithm. Q-table has state-action pairs in which each action is chosen corresponding to the current state. The reward is calculated for each action and Q-table is updated until the maximum reward is attained.

4.2   Policy Gradient

PG methods are based on the principle "Observe and act" [11, 12]. A policy is trained here, training makes actions with high rewards more probable or vice versa. PG holds what is working and dumps what is not. It focuses on optimizing the parameterized policies with respect to the anticipated cumulative reward by gradient descent. In this, the agent analyzes the scenario, and instinctively understands what to do with

Fig. 3 Q-learning flow chart
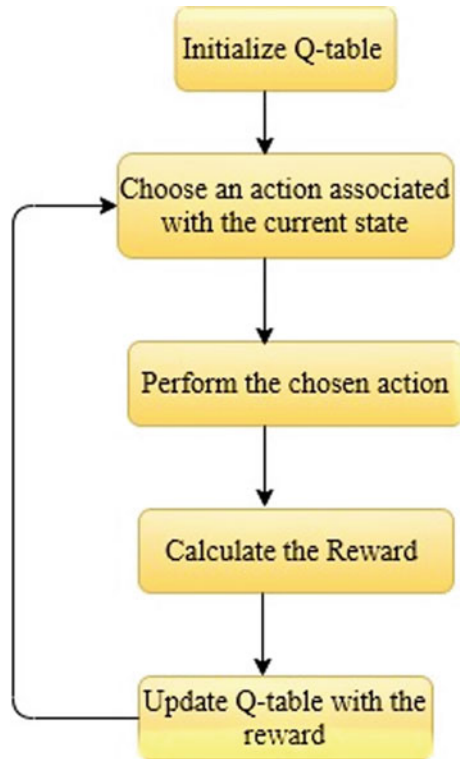
it. The more time of training enables the policy to maximize the rewards. PG aims to effectively optimize policy space. The neural network is used in PG to model the probability of action directly. When the agent interacts with the environment, we fine-tune the neural network parameters so that "good" actions are more likely to be sampled in the future. This process is repeated until the policy network converges towards an optimum policy.

5 Proposed Collision Avoidance Framework

The collision avoidance with and without DRL is conveyed here. The implementation starts with the collection of sensor readings from the chosen ultrasonic sensor and LM393 speed sensor. The sensors are mounted on the two-wheeler chassis which has 2 DC motors got fixed for the movement of chassis. L298N motor driver is used which enables to control speed and direction of 2DC motors. All these components are connected to the ESP32 microcontroller and this setup is taken as a vehicle node which is shown in Fig. 4, it is controlled by the code written into microcontroller using the Arduino IDE platform.
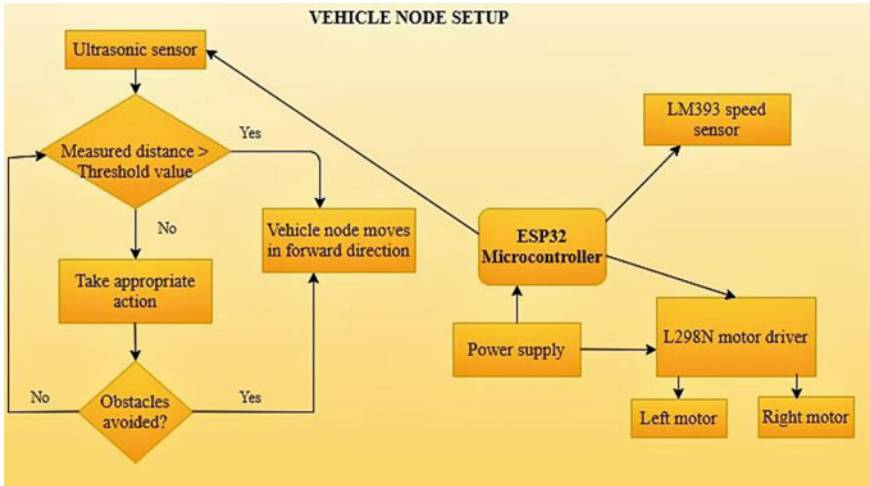
Fig. 4 Vehicle node setup for collision avoidance
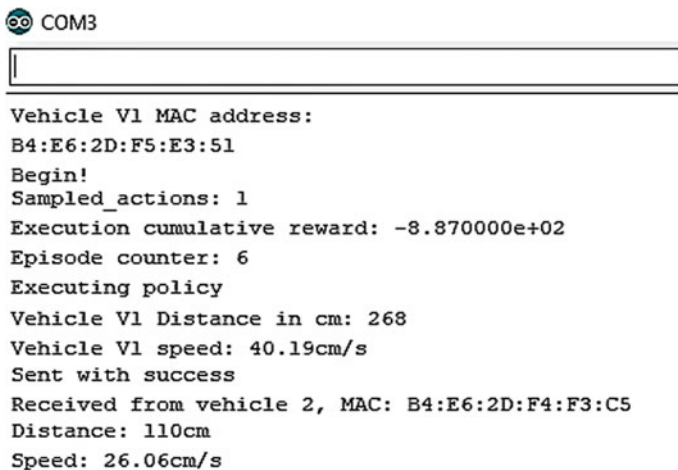
First, let's see how the simple collision avoidance code performs without any deep learning technique. For this method, we use a servo motor on which the ultrasonic sensor is mounted to rotate the ultrasonic sensor in the left and right directions from the center. A threshold distance value is set for the ultrasonic sensor, if ultrasonic measured distance > threshold value (i.e. obstacle is at safe distance from the sensor) vehicle node moves in forward direction. Otherwise (i.e., measured value $\leq$ threshold value) vehicle moves backward and checks the distance between sensor and obstacle in the left and right directions. The sensor compares the distance in both left and right directions and then the vehicle moves in the left or right direction which has more space to move.

Collision avoidance with DRL starts with designing of Q-learning framework by having state-action pairs, discrete states represent the distance range varying from 0 to 400 cms, actions are forward, backward, turn right and left, halting the vehicle. In this DRL method, servo motor is not used, the ultrasonic sensor is set to a fixed direction. Training is done in episodes. The PG network consists of an input layer of dimension 1×1. Distance to the nearest obstacle is taken as input, the hidden layer of 3×1 dimension with RELU (Rectified Linear Unit) activation, and the output layer of dimension 5×1, with soft max activation, corresponding to 5 possible directions at which the vehicle node can travel. PG network samples the 5 actions, performs them, stores, and updates the gained reward. This sampling process is encapsulated as the policy π, which is responsible for the implementation of action dependent on past environmental states $\{s\}$. PG agent here specifically determines an optimal strategy that enhances long-term rewards. By using both Q-learning and policy gradient, Q-values are determined from the policy preferred actions to which we apply Q-learning updates.

One can see the episode number, the number of actions sampled in the serial monitor of a particular vehicle node and this Deep reinforcement algorithm helps in maintaining the safe distance from the nearest obstacle. The vehicle node takes appropriate action based on the training when it finds an obstacle. In both with and without DRL algorithm methods, V2V communication is the same. It includes the transmission of vehicle's distance from the obstacle, speed of the vehicle which is sensed by the LM393 speed sensor and, also MAC address of the transmitted vehicle. These parameters are displayed on the serial monitor of all peer devices which are present in the ESP-NOW protocol range. ESP-NOW is responsible for V2V communication.

6 Results and Discussions

Figures 5 and 6 show the vehicle nodes V1 an V2 information display on arduino serial monitor which depicts the V2V communication. COM3 and COM10 show the display of V1 and V2 nodes respectively. COM3 showing the MAC address of the vehicle, the distance, and speed of Vehicle V1, the parameters of Vehicle V2 which are getting received from the Vehicle node V2 and vice-versa with the COM10. We chose to broadcast the vehicle's distance from the obstacle and speed to other vehicles that are present in the protocol range. With an open field range of up to 220 m, the exchange of sensor readings happens between two ESP32 boards. ESP-NOW holds its peers MAC addresses to broadcast the message. Every ESP32 has its own unique MAC address so that one can recognize every board from which the data is sent and received using ESP-NOW. Two-way communication is seen in our results, i.e., the

Fig. 5 Vehicle V1 displaying it's parameters and Receivingv2 parameters through ESP-NOW protocol

Fig. 6 Vehicle V2 displaying it's parameters and receiving V1 parameters through ESP-NOW protocol

vehicle V1 information is sent to the vehicle V2, and V2 information is received to the serial port COM3, vice-versa with the COM10.

The sampling of actions, reward, and episode counter are the terms of the DRL algorithm. As the training process is achieved in episodes. The episode counter is set, it increments until it reaches the maximum training value, after that counter resets to zero. Once the vehicle starts sensing the surrounding environment, if an obstacle is detected, the algorithm samples all the actions (i.e. forward, backward, stop, turn left, turn right), and optimized action which gets the highest reward is performed. The goal is to find an effective action that maximizes the total cumulative reward (sum of all the rewards received up to this point) of the agent that gets after execution in every episode.

The chosen algorithm makes the vehicle nodes to travel by having minimum distance from the obstacle. And ESP-NOW protocol enables in sending and receiving the information of vehicle nodes which tells the V2V communication. V2V communication and chosen algorithm allow the vehicle to maintain a safe distance from the obstacle and also to keep an eye on other vehicle movements that are present in the protocol range, which enables the driver to drive carefully to ensure collision-free environment.

7 Conclusion and Future Scope

VANETs are considered to be a significant research area and drawing the attention of a noticeable number of researches. It is a vast area where there are many challenges that are to be addressed. The training of ESP32 microcontroller with a deep learning

model is accomplished here using DRL. The proposed work is carried out with two sensors and two vehicle nodes, but V2V communication can be extended by adding GPS (Global Positioning System) module to the vehicle node setup which gives the location information of a vehicle and number of vehicle nodes can be increased to show the broadcasting of information effectively.

References

1. Kumar KV, Jayasankar T, Prabhakaran M, Srinivasan V (2017) Fuzzy logic based efficient multipath routing for mobile adhoc networks. Appl Math Inf Sci 11(2):449–455. https://doi.org/10.18576/amis/110213
2. Gopinath S, VinothKumar K, Jayasankar T (2019) Secure location aware routing protocol with authentication for data integrity. Springer Cluster Comput 22:13609–13618. https://doi.org/10.1007/s10586-018-2020-7
3. Ejmaa AME, Subramaniam S, Zukarnain ZA, Hanapi ZM (2016) Neighbor-based dynamic connectivity factor routing protocol for mobile ad hoc network. IEEE Access 4:8053–8064
4. Wang M, Liu D, Zhu L, Xu Y, Wang F (2016) LESPP: lightweight and efficient strong privacy preserving authentication scheme for secure VANET communication. Computing 98(7):685–708
5. Chim TW, Yiu SM, Hui LC, Li VO (2011) SPECS: secure and privacy enhancing communications schemes for VANETs. Ad Hoc Netw 9(2):189–203
6. Bayat M, Barmshoory M, Rahimi M, Aref MR (2015) A secure authentication scheme for VANETs with batch verification. Wirel Netw 21(5):1733–1743
7. Biswas SB, Iqbal MT (2018) Solar water pumping system control using a low cost ESP32 microcontroller. In: 2018 IEEE Canadian conference on electrical and computer engineering (CCECE). IEEE, pp 1–5
8. Kushnir V, Koman B, Yuzevych V (2019, September) IoT image recognition system implementation for blind peoples using ESP32, mobile phone and convolutional neural network. In: 2019 XIth international scientific and practical conference on electronics and information technologies (ELIT). IEEE, pp 183–187
9. Wang W, Min M, Xiao L, Chen Y, Dai H (2019, May) Protecting semantic trajectory privacy for VANET with reinforcement learning. In: ICC 2019–2019 IEEE international conference on communications (ICC). IEEE, pp 1–5
10. Dai C, Xiao X, Xiao L, Cheng P (2018, December) Reinforcement learning based power control for VANET broadcast against jamming. In: 2018 IEEE global communications conference (GLOBECOM). IEEE, pp 1–6
11. Luo Y, Chiu CC, Jaitly N, Sutskever I (2017, March) Learning online alignments with continuous rewards policy gradient. In: 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 2801–2805
12. Liu S, Zhu Z, Ye N, Guadarrama S, Murphy K (2017) Improved image captioning via policy gradient optimization of spider. In: Proceedings of the IEEE international conference on computer vision, pp 873–881
13. Vishnupriya E, Jayasankar T, Venkatesh PM (2015) SDAOR: secure data transmission of optimum routing protocol in wireless sensor networks for surveillance applications. ARPN J Eng Appl Sci 10(16):6917–6931

VBlock—Blockchain-Based Traceability in Medical Products Supply Chain Management: A Case Study in VietNam

Tuan Khoi Nguyen Huynh, Tuan Anh Dao, Trong Nhan Van Pham, Khang Hy Vuong Nguyen, Nghi Cong Tran, and Hoang Huong Luong

Abstract The COVID-19 is putting tremendous pressure on the medical facilities supply, as the demand for facilities has significantly outweighed the production capability. Several rogue traders have taken advantage of this issue to distribute counterfeit products. Moreover, some sellers advertise genuine products with unreasonably high prices. Our team believes that fake or overpriced facilities will significantly complicate the battle against COVID, thereby posing millions of lives to risk. That is why our team is developing V-Block. V-Block is a supply chain management system that harnesses the power of Blockchain. Its primary goals are to assist the government in tracking the product's distribution process and help customers avoid questionable deals.

Keywords COVID-19 · V-Block · Blockchain · Supply chain

1 Introduction

The world has been witnessing a tremendous shortage of Personal Protective Equipment (PPE) and other medical facilities since the beginning of the COVID pandemic.

T. K. N. Huynh · T. A. Dao · T. N. Van Pham · K. H. V. Nguyen · H. H. Luong (✉)
FPT University, Can Tho, Vietnam

T. K. N. Huynh
e-mail: khoinhtce140133@fpt.edu.vn

T. A. Dao
e-mail: tuandace140502@fpt.edu.vn

T. N. Van Pham
e-mail: nhanpvtce140137@fpt.edu.vn

K. H. V. Nguyen
e-mail: hynvkce140237@fpt.edu.vn

N. C. Tran
National Central University, Taoyuan, Taiwan

WHO estimated that the world needs up to 89 million medical face masks each month. Having acknowledged the scarcity of medical facilities, many rogue traders, and sellers distributed counterfeit products to the community and even hospitals to gain money. These facilities, as their quality is not guaranteed [1, 2], can pose significant threats to patients. Also, some stores advertise genuine products at unreasonably unreasonable prices. Our team believes that these illegal activities impede the progress in dealing with COVID. Therefore, they need to be eliminated.

Our team believes that the solution to rogue business is to use a proper supply-chain management platform. We define a proper supply-chain management platform as a platform that is: fast, immutable, and decentralized. As for speed, if the whole distributing process of a particular product is retrieved quickly, we can save much time. If the government detects a facility's problem, all entities related to that facility can be traced within seconds. About immutability, all sensitive data needs to be tamper-proof, meaning that rogue traders cannot alter what they have uploaded, and hackers cannot break havoc. We define sensitive data as business related information such as supply chain information, or companies' credentials. Such data is critical for the government to check business legality and products' origin. Decentralization is the last characteristic to mention. A decentralized platform does not store data in a central place, but copies and distributes it across the network. This ensures that all entities have a copy of the whole database, limiting the impact of a data loss at a node [3].

Many researchers have suggested the use of Blockchain to develop such a platform. Blockchain is a database system, commonly known as a distributed ledger. It includes traceable records or blocks that are linked via cryptography. Blockchain is intrinsically immutable and transparent [4]. Our team agrees with the researchers' perspective, so we want to develop a medical supply chain management system using Blockchain, which we name "V-Block".

In this paper, we will introduce V-Block and explain how it can stop fake and overpriced products. We will begin by describing some previous work and documents on implementing Blockchain for the medical supply chain. We will discuss their strengths as well as their shortcomings. From that analysis, we can suggest our ideas on improving the implementations. After pointing out our ideas, we will present our system. There are three main identities in our system: customers, members, and agencies. Each identity will have distinct functions that suit their needs and help enforce data safety. Our paper is expected to have three main benefits. First of all, current Blockchain implementations are analyzed, and their shortcomings are identified. Additionally, we propose our own system, which includes improvements. Finally, our system can be used to assist the medical industry in the battle against COVID-19.

The rest of the paper is organized as follows. Section 2 contains previous Blockchain applications in various areas, especially in the supply chain. Then, in Sect. 3, we briefly explain our Blockchain system. We will also get insight into the system's functions in Sect. 4 before describing experiments in Sect. 5. We will summarize our working Sect. 6.

2   Related Research

Today Blockchain is applied in many areas such as mobile payments [5] introduce a device that was made up of an Arduino ATMEGA256 board, A TFT-LCD touch screen, a smartcard socket, and a Bluetooth Low Energy (BLE) module named Bluetooth Blockchain payment terminal (PBT). PBT is a new generation of open payment terminals dedicated to Blockchain transactions. A removable secure element is attached to the BPT to keep the card safe from the internet-connected device. They inserted a smartcard with a level of security up to EAL5+ to store private keys and compute signatures to the PBT to make transactions via BLE. Medical [6] this study mentions some of the benefits of Blockchain in the medical field such as: Firstly, smart healthcare management. It used a continuously updated decentralized database to give all medical professionals and doctors access to a patient's treatment history with previous doctors as well as disease health information to get more information and to have the right treatment for that patient. In addition, the decentralized database helps to solve problems such as information that is not being updated, which causes time to verify when needed. Secondly, collect medical information for research purposes: Build a Blockchain system to help users update their health information (heart rate, blood pressure, weight, sleep, used medicine, …). Researchers can buy that information from users to research and find ways to cure new diseases. Finally, Blockchain prevents counterfeit medicine: The drug is not in the correct dose as provided, or the drug is produced in an impure way that can be dangerous to the patient. Blockchain can be used to provide accurate information about the medicine. Blockchain can be used to store transaction records in dental care [7] so that participants can access and control authority. It has the transaction id number generated by the hash number mechanism provided by the Hyperledger editor. This application creates a relationship of trust to establish the set of participants in an ecosystem to share information and interact between clinics so that it can save time and a variety of costs for the patients. Another application of Blockchain in security is the recommended Blockchain-based framework (B2MDF) [8] For detecting malicious apps in the App Store before being downloaded by a user. They proposed using two private Blockchains inside and outside to form a double private Blockchain to make the correct decision. Internal Blockchains store feature blocks that are analyzed by static and dynamic analysis programs to increase malware detection rates. External Blockchains store detected results as blocks for current versions of the application. B2MDF shares information with third parties to help security companies deliver anti-malware. B2MDF does not limit implementation to any particular machine learning algorithm.

Blockchain plays an important role in the supply chain [9]. We have referenced research on Blockchain applications in the supply chain "A Blockchain-based Supply Chain Quality Management Framework" [10]. In this paper, we study the Blockchain-based SCQI framework for improving supply chain quality management. It consists of 4 layers as follows: The first layer is the IoT Layer. It uses IoT sensors to collect Logistics Process Selling Process Manufacture and Quality Inspection Process data.

The information collected is captured using RFID technology or a bar code. The second layer is the Data Layer. Blockchains and secure distributed ledgers are in this layer. The distributed ledger contains the information Quality Data, Logistics Data, Transaction, Data, Assets Data. The third layer is the Contract Layer. This layer contains the data to be protected. So, this layer's data needs to be protected by means of digital identity access control. The last layer is the Business Layer. This layer contains the business operations of the enterprise. With Blockchain, businesses manage product quality in their supply chains more efficiently. "Supply Chain Management using Blockchain" [11] from this paper, we study how to apply Blockchain to product tracking from the raw material until it reaches the consumer, and how to solve problems such as inaccuracies in profile, fraud problem, payment problem. This study also shows that using Blockchain in product tracking can reduce the cost, time, and manpower in monitoring and controlling raw materials and products. Besides, this study also suggests the use of smart contracts to handle product quality problems such as handling compensation when factors are affecting product quality such as temperature, humidity, not in accordance with smart contract requirements. We also looked at the case of Blockchain use in the agricultural supply chain, "Blockchain-based Traceability in Agri-Food Supply Chain Management: A Practical Implementation" [12]. This article is about AgriBlockIoT, a fully decentralized application used to monitor agricultural and food products from seeding to consumption with the desire to monitor to improve food safety and maintain trust and reliability throughout the supply chain. After reviewing the feasibility study for application in Vietnam, we realized that AgriBlockIoT has some disadvantages when it comes to an application as follows: farmers must be trained to store information into the system, causing feeling constrained for farmers when being monitored by IoT sensors; Only some common values, such as weight, were measured, but special indicators such as the number of pesticides that were applied to agricultural products were measured. To measure the plant protection drug index, specialized sensors are required with a high cost and not suitable for practical implementation; IoT sensors cannot be widely equipped because of the high cost of setting and maintaining. That causes false data when farmers can affect agricultural products in areas without sensors. Assuming the case of sensors being placed throughout the garden, the farmer can still cheat by masking the sensors when using a pesticide so that the system recognizes "clean" and high yielding produce.; The last problem, the sensor fails to measure correctly, but the data is still within the allowable range and will not be detected and cause an incorrect data supply. We have looked at the challenges of applying Blockchain to the supply chain [13]. This article outlines the opinion that supply chain management (SCM) is the key factor in achieving financial, environmental, and social benefits in the supply chain industry. But the traditional SCM mechanisms are outdated and have many weaknesses in information sharing, data retrieval, and reliability in finding products. These problems can be solved by Blockchain due to Blockchain features such as immutability, transparency, and decentralization. What have we learned from this article? These are the requirements of supply chain management when adopting Blockchain, the technical challenges of applying Blockchain to the supply chain,

potential opportunities, and principles of supply chain management system design based on Blockchain.

3 System Overview

V-Block is a system used to store and retrieve information about medical products' supply chain. It differs from traditional systems in terms of data storage, as it uses Blockchain technology. This storage implementation is expected to improve data security, thereby assisting violation detection, and increasing business transparency.

In general, there are three main entities, including customers, manufacturers/distributors, and agencies. Manufacturers or distributors will upload data about the origin and prices of products. Agencies and customers, when in need, can query data and analyze it for anything questionable. Agencies have the right to view data without restrictions.

The variety of users, along with different levels of rights and permissions, means that V-Block requires a layer of security over data access and performed activities. For this reason, our team develops V-Block to be a permissioned Blockchain. Permissioned Blockchain is a well-known model, acknowledged by enterprises and researchers worldwide [14, 15]. The system has three main parts: users, access control, and data storage.

3.1 Users

There are three kinds of users, namely: customers, members, and agencies. Customers are the least-privileged identities in the system. They are hospitals or any other institutions that purchase a medical product and want to check for its authenticity. Customers are only able to query information about a particular product. Members are manufacturers and distributors of a product. They are allowed to submit data to the system and also view their uploaded data. Agencies refer to medical institutions from the government. They are granted the most rights and permissions. They are permitted to view any information as well as new changes made to the system. This enables them to check the validity of a particular item or a manufacturer.

3.2 Access and Data Control

We acknowledge the importance of keeping uploaded data safe and private in supply chain management [16]. That is why we implement different controls to ensure that only allowed individuals can access private data. Normal customers do not have to login or sign up to use our system, as they only want to query information about a

specific product. However, members must prove their identity before proceeding with any operations. Upon registration, they first need to fill a form, in which they include some personal details and business licenses. After this step, they will have to wait to be checked and approved by agencies. What users can query is also controlled. For example, only agencies can view all information about the system, while members can only ask for information about their own products. This is done by implementing smart contracts for searching and querying data [17]. Additionally, agencies can create full nodes, but members only have light nodes. Having full nodes means that a user can save everything on the system into their machines, as opposed to light nodes, which only stores limited information.

3.3 Data Storage

Any data created will be stored on Blockchain. New data, after verified by peers, will be kept in a block, which then is broadcasted via multiple network protocols [4]. Data input and output will be handled by different consensus algorithms [18] and smart contracts. This serves the system's functions and ensures the validity of data being stored and submitted.

4 System Functions

The functions of our system can be summarized as below (Fig. 1).

Members (manufacturers and product distributors) first need to register with the system before interacting with the database. Interaction with the data will be controlled by smart contracts. When manufacturers produce a product, they will upload the information about that product to the system. When they sell it to a distributor or a reseller, that transaction is also uploaded to the Blockchain. By doing this, the whole supply chain of a facility will be recorded (Fig. 2).

Customers can directly query information about the origin of a product from our system. If they find some questionable trades, they can report to the agencies. Agencies are in charge of verifying new members and handling reports from the customers. They are able to retrieve any necessary information from the system.

Our system is expected to deal with illegal activities using the following methods.

Counterfeit Facilities

The V-Block system aims to combat this issue in two ways: assisting customers in considering wisely before purchasing and supporting agencies to detect and punish rogue traders.

Customers, before purchasing, can access our system and have a quick analysis of the validity of the product. They can submit the code to the website by scanning the QR code and then receive the distribution process of that product. Therefore,
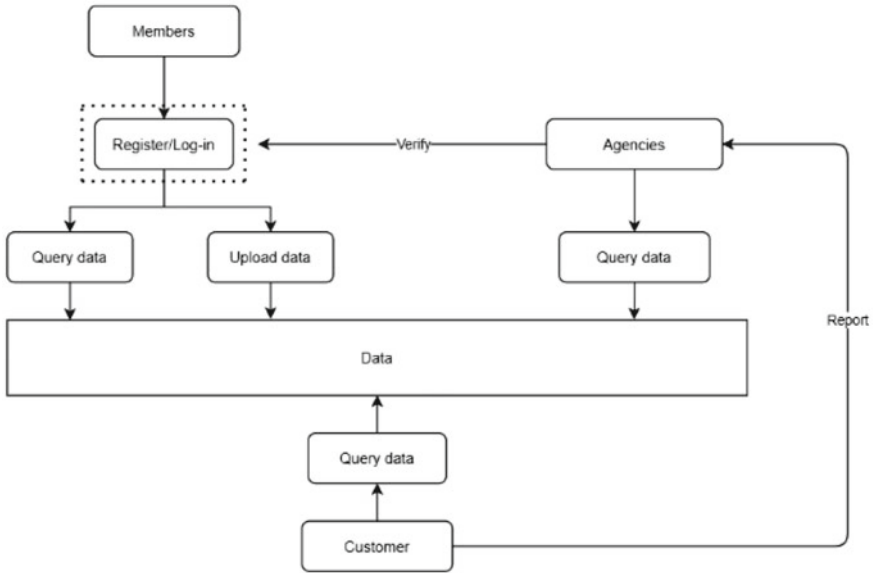
Fig. 1 System functions overview
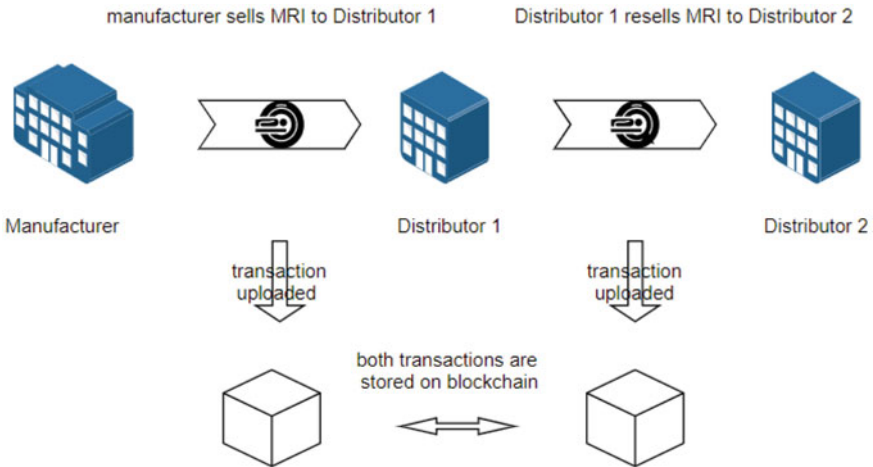
Fig. 2 Chain formation

customers are able to know who the manufacturers are and where the facilities are produced within a limited amount of time. If they find something suspicious about the supply chain of any facility, they can report it to the agency. The report includes: the name of the business, the name and the code of the product. Reports will be kept away from companies, which means companies cannot know the reporters.
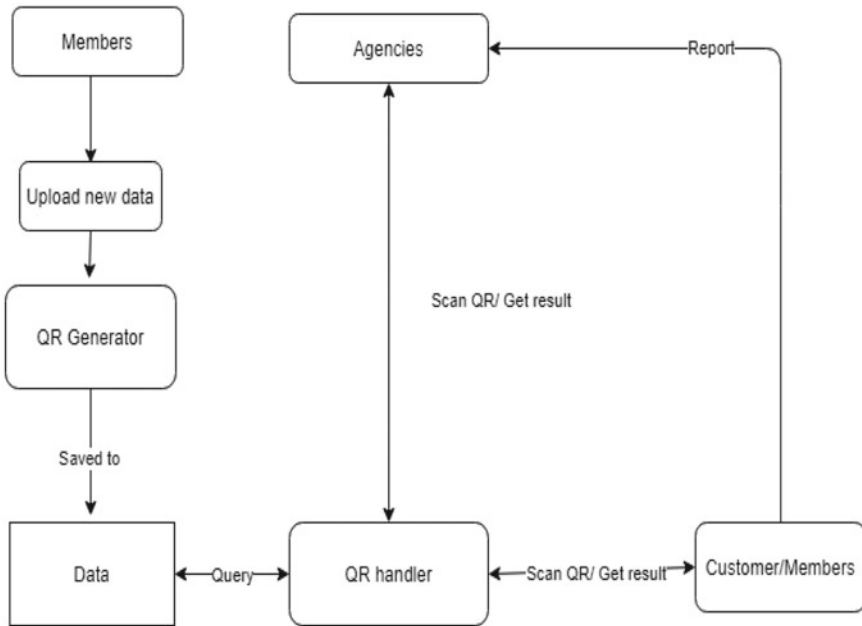
Fig. 3 Counterfeit facilities

If the agencies are reported about a potentially fake medical facility, they can use our system as a great tool. The agencies can quickly get all information about individuals and groups involved in the supply chain thanks to the origin-tracing feature of V-Block. All they need to do is to submit the reported product's code and wait for the result. This will greatly speed up the investigation. Additionally, as all data stored on our system is tamper-proof, agencies can use it as evidence against any illegal trades (Fig. 3).

Overpriced facilities

V-Block allows users to analyze fluctuations in products' prices. Manufacturers and distributors will upload the product's new cost after each stage in the supply chain. As a result, price changes can be recorded and retrieved whenever needed. Agencies and consumers can query the data and check if a distributor or a seller has unreasonably increased the cost.

Furthermore, our system attempts to block suspicious cost-related data upload. If a facility's cost is raised 20% after only one stage in the supply chain, it will be flagged, and the upload will be stopped (Fig. 4).
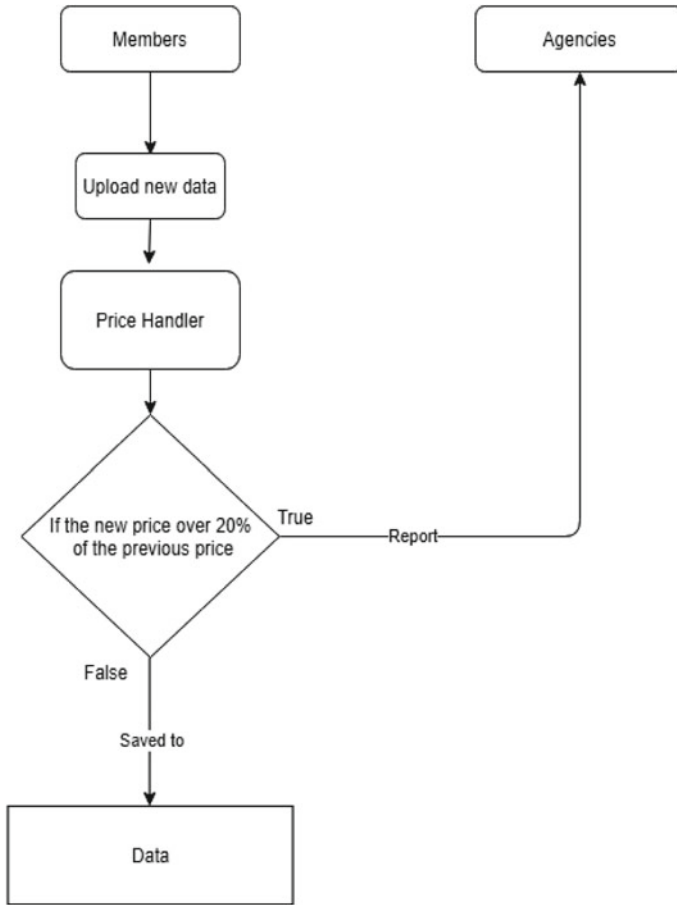
Fig. 4 Overpriced facilities

5 Experiment

5.1 Introduction

We developed our Blockchain system using Ganache. We tested it on a device using Ubuntu 16.04LTS, 4-core CPU, 8 GB RAM, and 50 GB HDD. The system was analyzed in three scenarios: a user checking a product's information, an agency checking a product's information, and a seller trying to upload an unreasonable price for a product.

In order to start the test, we uploaded data about a product whose name was 'product1'. It was manufactured by Tank Group. All data about individuals and organizations involved were unreal and for testing purposes only. We also generated

one account for a customer, another for an agency, and the remaining one for a seller (member).

5.2 Tested Scenarios

A user checks a product

In the first scenario, a user considered whether to buy a product. He scanned the QR code to see who the manufacturer is and when the product was made. Our system then processed the QR code to return data about the desired product. Besides displaying basic information, V-Block also showed if the product was verified by the agency. Other detailed data is hidden because the system knew that the person was just a customer (he did not sign in) (Figs. 5 and 6).

An agency checks a product

In the second scenario, V-Block processed a query from the agency. In order to receive the information, the agency needed to log in and then, like customers, had to scan the QR code. The system returned detailed information about the facility (Fig. 7).
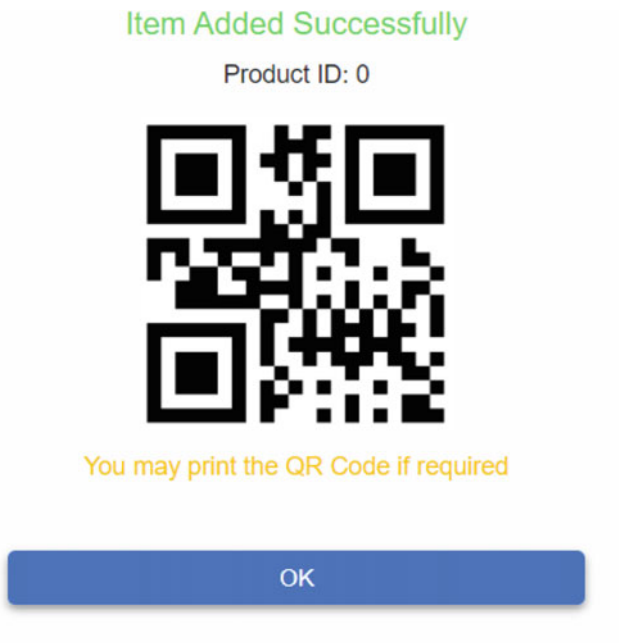
Fig. 5 Example of QR

ID	Product Name	Manufacturer	Manufacture Date	Approved
0	product1	TankyGroup	02-10-2020, 17:28	Approved

Fig. 6 Example of product which was verified

Information ID 0   ✕

Approved
Product Name: product1
Manufacturer: TankyGroup
Manufacture Date: 02-10-2020, 17:28
Price: 1000
Certificate: cert1

Date: 02-10-2020, 17:29
Location: 10.0440341, 105.77034019999999
Price: 1100

Fig. 7 Example of showing a product's information

A seller upload unreasonable price

The last scenario was a special case, as our system had to deal with an illegal attempt. A seller wants to sell a product at a hefty price (which was 20% higher than permitted). This will put a great financial strain on consumers, so it should be stopped promptly. When our system processed such upload, the smart contract detected the violation and showed a notification. V-Block successfully took the necessary action. The 20% rate can be modified if required by the agencies (Fig. 8).

6 Conclusion

In this paper, we explained the issue of counterfeit and overpriced medical facilities. We also described our solution to that issue, which is to develop a supply chain management system using Blockchain, which we call 'V-Block.'
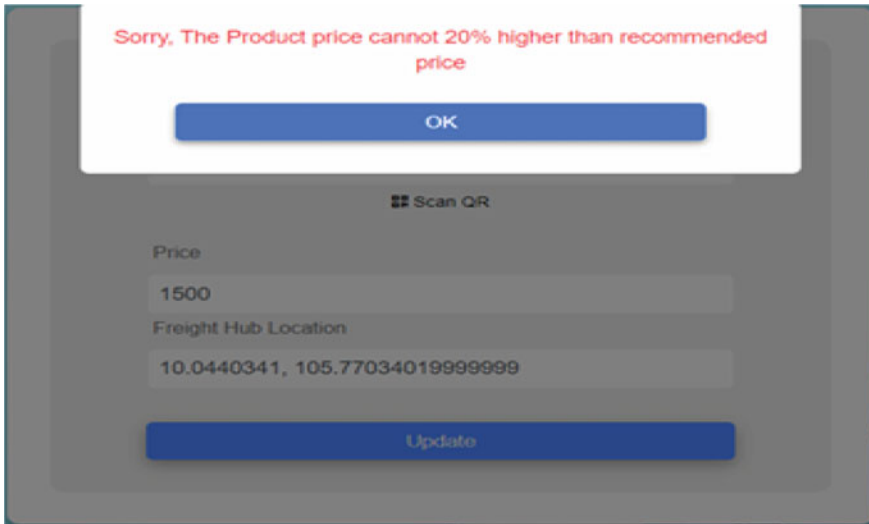
Fig. 8 Example of reject because of the fault price

V-Block system will be the pioneering system for Blockchain application to support the battle against the COVID pandemic. It helps users ensure medical integrity, authenticity and provides an easy way to search for essential medical information of medical manufacturers.

References

1. Cheng M (2001) A guide for the development of medical device regulation. Washington DC
2. Corciovă C, Andritoi D, Ciorap R (2013) Elements of risk assessment in medical equipment. In: 2013 8th international symposium on advanced topics in electrical engineering (ATEE)
3. Xiao T, Choi TM, Edwin Cheng TC (2018) Pricing and benefit of decentralization for competing supply chains with fixed costs. IEEE Trans Eng Manag 65(1)
4. Antonopoulos AM (2017) Mastering bitcoin, 2nd ed., June 2017
5. Urien P (2020) High security bare metal Bluetooth Blockchain payment terminal for trusted ethereum transaction. In: 2020 IEEE 17th annual consumer communications and networking conference (CCNC)
6. Mettler M (2016) Blockchain technology in healthcare: the revolution starts here. In: 2016 IEEE 18th international conference on e-health networking, applications and services (Healthcom)
7. Wutthikarn R, Hui YG (2018) Prototype of Blockchain in dental care service application based on Hyperledger composer in Hyperledger fabric framework. In: 2018 22nd international computer science and engineering conference (ICSEC)
8. Homayoun S, Dehghantanha A, Parizi RM, Choo KKR (2019) A Blockchain-based framework for detecting malicious mobile applications in app stores. In: 2019 IEEE Canadian conference of electrical and computer engineering (CCECE)
9. Caro MP, Ali MS, Vecchio M, Giaffreda R (2019) Blockchain-based traceability in agri-food supply chain management: a practical implementation. In: 2018 IoT vertical and topical summit on agriculture—Tuscany (IOT Tuscany) Wu H, Cao J, Yang Y, Tung CL, Jiang S, Tang B, Liu Y,

Wang X, Deng Y (2019) Data management in supply chain using Blockchain: challenges and a case study. In: 2019 28th international conference on computer communication and networks (ICCCN)

10. Chen S, Shi R, Ren Z, Yan J, Shi Y, Zhang J (2017) A Blockchain-based supply chain quality management framework. In 2017 IEEE 14th international conference on e-business engineering (ICEBE)

11. Bhalerao S, Agarwal S, Borkar S, Anekar S, Kulkarni N, Bhagwat S (2019) Supply chain management using Blockchain. In: 2019 international conference on intelligent sustainable systems (ICISS)

12. Caro MP, Ali MS, Vecchio M, Giaffreda R (2018) Blockchain-based traceability in agri-food supply chain management: a practical implementation. In: 2018 IoT vertical and topical summit on agriculture—Tuscany (IOT Tuscany)

13. Wu H, Cao J, Yang Y, Tung CL, Jiang S, Tang B, Liu Y, Wang X, Deng Y (2019) Data management in supply chain using Blockchain: challenges and a case study. In: 2019 28th international conference on computer communication and networks (ICCCN)

14. Mitani T, Otsuka A (2019) Traceability in permissioned Blockchain. In: 2019 IEEE international conference on Blockchain (Blockchain)

15. Islam MdA, Madria S (2019) A permissioned Blockchain based access control system for IOT. In: 2019 IEEE international conference on Blockchain (Blockchain)

16. Roy A, Gupta AD, Deshmukh SG (2012) Information security in supply chains—a process framework. In: 2012 IEEE international conference on industrial engineering and engineering management

17. "SmartContracts". https://solidity.readthedocs.io/en/v0.7.1/introduction-to-smart-contracts.html

18. Chaudhry N, Yousaf MM (2018) Consensus algorithms in Blockchain: comparative analysis, challenges and opportunities. In: 2018 12th international conference on open source systems and technologies (ICOSST)

Development of Real-Time Internet of Things (IoT) Based Water Quality Monitoring System

Huzein Fahmi bin Hawari, Mohamad Nor Syahid bin Mokhtar, and Sohail Sarang

Abstract This research involves discovering on real-time Internet of Things (IoT) based water quality monitoring system using three sensor parameters which are temperature, turbidity and pH. Using this, people can monitor the water quality, collect all relevant data and information, decide and perform suitable water treatment immediately. By performing comprehensive data analysis, people can measure water pollution level referring to Water Quality Index (WQI). Other than that, this project also involves cloud computing storing all data transferred remotely from the device at anytime and anywhere. Furthermore, this project applies good power management method to reduce power consumption and extend the battery life of the device within a single charge for up to 31 days. The system also able to be deployed at 7 different location in UTP and able to clarify monitoring data based on WQI. On the other hand, mobile application is developed to monitor relevant data. Lastly, this project has high reliability and practicality as it is applied with comprehensive techniques for every step used and good response time to transmit data for water quality monitoring in real-time.

Keywords Internet of Things · Water quality index · Pollution index

1 Introduction

The water quality status of rivers and lakes in Malaysia has always been a reason of concern for various local authorities, government agencies and the public at large scale. Water quality monitoring system is important for various applications that requires a real time monitoring to determine the contents and substances present inside the water. According to Departmental of Environmental (DOE) Malaysia, it is

H. F. Hawari (✉) · M. N. S. Mokhtar
Department of Electrical and Electronic Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Tronoh, Malaysia

S. Sarang
Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovi´ca 6, 21000 Novi Sad, Serbia

stated that two main parameters are most likely to affect the river water quality that are urbanization and development [1]. WQI is a method or a tool used to measure the pollution level of water using SI calculation whereas National Water Quality Standard (NWQS) is the national standards used to indicate the level of pollution after measuring the water quality using WQI [2, 3].

IoT is defined as internetworking physical objects embedded with electronics, sensors, network connectivity and actuators which can be used to collect and exchange the data [4–6]. IoT can connect billions of devices over the network which will eventually create much more need for data storage and much larger traffic to control the connectivity of those devices using TCP/IP protocol stack to communicate between network hosts for today's Internet [7].

In this proposed IoT system, users can monitor the quality of water parameters in real time using mobile application developed using suitable software and alert the user for water quality alarm system. This alert system is important for the user to alert them when the value of certain parameters exceed the threshold set. According to UN Water, it is stated that around 780 million of people did not manage to get safe and clean water while around 2.5 billion of people do not have suitable adequate clean drinking water provided by UN-Water [8].

The problem with existing technology is water quality monitoring involves the manual intervention of collecting water sample in different locations. This is followed by further laboratory analytical techniques to characterizes the water quality based on measurement parameters of WQI. This manual approach takes longer time and is considered as less efficient for water quality monitoring. It is also not considered as real-time as the person responsible must take samples from particular locations from time to time.

Therefore, this project is implemented to develop and implement a real-time application based on IoT for water quality monitoring system, to integrate water quality sensors together with IoT communication platform for an efficient real-time water quality monitoring system and to develop mobile application suitable in analysing the water quality conveniently.

2 System Design and Architecture

2.1 Central Processing Unit

Arduino Nano is an open-source compact development board contains embedded electronics parts such as micro-controller, voltage regulator, LED indicators and power controller. This board is powered up using 7.4 V lithium-polymer battery, operates at 5 V with 32 KB of flash memory, 16 MHz of clock speed and 19 mA of current consumption (during active).

2.2 Sensor Unit

A sensor unit consists of several sensors integrated together in a single system to detect the predetermined water quality parameters to indicate water quality level. In this project, three water quality parameters were selected as sensors unit which are temperature, turbidity and pH to measure the water quality. Figure 1 shows the block diagram of sensor unit constructed.

In marine ecosystems, climate changes will eventually affect the aquatic organisms and its surrounding that lead to the quality of water for temperature [9]. Turbidity is defined as the water cloudiness caused by many types of particles and suitable key parameter in drinking water analysis and the presence of any suspended particles that can cloud the water [10, 11]. pH is the measurement to describe the acidity and alkalinity of the water. A microbial facilitated process called denitrification in which the nitrate is reduced and produces molecular nitrogen through intermediate series of gaseous nitrogen oxide products [12].

All the sensors and micro-controller use battery for their operation. The data of water quality are being sensed by each sensor accordingly. Then, the sensor data are processed in micro-controller by managing the real data that is understandable by human. Table 1 shows the sensor response time and power consumption for each sensor used and microcontroller (Arduino Nano).

Fig. 1 Block diagram of sensor unit
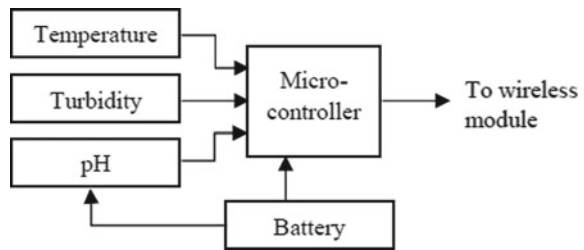
Table 1 Current consumption and response time for sensors and Arduino Nano

Components/Sensors	Current consumption (mA)	Response time (ms)
Arduino Nano	19	0.9
Temperature	10	0.8
Turbidity	14	0.3
pH	30	0.4
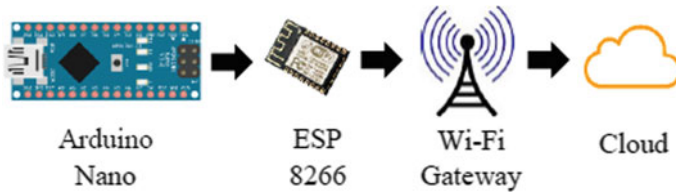
Fig. 2 Wireless communication architecture

2.3 Wireless Communication

The data from sensors are transferred over the network using Wi-Fi connection. A Wi-Fi module (EP8266) is integrated with Arduino Nano as controller to provide the connectivity of sensor device with cloud computing by connection to a Wi-Fi gateway to connect to Internet. It supports 802.11 b/g/n/e/i of network protocol with Wi-Fi Direct (P2P). It is advantageous of using this module as it supports standard network protocol, has antenna diversity and selection (to extend wireless range) and relatively low power consumption which is about 170 mA during Transmit Mode (Tx Mode). Figure 2 shows wireless communication architecture for this project.

2.4 Cloud Computing and Mobile Application

In the cloud, all data collected in the microcontroller are stored in the database through cloud computing system. The cloud used for this project is Ubidots which provide simple IoT implementation in cloud computing system. This cloud can provide sufficient features needed to implement this project successfully. To enhance the portability and mobility of data monitoring system, mobile application for tablets and smartphones is designed to capture those data stored in the cloud and visualized in charts and gauges representation for water quality monitoring purposes. Real time data will be measured and saved in the cloud to produce historical data for certain period.

2.5 Power Management

This prototype uses two lithium-polymer batteries as the power sources to all electronic components involved. To manage the power consumption, 1st battery will provide power to ESP8266 to set up the module and network. Once finished, 1st battery is turned-off while 2nd battery is switched on providing power to sensors to get measured sensor data from water. After all sensors data have been achieved, 2nd battery is switched off while turning on 1st battery to send sensors data to cloud via
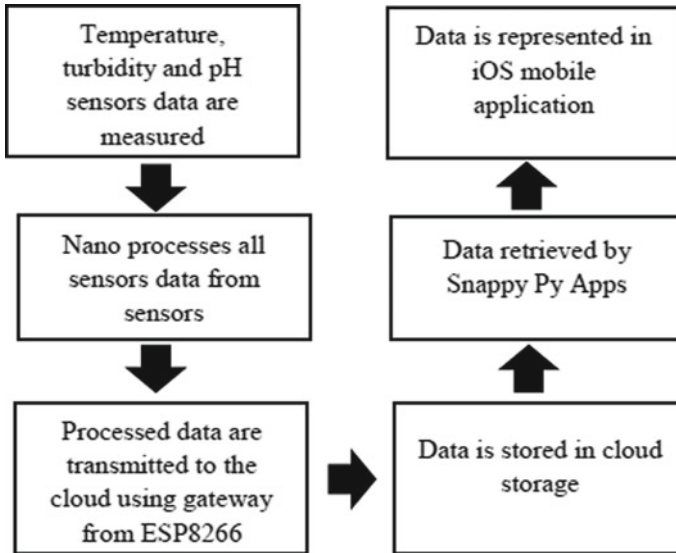
Fig. 3 Process flow on device operation

standard wireless communication protocol. The power consumption estimated for overall system is 159 mA.

2.6 *Operation of Real-Time Water Quality Monitoring*

Figure 3 shows the operation of prototype in water quality monitoring.

Sensors measure water quality from lakes and send to Nano. The sensor data will then be processed by Nano and transmit them to cloud via Wi-Fi connectivity. Then, data sent will be stored in cloud storage and will be used by mobile application in iOS to display all data in several representations for water quality monitoring.

2.7 *Overall System Architecture*

Figure 4 shows the system architecture of the system starting with sensors getting all the data from lakes and transmit to cloud via wireless connectivity and monitored in mobile application.
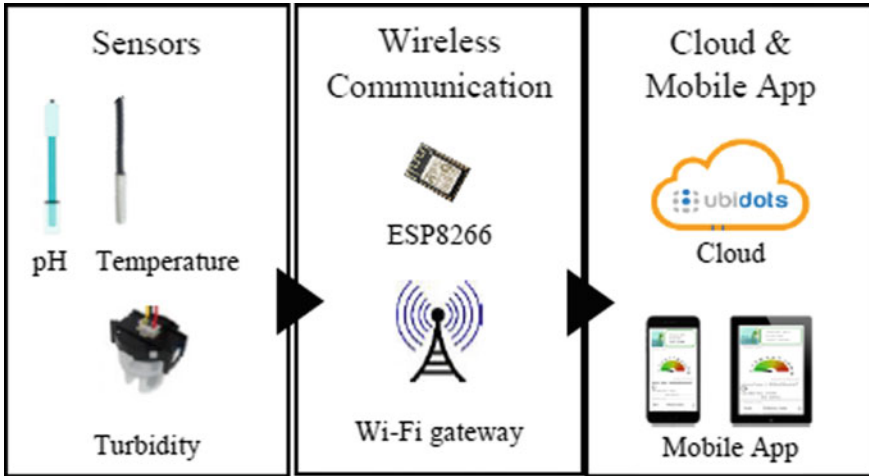
Fig. 4 System architecture

2.8 Prototype Deployment in UTP Lake

The prototype is placed on top of water supported by three pillars anchored in the water attached with stretchable strings. This is to prevent the device to flow away from desired location. Figures 5 and 6 shows the location of prototype development in UTP lakes and deployment technique of the device in UTP lake respectively:

Fig. 5 Prototype deployment in UTP lake

Fig. 6 Deployment location in UTP lake

3   Results and Discussion

Data for water quality parameters include turbidity, pH value and temperature were taken in two different locations in UTP. The duration for each data measurement is 1 h with 14 h of total data measurement per day for 7 days in a week.

3.1   Water Pollution Index

An equation is used to calculate water pollution index:

$$\text{Pollution index} = (0.152 \times \text{Temperature} + 0.002 \times \text{Turbidity} + 0.714 \times \text{pH})/3 \tag{1}$$

The sensor data are inserted into the formula to get water pollution index value. Table 2 shows the pollution index value for each pollution level.

The value of water pollution index in UTP lake is calculated based on formula constructed with the value of temperature, pH and turbidity measured. Data for WPI is calculated for overall 4 weeks of data collection to indicate the level of pollution of UTP lake. From Fig. 7, we can observe that the trend of WPI is stable during 1st

Table 2 Pollution index table

Pollution level	Water pollution index value
Polluted	0.00–1.40
Slightly polluted	1.41–2.70
Clean/Normal	2.71–5.90
Slightly polluted	5.91–7.30
Polluted	7.31–10

Fig. 7 Pollution index of UTP lake in location A and B
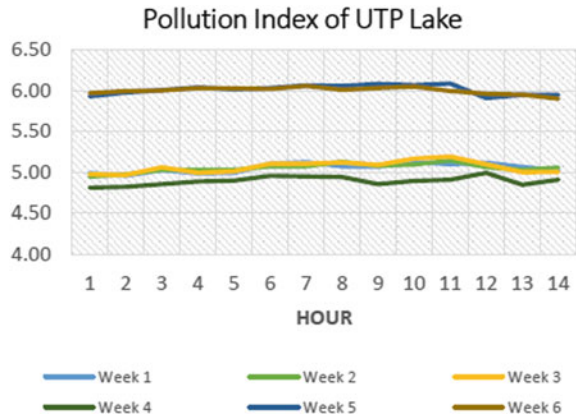
week until 3rd week, but there is slight drop in WPI value in 4th week. This is due to lower value of turbidity which contributes to lower WPI value in the calculation. WPI increases during 5th week to 6th week because it has been measured in location C where the water quality is found to be slightly polluted.

3.2 Temperature

Figure 8 shows the data for water temperature in UTP lakes for 6 weeks of data collection:

The average of temperature value for location A is 29.50 °C, for location B is 29.10 °C and for location C is 29.45 °C. During 1st week until 6th week of data collection, water temperature of location A, B and C shows nearly the same value except for 4th week. During 1st week to 3rd week, the mean average water temperature is higher than during 4th week due to warmer ambient condition. This increases
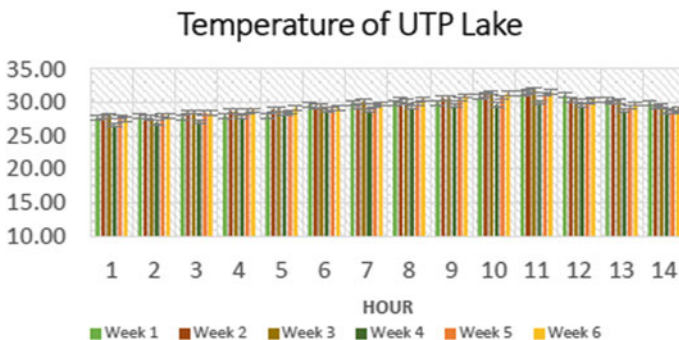
Fig. 8 Water temperature in UTP lakes for 6 weeks

the solubility and toxicity of certain compounds in the water. These compounds include heavy metals elements such as zinc, lead, cadmium and ammonia. With higher water temperature (more than 30 °C), oxygen consumption, metabolic rate and tissue of aquatic organism permeability increases. This will reduce the survival rate of aquatic organisms in the water as the lethal concentration 50 (LC50) drops for a long period. The lower the value of LC50, the lower the concentration of chemical inside the water, causing lower life time for aquatic organisms.

3.3 PH

Figure 9 shows the data for water pH in UTP lakes for 6 weeks of data collection:

The average pH value for location A is 6.84, for location B is 6.83 and for location C is 8.65. Based on WQI, the permissible range of value for normal lake water pH is 6.01–8. As the average of pH value for each week and total average of pH value for 1st week until 4th week falls in between the permissible range of value, the water pH measured contributes normal value of water pollution index of UTP lake.

Unlike water pH in location C where the average pH value is 8.65, it is included in slightly polluted range. This value has possibility to contribute slightly polluted water condition for water pollution index. When the pH value is below or higher than the optimal value or permissible range of value, aquatic life become susceptible to fungal infections and other physical damages thus it is not suitable for aquatic organisms to stay in water that has non-optimal pH value.

This can harm those organisms and reduce water quality. Related to spectrum, high pH levels can damage the gills and skin of aquatic organisms and can cause death at pH level higher than 10.0. While the mean pH value of location C is still below that pH level, the aquatic organisms are safe, but it also can harm those organisms inside the water.
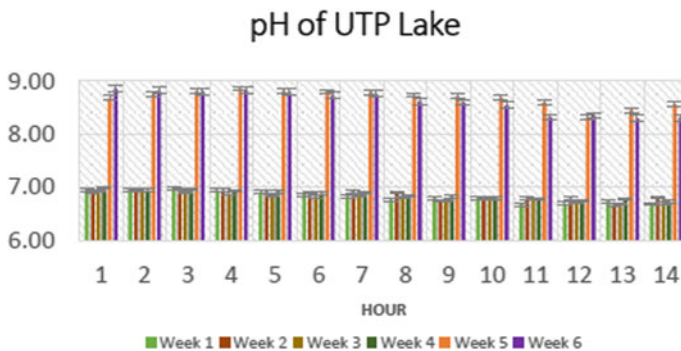
Fig. 9 Water pH in UTP lakes for 6 weeks

3.4 Turbidity

Figure 10 shows the data for water turbidity in UTP lakes for 6 weeks of data collection. The average of turbidity value for location A and B contributes to normal water pollution index while the average of turbidity value for location C contributes to slightly polluted water pollution index. The lower the value of turbidity, the higher the water quality based on water pollution index. This can be achieved by reducing the Total Suspended Solids (TSS) present in the water which higher total suspended solids causes the water to be turbid. TSS are particles that are larger than 2 μ, therefore any particles in the water that are smaller than 2 μ are considered as Total Dissolved Solids (TDS). TSS is contributed by algae and bacteria inside the water that can cause water to be more turbid.

Usually, clear water is considered as clean water. Excessive value of turbidity can impair the water quality for aquatic and human life. Other than that, higher water turbidity will increase water temperature and reduce Dissolved Oxygen (DO). Warmer water cannot hold DO more than cold water. Based on the data from those locations, location C is showing least DO quantity compared to the other 2 locations as it has highest water turbidity and included in slightly polluted range based on WQI. This lead to higher heat absorption from solar radiation and transferred to surrounding water by conduction. Therefore, location C is considered as slightly polluted and not suitable for aquatic life and human usage.

On the other hand, pollutants are most likely to be found in high water turbidity. The contaminants include bacteria, mercury, lead and other material that encourage the development of harmful substances inside the water. Since those 3 locations are used for recreational use, as turbidity reduces visibility of underwater structure, it negatively affects the body's recreational use for users in UTP.
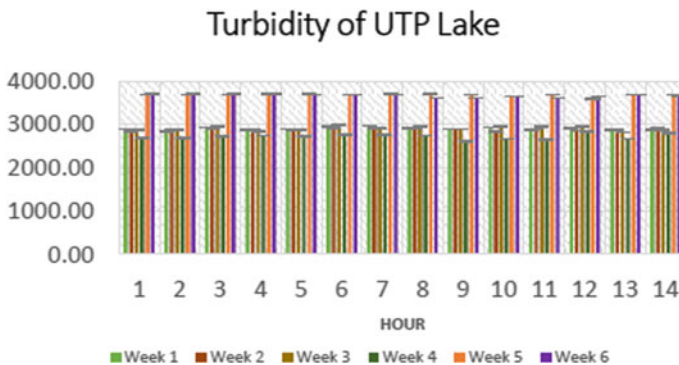
Fig. 10 Water turbidity in UTP lakes for 6 weeks

Fig. 11 Overall current
consumption for device
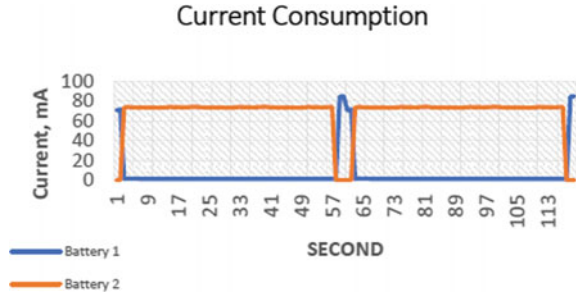
3.5 Power Consumption

Figure 11 shows the overall current consumption for device used in water quality monitoring:

Analysis is made to calculate the battery life for both batteries (battery A and B) using the formula shown below:

$$\text{Capacity of battery A}: 2200\,\text{mAh}$$

$$\text{Capacity of battery 2B}\ 1200\,\text{mAh}$$

$$\text{Battery}\ 1 = 2.2Ah/[(122/3600 \times 73) + (3478/3600 \times 0.004)]$$
$$= 37.05\,\text{days}$$

$$\text{Battery}\ 2 = 1.2Ah/[(2/3600 \times 75) + (2/3600 \times 84) + (3596/3600 \times 1.5)]$$
$$= 31.53\,\text{days}$$

From the calculation made, the battery life for the prototype is estimated up to 30 days.

4 Conclusion

In conclusion, Internet of Things (IoT) assist and help people in deciding the best solution for any problems faced in daily life especially water quality because water is the main element required by human beings for their life every day. Without good quality of water, people tend to have problems with their healthcare and welfare because everyone need water to survive in any situation. Development of innovative technology help them to monitor the water quality especially in the river as it is the main source of water supply for users in entire world. Once water quality can be

monitored in real time, it is easy for the department responsible in maintaining the water quality to get all required data of water quality, visualize them in most effective representation and decide the best solution for water treatment. The performance of this system is comprehensive with good techniques and methods used for every scope included. The system has good response time sending all data successfully to cloud with minimum communication error. This is a good practice as it increases reliability and practicality of this project. Without this technology, people are forced use manual intervention continuously by taking samples from site locations regularly and analyse them in the laboratory using effective laboratory analysis methods. This is time consuming and not considered as efficient water quality monitoring.

References

1. Suratman S, Sailan MI, Hee YY, Bedurus EA, Latif MT (2015) A preliminary study of water quality index in Terengganu River basin, Malaysia. Sains Malaysiana 19(44):67–73
2. Hossain MA, Sujaul IM, Nasly MA (2013) Water quality index: an indicator of surface water pollution in eastern part of Peninsular Malaysia. Res J Recent Sci 10(2):10–17
3. Gazzaz NM, Yusoff MK, Ramli MF, Aris AZ, Juahir H (2012) Characterization of spatial patterns in river water quality using chemometric pattern recognition techniques. Department of Environmental Science, Faculty of Environment Studies
4. Khan R et al (2012) Future internet: the internet of things architecture, possible applications and key challenges. DITEN Department University of Genova (UNIGE), 16145 Genova, Italy
5. Hawari HF, Zainal AA, Ahmad MR (2019) Development of real time internet of things (IoT) based air quality monitoring system. Indonesian J Electr Eng CS 13(3)
6. Chung Hua BD, Fahmi H, Yuhao L, Kiong CC, Harun A (2018) Internet of things (IOT) monitoring system for elderly. In: International conference on intelligent and advanced system (ICIAS), p 201
7. Minerva R, Biru A, Rotondi D (2015) Towards a definition of the Internet of things (IoT). IEEE Internet Initiative. Telecom Italia, Torino
8. UN-Water (2013) An increasing demand, facts and figures, coordinated by UNESCO in collaboration with UNECE and UNDESA. Available: http://www.unwater.org/water-cooperation-2013/en/
9. Vinegre C, Madeira D, Narciso L, Cabral NH, Diniz M (2012) Effect of temperature on oxidative stress in fish: lipid peroxidation and catalase activity in the muscle of juvenile seabass, Dicentrarchus labrax
10. Schveitzer R et al (2013) Effect of different biofloc levels on microbial activity, water quality and performance of Litopenaeus vannamei in a tank system operated with no water exchange. Aquacult Eng 56:59–70
11. Jonsson et al M (2013) Foraging efficiency and prey selectivity in a visual predator: differential effects of turbid and humic water. Can J Fish Aquat Sci
12. Norshidah B et al (2014) Characterization of spatial patterns in river water quality using chemometric techniques. Sains Malaysiana 9(43):1355–1362

Edge Computing Technology: State-of-the-Art and Challenges in an Internet of Things

Deepa Raghunathan and M. Krishnamoorthi

Abstract The Internet of Things (IoT) is a network of expedients that can intuiting, collecting, managing and communicating data over the Internet without human intervention. It is a concept that connects to nearby objects without user intervention, using a wired or wireless network. In IoT, objects interact and share information to provide the user or users with advanced, intelligent services. In the field of edge computing, data engendered by IoT devices can be processed closer to where it was created, and as an alternative to sending data to data centers or data centers in more ways. New research and customer support in selecting seemly edge computing structures for particular solicitations. This paper gives a complete overview of the methods of introduced power adding structures, comparing Open Source devices through their survival method. Finally, this work sheds light on smart home application tools using Edge Computing Systems and also explores the challenges posed by Edge Computing Systems.

Keywords IoT · Edge computing · WAN · Fog · Cloudlets

1 Introduction

Kevin Ashton first coined the term Internet of Things in the presentation tile, which introduced sensor tagging for large consumer company products in 1999 to strengthen supply chain management. Things to the Internet describe the ecosystem, where sensors and actuators are connected to the Internet. All objects in the IoT system speak the same language and understand and operate without human intervention. The Internet of Things has been labeled too late, the concept of connected networks is not

D. Raghunathan (✉)
Research Scholar, Department of Computer Science and Engineering, Dr.N.G.P Institute of Technology, Coimbatore, India

M. Krishnamoorthi
Professor, Department of Computer Science and Engineering, Dr.N.G.P Institute of Technology, Coimbatore, India
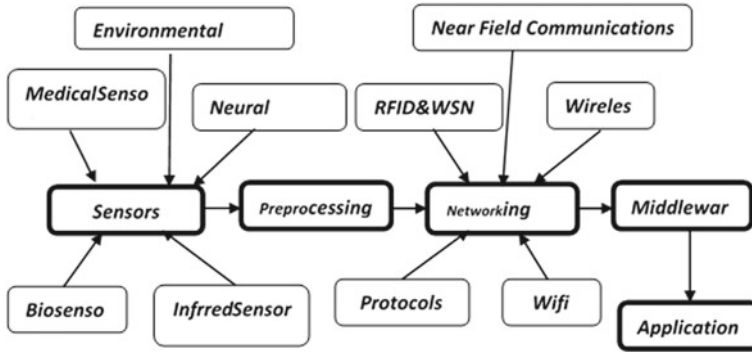
Fig. 1 Taxonomy of IoT

new and there are many developments like this in progress [1]. In the 1980s, the Coca-Cola machine was connected to the Internet and monitored for availability, which reduced unnecessary visits to vending machines. Sethi et al. The following statistics review basic IoT systems, architectures, available protocols, and applications of IoT.

The Edge Computing model has gained a significant reputation in the academic and industrial sectors over the past few years. It serves as the key to many future technologies, such as vehicle-to-vehicle communication using 5G, IoT and cloud computing services for future customers. It gives low-slung invisibility versatility and arrangement mindfulness bolster for delay-sensitive solicitations. In the context of the latest developments in cloud computing, mist computing, and versatile edge computing, basic investigation within the space of edge computing, pioneered by providing additional solutions to researchers in existing solutions and future applications [2]. Due to the characteristics and features of IoT services, smart home applications have also become a major problem in the IoT application areas (Fig. 1).

Edge Computing speaks to an unused concept within the computing extract that fetches the administrations and practicalities of cloud calculating closer to the conclusion patron. It is characterized by two variables, such as quicker handling and speedier application reaction time. Right now created Internet-facilitated applications such as observing, cybernetic authenticity and instantaneous activity observing involve quicker administration, speedier reaction times [3]. Conclusion clients ordinarily run these applications on their portable gadgets whereas performing center administrations and preparing on cloud servers. Cloud administrations are influenced by versatile gadgets, coming about in tall idleness and versatility issues [4]. Gives an edge computing application as per the prerequisites by moving the handling to the arrange edge. Mist computing concerns can be illuminated by three distinctive authority figuring simulations, such as cloudlets [5], mist computing [6] and versatile edge computing [7].

This paper describes Edge Computing methods and models. Section 2 describes the literature review of an IoT edge computing, fog computing, and mist computing. Section 3 gives a brief computing model. Sections 4 and 5 provides conclusions

about the availability of equipment and, finally, Sect. 6 grants the supposition on this work.

2 Literature Review

According to statista, Edge computing may be a fast-growing showcase, with the figure worldwide income set to reach nine billion U.S. dollars by 2024. The capital use on edge computing gadgets and framework is anticipated to sum to an amazing 146 billion U.S. dollars by 2028. A few businesses such as budgetary administrations and fabricating have as of now made broad utilize of edge computing in their day-to-day businesses, and nations like China are full-heartedly grasping the potential that the modern innovation holds for them.

Granjal et al. [8] demonstrate a complete analysis of security and privacy according to the protocols in each layer of the OSI model, and their results for the common IoT domains. Batalla et al. [9] The Smart Home Domain focuses on this topic and gives a detailed report on the most common threats to using the cloud platform in smart home systems, and key concepts such as protocol analysis and security implications. Jose has been introduced as an algorithm to get every location in a smart home and tested in a factual environment with successful results. We often look at smooth homegrown solutions for security issues, followed by instability, the costs we have limited, which is an important aspect for consumers [10].

Tiwari et al. [11] asserted that there is greater clarity in IoT, enabling Artificial Intelligence, the IoT features to build more LinkedIn systems, and the Active Engagement System customer interaction, technology optimization, waste reduction and improved data collection. In addition, he noted that security, privacy, complexity, flexibility and compliance are the main disadvantages of the Internet of Things.

Sharma et al. [12], Improved Security and Surveillance, Disability Assistance, Automated Industrial Control, Smart Agriculture, Home and Building Automation are important applications of the Internet of Things. However, it is not limited and is currently the smart city for all IoT applications and has become an important area of study.

The detailed review on challenges and issues of the Internet of things and categorize the problems in global use, standards and protocols, operational and technical and data and software. Privacy and security were the most significant challenges when IoT enabled for universal use. Data integrity, Data Authentication, Data Confidentiality, and Trust and Governance affects the global use of the Internet of things. Also, detailed how interoperability, mobility, scaling, human–machine interaction, and dependability are the significant operational and technical challenges of IoT presented by Sarhan [13].

A periodical of the applications and remunerations of authority adding in a variety of fields and also challenges such as service management, programmability, security, naming, data capture, optimization metrics, and privacy presented by Zhang [14].

Assuncao et al. [15] describes emerging technologies for distributed data stream processing, as well as cloud resource management techniques for elasticity management. But it is difficult to develop and implement stream processing applications under different distribution structures.

Bilal et al. [16] offered an outline of edge computing motivation and edge technology. Concerns in this paper include resource management and allocation, resource scalability, methods, data capture, fault tolerance, and privacy in service quality, conservation and edge computing. Table 1 shows some of the detailed reviews about edge computing in different application domains, also with future directions.

This review commends of an Edge Computing worldview is silent at a primary organize, and the logical communal requirements additional period to get it in what way to utilize its preferences to provide the administrations accessible to end-user clients. In expansion, the study centers as it were on the focal points of a superiority computing deprived of tending to great contests, such as the determination of the finest procedures that IoT operators got to take after in arrange to attain great benefit. Edge registering brings information handling power at the edge of the system, closer to the wellspring of the information.

3 Edge Computing Models Enabling IOT

Edge computing is one way to improve data processing in cloud-based computing systems at the network edge, which is closer to the source of the data. There are many advantages to using sophisticated computing technology to help with latency issues and bring bandwidth-intensive content closer to the user, as well as support future network infrastructure. Let's first look at the various state-of-the-art computing technologies that have been implemented to help promote and monitor growth in the use of IoT and 5G.

3.1 Portable Edge Computing

Portable edge computing may be an arrange design that empowers computing and capacity assets to convey substance over the radio get to organize, and to conclusion clients to make strides the viability of the network [25]. To do this, the MEC adjusts the load on the radio connection to increase network performance and reduce the need for long-range backhauling. It is hoped that network requirements will further spread to IoT and 5G compliant technologies and devices. Mobile Edge Computing enables operators to more efficiently meet traffic and resource demands while laying the foundation for the future smart, next-generation networks. Mobile Edge Computing supports services such as augmented reality, augmented reality, and the Internet of Things, to give startups and startups the opportunity to adapt to emerging technologies.

Table 1 Summary of edge computing reviews

Author	Methodology	Result and discussion	Future scope
Alonso et al. [17]	Developed a new edge computing engineering dwell in safekeeping and secrecy by utilizing block chain innovation	The place engineering is tried with IoT stage in keen agro industry to decrease transfer speed taken a toll between edge and cloud	To implement RA in different scenarios like IoT records in immense platform plus smart cities and brainy supervision also in healthcare system
Luo et al. [18]	Computation offloading method (COM)	The COM approach proposed was effective in solving multi objective optimization problems	Discover an offloading procedure to realize most extreme vitality utilization sparing in versatile gadget
Alamgir Hossian et al. [19]	Layered approach	A few quantitative and subjective thinks about are executed to preparing crude IoT information, which is successful in terms of inactivity and giving circumstance mindfulness choice to clients	Executing circumstance mindfulness (SA) framework in an IoT based shrewd city environment
Ray et al. [20]	Novel edge IoT based architecture	Created the novel edgent empowered edge IoT framework for e-healthcare and illustrated with test bed. The test was performed on two situation like client with fever and client without fever	To overcome the taking after issues like edge-IoT engineering usefulness and operational issues, prerequisite capability and determination criteria
Dhavijay et al. [21]	Wireless body area network (WBAN)	Gives the survey almost WBAN based IoT healthcare framework and different advances of IoT arrange frameworks	WBAN is the effective patient session technology so it will be used in the IoT and delivering home health also monitoring

Table 1 (continued)

Author	Methodology	Result and discussion	Future scope
Zamora et al. [22]	Precision agriculture (PA) with three tier platform	Developed an open source three tier platform like local plane (cyber physical system) to gather data, edge plane is for monitoring and managing PA tasks, finally cloud plane will collect the current and past data records. The system was implemented for tomato plants to monitoring growth of the crop	Experience the lack of modularity and delocalization. For this reason future work aims to use modular turnkey system to falling industrial time in situation active and conservation
Sha et al. [23]	Review based on IoT security at edge layer	Checked on almost edge centric IoT, edge based security plans and inquire about issues	The virtual machine and hypervisors are basic for the security in entirety edge layer. So these virtual machine to be lightweight conjointly require secure working framework like SeL4
Khan et al. [24]	Review edge computing techniques	Categorized the up-to-date in edge computing based on the application domain like cloudlet, mist and portable edge computing	Tended restriction of edge computing worldview. The restrictions are energetic charging component, genuine time application bolster, security, joint trade show, repetition, adaptable design, come up short over capabilities and asset administration

3.2 Fog Computing

Fog networking, or "fogging" directly, is a term used to describe a decentralized computing system. Mist computing takes an edge of the system, retaining data, handling, loading and submissions in a highly consistent and productive position concerning the cloud network and the source of data [26]. Fog Intelligent aims to expand mist computing and amenities on the control of the web and seek to

reduce the transmission, analysis and storage of data transmitted to the cloud. Data collected from IoT sensors and other devices is conveyed to the mist for review and dispensation. However, these sensors and devices are far from responsive, and cloud calculations allow short-term analytics and processes at the arrange edge to shrinkage the sum of information shown to the cloud [27].

3.3 Cloudlets

Cloudlets represent small-scale mist data hubs on the authority of the system and the second tier of the three-tiered hierarchy: cloud-to-cloud phone or smart device. Cloudlets aims to development resource-intensive and intelligently versatile applications by giving versatile gadgets with less idleness and more proficient computing assets in geographical proximity [28]. This reduces the latency delay usually associated with cloud computing in the WAN. While the 5G network is moving up on the horizon, cloudlets have evolved to support resource-intensive applications such as machine learning, speech recognition, and language processing applications. Due to the increased demand, they provide the necessary support for the 5G network that comes with them. Cloudlets need to be decentralized and spread widely to provide the greatest amount of network coverage possible. It helps to get power from nearby mobile computers.

3.4 Micro Data Centers

Microdata centers are small, access-level systems that provide all the necessary components of a traditional data center. Microdata centers are more suitable for some edge computing applications than traditional data centers because they are usually smaller in size and can be deployed indoors and outdoors in harsh conditions. This makes them an ideal solution for edge computing because they can be used locally in a data source to suit the needs of those who choose to implement them [29]. Microdata centers may be more attractive to small and medium-sized ones—they may not have data centers because large companies have more resources and do not need such a clarification. While the growth of the Internet of Effects is driving the expansion of new shrewd policies and IoT sensors, the souk for the micro data center is estimated to be worth $32 billion over the next two-and-a-half years.

Some of the edge computing techniques discussed in the next generation network infrastructure are discussed above. The Internet of Things and the continued advancement of the 5G wireless networking network are viable when changes in other areas, such as versatile edge and haze computing, are ready and provide a secure platform for innovation. MEC, Fog Computing, Cloudlets and Micro Data Center play important roles when it comes to the future of network computing, but their first test is when 5G is ready for roll-out.

In this segment we review a series of contemporary developments in edge computing systems and design, programming models and applications [30]. Five different application scenarios are discussed: general usage, vehicle information analytics, shrewd domestic, video stream analytics and fundamental realism. In this work, widespread use and smart home edge computing systems and equipment are described below.

4 General Utility Applications

In general, the application landscape consists of ten different edge computing systems with different end devices such as IoT devices, mobile devices, firewall nodes and edge, as well as cloudlets, mobile devices, local servers, PCs, home gateway, router, server, local cluster, FOG, MEC and seaport.

4.1 Cloudlet

Cloudlet could be a trusted, clever computer or computer cluster that is well connected to the Web and is accessible for versatile gadgets within the encompassing region [31]. It upgrades "Cloud to Mobile Device" from "Cloud to Mobile Device", the basic two-stage architecture of Mobile Cloud Computing. Cloudlet can support users as a stand-alone cloud, which is currently becoming a "minor server" or "statistics focus in the case". However, cloudlet has not been developed under the name Edge Computing, and its design and concept of edge computing are due to three key features, such as softness, great resources, and proximity to customers. As an edge computing system.

4.2 CloudPath

Cloudpath reduced the difficulty of designing and implementing Path Computing, in order to maintain the most popular RESTful development model of the cloud application. Cloudpath is based on the first observation that decomposition of RESTful stateless properties is possible by the existence of a specific storage layer. Cloudpath has facilitated the creation and deployment of computing applications by extending traditional storage abstractions to the Data Center hierarchy embedded within the geographic range of networked storage [32].

4.3   PCloud

PCloud [33] coordinating edge computing and capacity assets to bolster consistent portable administrations. Basically, these assets are virtualized through the primary virtualization layer called STRATUS [34]. Making a bunch of distributed resources that can hunt for modern assets to memorize almost changes in assets. The runtime method is capable for asset determination and allotment through the asset choice interface; the runtime instrument chooses and interfaces fitting resources based on diverse choice prerequisites. Beneath the Asset Get to Control Arrangement, the reserves will be combined to supply comparative administrations for outside applications and make a modern worldview. In expansion, the PCLoud framework moreover gives basic administration administrations such as authorization administration and client information accumulation to control access to assets for other clients.

4.4   ParaDrop

The University of Wisconsin—Madison WiNGS Laboratory developed a computing system called Paradrop [35]. It makes computing and capacity assets closer to portable gadgets and makes information assets accessible to third-gathering designers. The objective is to carry insights to the superiority of the network. Paradrop Edge replaces the actual entry point for a computer network, which supports standard server separation devices and utilities such as a multi-tenancy device. In addition, Paradrop leverages lightweight container virtualization technology.

4.5   SpanEdge

SpanEdge [36] may be a investigate venture of an Institute in Sweden. It interfaces central mist hub with the dominant edge hub to decrease arrange inactivity in WAN associations and the programming environment. This permits the framework to run close the source of the information. Designers ought to center on building gushing applications without considering the area and dispersion of information sources. The information midpoint in Span Advantage is made up of two planes: the mist information center to begin with equal, and the edge information center moment level. Fractional gushing preparing errands run on the internal edge to diminish latency and increment execution.

Spilling handling is one of the foremost vital sorts of applications in edge computing, where information from distinctive topographical sources are produced by diverse information sources and persistently transmitted in stream shape. All crude information is ordinarily transmitted to the WAN's information middle waiter, and brook handling frameworks such as Apache Stimulus and Flink are made and

arranged for a consolidated information center. In any case, this strategy does not switch enormous information created by numerous gadgets on the edge of organize impacts, and when the application requires less idleness and forecast, the circumstance gets to be more awful.

4.6 Cloud-Sea Computing Systems

The Chinese Academy of Sciences is launching the Cloud-Sea Computing System [37] in 2012. In cloud-sea figuring schemes, the term "cloud" denotes to data centers and "sea" mentions to the fatal. The plan of this venture is spoken to at three stages: the global framework design equal, the information center server and the capacity framework near, and the computer chip-near [38]. The architecture of cloud computing in the ocean has four distinctive structures, such as ternary computing by sea equipment, cooperation with locality, scalability for ZB and trillion devices, and minimal deployment to existing ecosystems.

4.7 Smart Home Applications

Another scenario that plays an imperative character in edge reckoning is the management of IoT strategies in insolent home environments. One common factor is the privacy issue of a variety of home appliances. There are two different edge computing systems, Vigilia and HomePad.

Vigilia

In [39], Vigilia proposed tightening the smart home system by limiting network access to system devices. An authentication framework has been introduced for applications that reject default access rules and implement authentication at the network level for API-granularity. Only check the time spent on the router for those communications, thereby helping customers secure their home appliances.

HomePad

The HomePad system in [40] also proposed setting IoT applications and presenting privacy-aware centers to diminish security issues. Home Pad permits clients to characterize protection approaches for how apps get to and handle their information. By requiring applications to utilize clear data streams, they can utilize Preamble rules to confirm that they have the potential to damage the security approach indicated amid establishment.

5 Open Challenges in Edge Computing

The complication of these edge computing frameworks has driven to numerous specialized issues such as versatility administration, security and security, benefit revelation, benefit conveyance and versatility, scalability, versatility, low-cost tolerant deployment model reliability and resource management. Still, some important challenges need to be addressed.

Xiao et al. [41] described an edge computing applications are affected by the security attacks named as Distributed Denial of service (DDos) attacks, Cross Station attacks, Malware inoculation attacks, Confirmation and Authorization attacks, Man in the central bouts, Bad data inoculation attacks are pointing edge computing substructures in real world. The idea behind this is that the author proposed direction to secure the Edge Computing System with an integrated outline. It consists of 3 layers, designated as External Fine Granular Access Control Layer, Middle Sanctuary Functional Sheet and Internal Hardware Discrete OS Layer.

Liu et al. [42] reviewed an Edge Computing security attacks targeting autonomous vehicles that pose a threat to personal and public safety. Autonomous vehicle protection includes sensor protection, functioning scheme safety, and controller scheme safety and communiqué security. This Autonomous Driving Edge covers various layers of computing stack.

5.1 IOT and Edge Computing Challenges

IoT and Edge devices are used by lightweight Operating Systems and limited storage capabilities projected for narrow, special tasks. For example AWS Lambda at the edge platform the content delivery network (CDN) feature within Amazon Cloud Front with benefits in the form of speed and the cost by pushing code closer the users through limited edge compute executions.

Cost and Connectivity

IoT sensors and devices services are managed locally compared to the edge devices. The edge devices are handled large services inside the infrastructure because of more number of transactions are processed locally so that less number of data traverses to and from of the cloud. The connectivity is a main problem while reaching out to an on-premises data to the cloud.

Security

In distributed data processing, edge computing is the major challenge. The security in edge computing usable for one edge computing application it's not applicable directly migratable to another situation because of variety of reasons for example heterogeneity of edge devices and communication protocols. In this section, illustrate that the Edge computing Challenges and Security issues in Table 2 [2].

Table 2 Challenges and security issues

Challenges	Security issues
Confidentiality and integrity	Achieving confidentiality and integrity of data in transit to and from the cloud-edge-device provider
Power consumption	Mist usage added knobs Power ingesting is higher than central cloud
Data security	Data producing knobs are spread Provided that confirmation and sanction scheme for the total nodes is not an informal task
Fault tolerance	Failure of a node should be immediately fixed
Architecture	Programming architecture

6   Conclusion and Future Work

Edge Computing is a new model that gives users the ability to network and store from remote cloud to IoT and 5G context. The Superiority Computing principle is auspicious in providing a better user experience for better services and applications. Newly developed edge computing systems and technologies normally decrease overhead records dispensation and storage and increase the reliability and effectiveness of transportable data analytics. This paper also discusses the basic concepts and systems and tools related to edge computing and classifies them as advanced computing such as cloudlets, mist, and portable edge computing and micro-data center application domains. This work gives future researchers an excellent idea to understand the superiority computing pattern and advance investigation to solve unexpected problems. Future research should explore investigation tendencies in edge-fog computing systems.

References

1. Hassan N, Gillani S, Ahmed E (2018) The role of edge computing in internet of things. IEEE Commun Mag 99:1–6
2. Yaqoob I, Ahmed E, Gani A (2016) Mobile ad hoc cloud: a survey. Wirel Commun Mob Comput 16(16):2572–2589
3. Liu M, Yu FR, Teng Y (2019) Distributed resource allocation in blockchain-based video streaming systems with mobile edge computing. IEEE Trans Wireless Commun 18(1):695–708
4. Pace P, Aloi G, Gravina R (2019) An edge-based architecture to support ancient applications for healthcare industry 4.0. IEEE Trans Ind Inf 15(1):481–489
5. Shaukat U, Ahmed E, Anwar Z (2016) Cloudlet deployment in local wireless networks: motivation, architectures, applications, and open challenges. J Netw Comput Appl 62:18–40
6. Bao W, Yuan D, Yang Z (2017) Follow me fog: toward seamless handover timing schemes in a fog computing environment. IEEE Commun Mag 55(11):72–78
7. Ahmed E, Akhunzada A, Whaiduzzaman M, Gani A (2015) Network centric performance analysis of runtime application migration in mobile cloud computing. Simul Model Pract Theor 50:42–56

8. Granjal J, Monteiro E, Silva JS (2015) Security for the internet of things: a survey of existing protocols and open research issues. IEEE Commun Surv Tutorials 17(3):1294–1312
9. Batalla JM, Vasilakos A, Gajewski M (2017) Secure smart homes: opportunities and challenges. ACM Comput Surv 50(5):75:1–75:32
10. Jose AC, Malekian R (2017) Improving smart home security: Integrating logical sensing into smart home. IEEE Sens J 17(13):4269–4286
11. Tiwary A (2018) Internet of Things (IoT) research, architectures and applications. Int J Future Revolution Comput Sci Commun Eng 4(3):23–27
12. Sharma V, Tiwari R (2016) A review paper on IOT and it's smart applications. Int J Sci Eng Technol Res (IJSETR) 5(2):472–476
13. Sarhan QI (2018) Internet of things: a survey of challenges and issue. Int J Internet Things Cyber-Assur1(1):40–75
14. Shi W, Cao J, Zhang Q, Li Y, Xu L (2016) Edge computing: vision and challenges. IEEE Internet Things J 3:637–646
15. de Assuncao MD, da Silva Veith A, Buyya R (2018) Distributed data stream processing and edge computing: a survey on resource elasticity and future directions. J Netw Comput Appl 103:1–17
16. Bilal K, Khalid O, Erbad A, Khan SU (2018) Potentials, trends, and prospects in edge technologies: fog, cloudlet, mobile edge, and micro data centers. Comput Netw 130:94–120
17. Alonso RS, Corchado JM (2019) A review of edge computing reference architectures and a new global edge proposal. Futur Gener Comput Syst 99:278–294
18. Luo Y, Peng K (2019) A computation offloading method over big data for IoT-enabled cloud-edge computing. Futur Gener Comput Syst 95:522–533
19. Alamgir Hossain SK, Anisur Rahman MD (2018) Edge computing framework for enabling situation awareness in IoT based smart city. J Parallel Distrib Comput 122:226–237
20. Ray PP, Dash D (2019) Edge computing for Internet of Things: A survey, e-healthcare case study and future direction. J Netw Comput Appl 140:1–22
21. Dhanvijay MM, Patil SC (2019) Internet of Things: a survey of enabling technologies in healthcare and its applications. Comput Netw 153:113–131
22. Zamora Izquierdo MA, Santa J (2019) Smart farming IoT platform based on edge and cloud computing. Biosyst Eng 177:4–17
23. Sha K, Yang TA (2019) A survey of edge computing based designs for IoT security. Dig Commun Netw (DCN) 1–12
24. Khan WZ, Ahmed E (2019) Edge computing: a survey. Futur Gener Comput Syst 97:219–235
25. Ahmed A, Ahmed E (2016) A survey on mobile edge computing. In: 10th International conference on intelligent systems and control (ISCO). pp 1–8. https://doi.org/10.1109/ISCO.2016.7727082
26. Yi S, Li C, Li Q (2015) A survey of fog computing: concepts, applications and issues. In: Proceedings of the workshop on mobile big data ACM. pp 37–42
27. Bao W, Yuan D, Yang Z, Wang S, Li W, Zhou BB, Zomaya AY (2017) Follow me fog: toward seamless handover timing schemes in a fog computing environment. IEEE Commun Mag 55(11):72–78. https://doi.org/10.1109/MCOM.2017.1700363
28. Shaukat U, Ahmed E, Anwar Z, Xia F (2016) Cloudlet deployment in local wireless networks: motivation, architectures, applications, and open challenges. J Netw Comput Appl 62:18–40
29. Yaqoob I, Ahmed E, Gani A, Mokhtar S, Imran M, Guizani S (2016) Mobile ad hoc cloud: a survey. Wirel Commun Mob Comput 16(16):2572–2589
30. Liu F, Tang G (2019) A survey on edge computing systems and tools. Proc IEEE 107(8):1–24
31. Satyanarayanan M, Bahl V, Caceres R, Davies N The case for vm-based cloudlets in mobile computing. IEEE Pervasive Comput
32. Mortazavi SH, Salehe M, Gomes CS, Phillips C, de Lara E (2017) Cloudpath: a multitier cloud computing framework. In: Proceedings of the second ACM/IEEE symposium on edge computing. ACM, p 20
33. Jang M, Schwan K, Bhardwaj K, Gavrilovska A, Avasthi A (2014) Personal clouds: sharing and integrating networked resources to enhance end user experiences. In: Proceedings of IEEE INFOCOM. IEEE, pp 2220–2228

34. Jang M, Schwan K (2011) Stratus assembling virtual platforms from device clouds. In: Proceedings of the IEEE international conference on cloud computing. IEEE, pp 476–483
35. Liu P, Willis D, Banerjee S (2016) Paradrop: enabling lightweight multi-tenancy at the networks extreme edge. In: IEEE/ACM symposium on edge computing (SEC). pp 1–13
36. Sajjad HP, Danniswara K, Shishtawy A, Vlassov A (2016) Spanedge: towards unifying stream processing over central and near the-edge data centers. In: IEEE/ACM symposium on edge computing (SEC). IEEE, pp 168–178
37. Xu Z-W (2014) Cloud-sea computing systems: towards thousand-fold improvement in performance per watt for the coming zettabyte era. J Comput Sci Technol 29(2):177–181
38. Fielding RT, Taylor RN (2000) Architectural styles and the design of network-based software architectures, vol 7. University of California Irvine Irvine, USA
39. Trimananda R, Younis A, Wang B, Xu B, Demsky B, Xu G (2018) Vigilia: securing smart home edge computing. In: Proceedings of the IEEE/ACM symposium on edge computing (SEC). IEEE, pp 74–89
40. Zavalyshyn I, Duarte NO, Santos N (2018) Homepad: a privacy-aware smart hub for home environments. In: Proceedings of the IEEE/ACM symposium on edge computing (SEC). IEEE, pp 58–73
41. Xiao Y, Jia Y, Liu C (2019) Edge computing security: state of the art and challenges. Proc IEEE 107(8)
42. Liu S, Liu L, Tang J (2019) Edge computing for autonomous driving: opportunities and challenges. Proc IEEE 107(8)

Comparison of Rectangular Tunnel with Shield Jacking Support and Pipe Roof Support in Finite Element Method

M. Y. M. Nasir, H. Mohamad, and H. Alarifi

Abstract Methods of the Gaussian distribution curve and Finite Element Method (FEM) were applied in predicting ground settlement however, those equations are still incomplete with a trough width parameter, *K* of soil condition that has evolved from the developing of Kenny Hills tunnel in Malaysia. Therefore, this paper of research analyses pipe roofing and shield support of rectangular tunnel using FEM by PLAXIS 2D software to estimate in occurrences of ground deformations. Charts of maximum ground surface settlement and trough width parameter are developed to show the difference between both lining supports through parametric study comprises variation of soil properties. Results show that pipe roofing support is suitable to be adopted in weak soil condition while shield support is suitable in soil condition with higher stiffness value. *K* can be concluded with value equal to 0.6 for the rectangular pipe roofing support and 0.7 for rectangular shield support in Kenny Hills soil formation.

Keywords Rectangular · Tunnel · Shield · Pipe · Support

1 Introduction

Rectangular tunnel is becoming popular and has been used in constructions especially by developed country. Hence, for the past few decades improvements have been done on its support systems in order to prevent deformation of ground especially on a weak ground condition. Generally, shield and pipe roofing support were used for constructing rectangular shaped tunnel.

Fang et al. [1] have go through on shield support and Koyama [2] has discussed further on the benefits of the shield support. Koyama [2] found shield support can be applied in different type of soil condition [2] due to rigid structure of shield while pipe roofing support can prevent the settlement of ground above the tunnel using a

M. Y. M. Nasir · H. Mohamad (✉) · H. Alarifi
Civil and Environmental Engineering Department, Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia
e-mail: muhammad_18003578@utp.edu.my
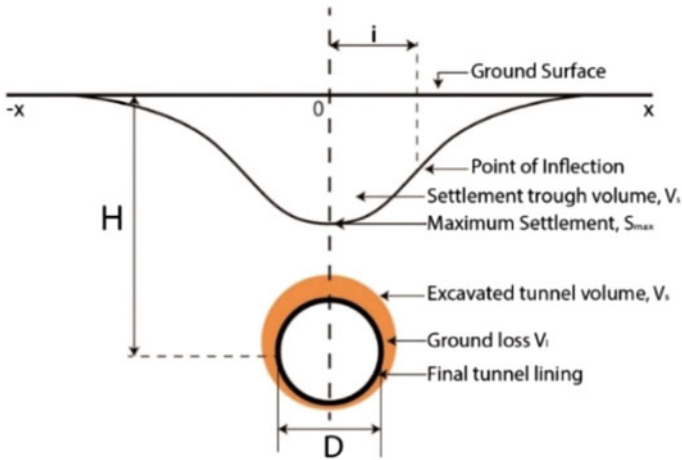
Fig. 1 Settlement trough of Gaussian form

series of steel pipe by consolidate the ground stress and disperse the ground stress to reduce the excavation stress during the tunnel excavation work [3]. The design of the pipe roof support may vary between the project because it solely based on experience [4].

Ground deformation is a major problem that crucial to be solved prior to develop the underground systems which is not able to be seen directly through underground layout. Furthermore, different properties and different characteristics of underground soil at different coordinate makes prediction on ground deformation hardly to be calculated manually. Therefore, simulation will be done to predict amount of risk and to identify the affected surrounding and structure constructed in underground space.

In order to determine the settlement above the tunnel, Peck [5] proposed an equation to develop transverse settlement above the tunnel known as Gaussian distribution curve as shows in Fig. 1.

Equation 1 give a settlement at various point of the trough,

$$S_x = S_{\max} e^{-\frac{x^2}{2ix^2}} \tag{1}$$

where:

S_x the settlement profile at the surface
S_{max} the maximum vertical settlement
i_x the trough width parameter which, physically, is the distance from the tunnel axis to the point of inflection of the curve.

Volume of the surface settlement profile is given by,

$$V_s = \sqrt{2\pi} i_x S_{\max} \tag{2}$$

$$i_x = K z_0 \tag{3}$$

where

K constant value

z_0 depth of the tunnel axis.

The width through parameter, K, varies with the type of soil. Kimura and Mair [6] suggested that K values must be equal to 0.5 for clay soil. Meanwhile, O'Reilly and New [7] recommend that K values must be between 0.2 and 0.3 for granular soil with condition of tunnel depth less than 10 m, 0.4–0.5 for stiff fissured clay, 0.5–0.6 for Glacial deposits and 0.6–0.7 for silty clay deposits. O'Reilly and New [8] also had suggested K values must be 0.4 for stiff clays, 0.7 for soft silty clays. In contrast, Mair and Taylor [9] suggested differently in K values where K values that they emphasized were 0.5 for all clay soil and 0.35 for granular soils. Khoo et al. [10] stated that an appropriate K to be used was 0.5 for soil type encountered in Klang Valley of Malaysia. On the other hand, Yeates [11] suggested that K value should be from 0.2 to 0.3 for granular material above the water table. Moreover, Rankin [12] had proposed K value which were 0.4–0.5 for stiff fissured clay, 0.5–0.6 for glacial deposits, 0.6–0.7 for silty clay. Overall K is depending on soil properties that significant to consider even though K value does not include in the Gaussian curve distribution that mostly used by researchers, due to none of the research had found K value that specifically use for rectangular or box tunnel. In addition, all trough width parameter is produced by analyze the circular tunnel cross-section. Hence, this research study aims at producing charts for the tunnel designer in estimating the maximum settlement and searching tough width parameter value for the rectangular tunnel with either shield support or pipe roof support located at Kenny Hill soil condition.

2 Kenny Hills Soil Formation

Mostly, Kuala Lumpur, Malaysia covers Kenny Hill soil formation. This type of soil consisting of interbedded shales, mudstone, siltstone, and sandstones. Kenny Hills soil has undergone some metamorphic events resulting in changes of sandstone/siltstone to quartzite and schist/phyllite respectively. As stated by Ooi [13], the depth of Kenny Hills soil layer will be more than 10 m below the existing ground level and the soil formation becomes very hard with SPT greater than $N = 50$.

The investigation of the Kenny Hills engineering properties has been done by Refs. [14, 15]. From their study, the measured bulk unit weight mostly ranged from 15.8 to 21.9 kN/m^3 for residual soil and for highly weathered rock (Grade IV) is 24.0 kN/m^3.

For the effective shear strength parameters, [15] has stated that for the residual soil with SPT ≤ 100, the range for cohesion, c' is from 5 to 10 kN/m^2 and for angle

of friction, φ' is 28° while for the soil which has SPT greater than 100, the cohesion, c' is 15 kN/m^2 and angle of friction, φ' is 29°. For highly weathered rock (Grade IV), the equivalent Mohr–Coulomb strength parameters are 30 kN/m^2 for cohesion, c' and 34° for an angle of friction, φ' [16].

3 Method of Analyses

The simulation is based on greenfield model with setting as follows:

Model Dimension = 40 m × 40 m
Surcharge Load = 0 kPa
Soil Layer = Homogenous Soil
Tunnel Depth = 15 m.

The boundary condition is fixed by standard fixities, where the side vertical boundaries are fixed in horizontal x-direction but free to move vertically, while the bottom boundary is restrained from any movement in all directions.

For lining support, the steel pipe will be used in the simulation of pipe roof support with an outer diameter of 813 mm with 16 mm thickness. The steel pipe will be used as a pipe roof with grade S275. The tunnel shield support specification will set as in Table 1.

3.1 Parametric Study

Each simulation in PLAXIS 2D software will have one variable parameter and constants for other parameters by Hardening Soil method in as in Table 2.

Table 1 Shield support properties

Parameter	Value
Behavior type	Elastic
Thickness (m)	0.35
Stiffness (MPa)	21,000
Specific weight (kN/m^3)	38
Poisson's ratio	0.3

Table 2 A typical range of Kenny Hill soil for simulation

Soil parameters	Study case	Constant value	Variable value
General properties	γ_{unsat} (kN/m³)	$\gamma_{sat} = 20$ kN/m³ $c' = 5$ kPa $\varphi' = 28$ $E = 25$ MPa $v = 0.3$	$\gamma_{unsat} = 15, 15.5, 16, 16.5, 17, 17.5, 18$
	γ_{sat} (kN/m³)	$\gamma_{unsat} = 19$ kN/m³ $c' = 5$ kPa $\varphi' = 28$ $E = 25$ MPa $v = 0.3$	$\gamma_{sat} = 20, 20.5, 21, 21.5, 22$
Strength parameter	c' (kPa)	$\gamma_{sat} = 20$ kN/m³ $\gamma_{unsat} = 19$ kN/m³ $\varphi' = 28$ $E = 25$ MPa $v = 0.3$	$c' = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25$
	φ' (º)	$\gamma_{sat} = 20$ kN/m³ $\gamma_{unsat} = 19$ kN/m³ $c' = 5$ kPa $E = 25$ MPa $v = 0.3$	$\varphi' = 20, 22, 23, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 40$
Stiffness	E (MPa)	$\gamma_{sat} = 20$ kN/m³ $\gamma_{unsat} = 19$ kN/m³ $c' = 5$ kPa $\varphi' = 28$ $v = 0.3$	$E = 20, 30, 50, 75, 100, 150, 200, 250$
	v	$\gamma_{sat} = 20$ kN/m³ $\gamma_{unsat} = 19$ kN/m³ $c' = 5$ kPa $\varphi' = 28$ $E = 25$ MPa	$v = 0.1, 0.15, 0.2, 0.25, 0.3$

4 Result

4.1 Relationship Between Maximum Settlement and Soil Parameter

The amount of soil settlements affects the degree damage of the ground surface. Therefore, factors of settlement need to be identified at first in order to ensure least ground deformation. Water content within the soil is one of the factors that determine the level of settlement. Figure 2 proven a higher level of water in soil properties that cause greater settlement from applying for rectangular shield supports. However, rectangular pipe roof supports show a striking effect of constant settlement although water content level is increasing. This condition occurs because pipe roof support provide advance protection for soil deformation caused by tunnel excavation work
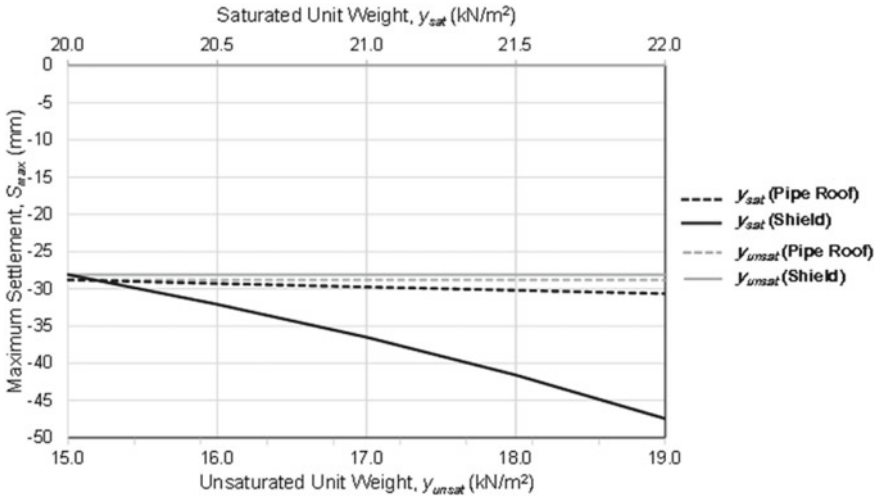
Fig. 2 Relationship between maximum settlement and general properties

while shield support provide soil protection just before the excavation. Next, cohesion and angle of friction as illustrated in Fig. 3 also affecting soil interaction when the pattern of settlement is decreasing inversely to cohesion and angle of friction of the soil with rectangular shield support. On the other side, rectangular pipe roofing support will be maintaining a constant settlement with increasing value of cohesion and angle of friction. Hence, constructing a rectangular pipe roofing support can
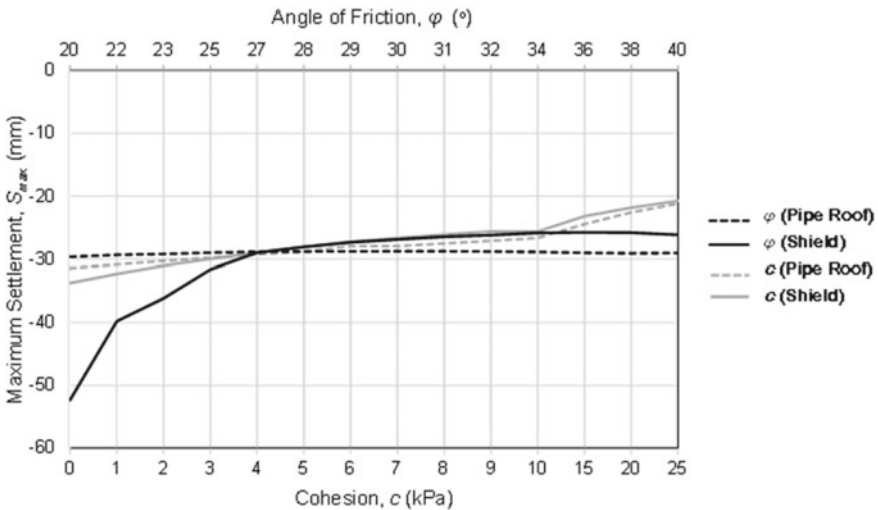
Fig. 3 Relationship between maximum settlement and shear strength properties
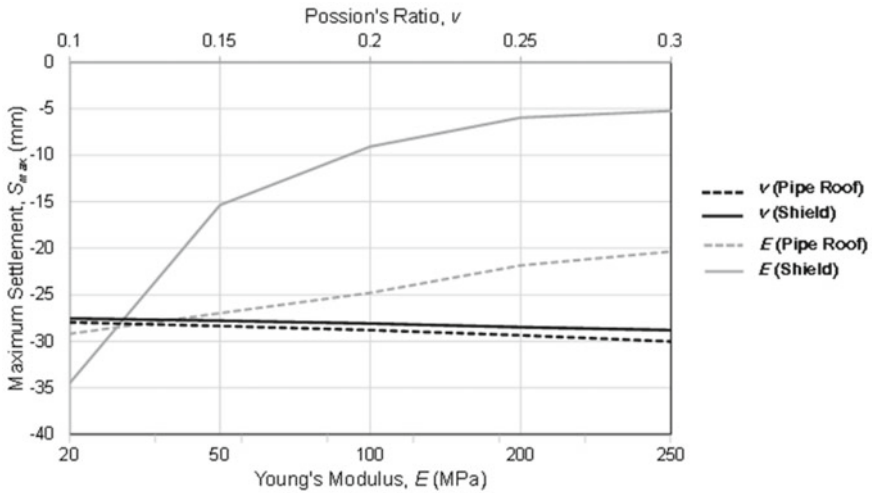
Fig. 4 Relationship between maximum settlement and stiffness properties

minimize the soil interaction thus reduce soil deformation at any degree of cohesion and angle of friction. Figure 4 illustrated a comparison between Young's modulus and Poisson's ratio towards soil settlement and results show reducing settlement with increasing Young's Modulus value on rectangular pipe roofing support which similar to the result of rectangular shield support. However, pipe roof supports produce less settlement reduction compared to shield support because of steel pipe installation for pipe roof support has disturbed the soil stiffness around the support's perimeter. Poison's ratio shows no effect toward the ground surface settlement for both supports.

4.2 Relationship Between Trough Width Parameter and Soil Parameter

By fitting the Gaussian distribution curve graph into the FEM graph, the value of the trough width parameter can be produced as a reference to estimate ground surface settlement in constructing any rectangular tunnel project which has approximate similar soil condition. Hence, investigations to determine a trough width parameter, K were done and mostly K value for rectangular pipe roofing support is 0.6 and K value for rectangular shield support is 0.7. In this study, the range of the trough width parameters are between 0.4 and 0.7 effected by saturated unit weight, angle of friction, cohesion and Young's modulus values used. Figure 5 illustrated the increment of water level within soil due to lesser in K value for rectangular shield support, while Fig. 6 resulted in inconsistent K value which between 0.6 and 0.7 for both angle of friction and cohesion properties. Hence, it is significant to further study in angle of friction and cohesion properties of soil that affected the K value inn understanding
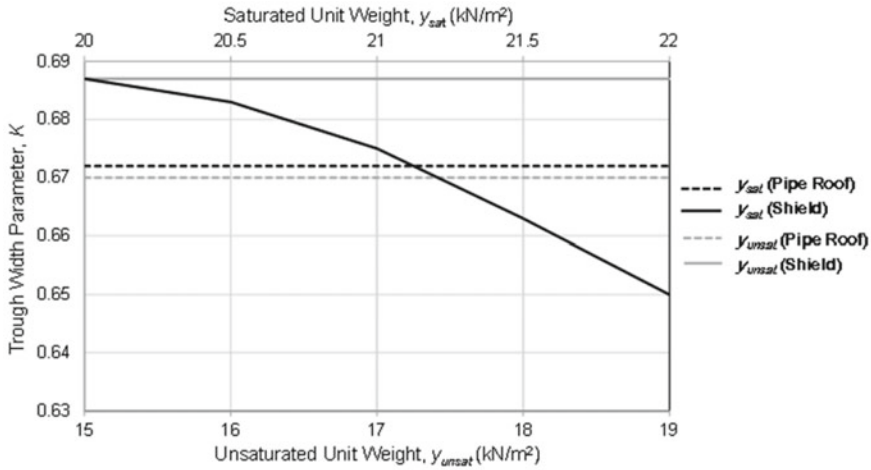
Fig. 5 Relationship between trough width parameter and general properties
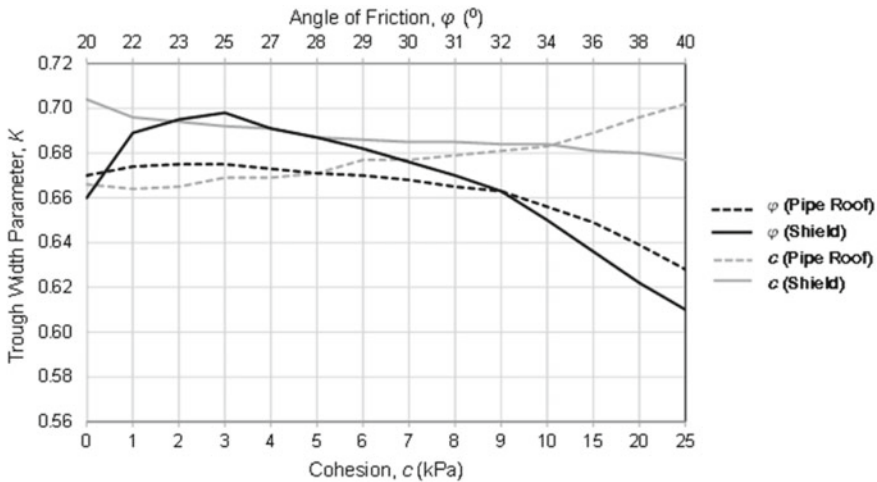
Fig. 6 Relationship between trough width parameter and shear strength properties

the stress of soil. Meanwhile, Fig. 7 illustrated reduction on K value from 0.7 to 0.4 when the value of Young's modulus increases for both types of supports with approximately constant K value on Possion's ratio properties. It can be concluded that unsaturated unit weight and Poisson's ratio properties give no effect to K value for rectangular tunnel with shield of pipe roof support.
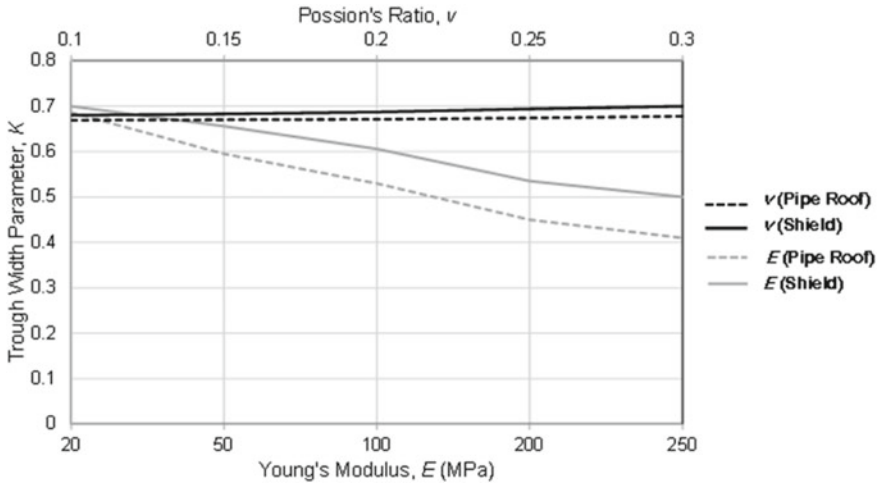
Fig. 7 Relationship between trough width parameter and stiffness properties

5 Conclusion

Previous literatures are yet discovering the trough width parameter use in constructing a rectangular tunnel cross-section. Hence, this simulation and analyses were done to investigate on the maximum settlement and trough width parameter for rectangular pipe roofing and shield support using Kenny Hill soil condition. The outcome of applying rectangular pipe roofing resulted in the suitability of constructing this type of support within the soil with properties of high-water content can still ensure the reduction of ground surface settlement. Furthermore, the result obtains also shows that by constructing a rectangular tunnel with pipe roof support, it is only suitable to be applied for low soil stiffness because the affected area by the pipe excavation in higher soil stiffness will increase soil plasticity area, hence reduction of ground surface settlement is lower. Meanwhile, a rectangular tunnel with shield support gives better respond in reducing the ground surface settlement in soil with high stiffness as shield support only involves with jacking process. Besides that, stiffness of soil is important to be considered prior to all construction works because it will affect the soil deformation during the excavation either rectangular pipe roofing or shield support. Finally, the trough width parameter value can be concluded with K equal to 0.6 for the rectangular pipe roofing support and 0.7 for rectangular shield support in Kenny Hills soil formation.

References

1. Fang Y, Chen Z, Tao L, Cui J, Yan Q (2019) Model tests on longitudinal surface settlement caused by shield tunnelling in sandy soil. Sustain Cities Soc 47:101504. https://doi.org/10.1016/j.scs.2019.101504
2. Koyama Y (2003) Present status and technology of shield tunneling method in Japan. Tunn Undergr Space Technol 18(2–3):145–159. https://doi.org/10.1016/S0886-7798(03)00040-3
3. Wu K, Shao Z (2018) Effects of pipe roof support and grouting pre-reinforcement on the track settlement. Adv Civil Eng.https://doi.org/10.1155/2018/6041305
4. Volkmann GM, Schubert W (2007) Geotechnical model for pipe roof supports in tunneling. In: Proceedings of the 33rd ITA-AITES world tunneling congress. pp 755–760
5. Peck RB (1969) Deep excavations and tunneling in soft ground. In: Proceedings of the 7th international conference soil mechanics and foundation engineering, Mexico City, State of the art volume, pp 225–290
6. Kimura T, Mair RJ (1981) Centrifugal testing of model tunnels in soft clay. In: Proceedings of the 10th international conference on soil mechanics and foundation engineering, Stockholm. pp 319–322
7. O'Reilly MP, New BM (1982) Settlement above tunnels in the United Kingdom—their magnitude and prediction. Tunnelling 82:173–181
8. New B, O'Reilly M (1991) Tunnelling induced ground movements: predicting their magnitudes and effects. In: 4th International conference on ground movements and structures. Pentech Press, Cardiff, pp 671–697
9. Mair R, Taylor R (1997) Theme lecture: board tunneling in the urban environment. In: Proceedings of the 14th international conference on soil mechanics and foundation engineering. Balkema, Hamburg, pp 2353–2385
10. Khoo CM, Idris NISI, Mohamad H, Rashid ASA (2018) Numerical evaluation of settlement trough width parameter. MATEC Web Conf 203:1–9. https://doi.org/10.1051/matecconf/201820304010
11. Yeates J (1985) The response of buried pipelines to ground movements caused by tunnelling in soil. In: GEDDES JDF (ed) Ground movements and structures. Pentech Press, Plymouth, pp 145–160
12. Rankin WJ (1988) Ground movements resulting from urban tunnelling: predictions and effects. Geol Soc Eng Geol Spec Publ 5(5):79–92. https://doi.org/10.1144/GSL.ENG.1988.005.01.06
13. Ooi TA (1986) Design and construction problems of foundation for the high rise structure in the Kuala Lumpur Areas. In: Proceedings of IEM-ISSMFE joint symposium on geotechnical problem, Kuala Lumpur. pp 112–122
14. Toh CT, Ooi TA, Chiu HK, Chee SK, Ting WH (1989) Design parameters for bored piles in a weathered sedimentary formation. In: Proceedings of the 12th international conference on soil mechanics and foundation engineering, Rio de Janeiro, vol 2. pp 1073–1078
15. Wong J, Sing M (1996) Some engineering properties of Kenny Hill formation in Kuala Lumpur. In: Proceedings of the 12th South-East Asian geotechnical conference, Kuala Lumpur. pp 179–187
16. Hoek E, Brown ET (1997) Practical estimates of rock mass strength. Int J Rock Mech Min Sci 34:1165–1186

The Effect of Skill Type on Skill-Gap in the Nigerian Construction Industry

M. S. Aminu, S. U. Kunya, I. Y. Mohammed, and S. A. Bustani

Abstract Skills gap and its effects has become a serious problem requiring urgent attention in the Nigeria construction industry. Basic literatures on construction practice left no doubts that the successful realization of construction projects requires the committed action of professionals and vocational trade operatives. The objective was to identify effects of skills-type on skills-gap related to Building projects in Abuja, Nigeria. Field survey of some building projects sites in Abuja compliment the secondary information on the existence of skill gap. The sample comprises both public and private projects, using both criterion-based and stratified random sampling techniques. 254 questionnaire was administered, 215 were returned, resulting to a response rate of 85%. The result shows that more than 80% of the Nigeria construction professionals are all male with years of experience ranges between 16 and 20 years, with academic qualification ranges from bachelor degree to Doctorate Degree respectively. Number of employees ranges within 10–50 people in most of the construction sites. From the result it shows that stake holders participation on training and development was not encouraging resulting to Poor workmanship, Construction waste, Rework as some of the major effects of skills-gap. The study therefore recommend that policy makers should develop skill acquisition programmed which include public private partnership, developmental training, on the job training. Government should come up with special scheme for the improvement of skills through the vocational and certification program to assist and encourage younger generation to various apprenticeships to reduce occupational segregation, provide the necessary requirement and standard training for both young men and women. The national directorate of employment (NDE), the national skills qualification framework (NSQF),

M. S. Aminu
Department of Building Technology, Jigawa State Polytechnic Dutse, Dutse, Nigeria
e-mail: mubarakaminu@jigpoly.edu.ng

S. U. Kunya (✉) · I. Y. Mohammed (✉)
Department of Building Technology, Abubakar Tafawa Balewa University, Bauchi, Nigeria
e-mail: sukunya@atbu.edu.ng

S. A. Bustani (✉)
Department of Quantity Surveying, Bayero University Kano, Kano, Nigeria
e-mail: sbahmad.qs@buk.edu.ng

479

the industrial training fund (ITF) and national board for technical education (NBTE) should ready to face the challenges of the recent skills-gap economy as we enter the third decade of the millennium for sustainability and productivity.

Keywords Construction industry · Effects · Nigeria · Skills-gap · Types

1   Introduction

The poor quality of project delivery can be attributed to the gradually diminishing in size of competent skilled construction workers and the influx of unskilled, inefficient and dissatisfied workers who see the sector as a last resort. There are indeed a few that see construction crafts as career worth coming to. This scenario will inevitably result in skill-gap, an issue of major concern in view of its invariable impact to the construction industry, citizens and national economy [12]. There are no exact data statistics of trade skill-gap in the construction industry, however, Medugu et al. [13] established that, there is skilled craftsmen shortage that needed urgent attention in the Nigerian construction industry.

Odediran and Babalola [15] observed that, the most needed large quantity manpower for building construction in Nigeria and globally are artisans and labourers. Okuntade [16] also observed that, despite the importance of craftsmen to the construction industry, many of them still remain untrained, even though the construction industry, all over the world have been implementing skills acquisition programmed to meet the demand of occasional change in technologies. Bilau et al. [3] suggested that, construction craftsmen are operatives who contribute skillfully with their hands in the practical realization of projects in a construction industry.

Siboe [18] stated that, training in the construction industry has a specific purpose, it should provide experience which develop the labor of employees in the area of skills, knowledge and attitude. Overtime, training has been a precursor to human resource planning which is a process for determining and ensuring that an organization will have enough number of skilled and experience persons available at the right time, quantity/quality and place, carrying out jobs that can meet the desires of the organization, which provide satisfaction for workers involved [18].

2   Statement of the Research Problem

Skill shortage is the major problem due to the rapid urbanisation in developing nations as a result to substantial increase in construction activities. This has tremendous consequences resulting to poor workmanship [10]. Lack of adequate training centers and training programmed in Nigeria, dominantly utilizing unskilled labour to maximize profit, resulting to constant rework leading to time and cost overruns, and

construction waste. Ameh and Itodo [2] assert that, most managers of the construction industry put little emphasis on the effects of construction waste contributed by unskilled labour to construction projects.

The shortage of skilled craftsmen in Nigeria has also been attributed to boom in the agricultural sector in recent times. Most craftsmen engage themselves on farming being a lucrative business, living closer to their families, making them happier than working for a long period of hours with a meager amount of pay. The quality shortage was as a result of lack of passion for the job, seeing the sector as a last resort that doesn't get promoted, eager of finishing work on time irrespective of its quality. This has drastic consequences on productivity in Nigeria. Construction waste sent to landfills has a drastic effects of polluting the environment, reducing the quality of construction, delay in construction activities leading to time and cost overruns in the study area.

In spite of the contributions of skilled craftsmen to the development of the construction industry, little efforts is being put on the issues that relate to the types of skill-gap and skill-gap effects in the industry. As a result, this is rarely been addressed and is little understood. There is need to address this problem by providing a clear theoretical and conceptual framework to understands of the basics and related concepts of skills-gap and its effects in the construction industry in Abuja Nigeria.

3 Research Aim and Objective

The aim of this research was to investigate the effects of skills-gap on types of skills-gap in the construction industry in Abuja Nigeria.

To achieve the above aim, the objective was:

(i) To study the effects of skill-gap on types of skills-gap in the Nigerian construction industry.

4 Skills-Gap and Its Effects in the Nigerian Construction Industry

According to Kurushi [12], skills is what an individual possesses which may be learnt informally and/or on-the-job. Skill-Gap is the difference in the skills required to do the job and the actual skills possessed by the employees. A "Skill-Gap" arises when a worker does not have all the skills necessary to do the job effectively. Therefore, skill-gap refer to the situation where a company has employees but they are not skilled enough to meet the organization's objectives.

4.1 Types of Skill-Gap in the Construction Industry

The types of skill-gap in an organization can be attributed to Skills shortages of masonry, Skills shortages of carpentry, Skills shortages of iron-bending, Skills shortage of plumbing, Skills shortage of electrician, Skills shortages of tiling, Skills shortages of painting, skills short-age of mixing/batching skills shortage of finishes, and skills shortage of scaffolding [12].

4.2 Effects of Skills-Gap

Skills-gap has a drastic effects on construction resulting to: poor workmanship, construction waste, rework, and client dissatisfaction, decrease in profitability of projects, and so on.

4.2.1 Construction Waste

Most construction projects has frequent changes in design often result in construction waste, time delays, cost overruns, quality defects and other negative impacts on construction. In recent years, many researchers investigated the effects of design change and more importantly the common causes behind the changes from different perspectives. Reviewing the existing literature on design changes, causes and effects, developing two taxonomies for change causes and change effects; and illustrating how the taxonomies can be used during the design change management process to reduce construction waste [20].

4.2.2 Rework

One of the most disturbing issues facing organizations in the construction industry is their inability to become quality focused. As a result sub-standard products and services often emanate, which inadvertently result in rework. Typically, rework is caused by errors made during the design and construction process. These errors appear downstream in the design process and therefore have a negative impact on a project's performance [17]. The lack of attention to quality, especially during the design and construction process, has meant that rework has become an inevitable feature of the construction work, and the costs have been found to be as high as 12.4% of total project costs. Such costs could be even higher because they do not represent schedule delays, and other intangible costs of poor quality. To reduce the cost and effect of rework, an understanding of its causal structure is needed so that effective prevention strategies can be identified and the effects of rework reduced or eliminated [17].

4.2.3 Client Dissatisfaction

The rate of client dissatisfaction is adversely affecting the timely delivery of construction projects. This assesses construction stakeholders' perception of the causes of client dissatisfaction and its effects on project delivery in a bid to proffer solution in minimizing its occurrences. This suggest that client's dissatisfaction related problems are the main causes of delays while time and cost overruns are the major identifiable effects of dissatisfaction in construction projects [1].

4.2.4 Decrease in Profitability of Projects

Change of design or nature of the work is one of the most significant causes of variation in construction projects, the client financial status can affect project progress, and this status often leads to change in work schedules, causing delay, affecting the quality of the project. Hindrance to prompt decision-making process leading to decrease in productivity and profitability [22].

4.3 Conceptual Framework of the Study

A concept is a plan, vision, or a symbolic representation of an abstract idea. A conceptual framework in research shows the researcher's position on the research problem, which gives direction to the study, and further shows the relationships that exist between different constructs which the study investigate. It may be an adoption of a model used in a previous study with modifications to suit the present investigation [10] (Fig. 1).

5 Sample Size

Using the yiros formula of sample size calculation

$$S = \frac{z^2 \times P^2}{E^2}$$

where

Z   is the z-score having a constant value of 1.96 at 95% confidence level [21].
P   is the standard deviation ranges from 20 to 100 distributions (my $P = 56$).
E   is the margin of error having a value of 4.

From the equation above
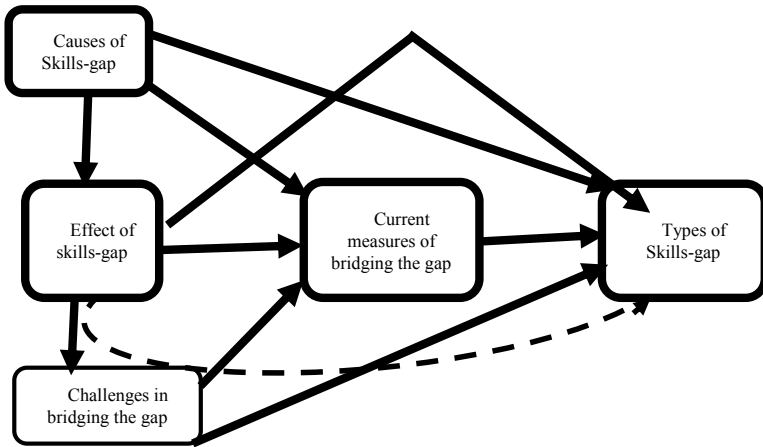
Fig. 1 Conceptual framework of the study

$$S = \frac{1.96^2 \times 56^2}{4^2} = \frac{3.84 \times 3136}{16} = \frac{12,042}{16} = \mathbf{752}$$

Sample size is **752** [11]; which was translated to **254** respondents and sampling technique was simple random sampling method.

Therefore according to Krejcie and Morgan [11] the sample of this population is 254. The sample comprises both public and private projects, using both criterion-based and stratified random sampling techniques. The choice of targeting building construction sites and professionals in this area is done for the following reasons: The proximity of the researcher's state of residence to the study area (FCT, Abuja), Abuja has the highest population of built environment professionals in the country, and it has many on-going building construction projects, The location of the construction activities in the same area makes the study more accurate in terms of cost and time.

Knowing the number of score and Ranking then in order from lowest to highest the formula

$$R = \frac{P}{100(N+1)}$$

to calculate the percentile Rank [21].

6 Result and Discussions

Table 1 represent the respondents educational qualification, respondent with first degree constitute the highest percentage (46.5%) while (20.9%) has master's degree and the least are those with PhD having (4.7%). Table 5 provides the results of

Table 1 Percentage of educational qualification of respondents

Qualification	Number of respondents	Percentage
National diploma	60	27.9
First degree	100	46.5
Master's degree	45	20.9
PhD	10	4.7
Educational specialization	Number of respondents	Percentage
Builders000	85	39.5
Architects	40	18.6
Quantity surveyors	20	9.3
Civil engineers	60	37.2
Others	10	4.6
Years of experience	No of respondents	Percentage
1–5 years	25	11.6
6–10 years	30	13.9
11–15 years	70	32.5
16–20 years	70	32.5
Above 20 years	20	9.3
Organizational position	No of respondents	Percentage
CEO/owner	15	6.9%
Projects manager	90	41.8%
Site engineer	80	37.2%
Others	30	13.9%
Number of employee	No of respondents	Percentage
Less than 10	0	0.0
10–50	140	65.1
50–200	50	23.3
Above 200	25	11.6

the professionals that participated in the survey. Where Builders has the highest percentage (39.5%), Architect (18.6%), Civil engineers (37.2%), Quantity surveyors (9.3%) and other Professionals (4.6%).This indicates that the respondents that participated in the survey are all professionals. Table 1 identify the working experience of respondents are within the range of 11–15 years has the highest percentage (32.5%), and those within (16–20) year's also. and above 20 years (9.3%). Despite the highest percentage goes to the group of 11–15 and 16–20 years of experience, therefore, the perception of the experienced Expertise is respected in this research. Table 5 further represent the organizational position of respondents in their various companies, projects manager has (41.8%), (37.2%) for site engineer, and other positions has 13.9%. Table 5 further shows the number of employees in their various organization,

with (65.1%) between 10 and 50 workers, (23.3%) between 50 and 200 workers and (11.6%) above 200 workers.

6.1 Research Objective: To Study the Effect of Skills-Gap on Types of Skills-Gap in the Nigerian Construction Industry

The objective was achieved by evaluating the measurement and structural models using the Partial Least Squares Structural Equation Modeling (PLS-SEM) analysis. The PLS-SEM approach involves a two-staged process of framework evaluation that involved test of convergent validity, discriminant validity, collinearity diagnostic, determining R^2 and path coefficients [9]. The output of these criteria were gauged against the recommended benchmarks provided in the extant literature. Table 2 shows the recommended bench- marks for decision on the parameters.

6.2 Measurement Model Assessment

Figure 2 shows the measurement model for effects of skills-gap on types of skills-gap. The model consist of five constructs measured in reflective mode [8]. The constructs are linked with one-headed arrows which indicate causal effect between them. Being a reflective model, the assessment of the measurement model involves testing the validity of the model, the indicator loadings, Average Variance Extracted (AVE), composite reliability and discriminant validity needs to be computed [8]. The main purpose of assessing measurement model is to test for construct validity. According to Hair et al. [9] construct validity is the degree to which the indicators of a latent construct represent the underlying latent construct.

Construct validity is measured by testing the convergent and discriminant validity. The convergent validity indicates the extent of correlation among the indicators of a latent construct. On the other hand, discriminant validity test the degree of uniqueness of a construct relative to other constructs in the research model [19].

6.3 Construct Validity/Convergent Validity

According to Hair et al. [9] convergent validity is the extent of correlation among the indicators of a latent construct. Hair et al. [8] recommended the use of indicator loading and AVE as criteria for judging convergent validity.

Table 2 PLS-SEM model evaluation criteria

Criteria		Description	Acceptable/Recommended threshold
Measurement model assessment criteria			
Convergent validity	Indicator reliability	Measures the variance explained by the observed variable towards underlying latent construct	Value ≥ 0.708 [9]
	Cronbach's alpha	Measures the indicators uni-dimensionality (inter-correlation) with their latent construct	$\alpha \geq 0.70$ [8, 14]
	Composite reliability	It is a measure of internal consistency	$\rho_c > 0.70$ [8, 14]
	Average variance extracted (AVE)	Average percentage of variation explained among the items of a construct	AVE > 0.50 [6]
Discriminant valid	Fornell-Larcker criterion	Measures the degree to which two conceptually similar concepts are distinct (in such a way that each latent variable shares more variance with its own block of indicators than with another latent variable)	$\sqrt{AVE}$ > latent variables correlations [6]
	Cross-loadings	A situation where a variable has two or more factor loadings exceeding the threshold value deemed necessary for inclusion in the factor interpretation process	Variable loading on any two or more factors at > 0.40
	Heterotrait-mono-trait (HTMT) ratio of correlations	Measures the degree to which two conceptually similar concepts are distinct (in such a way that each latent variable shares more variance with its own block of indicators than with another latent variable)	Correlation between any pair of construct should be < HTMT. [19]
Structural model evaluation criteria			
Collinearity		Measures the relationship between two or more independent variables	VIF < 5 is considered mild case of collinearity

(continued)

Table 2 (continued)

Criteria	Description	Acceptable/Recommended threshold
Measurement model assessment criteria		
R^2 (Coefficient of Determination)	Amount of variance explained in endogenous latent variable by all exogenous latent variable in structural model	$R^2 = 0.20$ is considered high in consumer behaviour studies [9] $R^2 = 0.25, 0.50, 0.75$ are considered weak, moderate and substantial respectively in [8] $R^2 = 0.19, 0.33, 0.67$ are considered weak, moderate and substantial [5]
β coefficient	Standardized regression coefficient that allows for a direct comparison between coefficients as to their relative explanatory power of independent variable	Higher magnitude indicates more explanatory power [7, 9]

Source Sinkovic et al. [19]
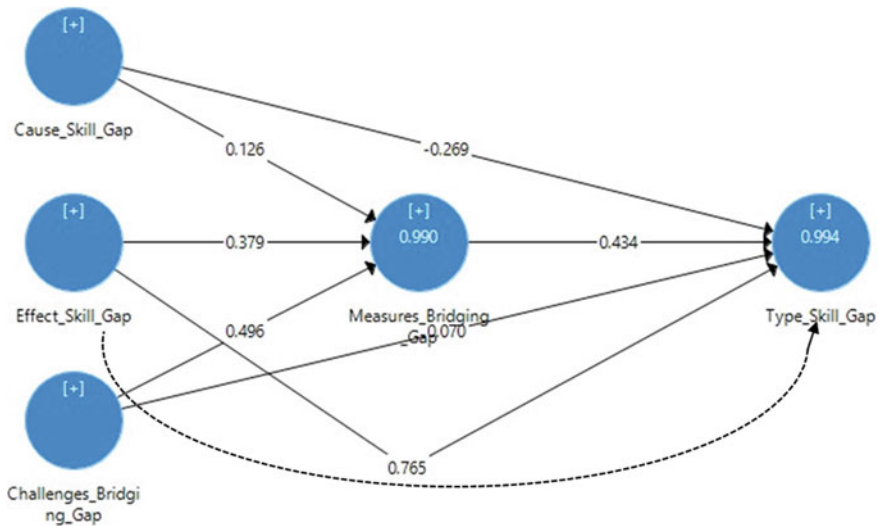
Fig. 2 Measurement model for effects of skills-gap on types of skills-gap

6.3.1 Indicator Reliability

Indicator reliability measures the absolute importance of individual measurement items in defining the latent construct. It shows the relationship between the latent construct and the measurement items [9]. A factor loading of 0.70 is considered sufficient [4] although [9] pointed that values less than 0.70 are also acceptable provided that the AVE is achieved.

6.3.2 Average Variance Extracted (Ave)

The second condition for ensuring convergent validity is the AVE. It is defined as the overall mean value of the squared loadings of the indicators of a latent construct [9]. The AVE measures the amount of variance a latent construct explained the measurement items with respect to the amount explained by measurement error [6]. Hair et al. [9] recommended 0.50 and above as suitable to ensure that at least two-third (2/3) of the variance in the indicator is explained by the construct.

Composite reliability (CR) measures the internal consistency of the measurement model. Hair et al. [9] described it as the degree of inter-correlations among indicators of a latent construct. The recommended value is 0.70. In an exploratory research, however, Hair et al. [8] pointed that 0.60 is equally acceptable.

Table 3 shows the outer loading of the indicators, AVE and composite reliability of the measurement model. As shown in the Table 3, the outer loadings, the AVE and CR for the two constructs all fall within the recommended thresholds of the respective parameter desired for convergent validity. The AVEs reported 0.884 while the CR reported 0.991 for causes of Skill Gap. The AVEs reported 0.881 while the CR reported 0.991 for challenges in bridging the gap. The AVEs reported 0.887 while the CR reported 0.992 for effects of skills-gap. The AVEs reported 0.895 while the CR reported 0.989 for measures of bridging the gap. The AVEs reported 0.886 while the CR reported 0.987 for types of skills-gap. The values were well above the recommended minimum of 0.50 and 0.70 for AVE and CR respectively. Similarly, all outer loadings of the indicators measuring the constructs reported values above 0.70 which further indicate the achievement of convergent validity.

6.4 Discriminant Validity

Discriminant validity measures the uniqueness of latent construct with respect to other constructs in the model. Discriminant validity is assessed using the Hetrotrait-Monotrait (HTMT) criterion. According to Henseler et al. (2015) the HTMT is the most conservative measure of discriminant validity when compared with the Fornell-Larcker or cross loading criteria. The criterion is to compare the calculated result with a value of 0.85 and 0.90 in such a way that any value in the table above the recommended threshold implies that discriminant validity is not achieved. Table 4

Table 3 Outer loadings construct validity and average variance explained

Construct	Measurement items	Outer loading	T statistics	P Values	CR	AVE
Causes of_Skill_G ap	CSGI	0.970			0.991	0.884
	CSGII	0.969				
	CSGIII	0.935				
	CSGIV	0.974				
	CSGIX	0.955				
	CSGV	0.928				
	CSGVI	0.957				
	CSGVII	0.961				
	CSGVIII	0.914				
	CSGX	0.876				
	CSGXI	0.913				
	CSGXII	0.967				
	CSGXIII	0.944				
	CSGXIV	0.894				
Challenges in_Bridging the_Gap	CBGA	0.936			0.991	0.881
	CBGB	0.965				
	CBGC	0.959				
	CBGD	0.943				
	CBGE	0.913				
	CBGF	0.913				
	CBGG	0.939				
	CBGH	0.951				
	CBGI	0.970				
	CBGJ	0.975				
	CBGK	0.936				
	CBGL	0.962				
	CBGM	0.969				
	CBGN	0.814				
	CBGO	0.923				
Effect of_Skill Gap	ESGA	0.965			0.992	0.887
	ESGB	0.953				
	Q8C	0.896				
	ESGD	0.956				
	ESGE	0.878				
	ESGF	0.953				

(continued)

Table 3 (continued)

Construct	Measurement items	Outer loading	T statistics	P Values	CR	AVE
	ESGG	0.959				
	ESGH	0.935				
	ESGI	0.911				
	ESGJ	0.948				
	ESGK	0.962				
	ESGL	0.938				
	ESGM	0.969				
	ESGN	0.958				
	ESGO	0.930				
	ESGP	0.956				
Measures of _Bridging the_Gap	MBGA	0.949			0.989	0.895
	MBGB	0.918				
	MBGC	0.934				
	MBGE	0.971				
	MBGF	0.956				
	MBGG	0.926				
	MBGH	0.962				
	MBGJ	0.934				
	MBGK	0.935				
	MBGL	0.959				
	MBGM	0.959				
Type of_Skill_Gap	TSGI	0.945			0.987	0.886
	TSGII	0.916				
	TSGIII	0.947				
	TSGIV	0.922				
	TSGIX	0.952				
	TSGV	0.930				
	TSGVI	0.959				
	TSGVII	0.958				
	TSGVIII	0.921				
	TSGX	0.961				

shows the result of the HTMT test as effects of skills-gap. The correlation matrix indicates that the HTMT value reported is below 0.90 that is (0.889 and 0.892) which implies that discriminant validity was achieved.

Table 4 Discriminant validity HTMT

	Cause_Skill_Gap	Challenges_Bridging_Gap	Effect_Skill_Gap	Measures_Bridging_Gap	Type_Skill_Gap
Cause_Skill_Gap					
Challenges_Bridging_Gap	0.891				
Effect_Skill_Gap	0.900	0.854			
Measures_Bridging_Gap	0.900	0.900	0.889		
Type_Skill_Ga p	0.886	0.897	0.892	0.870	
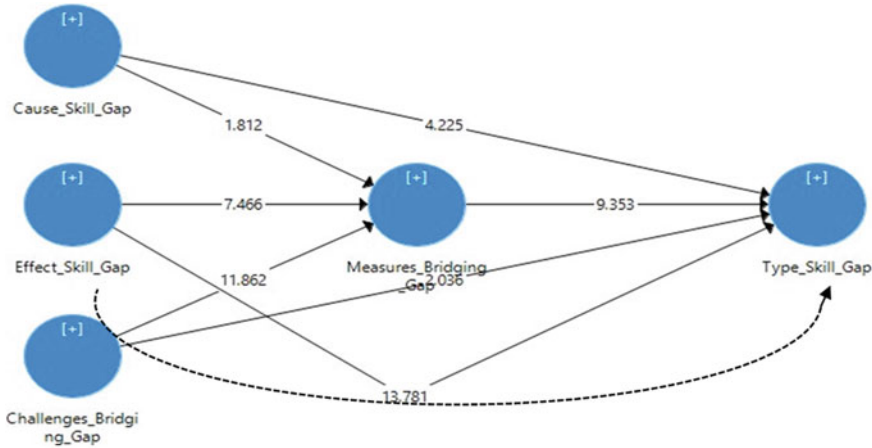
Fig. 3 Structural model extracted from the PLS software

6.5 Structural Model Evaluation

After assessing the measurement model, next stage is to evaluate the effects of skills-gap on types of skills-gap. The structural model is evaluated using collinearity assessment, test of significance of path coefficients determination of R^2 and effect sizes [9]. Figure 3 shows the structural model extracted from the PLS software .

6.6 Path Coefficient and R^2

The path coefficient measured the relative significance of the independent variables in a model. According to Hair et al. [8], path coefficient ranges from $+1$ to -1 where values tending towards 1 indicate stronger relationship. Apart from magnitude of the path coefficient, it is equally important for the estimates to be statistically significant.

On the other hand, R^2 is considered the overall measure of model performance. It is an indicator of the combined effects of all the independent variables in explaining the variation in the dependent variable in a model. Hair et al. [9] provided a threshold for interpreting the R^2 value where R^2 of 0.75, 0.50 and 0.25 are considered to be of substantial, moderate and weak effect respectively. However, in behavioural research, consumer satisfaction studies and other related areas in management and social science fields, R^2 value of 0.20 are considered of high effect [9]. Table 5 presents the path coefficients for the relationships among the two constructs in the model. The reported path coefficients showed a very strong positive and statistically significant effect of the independent variable (Effects of skills-gap) on the types of skills-gap ($\beta = 0.765$, $t = 13.781$, $p = 0.000$). As shown in the Table 5 explained about 97% variation in Measures of bridging the skills-gap.

Table 5 Path coefficients and R^2

Path	Beta	Standard deviation	T-statistics	P-values	R^2
Direct relationship					
Cause_Skill_Gap → Measures_Bridging_Gap	0.126	0.070	1.812	0.070	0.990
Challenges_Bridging_Gap → Measures_Bridging_Gap	0.496	0.042	11.862	0.000	0.994
Effect_Skill_Gap → Measures_Bridging_Gap	0.379	0.051	7.466	0.000	
Cause_Skill_Gap → Type_Skill_Gap	−0.269	0.064	4.225	0.000	
Challenges_Bridging_Gap → Type_Skill_Gap	0.070	0.034	2.036	0.042	
Effect_Skill_Gap → Type_Skill_Gap	0.765	0.055	13.781	0.000	
Measures_Bridging_Gap → Type_Skill_Gap	0.434	0.046	9.353	0.000	
Indirect relationship					
Cause_Skill_Gap → Measures_Bridg- ing_Gap → Type_Skill_Gap	0.055	0.032	1.709	0.088	
Challenges_Bridging_Gap → Measures_Bridging_ Gap → Type_Skill_Gap	0.215	0.028	7.561	0.000	
Effect_Skill_Gap → Measures_Bridg- ing_ Gap → Type_Skill_Gap	0.164	0.026	6.408	0.000	

6.7 Effect Size

Indicator reliability measures the absolute importance of individual measurement items in defining the latent construct. It shows the relationship between the latent construct and the measurement items [9]. A factor loading of 0.70 is considered sufficient [4]. This shows that the relationship between the latent independent variable (effects of skills-gap) and the measurement item (types of skills-gap) was considered sufficient having a factor loading of 0.95 and the result was successfully achieved (Table 6).

7   Conclusions and Recommendations

The findings from the research indicated that the skills shortage has directs effects on all the types of skills-gap. Major effects of these skill-shortage were poor workmanship, construction waste, rework which has definitely affects the productivity and profitability of projects in the study area. Some current measures of bridging the skills-gap include Establishing industry base skills acquisition centers, good wages and remuneration to skilled artisans. The introduction of apprenticeship training under the new national skills qualification framework (NSQF) across the country for the informal skills sector should be upgraded, polytechnic should lead this National Skills Qualification Framework intervention as trainers.

Poor workmanship apparent in buildings across the country make up skilling existing Artisans and craftsmen imperative. On the job training, developmental training will help to improve the situation. However, appropriate framework for setting up vocational training, building trade schools/centers devoted to skill training development no doubt will enhance workforce development.

Performance Enhancement: Employers should develop standard for improving productivity of workforce with appropriate incentives. In addition, they need to have in place clear and definite strategies, goals and performance metrics.

The Nigerian construction industry should diversify its awareness by establishing a special link with their foreign counterparts in creating the best ways of handling sustainable skills development internationally.

The federal government should formulate a regulation mandating all construction project stake-holders to attend a compulsory workshop on the issues leading to skills development, as well as their management principles.

The national directorate of employment (NDE), the national skills qualification framework (NSQF), the industrial training fund (ITF) and national board for technical education (NBTE) should ready to face the challenges of the recent skills-gap economy as we enter the third decade of the millennium for sustainability and productivity.

Table 6 Effect size

	Cause_Skill_Gap	Challenges_Bridging_Gap	Effect_Skill_Gap	Measures_Bridging_Gap	Type_Skill_Gap
Cause_Skill_Gap				0.012	0.092
Challenges_Bridging_Gap				0.846	0.015
Effect_Skill_Gap				0.163	0.950
Measures_Bridging_Gap					0.314

References

1. Akinsiku O, Akinsulire A (2012) Stakeholder' perception of the causes and effects of construction delays on project delivery. J Constr Eng Proj Manag 2. https://doi.org/10.6106/JCEPM.2012.2.4.025
2. Ameh JO, Itodo ED (2013) Professionals' views of material wastage on construction sites. Organ Technol Manage Constr Int J 5(1):747–757
3. Bilau AA, Ajagbe AM, Kigbu H, Sholanke AB (2015) Review of shortage of skilled craftsmen in small and medium construction firms in Nigeria. J Environ Earth Sci 5(1):35–42
4. Becker J, Klein K, Wetzels M (2012) Hierarchical latent variable models in PLS-SEM: Guidelines for using reflective-formative type Models. Long Range Plan 45:359–394. https://doi.org/10.1016/j.lrp.2012.10.001
5. Chin WW (1998) The partial least squares approach to structural equation modeling. Mod Methods Bus Res 295(2):295–336
6. Fornell C, Larcker DF (1981) Structural equation models with unobservable variables and measurement error. Algebra Stat
7. Field A (2009) Discovering statistics using SPSS. 3rd Edition, Sage Publications Ltd., London
8. Hair J, Sarstedt M, Ringle C, Mena J (2012) An Assessment of the use of partial least squares structural equation modeling in marketing research. J Acad Mark Sci 40:414–433. https://doi.org/10.1007/s11747-011-0261-6
9. Hair JF, Hult GTM, Ringle CM, Sarstedt M (2014) A primer on partial least squares structural equation modeling (PLS-SEM). SAGE Publication, Inc., Thousand Oaks, California
10. Ibrahim S (2016) Management of material waste and cost overrun in the Nigerian construction industry: Unpublished PhD thesis submitted to Nelson Mandela Metropolitan University
11. Krejcie RV, Morgan DW (1970) Determining sample size for research activities. Educ Psychol Measur 30(3):607–610
12. Kurushi P (2015) Construction practices and skills gap in the construction industry. In: Proceedings of the 45 builder's conference of the Nigerian institute of building
13. Medugu NI, Majid MR, Bustani SA, Bala K, Abdullahi U, Mbamali I (2011) Craft skills availability in the Nigerian construction industry: perception of contractors and consultants. IUP J Infrastruct 9(3):63–69
14. Nunnally JC, Bernstein IH (1994) The assessment of reliability. Psychom Theory 3:248–292
15. Odediran SJ, Babalola MO (2013) Assessment of business development strategies in the Nigerian construction industry. J Bus Manag 2(1):34–45. https://doi.org/10.12735/JBM.V2I1P34
16. Okuntade TF (2014) Building construction technician training: it's relevance to modern construction industry in Nigeria. Int J Technol Enhancements Emerg Eng Res 2(3):234–4289
17. Proverbs DG, Holt GD, Cheok HY (2000) construction industry problems: the views of UK construction directors. In: 16th annual ARCOM conference, pp 6–8
18. Siboe IW (2016) Investigating the adequacy of construction planning in Kenya. Doctoral dissertation, Jomo Kenyatta University of Agriculture and Technology
19. Sinkovics RR, Henseler J, Ringle CM, Sarstedt M (2016) Testing measurement invariance of composites using partial least squares. Int Mark Rev
20. Sun M, Meng X (2009) Taxonomy for change causes and effects in construction projects. Int J Project Manage 27(6):560–572
21. Vogt WP, Johnson RB (2011) Dictionary of statistics and methodology: a nontechnical guide for the social sciences. Sage
22. Zaghloul SM (2006) Effect of poor workmanship and lack of smoothness testing on pavement life-cycle costs. Transp Res Rec 1539(1):102–109

Propagation Model of Molecular Communication Based Targeted Drug Delivery for Atherosclerosis Disease Therapy

Pradeep Murugesan, S. Prabakar, K. Porkumaran, and R. Karthikeyan

Abstract Atherosclerosis is one of the major cardiovascular disorder which causes severe health issues in human body. It also ends with patient death, when improper therapy takes place. Conventional therapy measures includes oral statin therapy, skin mode application, Inhalation mode of therapy etc., has their own advantages. But they cannot fully exploit the inflammation because of lower systemic bioavailability. Nowadays, Nano molecular communication provides numerous solutions in the field of targeted drug delivery system. The aim of this article is to propose a new analytical model for the propagation process of molecules based on Brownian motion mechanism by formulating the probability density of the Latency in blood medium. This model is analyzed based on crucial parameters such as radius of the propagating molecules, blood viscosity, drift velocity, distance between Nano-Tx and Nano-Rx, temperature of the fluid medium with respect to various blood shear rates. Based on simulation results, the latency is highly affected with molecular radius, distance temperature, shear rate and drift velocity. Our future work is to apply this model for various drug carrying molecules used in the treatment of the cardiovascular diseases and to assess its propagation capacity under various conditions of blood medium.

Keywords Atherosclerosis · Molecular communication · Drug delivery · Nano-Tx · Nano-Rx

P. Murugesan (✉)
Department of BME, Dr.N.G.P Institute of Technology, Coimbatore, India

S. Prabakar
Sona College of Technology, Salem, India

K. Porkumaran
Sri Sairam Engineering College, Chennai, India
e-mail: porkumaran@gmail.com

R. Karthikeyan
Department of ECE, Dr.N.G.P Institute of Technology, Coimbatore, India

1 Introduction

The research and development of Body Area Networks (BAN) has gained maturity in recent years. Hence the researchers are now working for the In body networks using Nano machines [1]. The first useful applications of Nano machines are in medical technology, where they could be used to identify pathogens and toxins from the samples of body fluid. The reduced size of Nano machines translates into higher operational speed. They are designed according to their applications. Specialized Nano machines called nanobots might be designed not only to diagnose, but to treat the disease conditions, perhaps by seeking out invading bacteria and viruses and destroying them. In the field of biomedicine, drug delivery and health monitoring systems are achieved by biological Nano machines. Targeted Drug Delivery System (TDDS) is an emerging technology that plays a major role in specialized thera-peutic purposes. It promises to tackle the conventional hurdles in order to achieve a controlled rate of drug release with precise location [2]. A new network Paradigm called IoNT is developed for providing solutions in the drug delivery system. There are four different types of communication in IoNT. The first type is Electromagnetic communication. It enables Nano machines to communicate over electromagnetic waves. The second type is Acoustic, where Acoustic energy is used for communi-cation. The last type is called as Molecular communication. Here, the molecules are used for communication between each other. This MC plays a major role in Nano networks. In this article, focus is put forward on analysis of Diffusion based molec-ular communication, its propagation model and finally numerical analysis is carried out for MC-TDDS for Atherosclerotic disease therapy.

2 Methodology

2.1 General Propagation Model

Molecular communication takes place between Nano-transmitters and Nano-receivers. The information molecule is propagated in the blood medium via different process namely walkway-based, flow-based, and diffusion-based [3]. In walkway based, there are pre-defined paths transmitting molecules to the communicating transmitter and receiver. In flow-based and diffusion-based techniques, molecules propagate over diffusion in a fluid medium. Most of the studies in the field of molec-ular communication has been carried out by the diffusion propagation model based on Brownian motion or Brownian motion with drift [4]. The diffusion process is explained by Fick's equations [5]. In the propagation process, propagation medium is a fluid medium (blood) and the propagation path is from sender bio-nanomachine to receiver bio-Nano machine. The propagation of messenger molecules is random, that is to say, in all directions in the fluid medium. These random movements of propagating molecules are modeled by the Brownian motion. The Brownian motion

is an important physical process that undergoes random movement of particles in the fluid medium resulting from their collision with the fast moving molecules in the fluid. This motion is generally approximated as a Gaussian model in accordance with the probability theory [5].

2.2 Mathematical Modeling of Propagation Process in MC-TDDS

Mathematical models are developed in order to compare the quality of the molecular communication. Propagation period is calculated on the probability basis using Gaussian distribution [5].

For Brownian motion in the blood medium, we apply the Gaussian distribution as follows,

$$f(t) = \sqrt[1/2]{4\pi Dt} \exp\left(\frac{-d^2}{4Dt}\right) \tag{1}$$

This is known as the PDF of first hitting time (i.e., time when the molecule first hit the Nano-receiver or absorption time).

Where D is the diffusion coefficient of the propagating molecules during the propagation process and d is the distance between the Nano-transmitter and Nanoreceiver in a one dimensional interval $(-\infty, d]$. Diffusion coefficient represents the inclination of the propagating molecules during the propagation process through the fluid medium and it can be obtained by the following formula [6]

$$D = \frac{K_b * T}{b} \tag{2}$$

where Kb is a fixed value called the Boltzmann constant, T is the temperature of the blood medium, and b is also a fixed value representing the drag constant of the molecule in the fluid medium. The drag constant b is derived for two conditions based on the size of the drug carrier molecule S_{cm} and the size of the propagating molecule S_{fluid} in the fluid (blood) medium [6],

When $S_{cm} = S_{fluid}$

$$b = 4\pi \eta r_{nm} \tag{3}$$

When $S_{cm} > S_{fluid}$

$$b = 6\pi \eta r_{nm} \tag{4}$$

η represents the viscosity of the fluid medium and r_{nm} is the radius of the propagating molecule in the fluid medium. According to author [7], liposomes play a major

role in drug delivery systems. At the systemic level, drug bioavailability is increased due to the high relative surface area of nano particles and it has been shown that liposomes around 150–200 nm in diameter remains in the blood stream larger than those with diameters less than 70 nm. Liposomes exhibit Brownian motion which is directly proportional to temperature and inversely to the liposome size.

Since the diseased vessel has a blood flow with increased velocity, i.e., fluid medium has a medium drift velocity, we go with Inverse Gaussian model [8] as follows,

$$f(t) = \sqrt[d/2]{4\pi D t^3} \exp - \left(\frac{vt - d^2}{4Dt} \right) \tag{5}$$

3   Numerical Analysis and Results

This part deals with the numerical analysis for the following conditions,
    PDF of latency for different distances in blood,

- PDF of Latency for different shear rates with different Drift velocity
- Comparison of different shear rates.

3.1   PDF of Latency for Different Distances in Blood

Condition 1: In our scenario, we consider three different distances i.e., distance between Nano-Tx and Nano-Rx as {2, 6, 10} in µm with drift velocity to be approximately 1 µm for three different cases. We have some constant values in the PDF. The Boltzmann constant (Kb) is equal to $1.3807 * 10^{-23}$. The radius of the liposome molecule is considered into two different cases 75 nm and 100 nm respectively. The temperature T reveals two different values 277 K as minimum human body temperature and 310 K as maximum human body temperature. The blood viscosity for two different shear rates $1s^{-1}$ and $1000s^{-1}$ have also been defined in the table.

> **Case 1:** PDF of Latency for different distances {2, 6, 10}µm, constant drift velocity 1 µm/s and $r_{nm} = 75$ nm.
> **Case 2:** PDF of Latency for different distances {2, 6, 10}µm, constant drift velocity 1 µm/s and $r_{nm} = 100$ nm (Figs. 1 and 2).

From the above graphs, it is very clear that when the distance between the nano-Tx and Nano-Rx increases, the PDF decreases accordingly. This increase in distance leads to the time delay in both the cases. i.e., the nano receiver should await for a longer duration for the molecules in a blood medium. The temperature and blood shear rate will also plays a major role in the propagation process. Blood medium with
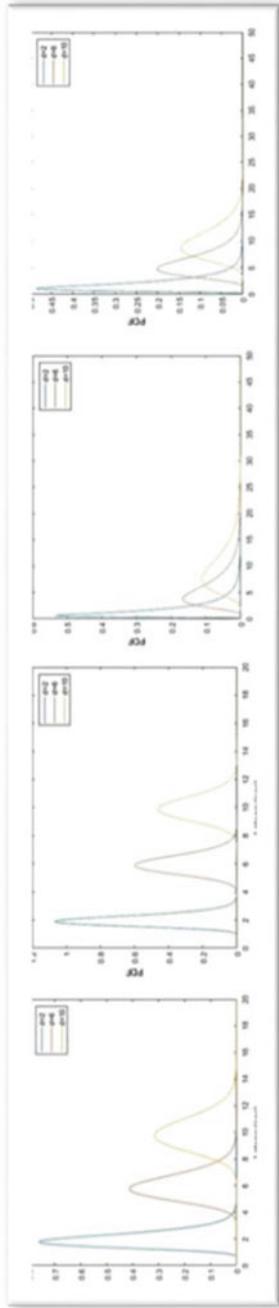
Fig. 1 PDF of Latency for different distances {2, 6, 10}μm, constant drift velocity 1 μm/s and $r_{nm} = 75$ nm
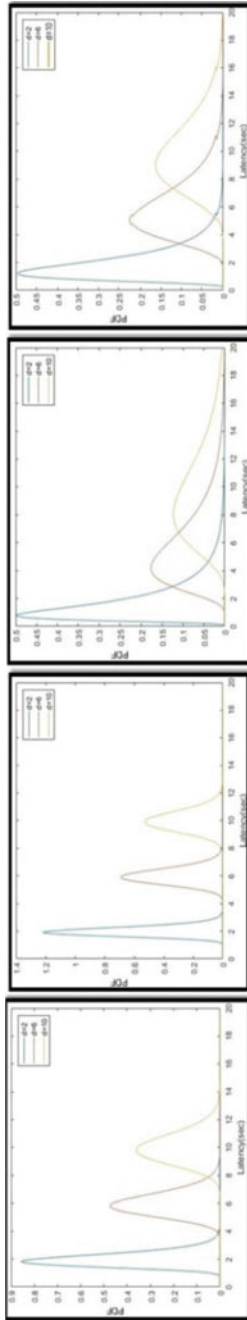
Fig. 2 PDF of Latency for different distances {2, 6, 10}μm, constant drift velocity 1 μm/s and $r_{nm} = 100$ nm

decreasing temperature and minimum shear rate has a greater chance for propagating the molecule with very less time delay.

3.2 PDF of Latency for Different Shear Rates with Different Drift Velocity

Condition 2: Minimum Distance:

The drift velocity inside the blood medium plays a major role in the propagation process. Increase in the drift velocity has a greater efficiency in Propagation of the drug carrying molecule. In this case, we have considered the latency in a blood medium with drift velocity {1, 2, 2.5, 3} and the minimum distance 2 μm.

Case 1: PDF of latency for different drift velocity {1, 2, 2.5, 3}μm/s, minimum distance −2 μm and $r_{nm} = 75$ nm.

Case 2: PDF of latency for different drift velocity {1, 2, 2.5, 3}μm/s, minimum distance −2 μm and $r_{nm} = 100$ nm (Figs. 3 and 4).

Condition 3: Maximum distance:

In this condition, we have considered the latency in a blood medium with drift velocity {1, 2, 2.5, 3} and the maximum distance.

Case 1: PDF of latency for different drift velocity {1, 2, 2.5, 3}μm/s, maximum distance −10 μm and $r_{nm} = 75$ nm.

Case 2: PDF of latency for different drift velocity {1, 2, 2.5, 3}μm/s, maximum distance −10 μm and $r_{nm} = 100$ nm (Figs. 5 and 6).

From both the conditions, it is very clear that when the drift velocity increases, the first hitting time decreases for the propagating molecules inside the blood medium with the same distance. In this condition, temperature and blood shear rates along with the drift velocity has a greater influence in the propagation process. Higher the drift velocity and lower temperature with minimum shear rate has a greater tendency for the molecule to travel without any time delay. From the distance point of view, though with the increased drift velocity in both cases, the PDF is high only in the case 1. This confirms that the minimum distance between Nano-Tx and Nano-Rx will be an suitable condition for a better propagation.

3.3 Comparison of Different Shear Rates

In this part, we compare different shear rates in a blood medium for the PDF of the latency. Here, we have considered the latency in the blood medium with different shear rates {1s^−1, 10s^−1, 50s^−1, 100s^−1, 1000s^−1}.
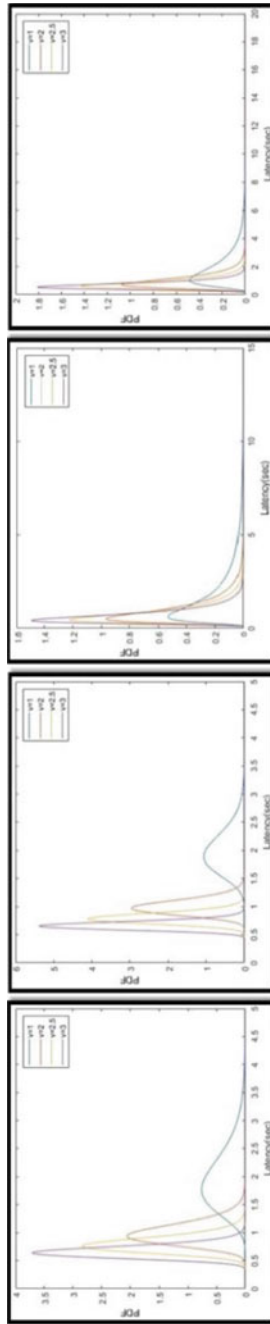
Fig. 3 PDF of latency for different drift velocity {1, 2, 2.5, 3} μm/s, minimum distance −2 μm and r_{nm} = 75 nm
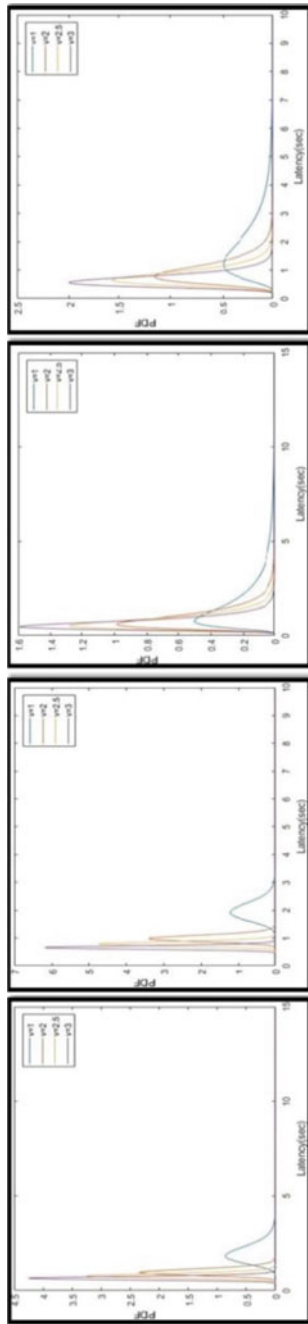
Fig. 4 PDF of latency for different drift velocity {1, 2, 2.5, 3}μm/s, minimum distance −2 μm and r_{nm} = 100 nm
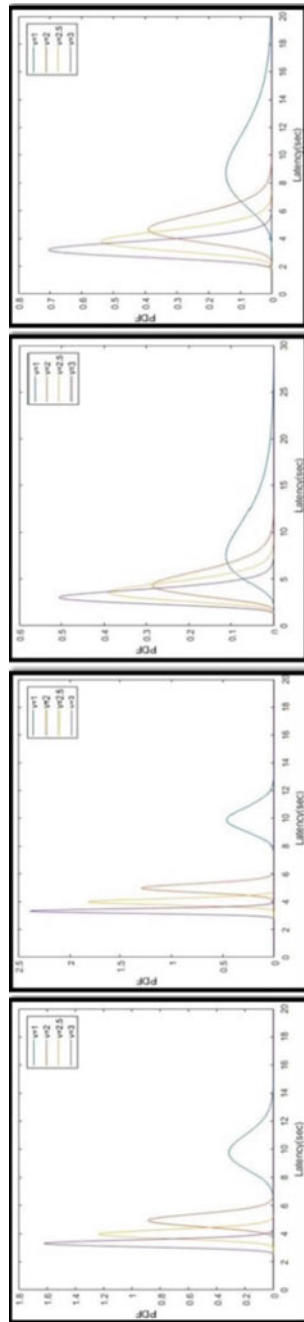
Fig. 5 PDF of latency for different drift velocity {1, 2, 2.5, 3} μm/s, maximum distance −10 μm and $r_{nm} = 75$ nm
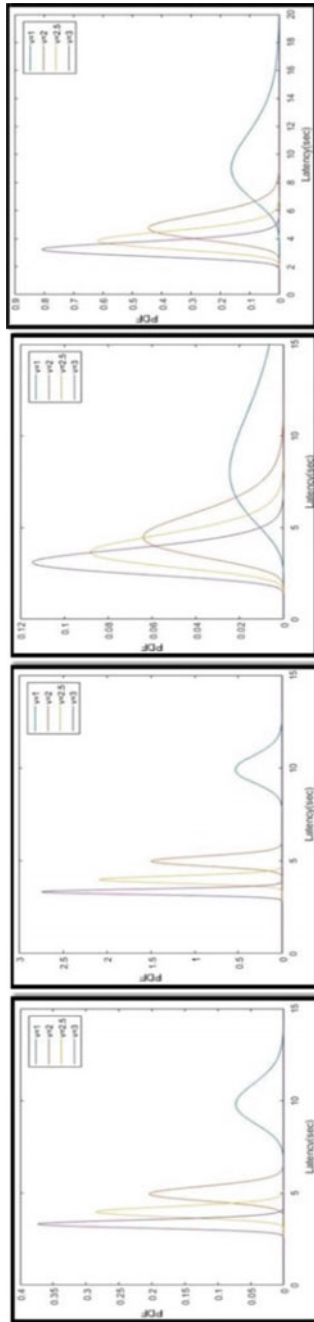
Fig. 6 PDF of latency for different drift velocity {1, 2, 2.5, 3}μm/s, maximum distance −10 μm and $r_{nm} = 100$ nm

Condition 4: Minimum distance $= 2\,\mu m$ and Minimum drift velocity $= 1\,\mu m$.
PDF of latency for different shear rates $\{1s^{-1}, 10s^{-1}, 50s^{-1}, 100s^{-1}, 1000s^{-1}\}$.

Case 1: Minimum distance $-2\,\mu m$, minimum drift velocity $= 1\,\mu m/s$ and $r_{nm} = 75$ nm.
Case 2: Minimum distance $-2\,\mu m$, minimum drift velocity $= 1\,\mu m/s$ and $r_{nm} = 100$ nm (Fig. 7).

Condition 5: Minimum distance $= 2\,\mu m$ and Maximum drift velocity $= 3\,\mu m$.
PDF of latency for different shear rates $\{1s^{-1}, 10s^{-1}, 50s^{-1}, 100s^{-1}, 1000s^{-1}\}$.

Case 1: Minimum distance $-2\,\mu m$, maximum drift velocity $= 3\,\mu m/s$ and $r_{nm} = 75$ nm.
Case 2: Minimum distance $-2\,\mu m$, maximum drift velocity $= 3\,\mu m/s$ and $r_{nm} = 100$ nm (Fig. 8).

Condition 6: Maximum distance $= 10\,\mu m$ and Minimum drift velocity $= 1\,\mu m$
PDF of latency for different shear rates $\{1s^{-1}, 10s^{-1}, 50s^{-1}, 100s^{-1}, 1000s^{-1}\}$.

Case 1: Maximum distance $-10\,\mu m$, minimum drift velocity $= 1\,\mu m/s$ and $r_{nm} = 75$ nm.
Case 2: Maximum distance $-10\,\mu m$, minimum drift velocity $= 1\,\mu m/s$ and $r_{nm} = 100$ nm (Fig. 9).

Condition 7: Maximum distance $= 10\,\mu m$ and Maximum drift velocity $= 3\,\mu m$.
PDF of latency for different shear rates $\{1s^{-1}, 10s^{-1}, 50s^{-1}, 100s^{-1}, 1000s^{-1}\}$.

Case 1: Maximum distance $-10\,\mu m$, Maximum drift velocity $= 3\,\mu m/s$ and $r_{nm} = 75$ nm.
Case 2: Maximum distance $-10\,\mu m$, Maximum drift velocity $= 3\,\mu m/s$ and $r_{nm} = 100$ nm (Fig. 10).

The latency is strongly affected by the shear rate of the blood medium. When the shear rate of the blood medium increases, propagation time is more for the propagating molecules. After the distance between the Nano-Tx and Nano-Rx increases, the propagation time is always more for each shear rate of the blood at the same temperature. From the above cases, it is very clear that case 2 has a greater.

Probability to reach the receiver at very low time delay. Hence blood is an efficient medium for propagation of molecules during lower temperature and minimum shear rate.

3.4 Comparison of Higher PDF Rate

In this section, the higher PDF rate obtained from different conditions using Inverse Gaussian distribution methodology is compared for two different Liposome radii.
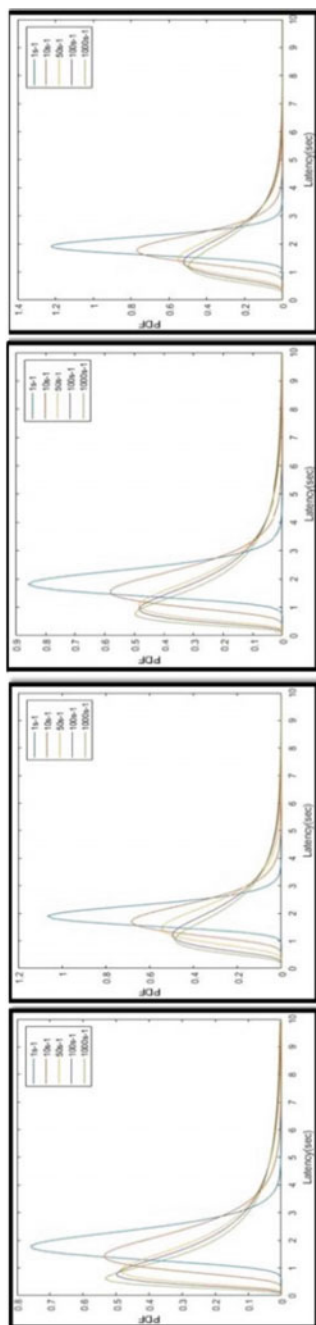
Fig. 7 PDF of latency for different shear rates $\{1\,s^{-1}, 10s^{-1}, 50s^{-1}, 100s^{-1}, 1000s^{-1}\}$ Minimum distance 2 μm and Minimum drift velocity 1 μm
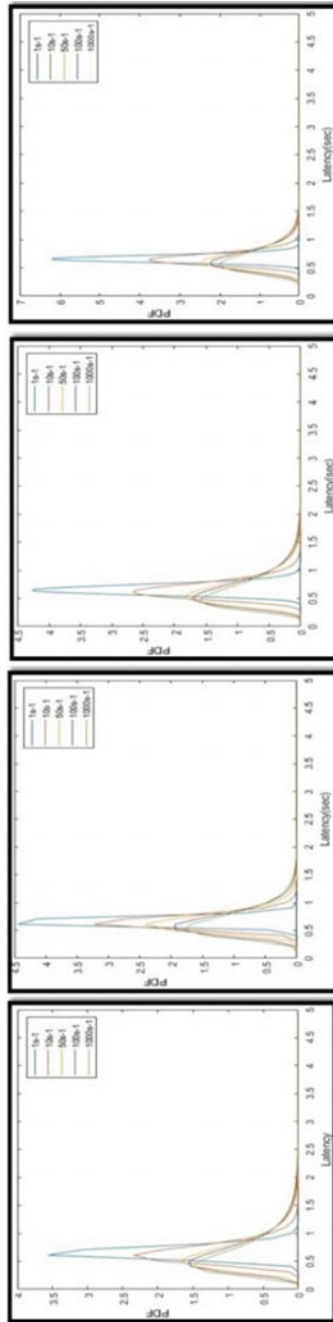
Fig. 8 PDF of latency for different shear rates $\{1s^{-1}, 10s^{-1}, 50s^{-1}, 100s^{-1}, 1000s^{-1}\}$ for minimum distance 2 µm and Maximum drift velocity 3 µm
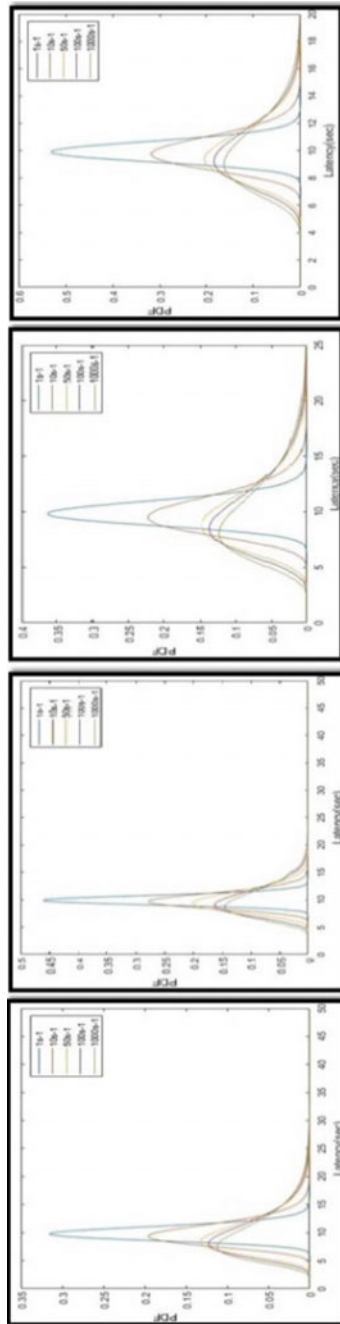
Fig. 9 Maximum distance = 10 μm and Minimum drift velocity = 1 μm

Fig. 10 Maximum distance = 10 μm and Maximum drift velocity = 3 μm

Table 1 Higher PDF rate comparison

Condition	Temp (K)	Shear rate	PDF value for case1	PDF value for case 2
1	277	$1s^{-1}$	1.066	1.128
2	277	$1s^{-1}$	5.378	6.192
3	277	$1s^{-1}$	2.378	2.747
4	277	$1s^{-1}$	1.065	1.217
5	277	$1s^{-1}$	4.446	6.192
6	277	$1s^{-1}$	0.4608	0.531
7	277	$1s^{-1}$	2.377	2.747

This comparison predicts the role of molecular radius inside the propagation medium. Let consider two different radius of the Liposome molecule 75 nm and 100 nm respectively.

From the Table 1, it is very clear that Liposome molecule with radius of 100 nm has a greater efficiency to propagate inside the blood medium with less time delay. By comparing the higher probability values for both cases, the higher most probability takes place in the case 2 for all the conditions. Hence, molecule with increasing radius will play a major role in the propagation process in targeted drug delivery system.

4 Conclusion

Molecular communication, in other words the use of molecules to deliver drugs among Nano machines is new communication paradigm. The most important expectation of molecular communication is to provide new solutions for the treatment of cancer. This thesis provided a novelty by applying molecular communication in Atherosclerosis disease therapy. The proposed mathematical model is evaluated under various conditions and compared with two different radii of the liposome molecules. The results shows that, the latency is highly dependent on parameters such as distance, molecular radius, blood viscosity, blood shear rate and drift velocity. Hence blood becomes an efficient medium at lower shear rates in decreased temperatures. Our future work is to apply this model for various drug carrying molecules used in the treatment of the cardiovascular diseases such as Lipid nanoparticles, polymeric nanoparticles Micelles, CNT etc., and to assess its propagation capacity under various conditions of blood medium.

References

1. Tiwari G, Tiwari R (2012) Drug delivery system—an updated review. Pharma Investig
2. Nasrdlahzadeh M, Sajadi SM, Sajjadi M, Issaabadi Z (2019) An introduction to nanotechnology. Interface Sci Technol
3. Akyildiz IF, Jornet JM, Pierobon M (2010) Propagation models for nano communication networks. In: Proceedings of the fourth European conference on antennas and propagation. pp 15–19
4. Kadloor S, Adve RS, Eckford AW (2012) Molecular communication using Brownian motion with drift. IEEE Trans Nanobiosci 11(2):89–99
5. Berg HC (1993) Random walks in biology. Press Princeton University Press, Princeton
6. Kuran MŞ, Yılmaz HB, Tugcu T, Özerman B (2010) Energy model for communication via diffusion in nanonetworks. Nano Commun Netw 1(2):86–95
7. https://www.pharmatutor.org/articles/liposomes-novel-drugdelivery-carrier
8. https://holisticprimarycare.net/topics/topicsn/healthy-aging/1299bloodviscosity-the-unifying-parameter-in-cardiovascular-diseaserisk.html

Adaptation of Machine Learning and Blockchain Technology in Cyber-Physical System Applications: A Concept Paper

Mujaheed Abdullahi, Hitham Alhussian, and Norshakirah Aziz

Abstract In recent years, Cyber-Physical Systems (CPS) have been adopted in various sectors such as smart cities, smart industries etc. These types of systems continuously generate a huge amount of data which increasingly attract cyber-crimes. There are several existing approaches produced to overcome these issues by using Blockchain Technology (BT) such as Public, Private, Construme, Hybrid Blockchain-based on CPS applications and Machine Learning (ML) such as Support Vector Machine (SVM), Linear Regression, and Decision Tree etc. With the rapid increase in data size affix with cyber-crimes, such approaches become less effective and therefore necessitate the invention of a more robust and self-trainable approach. In this paper, we presented brief details on ML and BT and how they can be adopted in CPS applications to solve security issues concerning cyber-crimes. The architecture was also presented to depict the proposed method. Moreover, technologies/techniques which can be implemented in CPS applications are discovered such as industrial automation, smart buildings, medical systems, and vehicular systems. We also have some future scope and conclusion.

Keywords Cyber-physical system (CPS) · Machine learning (ML) · Blockchain technology (BT) · Security · Privacy

M. Abdullahi (✉) · H. Alhussian · N. Aziz
Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia
e-mail: abdullahi_18001208@utp.edu.my

H. Alhussian
e-mail: seddig.alhussian@utp.edu.my

N. Aziz
e-mail: norshakirah.aziz@utp.edu.my

1   Introduction

From past few years, cybersecurity has become an increasing demand in computations devices, digital transformation in Cyber-Physical System (CPS) such as smart manufacturing, smart homes, smart transportation and smart factory, where trust has to be crucial aspect [1]. Those devices use wireless network as a means of communication to broadcast their data, hence it is easy to be attacked during normal network communication [2, 3]. Cybersecurity [4] was major concerned because the technology moving is moving to industry 4.0 [5] where a large amount of data are generated by CPS and others smart devices such as IoT [6] are usually used for real-time monitoring and predictions failure [7, 8]. Many cyber attacked has been accoutered in past few days which include Ukraine power plant in 2016 and Stuxnet worm which target a nuclear power plant, also there is cyber-attacked on Australia's Maroochy water services in 2002 [9]. Also, those CPS applications are generating a huge amount of data, storage this communications process is vulnerable against attack. To overcome this problem Machine Learning and Blockchain can be used, BT has a distributed database network with time series stamped of tamper-proof record, while ML will analysis the data traffic through effective, efficient designing of algorithms [10, 11].

Machine Learning (ML) in the field of Artificial Intelligence focus on building applications that learn through experience and they make self-decision without human interference such as Logistic Regression, Linear Regression, and Support Vector Machine etc. ML was applied to various field of discipline such as statistics, computational complexity [12, 13] smart devices [14] internet of things [15, 16] and artificial intelligence to machines are getting smarter [17, 18]. While Blockchain Technology (BT) are cryptography immutable set of records or distributed ledger which is link together for auditing, by linking its valid blocks together the current block contains the hash of the previous block. This process makes the Blockchain traceable and resistant to change [10, 19].

Cyber-Physical System (CPS) was introduced by Helen Gill in 2006 at National Science Foundation, they are integrated of both physical and cyber to enhance the control and operation which include computational, communication, sensing and actuation, in general, they are dynamics joint of computers, software, physical process, event processing [20] and networks. The CPS includes operating in real-time, transactional services and secure communication between components [21].

1.1   Adaptation of ML and BT in CPS

The ML processes can be applied to BT for CPS applications to improve their security, privacy, monitoring [22] and making them smarter. It also gives an opportunity with improvement in developing good models by using the decentralized architecture of

BT [23]. We also proposed architecture for ML adaptation in BT for CPS applications as shown in fig, whereby after the data was given which was generated by the CPS. The BT will be included as an integral of the applications. At this point, the ML will be applied for data analytics and real-time predictions [24] to reduce error and detect anomaly activities [25].

1.2 Research Contribution

There are several research works which address the importance of combining ML and BT, in this paper, we investigate how ML can be adopted in BT based on CPS applications. The research contribution for this paper include the followings:

- We find out that ML and BT can be adopted together and a brief discussion on ML with BT can be used together in CPS applications.
- We also proposed architecture on how ML and BT can be adopted and implemented together.

2 Core Concepts and Rationale

The core concept and rationale of ML algorithms with BT can be adopted and implemented in CPS applications to improve their performance in terms of security, monitoring and detection activities such as improved security and privacy were ML and BT covers CPS applications such as data protection, data encryption, instruction detection system (IDS), intrusion protection system (IPS). They provide an important role in the monitoring and detection of malware. Furthermore, BT increased the privacy between users for both private and public Blockchain through secure authentication of data accessibility.

Improved efficiency and scalability increasing ability in multiple of CPS applications from multiple devices that consist of huge datasets due to their multiple transactions through providing decentralize self-authorization, detection from ML. Scalability form BT also provides capable of increment between users.

Secure of data sharing in CPS Applications was the BT contains chains blocks of all data transaction for CPS applications by providing accountability, transparency within the data access. ML components were included with its algorithms to provide decentralised platform and removal of fraud fewer datasets.

3 Research Question and Methodology

In this section research question and methodology has been discussed, the research question was (Can Machine Learning and Blockchain Technology be adopted

together in Cyber-Physical Systems Applications?). The question has been analyses using a systematic literature review by using PRISMA method, we also provide a comparison study from previous research.

3.1 Systematic Literature Review

PRISMA guideline was used to structure the systematic literature review by creating a question to guide the review process (Can Machine Learning and Blockchain Technology be adopted together in Cyber-Physical Systems Application?). The search strategy used for this systematic review, we developed a search strategy to identify relevant literature. The search strategy was tailored by one database: Scope and search terms used were the followings: "Machine Learning for Blockchain Applications" and "Blockchain for Cyber-Physical Systems Application". All search spanned from the database inception were until 2020, and included journals articles with few review papers published in English only.

The selection criteria were mainly focused on the mapping on existing literature on Machine Learning for Blockchain Applications and Blockchain for Cyber-Physical Systems Application in the field of computer science, Decision Science and mathematics. The search span was from 2020 to 2018. All articles before 2018 below were excluded from the search. The also was mainly focused on the global search, not a specific country or region. A total of 95 research articles were excluded at this stage. There were 42 research article included and extracted at this stage. The quality assessment is based on the original research and a few review papers. For maintain the quality of the review, all duplicated were checked thoroughly.

3.2 Methodology

The methodology investigates process, steps to overcome this problem with rational for CPS applications, by using a combination of two technologies which are ML and BT with a proposed framework to support this research study (see Fig. 1).

4 Findings

Firstly, providing detection with prevention in CPS applications to handle data security and cybersecurity issues within its application. Those methods are intrusion detection system (IDS), intrusion prevention system (IPS) with ML classifiers. BT also add additional support through linked blocks using cryptography for prevention of datasets against attack.
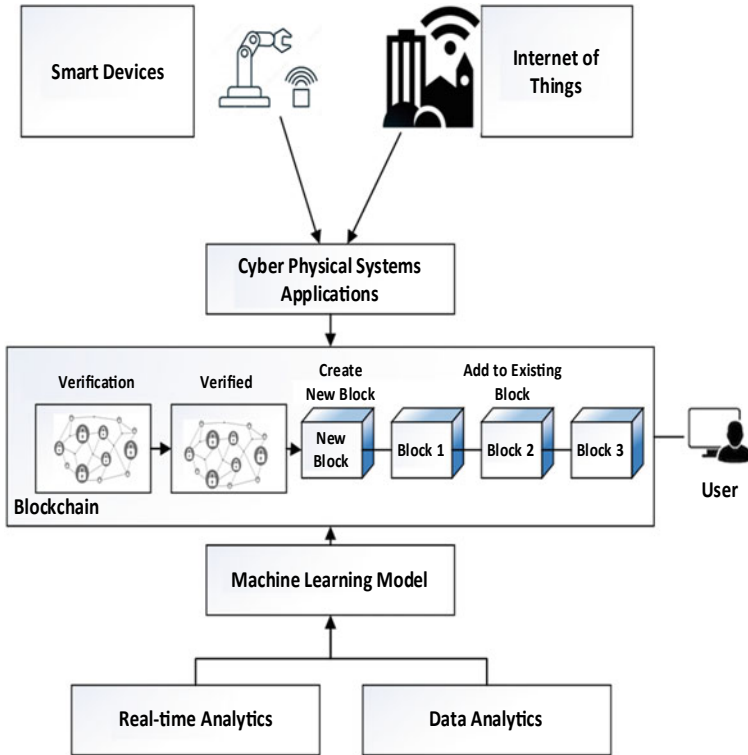
Fig. 1 Proposed architecture for ML and BT adaptation in CPS applications

Secondly, prediction with a response in CPS applications most methods used for prediction of ML algorithms whereby providing security, improving the right decision in datasets and data analysis. They also involve failures prediction of applications with taking security majors for necessary actions in terms of response. The ML response through anomalies activities. Finally, they allow the CPS applications to make self-decision to command and action using security policies.

Lastly, providing monitoring in CPS applications by using ML and BT through verification of modelled datasets and external control. The CPS applications network events is been analyzed which include input data. All records from events and datasets are monitored in real-time control.

5 Conclusion and Future Scope

We conclude, that recent development of CPS application with their huge increasing data generated this leads to the increase of cyber-attack. With the development of ML and BT are helping to overcome these issues by adopting them together in

various CPS applications such as power systems, medical systems, smart buildings and industrials automation. The research paper also presents brief details on ML and BT with how they can be adopted together in CPS applications using a proposed architecture. A study comparison has been made which shows methods, contributions and limitation. We also have some suggestion for further scope.

The future scope on this topic is therefore recommended to provide in details on how ML and BT can be adopted into specific CPS applications such as Industrial Automation, Power Systems and Medical Systems or Internet of Things devices.

References

1. Kochovski P, Gec S, Stankovski V, Bajec M, Drobintsev PD (2019) Trust management in a blockchain based fog computing platform with trustless smart oracles. Futur Gener Comput Syst 101:747–759
2. Hasan M, Islam MM, Zarif MII, Hashem MMA (2019) Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. Internet Things 7:100059
3. Li D, Deng L, Liu W, Su Q (2020) Improving communication precision of IoT through behavior-based learning in smart city environment. Futur Gener Comput Syst 108:512–520
4. Sri Devi R, Mohan Kumar M (2019) Cyber security affairs in empowering technologies. Int J Innov Technol Explor Eng 8(10):1–7
5. O'Donovan P, Gallagher C, Leahy K, O'Sullivan DTJ (2019) A comparison of fog and cloud computing cyber-physical interfaces for Industry 4.0 real-time embedded machine learning engineering applications. Comput Ind 110:12–35
6. Jia P, Wang X, Zheng K (2020) Distributed clock synchronization based on intelligent clustering in local area industrial IoT systems. IEEE Trans Ind Inform 16(6):3697–3707
7. Lv L, Wang W, Zhang Z, Liu X (2020) A novel intrusion detection system based on an optimal hybrid kernel extreme learning machine. Knowl-Based Syst 195:1–17
8. Junejo KN (2020) Predictive safety assessment for storage tanks of water cyber physical systems using machine learning. Sadhana—Acad Proc Eng Sci 45(1):1–6
9. Goh J, Adepu S, Tan M, Shan LZ (2017) Anomaly detection in cyber physical systems using recurrent neural networks. IEEE Comput Soc 36:1–7
10. Tanwar S, Bhatia Q, Patel P, Kumari A, Singh PK, Hong WC (2020) Machine learning adoption in blockchain-based smart applications: the challenges, and a way forward. IEEE Access 8:474–488
11. Gupta S, Sabitha AS, Punhani R (2019) Cyber security threat intelligence using data mining techniques and artificial intelligence. Int J Recent Technol Eng 8(3):6133–6140
12. Rahman M (2019) Cultural differences, strengths, weaknesses and challenges of Kentucky fried chicken (KFC)—a comparison between United Kingdom (UK) and Malaysia. Double Blind Peer Rev Int Res J Publ 19(1):7
13. Sharma R, Kamble SS, Gunasekaran A, Kumar V, Kumar A (2020) A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. Comput Oper Res 119:104926
14. Gupta M, Abdelsalam M, Khorsandroo S, Mittal S (2020) Security and privacy in smart farming: challenges and opportunities. IEEE Access 8:34564–34584
15. Shafiq M, Tian Z, Sun Y, Du X, Guizani M (2020) Selection of effective machine learning algorithm and Bot-IoT attacks traffic identification for internet of things in smart city. Futur Gener Comput Syst 107:433–442
16. Viejo A, Sánchez D (2020) Secure monitoring in IoT-based services via fog orchestration. Futur Gener Comput Syst 107:443–457

17. Bello A, Mahadevan V (2019) A cloud based conceptual identity management model for secured Internet of Things operation. J Cyber Secur Mobil 8(1):53–74
18. Al-Saud M, Eltamaly AM, Mohamed MA, Kavousi-Fard A (2020) An intelligent data-driven model to secure intravehicle communications based on machine learning. IEEE Trans Ind Electron 67(6):5112–5119
19. Islam A, Young Shin S (2020) A blockchain-based secure healthcare scheme with the assistance of unmanned aerial vehicle in Internet of Things. Comput Electr Eng 84:1–12
20. Roldán J, Boubeta-Puig J, Luis Martínez J, Ortiz G (2020) Integrating complex event processing and machine learning: an intelligent architecture for detecting IoT security attacks. Expert Syst Appl 149:1–22
21. Al-Hadhrami Y, Hussain FK (2020) Real time dataset generation framework for intrusion detection systems in IoT. Futur Gener Comput Syst 108:1–10
22. Zhong CL, le Li Y (2020) Internet of things sensors assisted physical activity recognition and health monitoring of college students. Meas J Int Meas Confed 159:1–9
23. Lu Y, Huang X, Dai Y, Maharjan S, Zhang Y (2020) Blockchain and federated learning for privacy-preserved data sharing in industrial IoT. IEEE Trans Ind Inform 16(6):4177–4186
24. Lima E, Vieira T, de Barros Costa E (2020) Evaluating deep models for absenteeism prediction of public security agents. Appl Soft Comput J 91:106236
25. Lin X, Li J, Wu J, Liang H, Yang W (2019) Making knowledge tradable in edge-AI enabled IoT: a consortium blockchain-based efficient and incentive approach. IEEE Trans Ind Inform 15(12):6367–6378

Application of Artificial Neural Network (ANN) and Adaptive Neuro Fuzzy (ANFIS) Techniques for the Modelling and Optimization of COD Adsorption Process

Hifsa Khurshid, Muhammad Raza Ul Mustafa, and Yeek-Chia Ho

Abstract Artificial neural network (ANN) and adaptive neuro fuzzy (ANFIS) modelling techniques have been applied in this study to model and optimize the chemical oxygen demand (COD) adsorptive removal in produced water. The models were well trained and showed minimum error values for predicted data when compared to experimental data. The error values were 0.4035 and 0.2886 for sum of squared error (SSE), 0.1628 and 0.0832 for mean square error (MSE) and 0.13 and 0.23% for average relative error (ARE) using ANN and ANFIS, respectively. Error analysis and coefficient of determination (R^2) of the models determined that ANFIS was better than ANN for the prediction of COD adsorption on the biochar. Also, ANFIS required minimum run time as compared to ANN. Both artificial intelligence (AI) based techniques well predicted the optimized values of adsorption process, when compared with the experimental values. It is concluded that the use of AI techniques can inevitably pave the way in the water treatment sector using adsorption for improved efficiency and process automation.

Keywords Adsorption · Artificial intelligence · Optimization · Automation · Water treatment · Green technology

1 Introduction

Adsorption is an efficient and economical method of elimination of pollutants from wastewater and can be further improved by automation. The automation of wastewater treatment systems is necessary to making it easier to run, saving both manpower and resources [1]. However, it is very difficult to build an automatic wastewater treatment plant when the attributes of commercial wastewater change significantly, resulting in a drastic shift in the performance of the treatment plant. In fact, its

H. Khurshid (✉) · M. R. U. Mustafa · Y.-C. Ho
Department of Civil and Environmental Engineering, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Perak Darul Ridzuan, Malaysia
e-mail: hifsa_18002187@utp.edu.my

complex properties are very difficult to handle when it comes to the adsorption treatment process [2]. Due to the interaction of several factors in adsorption process, the resulting correlations are strongly non-linear [3]. This problem leads to inadequate understanding and false quantification of interactions between process inputs and outputs, and it is thus difficult to explain by linear mathematical models. The standard regression methods are also not sufficient for mimicking the experimental adsorption data as the methodology does not grasp the physics [1]. For interrelating the input and output parameters of adsorption process, advanced computer-simulated models are therefore required [4].

Artificial intelligence (AI) methods e.g. fuzzy logic (FIS) [5], particle swarm optimization (PSO) [6], genetic algorithm (GA) [7] and artificial neural networks (ANNs) [8], have been applied recently in water treatment and adsorption optimization and automation systems. ANNs are particularly well known for their high capacity to research and arrange data [9]. Researchers have reported different studies using ANN for modelling and optimization of adsorptive removal of various pollutants e.g. chromium removal using cyanobacterial biomass [10], copper removal using light expended clay aggregate [11], removal of various dyes using nanoparticles and activated carbons [12, 13], removal of As (III) and various other metals [5]. Fuzzy systems also have some advantages over traditional approaches, particularly where ambiguous data or prior experience is involved [14]. Few studies have reported application of ANFIS in adsorption studies for modeling and optimization of adsorption process e.g. indium (III) [15], dyes and metals [5]. Meta-heuristic optimization algorithms e.g. GA and PSO, are the products of social actions or natural phenomena and used to optimize the various operations. The PSO is a well-known optimization algorithm motivated by the action of a flock of birds. It is now widely used in system's optimization because it is not conveniently stuck in a local minimum.

Most of the studies in literature are performed on synthetic waters or stock solutions containing single pollutant as an objective. These solutions do not completely represent the industrial effluents, which contain multiple types of pollutants together. Hence, practical application of the adsorptive materials and automation of treatment plants becomes limited considering single pollutant studies. Therefore, the purpose of this research work was to present the design, implementation, comparison and evaluation of the ANN and ANFIS approaches to the adsorption of chemical oxygen demand (COD) in produced water (PW). The wastewater obtained during oil and gas exploration is known as PW. The COD in the water is caused by various carcinogenic organic pollutants. Organic pollutants have been confirmed to be one of the top priority pollutants to date including PAHs, dyes, and pharmaceuticals etc. COD is an important parameter for interpreting the organic pollutants in wastewater and has been used widely for effluents quality measurement [16]. Various studies have reported efficient adsorption of COD in wastewater which also represents the elimination of organic pollutants in the water [17, 18].

To the best of our knowledge, no study has been found in literature containing the application of ANN and ANFIS for modelling and optimization of COD adsorption process for predicting organic pollutants removal in PW. Tea waste (TW) had been transformed into biochar for this purpose and was applied under the control of various

parameters (pH, dosage and contact time). To improve the automated adsorptive removal of COD from the wastewater sample, adsorption process was optimized using the ANN-PSO and ANFIS techniques. The detailed objectives of this study are: to establish ANN and ANFIS models capable of predicting adsorption efficiency for the removal of COD from PW; Optimizing the COD adsorption process using ANN-PSO and ANFIS; and to incorporate ANN and ANFIS models such that they can be used effectively for PW treatment system automation. This study will help the decision makers and wastewater treatment plant operators in decision making, strategic designing of system, automation, and optimization of adsorption process.

2 Materials and Methods

2.1 Materials

Tea waste was obtained from the nearby restaurants in Malaysia. High range (HR) COD vials were obtained from the laboratory stock of Avanti's Malaysia. About 98%, H_3PO_4 was purchased in Malaysia from local supplier Merck. The water sample was obtained from an oil and gas company operating in South East Asia.

2.2 Preparation of Adsorbent

The biochar was produced using the chemical activation technique. For 12 h, tea waste was immersed in 10% phosphoric acid solution and then left to dry at room temperature. Biochar was developed by pyrolysis at 700 °C using the dried waste in the tube furnace. Afterward, the prepared biochar was cleaned with filtered water and kept in the oven to dry. In this study, the biochar was only used for exploring the application of AI based techniques in automation of adsorption process and mechanism evaluation was not the objective. Therefor the material was not characterized and used only for the experimental purpose.

2.3 Adsorption Batch Experiments

In this analysis, the commonly documented adsorption process was used to observe the elimination of COD [19]. A significant segment of AI configuration is the selection of input variables from a variety of parameters that influence the mechanism [1]. To select different important factors affecting the adsorptive removal of COD from PW, batch adsorption tests were performed. From the experiments and literature review, it was found that initial pH of water, dosage of adsorbent and contact

time were most significant input parameters effecting the adsorption process. Therefore, the biochar was applied in varying dosage (25–300 mg/L) for 100 mL of PW at varying initial pH (3–12) and shook for various contact times (10–60 min) at temperature 20 ± 5 °C. The percentage of COD removal was detected by the formula given [20]:

$$\text{Percentage removal} (\%) = \frac{(C_i - C_o) \times 100}{C_i} \tag{1}$$

where, C_i and C_o represent the initial and final concentrations of COD in mg/L, respectively.

2.4 Artificial Neural Network (ANN)

A three-layer feed forward neural network was formed using the neural network toolbox using programming codes written in MATLAB R2019b mathematical software. The input layer was given three inputs i.e. pH, contact time and dosage of biochar and one output e.g. removal efficiency of COD as shown in Fig. 1. Number of neurons in hidden layer were selected based on hit and trial method using 1–10 neurons. Higher number of neurons were not tested to avoid the over fitting of model. Cascade forward backpropagation (CFBP), Levenberg–Marquardt backpropagation (LMBP), Bayesian Regularization (BR) and Scaled Conjugate Gradient (SCG) algorithms were tested to find the best suitable algorithm. The most suitable algorithm
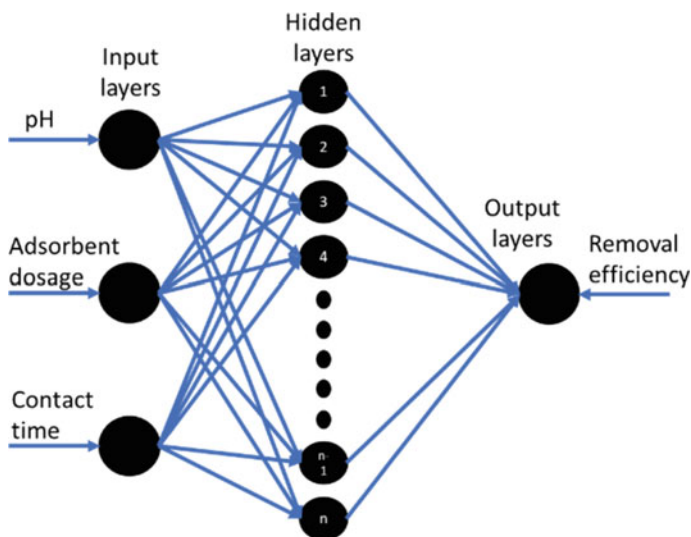
Fig. 1 Schematic diagram of ANN model

was selected based on minimum value of mean square error (MSE). For all data sets in ANN, the symmetric sigmoid transfer function (tansig) was used in the first layer and the linear transfer function (purelin) was used at the output node for the simulation and prediction of COD elimination. The backpropagation (BP-ANN) network was selected for the tuning of parameters. It is a first-order gradient descent technique to model experimental data [21]. Total 36 no. of data sets were taken through batch experiments. 70% data was used for training of model and 30% data was used for testing purpose. For validation of model and prediction of COD removal efficiency 13 more data values were taken through lab experiments using Box–Behnken Design (BBD) of experiments.

2.5 ANN-PSO Hybrid

ANN and PSO hybrid was used for the optimization of input variables using ANN predicted data with 1000 no of population, C1 = 1.5, C2 = 2.5 and 1000 iteration values. The schematic diagram of hybrid model is shown in Fig. 2. MSE was taken as function to be minimized. Tolerance value was taken as 1. Boundary violations were taken as 1.5 and −1.5. The algorithm was run for optimization further. The

Fig. 2 ANN-PSO hybrid
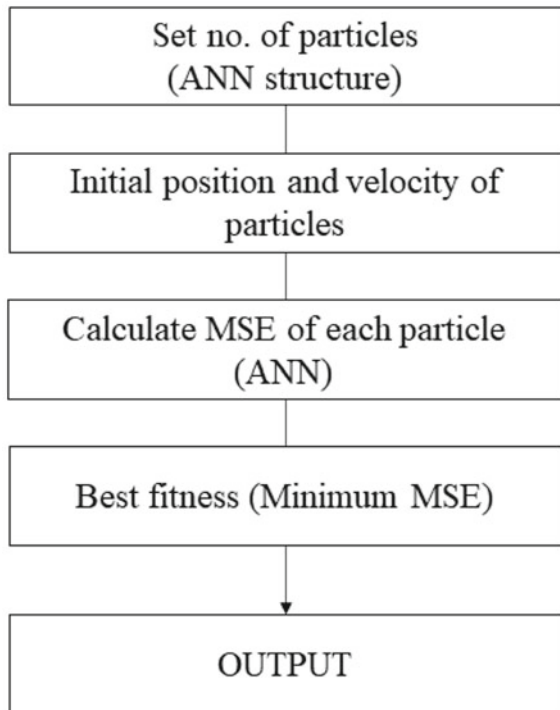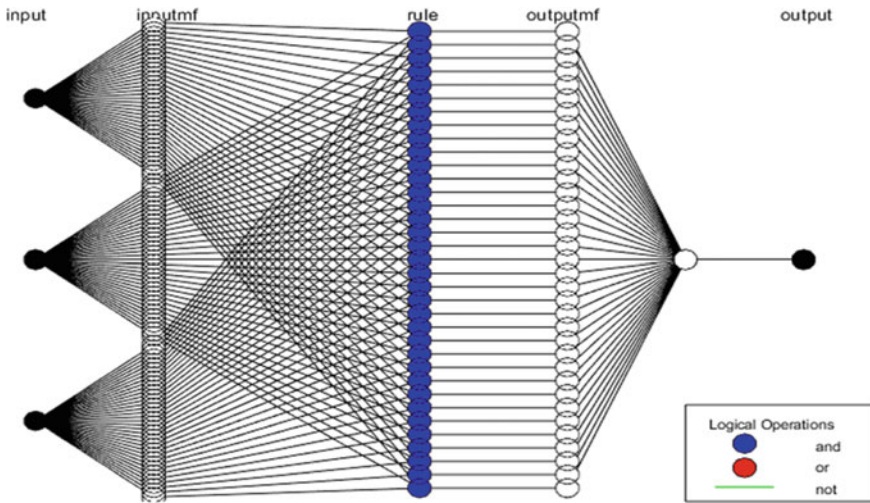
Fig. 3 Schematic diagram of ANFIS model

purpose was to obtain the input variables causing the maximum removal of COD in PW.

2.6 Adaptive Neuro-Fuzzy Interference Systems (ANFIS)

The fuzzy inference method (FIS) applied to ANFIS was developed in the MATLAB using Neuro-Fuzzy designer. The developed ANFIS was of the Sugeno type, which consist of four hidden layers [15]: (i) the fuzzification layer with two membership functions; (ii) the inference layer with weight rules; (iii) the de-fuzzification layer; and (iv) the output layer (Fig. 3). Data was randomly divided into training (70%) and testing data (30%). Three variables (pH, Dosage and contact time) were selected as inputs and removal efficiency of COD was taken as target. Range of influence was selected based on minimum MSE. To obtain the optimized results for COD adsorption using ANFIS, surface plots were generated.

2.7 Error Analysis

The performance of the ANN and ANFIS techniques for predicted data was evaluated using statistical equations i.e. (i) the coefficient of determination (R^2), (ii) the sum of squared errors (SSE), (iii) mean–squared error (MSE) and (iv) average relative error (ARE) [22]. Following equations were used to measure the errors [23]:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_{pred,i} - y_{\exp,i}\right)^2}{\sum_{i=1}^{n}\left(y_{pred,i} - y_m\right)^2} \tag{2}$$

$$SSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{pred,i} - y_{\exp,i}\right)^2} \tag{3}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_{pred,i} - y_{\exp,i}\right)^2 \tag{4}$$

$$ARE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_{pred,i} - y_{\exp,i}}{y_{\exp,i}}\right| \tag{5}$$

where, y_{pred} and y_{exp} denote the predicted and experimental values, respectively. n denotes the total number of values in data. Mean value of the response is denoted by y_m.

3 Results and Discussion

3.1 Modelling of ANN

The ANN structure was opted after analyzing the various parameters i.e. no. of neurons and training algorithm. The algorithm and no. of neurons were selected based on the best performance as shown in Fig. 4a, b. MSE was taken as performance criteria. Total 10 no. of neurons and LMBP algorithm were selected based on minimum MSE and simulation time. After selecting the structure of ANN, selected model was trained by using trainlm training function. It was observed that, after training, the output data had a correlation coefficient (R^2) of 0.99 with the experimental data (Fig. 4c). The MSE value was 0.0001 and training time was 400 s. The lower value of error indicated efficient performance of selected ANN model for the given dataset. For the validation and testing of model, correlation coefficients of 0.998 and 0.98 were obtained respectively (Fig. 4d).

Higher values of R^2 showed that the model was well trained and could be used effectively for the prediction of COD adsorption on biochar. The model was validated further and a correlation coefficient of 0.99 was obtained for the predicted data when compared with experimental data using 13 dataset points, as given in Table 1.
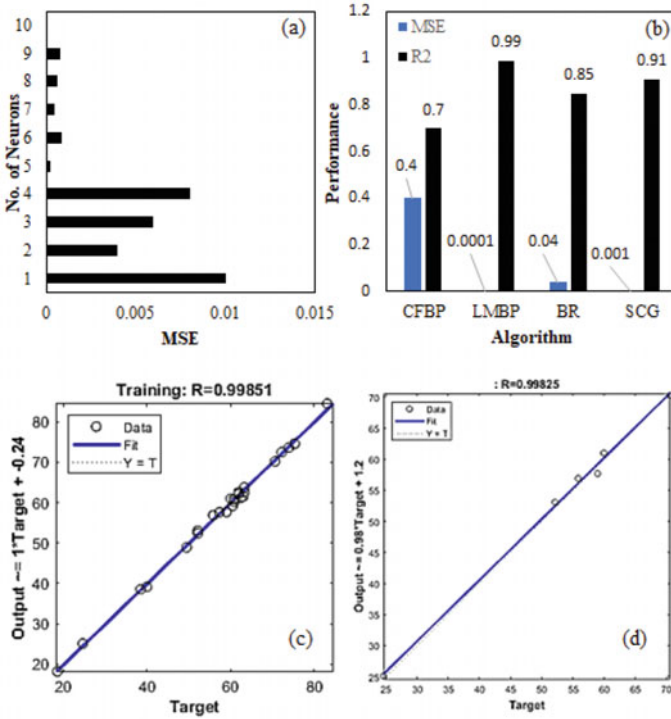
Fig. 4 Modelling of ANN; **a** relationship between MSE and no. of neurons in hidden layer, **b** relationship between CFBP, LMBP, BR and SCG algorithms and their performance, **c** ANN training, **d** ANN testing

3.2 Optimization of COD Adsorption Process Using ANN-PSO

The PSO technique was hybridized with the generated ANN model for optimizing process parameters with the goal of maximizing the COD removal efficiency. The performance of ANN-PSO model was analyzed on basis of MSE i.e., 5.27. Minimum value was obtained after 500 runs and it took time of 4000 s. The optimum conditions for the COD removal process were: 11.79 pH, 300 mg/L concentration of biochar and 60 min contact time. The COD removal efficiency obtained under optimum conditions was 89.80%. The value was confirmed in the lab and obtained as 89.3%. It indicated the good performance of ANN-PSO in predicting the COD adsorption optimized process parameters.

Table 1 Experimental and predicted results of COD removal efficiency using ANN and ANFIS

No	pH	Dosage	Contact time	COD removal efficiency (%)		
		mg/L	min	Actual	Predicted ANN	Predicted ANFIS
1	10	162.5	60	53.123	52.31	52.123
2	3	162.5	60	36.611	36.12	36.59
3	6.5	25	60	69.269	69.01	69.16
4	6.5	300	60	78.044	78.77	78.12
5	6.5	162.5	35	63.918	64.66	63.91
6	10	25	35	54.296	54.21	54.3
7	6.5	25	10	65.574	65.57	65.57
8	3	162.5	10	36.892	36.72	36.89
9	3	25	35	39.169	39.23	39.18
10	10	162.5	10	44.941	45.01	44.9
11	3	300	35	49.431	49.43	49.39
12	10	300	35	60.831	60.85	61.02
13	6.5	300	10	72.958	72.8	72.8
R^2					**0.9991**	**0.9995**
SSE					**0.4035**	**0.2886**
MSE					**0.1628**	**0.0832**
ARE					**0.13%**	**0.23%**

3.3 Modelling of ANFIS

Sugeno type sub-clustering ANFIS was generated for three inputs and one output. For all input variables membership functions were taken as 35 in number and guassmf type. Range of influence was selected as 0.00001 based on minimum MSE value, when tested from 0–1 as shown in Fig. 5a. Whereas, squash factor, accept ratio and rejection ratio were 1.25, 0.5 and 0.15, respectively. For training of generated FIS structure hybrid optimization method was taken as method type due to lower MSE (0.15721) as compared to backpropagation optimization method (2.32). Total 100 no of epochs were taken for training of ANFIS model using 70% dataset. The model was well trained and minimum MSE of 0.15721 was obtained at epoch 2 (Fig. 5b). The run time of ANFIS was only 10 s for 100 epochs which was much lesser than ANN. The model was further tested and checked using the 30% data. The MSE of 0.00017 and 0.0008 was obtained for the test and checked data, respectively (Fig. 5c, d). After training, the ANFIS model was used for prediction of COD removal efficiency using 13 input datasets. The predicted outputs were compared with the experimental data as given in Table 1. A higher correlation of 0.9995 and lower MSE of 0.0832 was obtained for the predicted data. It showed that generated ANFIS model could be used effectively for the optimization of COD adsorption process and selecting suitable input parameters.
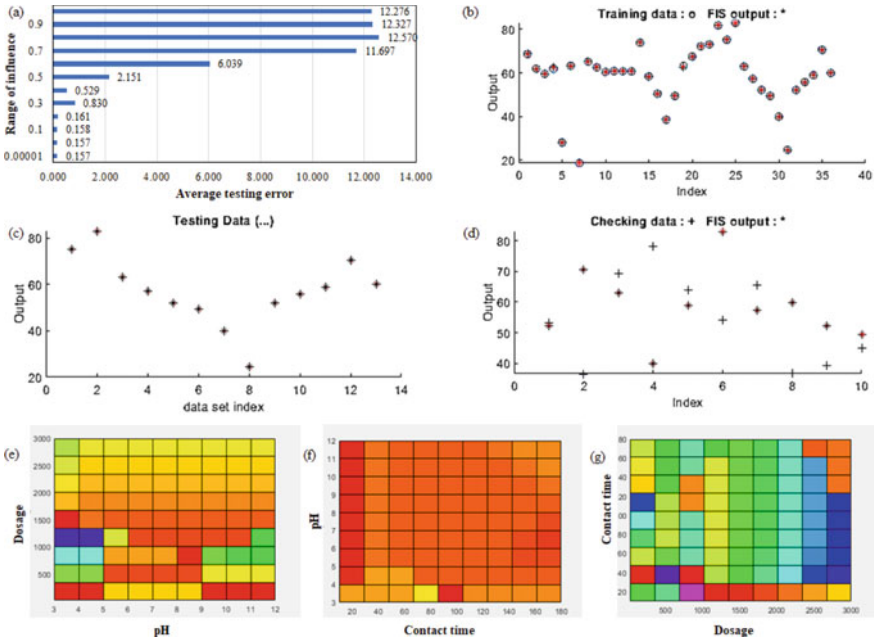
Fig. 5 Modelling and contour plots for COD adsorption on biochar using ANFIS; **a** error analysis, **b** ANFIS training data, **c** ANFIS testing data, **d** ANFIS checked data, **e** impact of pH and dosage on COD removal efficiency, **f** impact of pH and contact time on COD removal efficiency, **g** impact of dosage and contact time on COD removal efficiency

3.4   Optimization of COD Adsorption Process Using ANFIS

For the prediction of optimized values of adsorption data through ANFIS surface plots were generated for pH range 3–12, dosage 25–300 mg/L and contact time 10–60 min. Figure 5e–g shows that COD removal efficiency was highly affected by all the three variables. It also shows the correlation between each variable. Maximum value of COD removal was obtained at pH value of 12, dosage of 300 mg/L and contact time of 60 min. The COD removal efficiency achieved at the optimized parameters was 90.8%. The experimental value for the optimized input parameters was 89.3%. It showed that the model was able to predict the optimized values with significant accuracy.

3.5   Comparison of ANN and ANFIS

Developed ANN and ANFIS models were well trained for predicting the COD removal efficiency using pH, dosage of biochar and contact time as input variables. Both models well fitted the experimental data and predicted values were near to

Table 2 Optimized values for COD adsorption

Parameters	Experimental	ANN-PSO	ANFIS
pH	12	11.79	12
Dosage (mg/L)	300	300	300
Contact time (min)	60	60	60
COD removal efficiency (%)	89.30	89.80	90.80

experimental data (Table 1). Whereas ANN-PSO and ANFIS models also performed very well for the optimization of adsorption process (Table 2). Further performance of the models was analyzed using Eqs. 2–5 for error analysis of predicted data. The analysis is given in Table 1. The value of R^2 should lie between 0–1. Value near to 1 shows good correlation between the data sets. SSE has values range from 0–1, while best value is closer to 0 [4]. MSE minimum value is taken as better value [24], whereas ARE is %age error and minimum value is recommended [25]. Table 1 shows that ANN and ANFIS had R^2 values of 0.9991 and 0.9995, error values of 0.4035 and 0.2886 for SSE, 0.1628 and 0.0832 for MSE and 0.13% and 0.23% for ARE, respectively. Error analysis and coefficient of determination of the models determined that ANFIS > ANN in performance for the prediction of COD adsorption on the biochar. Also, for both methods no specific design of experiments was needed. It helped the models to analyze the correlation between inputs and outputs in broader range.

4   Conclusion

Tea waste biochar was synthesized and applied for treatment of produced water. Batch tests were performed using pH, dosage, and contact time as input variables. The experimental results were modelled and optimized using ANN and ANFIS models. ANN model was modelled using 3 layered, feed forward Levenberg–Marquardt backpropagation algorithm and 10 no. of neurons. For optimization it was hybridized with PSO. ANFIS was generated sugeno type using sub-clustering FIS type. The experimental values were successfully optimized using surface plots correlations. Optimization values were well matched with the experimental results. ANFIS required minimum run time (10 s for 100 epoch) as compared to ANN model (400 s for 1000 epoch). Error analysis and coefficient of determination of three models determined that ANFIS was slightly better than ANN in performance for the prediction of COD adsorption on the biochar. However, it can be stated that the ANN and ANFIS methods are promising techniques and can be used effectively for the automation of COD adsorption process for produced water treatment. Both methods have high capability to predict the adsorption efficiency of pollutants considering multiple factors and can be implemented for sustainable and automated water treatment systems.

References

1. Anupam K, Dutta S, Bhattacharjee C, Datta S (2016) Artificial neural network modelling for removal of chromium (VI) from wastewater using physisorption onto powdered activated carbon. Desalin Water Treat 57(8):3632–3641. https://doi.org/10.1080/19443994.2014.987172
2. Al Aani S, Bonny T, Hasan SW, Hilal N (2019) Can machine language and artificial intelligence revolutionize process automation for water treatment and desalination? Desalination 458:84–96. https://doi.org/10.1016/j.desal.2019.02.005
3. Khan T, Mustafa MRU, Isa MH, Manan TSBA, Ho YC, Lim JW et al (2017) Artificial neural network (ANN) for modelling adsorption of lead (Pb (II)) from aqueous solution. Water Air Soil Pollut Water 228(11). https://doi.org/10.1007/s11270-017-3613-0
4. Dutta S, Parsons SA, Bhattacharjee C, Bandhyopadhyay S, Datta S (2010) Development of an artificial neural network model for adsorption and photocatalysis of reactive dye on TiO$_2$ surface. Expert Syst Appl 37(12):8634–8638. https://doi.org/10.1016/j.eswa.2010.06.090
5. Dolatabadi M, Mehrabpour M, Esfandyari M, Alidadi H, Davoudi M (2018) Modeling of simultaneous adsorption of dye and metal ion by sawdust from aqueous solution using of ANN and ANFIS. Chemom Intell Lab Syst 181:72–78. https://doi.org/10.1016/j.chemolab.2018.07.012
6. Fan M, Hu J, Cao R, Xiong K, Wei X (2017) Modeling and prediction of copper removal from aqueous solutions by nZVI/rGO magnetic nanocomposites using ANN-GA and ANN-PSO. Sci Rep 7(1):1–14. https://doi.org/10.1038/s41598-017-18223-y
7. Bhatti MS, Kapoor D, Kalia RK, Reddy AS, Thukral AK (2011) RSM and ANN modeling for electrocoagulation of copper from simulated wastewater: multi objective optimization using genetic algorithm approach. Desalination 274(1–3):74–80. https://doi.org/10.1016/j.desal.2011.01.083
8. Mortula M, Abdalla J, Ghadban A (2010) Modeling phosphorus removal process using artificial neural network modeling approach. In: BALWOIS 2010—Ohrid, pp 1–7. Republic of Macedonia
9. Igwegbe CA, Mohmmadi L, Ahmadi S, Rahdar A, Khadkhodaiy D, Dehghani R (2019) Modeling of adsorption of methylene blue dye on Ho-CaWO4 nanoparticles using response surface methodology (RSM) and artificial neural network (ANN) techniques. MethodsX 6:1779–1797. https://doi.org/10.1016/j.mex.2019.07.016
10. Sen S, Nandi S, Dutta S (2018) Application of RSM and ANN for optimization and modeling of biosorption of chromium(VI) using cyanobacterial biomass. Appl Water Sci 8(5):1–12. https://doi.org/10.1007/s13201-018-0790-y
11. Shojaeimehr T, Rahimpour F, Khadivi MA, Sadeghi M (2014) A modeling study by response surface methodology (RSM) and artificial neural network (ANN) on Cu2+ adsorption optimization using light expended clay aggregate (LECA). J Ind Eng Chem 20(3):870–880. https://doi.org/10.1016/j.jiec.2013.06.017
12. Ghaedi AM, Karamipour S, Vafaei A, Baneshi MM, Kiarostami V (2019) Optimization and modeling of simultaneous ultrasound-assisted adsorption of ternary dyes using copper oxide nanoparticles immobilized on activated carbon using response surface methodology and artificial neural network. Ultrason Sonochem 51:264–280. https://doi.org/10.1016/j.ultsonch.2018.10.007
13. Gadekar MR, Ahammed MM (2019) Modelling dye removal by adsorption onto water treatment residuals using combined response surface methodology-artificial neural network approach. J Environ Manage 231:241–248. https://doi.org/10.1016/j.jenvman.2018.10.017

14. Sargolzaei J, Haghighi Asl M, Hedayati Moghaddam A (2012) Membrane permeate flux and rejection factor prediction using intelligent systems. Desalination 284:92–99. https://doi.org/10.1016/j.desal.2011.08.041

15. Franco DSP, Duarte FA, Salau NPG, Dotto GL (2020) Analysis of indium (III) adsorption from leachates of LCD screens using artificial neural networks (ANN) and adaptive neuro-fuzzy inference systems (ANIFS). J Hazard Mater 384:121137. https://doi.org/10.1016/j.jhazmat.2019.121137

16. Malakahmad A, Law MX, Ng KW, Manan TSA (2016) The fate and toxicity assessment of polycyclic aromatic hydrocarbons (PAHs) in water streams of Malaysia. Procedia Eng 148:806–811. https://doi.org/10.1016/j.proeng.2016.06.572

17. Mohammad-pajooh E, Turcios AE, Cuff G, Weichgrebe D, Rosenwinkel KH, Vedenyapina MD et al (2018) Removal of inert COD and trace metals from stabilized landfill leachate by granular activated carbon (GAC) adsorption. J Environ Manage 228:189–196. https://doi.org/10.1016/j.jenvman.2018.09.020

18. Bansode RR, Losso JN, Marshall WE, Rao RM, Portier RJ (2004) Pecan shell-based granular activated carbon for treatment of chemical oxygen demand (COD) in municipal wastewater. Biores Technol 94(2):129–135. https://doi.org/10.1016/j.biortech.2003.12.009

19. Ademiluyi F, Amadi S, Amakama N (2010) Adsorption and treatment of organic contaminants using activated carbon from waste Nigerian Bamboo. J Appl Sci Environ Manage 13(3). https://doi.org/10.4314/jasem.v13i3.55351

20. Azari A, Mahmoudian MH, Niari MH, Eş I, Dehganifard E, Kiani A et al (2019) Rapid and efficient ultrasonic assisted adsorption of diethyl phthalate onto FeIIFe2IIIO4@GO: ANN-GA and RSM-DF modeling, isotherm, kinetic and mechanism study. Microchem J 150:104144. https://doi.org/10.1016/j.microc.2019.104144

21. Olawoyin R (2016) Application of backpropagation artificial neural network prediction model for the PAH bioremediation of polluted soil. Chemosphere 161:145–150. https://doi.org/10.1016/j.chemosphere.2016.07.003

22. Dutta M, Basu JK (2013) Application of artificial neural network for prediction of Pb(II) adsorption characteristics. Environ Sci Pollut Res 20(5):3322–3330. https://doi.org/10.1007/s11356-012-1245-x

23. Khoshsang H, Ghaffarinejad A (2018) Rapid removal of lead (II) ions from aqueous solutions by saffron flower waste as a green biosorbent. J Environ Chem Eng 6(5):6021–6027. https://doi.org/10.1016/j.jece.2018.09.020

24. Dil EA, Ghaedi M, Ghaedi A, Asfaram A, Jamshidi M, Purkait MK (2016) Application of artificial neural network and response surface methodology for the removal of crystal violet by zinc oxide nanorods loaded on activate carbon: kinetics and equilibrium study. J Taiwan Inst Chem Eng 59:210–220. https://doi.org/10.1016/j.jtice.2015.07.023

25. Singh DK, Verma DK, Singh Y, Hasan SH (2017) Preparation of CuO nanoparticles using Tamarindus indica pulp extract for removal of As(III): optimization of adsorption process by ANN-GA. J Environ Chem Eng 5(1):1302–1318. https://doi.org/10.1016/j.jece.2017.01.046

An Evolutionary Stream Clustering Technique Outlier Detection in Medical Data

Nadilah Ayu Supardi, Said Jadid Abdulkadir, and Norshakirah Aziz

Abstract Clustering data streams from data mining become evident as one of the most well-liked studies among the researchers correspond to their evolutionary field. Numerous threats addressed by data streams on clustering like limited time, memory and single scan data. In generic terms, data streams are defining an infinite sequence of the element data and evolve without prior knowledge number of the clusters. Many factors such as noise (outliers) have appeared periodically have a negative impact on data streams environment. The density-based technique has demonstrated to be an astonishing method in clustering data streams. It is computationally competent to generate arbitrary shape clusters and detect noise instantaneously. The number of the clusters are not required in advance to set as a parameter during the assessment begin. In contradictory, traditional density-based clustering is not relevant to conduct in data streams due to its own characteristics. Mostly all traditional density-based clustering can be extended to the updated version of algorithms to achieve the objective of data streams research. The idea is emphasizing on the density-based technique in the clustering process dominate the restrain from data streams nature. The objective of this paper intends a preliminary result on a density-based algorithm, named evoStream to explore outlier detection on medical data sets, heart failure clinical records and gene expression cancer RNA-seq. In due course, extensive evoStream later to develop to optimize the model to detect outlier in data streams.

Keywords Data streams · Outlier · Density-based · evoStream · Data sets · Evaluation metrices

1 Introduction

Due to the fact, advance applications are practically implemented in related to various field such as medical research. Often, data streams relate with as an enormous,

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022
R. Ibrahim et al. (eds.), *International Conference on Artificial Intelligence for Smart Community*, Lecture Notes in Electrical Engineering 758,
https://doi.org/10.1007/978-981-16-2183-3_50

continuous and infinite data [1]. However, machine learning [2–4] address an issue of interpretation data due to ambiguity from data streams. Data streams evolve with time and complicate the stream processing. Data streams size increasing disproportionately and leads to infinite data size [5]. Consequently, the transmission data streams have influence memory space [6, 7]. Generally, the process to capture data streams are extremely difficult to determine the number of the cluster and the shape of the clusters where it considered unknown [8]. The outcome of clustering results is changing consistently as the data arrives in streams [9–11]. The clustering process and the shape of the clusters are mostly affected by outliers [11]. An outlier is defined generally by [12, 13] which appears inaccurate information from the data.

This paperwork highlights the importance for clustering technique in outlier detection due to following factors [7, 14, 15]; (i) an outlier may distort data structures and cluster result, (ii) the clustering methods can identify outlier detection as a small cluster, (iii) the hybrid clustering algorithms are exemplifying the result for outlier detection performs better than the existing algorithms.

In research studies, evoStream refers to data sets to conduct the clustering process in the online step during idle periods and to optimize the outcome of the cluster [16]. Thus, becomes a motivation to research problem. This research, however, was primarily aimed at evaluating the outlier identification of evoStream implementation with real datasets called clinical records as *preliminary results.*

The rest of this paper is arranged as follows: The related works is summarized in Sect. 2. Methodology described in Sect. 3. The assessment is seen in Sect. 4, and the result is outlined in Sect. 5.

2   Related Works

This section provides supplemental information literature review on density-based clustering techniques.

2.1   *Density-Based Clustering*

Density-based clustering technique is one of the prominent applied by the researcher. To identify density-based characteristics are between regions and density functions [13]. Comprehensively, this technique takes place when data point calculates referring to density area. Otherwise stated further refines where data points are in a low-density area perceives as outliers. Algorithms are mostly request parameter in order to determine the radius from data point at a minimum number in a cluster [9]. In most cases, the algorithms immediately tackle arbitrary-shaped clusters identification from the data stream. Nearly the entire algorithms are highly able to scrutinize clusters and form arbitrary shapes structure and detect outliers. Contrarily, this technique is limited to a single scan from raw data [9, 17].

According to the recent study [18] found in the survey has made with distinct approaches where density-based technique enhanced by practicing an efficient density-based clustering along with maintaining the evolving medical data clusters.

evoStream algorithm from density-based clustering technique [16, 19] is implemented on the following methodology in Sect. 3 to present the *preliminary study* on outlier detection. The term of evoStream comprises from evolutionary algorithms (EA) and DBStream [16]. evoStream algorithm brought together on pros and cons from [19]. The *advantage* of evoStream are; (i) the performance of the final cluster during online phase with high processing speed, (ii) removes computational overhead. On the contrary, the *disadvantage* of evoStrem; (i) the application is not suitable for multi-objective data stream, (ii) high dimensional datasets are impossibly executed in evoStream. The role of evoStream exclusively focusing on the best optimization solution for an online phase in the idle time to improve the final clusters. Hence, a new approach to the purpose of this algorithm is mainly focusing on investigating outlier detection from an inconsistency data set.

3 Methodology

An initial method to discover outlier detection in real data sets are implementing by an algorithm named evoStream [16]. The purpose of this algorithm is mainly focusing on investigating outlier detection from an inconsistency data object to differentiate from a categorical problem.

3.1 Proposed Methodology

Benchmarking data sets are retrieving collected from an open source *UCI Machine Learning Repository*. The results yielded from evoStream are the *preliminary result* for outlier detection. The process of evoStream is comprehensively scrutinizing through preprocessing data, calculating centers and weight of clusters, profiling clusters results, optimizing weight vector, develop as well as finalizing the validity of optimizing evoStream model.

The beginning process is data collection and preprocessing. The data sets imported into the R language and the data sets implemented in evoStream algorithm. The radius, r has been set to $r = 0.05$ and k which represents as macro-cluster has been fixed to $k = 3$ with *incremental generations* $= 1$ as well as *re-cluster generations* $= 1000$. Then, initialize evoStream package to insert observations, $n = 1000$ to process with calculating centers and weight for micro-clusters and macro-clusters. The plotting result is described according to the centers and weight of micro-clusters and macro-clusters. Additional generations, $n = 2000$ examine to observe the different to improve from the current plot result. The numbers assigned from micro clusters to macro-clusters according to the nearest neighbor of data points to cluster.

Weight vector optimization utilized to enhance the product of evoStream model by integrating the *Grey Wolf Optimization (GWO) algorithm* for an outcome which later implemented in due course. Measures weight vector will produce the optimization result. As a result, the data objects will be clusters accordingly in the respective groups.

3.2 Evaluation Metrics

Evaluation metrics are applied to evaluate the quality of statistical machine learning model, well-known as cluster validation with known as internal measures and external measures [8, 20]. The process cluster validation investigated under the measurement of an algorithm to generate excellent results in data mining, particularly in clustering techniques [20].

Seven (7) multitudes metrics *SSQ* and *silhouette* width are internal evaluation metrics while *purity, precision, recall, F-measure, NMI* are external evaluation metrics are selected to evaluate comparative analysis from [8, 10, 16]. Seven metrics referring to Mathematically, the metrics are indicating with the following equations except *SSQ*:

$$Silhouette = a(\boldsymbol{o}) = \frac{\sum o' \epsilon C_i, o \neq o' \ dist(o, o')}{|C_i| - 1} \tag{1}$$

$$NMI = \frac{\sum d_{h,l} \log\left(\frac{|\Omega|.d_{h,l}}{d_h c_l}\right)}{\sqrt{\left(\sum h d_h \log\left(\frac{d_h}{d}\right)\right)\left(\sum l C_l \log\left(\frac{c_l}{d}\right)\right)}} \tag{2}$$

$$F - Measure = \frac{1}{k} \sum_{i=1}^{k} Score C_i \tag{3}$$

$$\text{Purity} = \frac{1}{k} \sum_{i=1}^{k} Precision C_i \tag{4}$$

$$\text{PrecisionC}_i = \frac{V_{isum}}{n_{ci}} \tag{5}$$

$$\text{RecallC}_i = \frac{V_{isum}}{V_{itotal}} \tag{6}$$

4 Assessment

4.1 Experimental Setup

Programming language R used to evaluate the algorithm. All experiments are performed on Windows Machine with intel i5-4200U CPU @ 1.60 GHz 2.30 GHz. There are two real datasets utilize for this investigation.

The *gene expression cancer RNA-Seq data set* is a random extraction of gene expressions of patients having different types of tumor: BRCA, KIRC, COAD, LUAD and PRAD. This data set gathered from the University of Genoa. In addition, the *heart failure clinical records data set* consist of the medical records of 299 patients who had heart failure, collected during their follow-up period, where each patient profile has 13 clinical features and originally collected from Government College University, Faisalabad, Pakistan.

4.2 Result

In this section, Figs. 1 and 2 show the final improved result from the preliminary studies of two different data sets. Each of these benchmarking data sets employed evoStream algorithm to report the studies outlier detection. The initial observations of the online component $n = 1000$ required to complete at an early stage of the investigation. During the idle time, the algorithm generates a macro-cluster when there is no observation arrives, and the data point will move into their pertinent clusters. Note that the performance of evolutionary step will not slow down the processing time.

In general, evoStream did not employ re-clustering approach but applying re-clustering steps when macro-clusters are requested and not incrementally Evaluating generations does not affect underlying micro-clusters but does affect minor improvement when implementing with 2000 generations into re-clustering. Hence,

Fig. 1 Gene expression cancer RNA-Seq data set
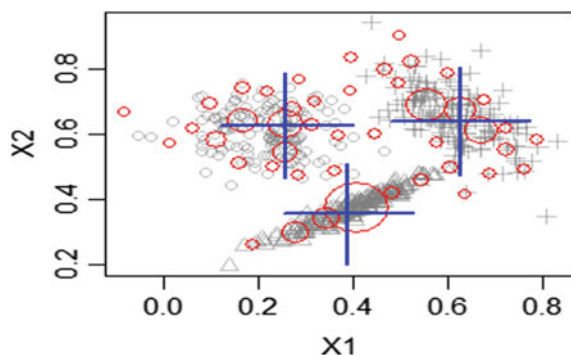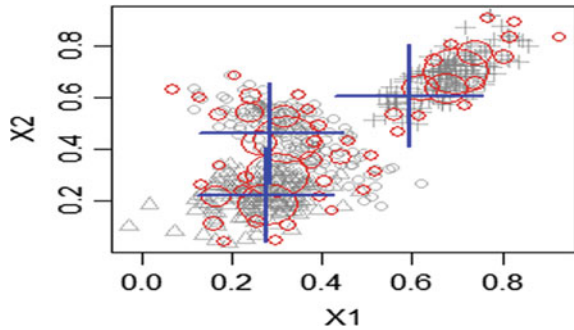
Table 1 Internal and external evaluation metrics: micro-cluster

Data set/Quality metrics	Gene expression cancer RNA-Seq data set	Heart failure clinical records data set
SSQ	0.11744971	0.1075193
Silhouette	0.20763806	0.2518121
Purity	0.97747748	0.9761905
Precision	0.98170732	0.9821429
Recall	0.09895513	0.1012891
F1-measure	0.17978783	0.1836394
NMI	0.55517022	0.5560689

re-clustering testified the slightest improvement because additional generations evaluated.

Moving on to the Tables 1 and 2 shown results from quality metrics of internal and external cluster validation with types micro-clusters and macro-clusters. The target minimum of median SSQ of micro-cluster achieved better results in *heart failure clinical*. Contradictory to the macro-cluster from internal metrics where the median SSQ achieve better result in *gene expression cancer RNA-Seq data set*. On the other

Table 2 Internal and external evaluation metrics: macro-cluster

Data set/Quality metrics	Gene expression cancer RNA-Seq data set	Heart failure clinical records data set
SSQ	1.2751618	1.7381368
Silhouette	0.5845640	0.5557643
Purity	0.9809140	0.9814815
Precision	0.9582071	0.9580764
Recall	0.9640463	0.9604450
F1-measure	0.9611179	0.9592593
NMI	0.9179111	0.9301648

hand, external metrics yielded different numbers according to their performance of data set.

The values reveal the measurement quality metrics from evoStream algorithm into two sample data sets. The finish result will be contributed for later development and validation of an optimize evoStream model.

5   Conclusion

The incoming data made stream processing is difficult to track down when data are processed in real-time and brings the challenge of understanding the hidden patterns from data. For that reason, the clustering technique introduced to unveil the nature of data streams, especially outliers. Thereby, density-based clustering technique is selected and has proven the most decent model in handling this investigation studies. The proposed methodology will be introduced to optimize an evoStream model to encounter the current problem. The purpose of initial investigation to analyze an optimum weight vector by clustering the data according to their characteristics. Clustering analysis outstandingly comes up with an accurate understanding of relevant variables in the clusters. All in all, determine a clustering model is validating the significance of the research question to achieve a valid result.

References

1. Twinkle JSD, Ankleshwaria B (2014) Mining data streams : a survey. 2(2):379–386
2. Abdulkadir SJ, Yong S-P (2013) Unscented kalman filter for noisy multivariate financial time-series data. In: International workshop on multi-disciplinary trends in artificial intelligence. Springer, pp 87–89
3. Abdulkadir SJ, Yong S-P, Marimuthu M, Lai F-W (2014) Hybridization of ensemble kalman filter and non-linear autoregressive neural network for financial forecasting. In: Mining intelligence and knowledge exploration. Springer, pp 72–81
4. Abdulkadir SJ, Yong S-P (2015) Scaled ukf–narx hybrid model for multi-step-ahead forecasting of chaotic time series data. Soft Comput 19(12):3479–3496
5. Ahmed KN, Razak TA (2014) A comparative study of different density based spatial clustering algorithms. Int J Comput Appl 99(8):18–25
6. Kumar A, Singh A (2017) Stream mining a review: tool and techniques. In: Proceedings of the international conference on electronics, communication and aerospace technology ICECA 2017, pp 27–32
7. Vijayarani S, Jothi P (2016) A hybrid clustering algorithm for outlier detection in data streams. Int J Grid Distrib Comput 9(11):285–396
8. Kokate U, Deshpande A, Mahalle P, Patil P (2018) Data stream clustering techniques, applications, and models: comparative analysis and discussion. Big Data Cogn Comput 2(4):32

9. Ghesmoune M, Lebbah M, Azzag H (2016) State-of-the-art on clustering data streams. Big Data Anal 1(1):13

10. Yeoh JM, Caraffini F, Homapour E, Santucci V, Milani A (2019) A clustering system for dynamic data streams based on metaheuristic optimisation. Mathematics 7(12):1–24

11. Agrawal LS, Prof A, Adane DS (2016) Models and issues in data stream mining. Int J Comput Sci Appl 9(1):6–10

12. Patel SP, Vala JAY (2015) A survey of outlier detection in data mining. Int J Adv Eng Res Dev 3(01):1–4

13. Bansal R, Gaur N, Singh SN (2016) Outlier detection: applications and techniques in data mining. In: Proceedings of the 6th international conference—cloud system big data engineering confluence, pp 373–377

14. Hassani M, Seidl T (2016) Clustering big data streams: recent challenges and contributions. IT—Inf Technol 58(4)

15. Yogita, Toshniwal D (2013) Clustering techniques for streaming data-a survey. In: Proceedings of the 3rd IEEE international advance computing conference IACC, pp 951–956

16. Carnein M, Trautmann H (2018) evoStream—evolutionary stream clustering utilizing idle times. Big Data Res 14:101–111

17. Mousavi M, Bakar AA, Vakilian M (2015) Data stream clustering algorithms: a review. Int J Adv Soft Comput Appl 7(3):1–15

18. Al-Shammari A, Zhou R, Naseriparsaa M, Liu C (2019) An effective density-based clustering and dynamic maintenance framework for evolving medical data streams. Int J Med Inform 126:176–186

19. Aljibawi M, Zakree M, Nazri A, Othman Z (2018) A survey on clustering density based data stream algorithms. 7:147–153

20. Palacio-Niño J-O, Berzal F (2019) Evaluation metrics for unsupervised learning algorithms

Analysis of Intelligent Agent and Its Components for the SLA Negotiation Process in Cloud Computing

Rishi Kumar, Mohd Fadzil Hassan, and Muhamad Hariz M. Adnan

Abstract Cloud service providers (CSP) are facing a competitive environment in the cloud-computing world and CSP is huge in demand with the growing field of IoT, fog, and cloud services. Consumers are always looking forward to acquiring services as per requirement. Due to the competitive field in the market of cloud computing, consumers face issues during the delivery of services. Consumer and CSP need an intelligent agent to mediate which performs the negotiations process on behalf of both parties and maintaining other responsibilities. An intelligent agent-based negotiation system (IANS) is required to perform the negotiation process and measure the KPIs during the process that helps to make quick decisions in minimum time. Researchers suggested many learning algorithm-based agents that maximize the success rate and speed of negotiation. IANS with behavior analysis of consumers and CSP helps to opt for more reliable and secure services. IANS uses negotiation metrics, performance factors, and evaluation measures for decision. Due to the dynamic environment of cloud services, IANS is recommended for the negotiation process.

Keywords Agent · SLA · Cloud services · Negotiation · Intelligent system · Cloud computing

R. Kumar (✉) · M. F. Hassan
Department of Computer and Information Sciences, Universiti Teknologi Petronas, Seri Iskandar, Malaysia
e-mail: rishi_18001163@utp.edu.my

M. F. Hassan
e-mail: mfadzil_hassan@utp.edu.my

M. H. M. Adnan
Computing Department, Faculty of Arts, Computing, and Creative Industry, Universiti Pendidikan Sultan Idris, Tanjung Malim, Malaysia
e-mail: mhariz@fskik.upsi.edu.my

1   Introduction

Cloud computing services provide solutions to many ventures in form of infrastructure, software, and platform, and business process [1]. A cloud is a large group of interconnected computers that provide the required services to any scale of the enterprise [2]. To obtain required services both consumers and providers must settle on pre-established contracts called Service Level Agreements (SLA) [3]. The resources assigned are based on SLA that contain Service Level Objectives (SLOs) for each SLA parameter, like price, availability, response time, reliability, CPU utilization, etc. [4]. Consumers and CSP have struggled over requirements; negotiation can provide a solution to established deliverable SLA between them [5].

Negotiation is a process of settlement of SLA requirements between consumers and CSP via the agent [6]. Consumer's requirements are dynamic and change according to time and demand, in this manner consumers cannot stick to one CSP [7]. To sustain those dynamic requirements with compound services, there is a necessity of designing an intelligent agent-based negotiation system (IANS) [7, 8].

IANS can provide negotiation, decision-making, resource management, interacting with other interconnected CSP also called inter-cloud, knowledge acquisition; analyze behavior, monitoring, maximizing success rate in minimum time [9–11]. IANS maintains the requirements for cloud services within minimum response time, maximum availability, reliable server, and services at the minimum price [3, 4].

2   Background Study

SLA is legal documents that are defined and modeled on which CSP and consumers must agree, approved, and signed after SLA negotiation [12]. SLA violations may lead to monetary penalties, credibility, and reduced reliability for services. SLA management describes complete assistance to configure SLA negotiation metrics information, automatic negotiation, and monitoring, speedy recovery over issues, and quick response over violation [2, 13].

Negotiation between CSP and a consumer based on metrics mainly price (P), availability (Av), the response time (RT), reliability (R) [3]. Performance of cloud services, negotiation strategy, and the process can be measure by a utility function (UF), negotiation speed (NS), success rate (S) [14]. Factors to evaluate the strategy of the negotiation process are the number of resources (NoR), Deadline (DL), degree of satisfaction (DoS) [3, 6, 15, 16].

In the Negotiation process, consumers want to fulfill maximum requirements and on the other hand, providers want to maximize sales of services. IANS focus on the cycle of negotiation, negotiation matrices, and performance metrics [17].

Negotiation metrics: Cloud consumers must understand the key performance indicators (KPI) for better negotiation. In Table 1, there are eight major high-level cloud service attributes and under each level, some major KPIs to evaluate and

Table 1 High-level cloud services attributes and KPIs [4]

S. No.	Cloud service attributes	Key performance indicators
1	Financial	**Price of service**, free features, cost of security
2	Security and privacy	Data center and location, encryption and decryption, privacy, trust
3	Performance	**Response time**, efficiency, latency, load balancing, network quality, robustness, throughput, scaling latency
4	Agility	Adaptability, elasticity, flexibility, portability
5	Usability	Accessibility, compatibility, controllability, functionality, maintainability, suitability, transparency, pooling
6	Accountability	Auditability, integration support, sustainability
7	Assurance	**Reliability, availability**, credibility, fault tolerance, service stability
8	Management	Data management, monitoring, value-added services

negotiate [4]. According to [3, 6, 11, 16], major negotiation metrics are price (P), availability (Av), the response time (RT), reliability (R).

3 Framework and Workflow of Ians

The proposed IANS design in Fig. 1. Follow the life cycle of SLA. IANS helps to finds out the suitable CSP concerning the consumer requirements [7]. Agent components allow us to understand the behavior of CSP by learning algorithms like reinforcement learning [15], which help to speed up the process of negotiation.

There are five steps for SLA satisfaction between CSP and consumer [2, 5].

1. **SLA Definition:** CSP provides SLA with available offers and resources. A CONSUMER SLA covers the requirements of the resource. Both types of SLA can customize later [2].
2. **Published and Discovery:** Consumer published requirement to IANS, which translates and analyzes the SLA and discover suitable CSP from its database. Discovery of CSP based on rank-based that depends upon security, trust, performance, financial, etc. IANS analyzes behavior property based on the functioning of KPIs [2, 4].
3. **Negotiation:** Negotiation process has five major steps, (i) assessment of SLA, (ii) allowance computation, (iii) achievable solution development, (iv) produce offer, and (v) negotiation termination [16].

   (a) **Assessment of SLA:** IANS assess the SLA and analyze the requirements of resources with price, response time, and availability. IANS computes UF, if the value of UF maximum, the agent gives positive feedback [16, 19].
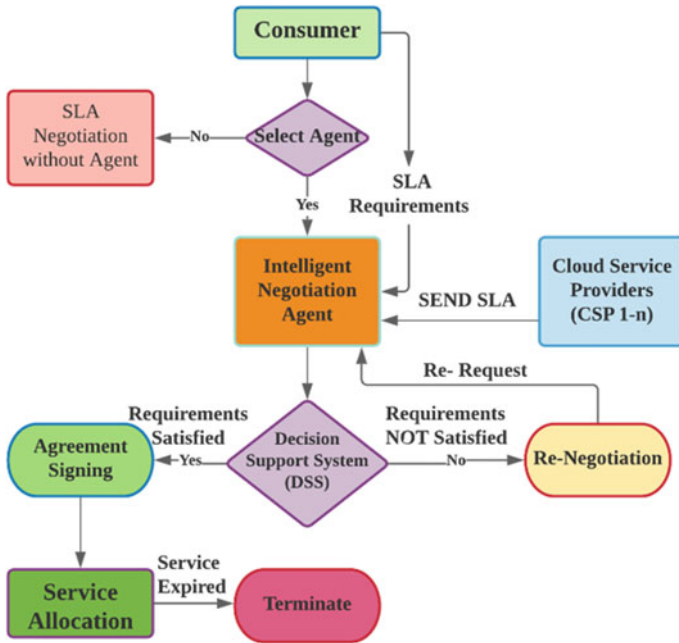
Fig. 1 The framework of intelligent agent-based negotiation system [18]

- (b) **Allowance Computation:** IANS computes discounts over the prices of different resources. IANS computes the difference between previous offers and current offers and produces satisfaction value [16].
- (c) **Achievable Solution Development:** IANS keep threshold value (TV) for KPI. If there is a vast difference concerning TV, then IANS produce negative feedback, in case of negligible difference, gives positive feedback, which reduces '*Ti*' and '*Nr*' and increases NS [11, 16].
- (d) **Produce Offer:** While comparing and analyzing the latest offer by CSP, IANS makes the final call to produce an SLA offer. Results will mainly depend upon the DoS of consumers and CSP for SLA and UF of consumers [11, 16].
- (d) **Negotiation Termination:** The termination of the negotiation process has two values success and failure. Successes, if the process meets with requirements before the deadline (DL) with a minimum number of negotiation rounds. Failure, when define deadline is over before settlement. In case of failure, the consumer offered a new SLA or new CSP for the further process [11, 16].

4. **Operationalization:** The decision support system (DSS) of IANS will finalize the agreement and start operation. Operationalization of SLA means negotiation completed successfully. Operationalization includes monitoring of secure service delivery, accounting of SLA, execution of SLA [2].
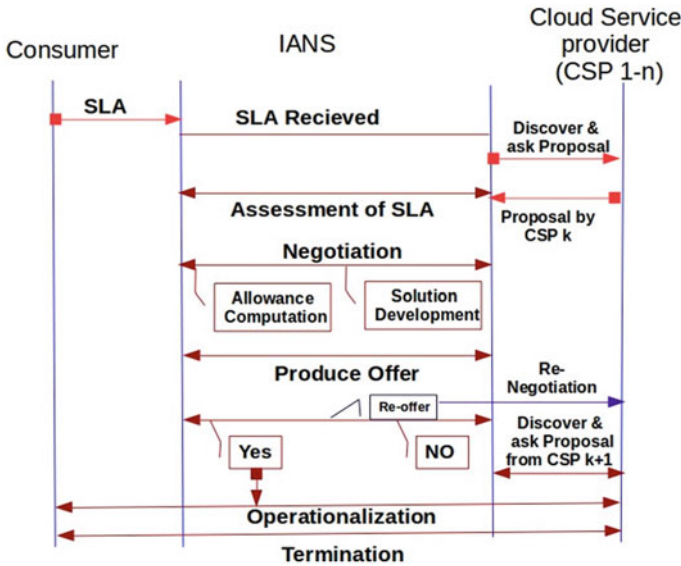
Fig. 2 Workflow diagram of negotiation process in IANS [16]

5. **Termination:** After the completion of the process, sharing of resources and all activities ended [2].

In Fig. 2 [16], Workflow Diagram of Negotiation process showing communication between consumer and CSP via IANS. IANS helps to make the process fast, secure, and provide the best options for the consumer from a massive number of CSP [12].

4   Conclusion and Future Work

In this research study, IANS functioning explored dependency factors of SLA negotiation process for cloud services. It is defined that cloud SLA negotiation process with IANS for negotiation will understand the requirement of the client and produces perfect match from the massive number of CSPs with maintaining KPIs price, availability, response time and reliable. The descriptive analysis of the agent, its key factor, and functioning is to develop an intelligent intermediary for the negotiation process. IANS can focus on the dynamic environment of the cloud market and consumers' requirements. Management of SLA described understanding the roles and responsibilities of SLA. The framework of IANS discussed the workflow diagram of the negotiation process. In future aspects, IANS may explore various learning algorithms to design agents, decision-making algorithms for optimizing solution, security, monitoring solution, and add more attributes for analysis.

References

1. Bahsoon R et al (2018) A manifesto for future generation cloud computing. ACM Comput Surv 51(5)
2. Voorsluys RBW, Broberg J (2011) Cloud computing : principles and paradigms table of contents
3. Shojaiemehr B, Rahmani AM, Qader NN (2018) Cloud computing service negotiation: a systematic review. Comput Stand Interfaces 55:196–206
4. Nadeem F (2020) A unified framework for user-preferred multi-level ranking of cloud computing services based on usability and quality of service evaluation. IEEE Access 1–1
5. Iyer GN (2016) Cloud testing: an overview
6. Shojaiemehr B, Rahmani AM, Qader NN (2019) A three-phase process for SLA negotiation of composite cloud services. Comput Stand Interfaces 64:85–95
7. Sim KM (2018) Agent-based approaches for intelligent interCloud resource allocation. IEEE Trans Cloud Comput 1
8. Elhabbash A, Samreen F, Hadley J, Elkhatib Y (2019) Cloud brokerage: a systematic survey. ACM Comput Surv 51(6):1–28
9. Rajavel R, Thangarathanam M (2016) Adaptive probabilistic behavioural learning system for the effective behavioural decision in cloud trading negotiation market. Futur Gener Comput Syst 58:29–41
10. Rajavel R, Iyer K, Maheswar R, Jayarajan P, Udaiyakumar R (2019) Adaptive neuro-fuzzy behavioral learning strategy for effective decision making in the fuzzy-based cloud service negotiation framework. J Intell Fuzzy Syst 36(3):2311–2322
11. Hsu CY, Kao BR, Ho VL, Li L, Lai KR (2016) An agent-based fuzzy constraint-directed negotiation model for solving supply chain planning and scheduling problems. Appl Soft Comput J 48:703–715
12. De la Prieta F, Rodríguez-González S, Chamoso P, Corchado JM, Bajo J (2019) Survey of agent-based cloud computing applications. Futur Gener Comput Syst 100:223–236
13. Paputungan IV, Hani AFM, Hassan MF, Asirvadam VS (2019) Real-time and proactive SLA renegotiation for a cloud-based system. IEEE Syst J 13(1):400–411
14. Mubeen S, Asadollah SA, Papadopoulos AV, Ashjaei M, Pei-Breivold H, Behnam M (2018) Management of service level agreements for cloud services in IoT: a systematic mapping study. IEEE Access 6:30184–30207
15. Rajavel R, Thangarathinam M (2017) ADSLANF: a negotiation framework for cloud management systems using a bulk negotiation behavioral learning approach. Turkish J Electr Eng Comput Sci 25(1):563–590
16. Li L, Yeo CS, Hsu CY, Yu LC, Lai KR (2017) Agent-based fuzzy constraint-directed negotiation for service level agreements in cloud computing. Cluster Comput 1–15
17. El-Matary DM, El-Attar NE, Awad WA, Hanafy IM (2019) Automated negotiation framework based on intelligent agents for cloud computing. In: Proceedings of the 2019 international conference on innovative trends in computer engineering ITCE. pp 156–161
18. Vallejo D, Castro-Schez JJ, Glez-Morcillo C, Albusac J (2020) Multi-agentarchitecture for information retrieval and intelligent monitoring by UAVs in known environments affected by catastrophes. Eng Appl Artif Intell 87:103243
19. Pouyllau H, Carofiglio G (2013) Inter-carrier SLA negotiation using Q-learning. Telecommun Syst 52(2):611–622

Early Detection of Myocardial Infarction Using Machine Learning with Maximum Accuracy

S. Abirami Manisa, B. J. Abarna, V. Geethanjali, G. V. Hari Venkat, and R. Karthikeyan

Abstract This paper presents a technique for detecting Myocardial Infarction (MI) using Machine Learning through the ECG data. In today's scenario, MI is one of the major causes of demise worldwide. MI occurs due to coronary heart disease and if detection or treatment is not done at appropriate time, the untreated MI may present with serious late complications. For treating MI in earlier stage several methods have been employed, but the parameters employed for those does not provide as much accuracy. So, a system which detects the abnormality with maximum accuracy is proposed. For that, Pan Tompkins algorithm is employed to filter and remove the noise in the acquired ECG signal and S-T segment is extracted from it. The features that are obtained from the extraction of S-T segment when given to the classifier provided better results. Based on the application of the Machine Learning classifiers, Naïve Bayes and Decision tree, the Decision tree classifier gave a higher accuracy of 98.5% compared to Naïve Bayes of 93.9%.

Keywords Myocardial infarction · Pan Tompkins algorithm · Machine learning · Decision tree classifier · Naïve Bayes classifier · S-T Segment · ECG

1 Introduction

An Electrocardiogram (ECG) records the heart's electrical activity, over a period of time, by placing the electrodes on the patient's body. From the captured ECG signal, cardiologists observe the rhythm and functioning of heart. They can also identify the cardiac abnormalities from ECG recordings [1]. Manual classification and detection have its own drawbacks, which paved a way for an automated ECG classification and earlier detection techniques. This paper deals with the detection of most pivotal cardiac abnormality known as myocardial infarction, or heart attack.

Today, major cause of increasing cardiac death is myocardial infarction. MI is the first symptom of coronary artery disease. Nearly 17.9 million people died in 2016

S. A. Manisa (✉) · B. J. Abarna · V. Geethanjali · G. V. H. Venkat · R. Karthikeyan
Department of ECE, Dr. N.G.P Institute of Technology, Coimbatore, India

Fig. 1 ECG waveform
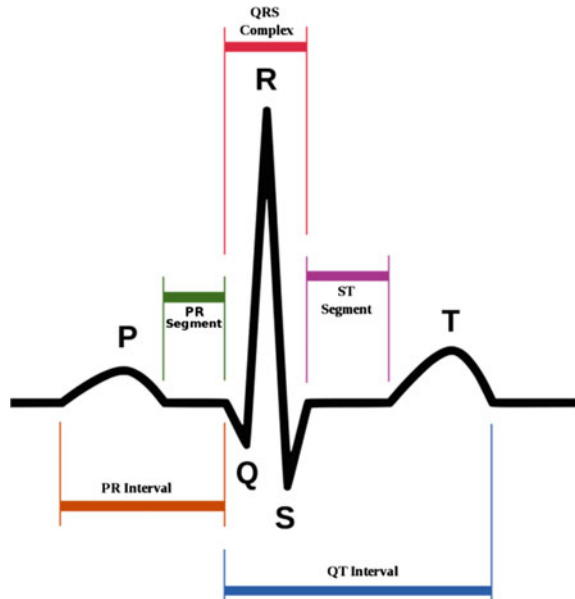
due to CVDs, which accounts for 31% of all global deaths. Out of these deaths, 85% are due to heart attack and stroke [2].

It is observed that ECG has three main components: The atrial depolarization is represented by P wave; the depolarization of the ventricles is represented by the QRS complex; and the repolarization of the ventricles is represented by the T-wave. It is shown in Fig. 1. In between the period of ventricular ion change, blood is supplied to the heart muscles, via coronary arteries. When the blood flow decreases or stops to a part of the heart, causes myocardial infarction. So, the salient point in early detection of MI is the ST segment. The ST segment duration indicates the blood supply to the heart. MI has three stages namely injury, ischemia and infarction [3]. An elevated ST segment occurs at the stage of ischemia. For detecting more precisely, Machine Learning is opted due to its less complexity and training time [4].

Previous methods [5–8], of MI detection are supervised learning approaches, in which they took only the parameters like Age, BP etc., They did not provide that much accuracy. Some methods [9–12] process the ECG recordings, but good beneficial results are not yielded because only the difference between labels of the entire ECG is taken. They did not take the difference between labels of heart beats within the ECG. By using these methods, a large data cannot be classified and it is to be overtrained. Even though these methodologies [13–15] gave good classification results, it had less accuracy.

As this paper aims at detecting MI by extracting ECG from a noisy signal, it requires accurate detection. For that, the filtering is done by Pan Tompkins algorithm and the proper extraction of the S-T segment is done and, that is given to classifier to detect the abnormality.

2 Proposed Methodology

The structure of MI detection in our work consists of three main phases, namely: Signal acquisition, ECG feature extraction and classification. The proposed method is as shown in Fig. 2.

The noisy raw input is denoised using Pan Tompkins QRS Detection algorithms. The extracted noiseless ECG signal is given to S-T segment detection and then the features are extracted. The classification is done using naive bayes and decision tree classifier and MI is detected.

2.1 ECG Signal Acquisition

The ECG signal is obtained using the ECG Sensor AD8232 and the Arduino Uno Controller. The ECG sensor which senses the electrical activity of the human heart. The three electrodes are attached to the left arm and right arm and right abdomen to detect the heartbeat. The sensor is very small in size, portable and accurate to measure the continuous heartbeat.

The Arduino controller gets the data from the ECG sensor and transmits the signals to the signal processing through the COM port.
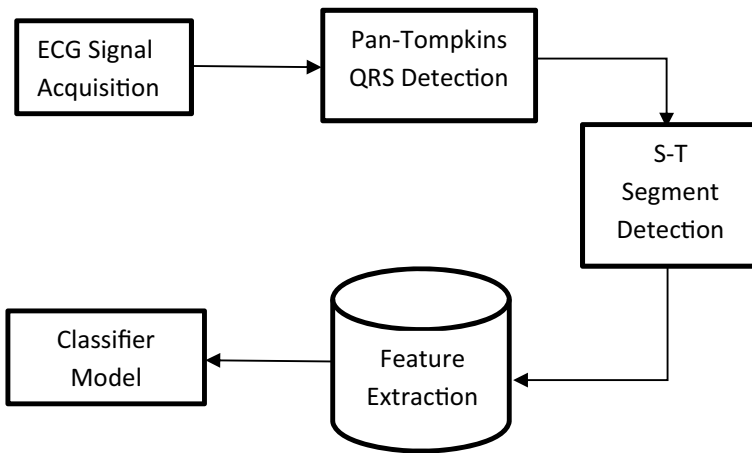
Fig. 2 Block diagram of proposed system
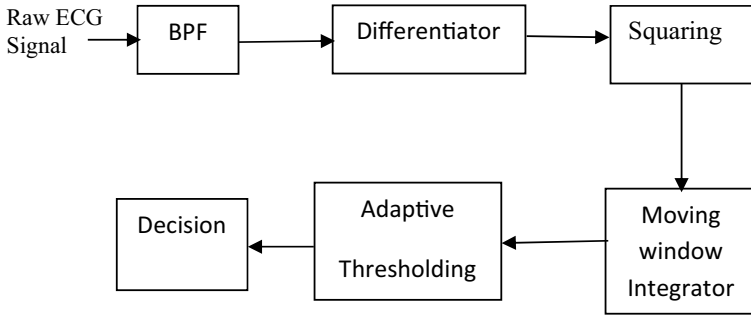
Fig. 3 Block diagram of Pan Tompkins algorithm

2.2 Pan Tompkins Algorithm

For a remote ECG monitoring, QRS detection is the preliminary step for detecting the heartbeat for the subsequent rhythm classification [16]. It is the most significant part of the ECG analysis algorithm. The Pan Tompkins algorithm uses the bandpass filter having low pass filter and high pass filter which reduces the noises. It has derivative function that highlights the R wave and then a quadratic function is applied to enhance the high frequency features of the complexes (Fig. 3).

2.3 S-T Segment Detection

In ECG, initially the PQRST points are obtained by setting the threshold values and from which the S-T segments are detected. This is the segment where the depression or elevation of it distinguishes the normal and abnormal ECG. The isoelectric period during which the ventricles are in between depolarization and the repolarization is represented by the S-T segment.

2.4 Feature Extraction

The statistical features such as mean, standard deviation, skewness and kurtosis., are extracted from the MATLAB Tool Diagnostic Feature Designer [17] through the data sets obtained from PTB Diagnostic Database, MIT-BH Arrythmia Database and over 330 data were collected and features were extracted and tabulated as shown in Fig. 4.
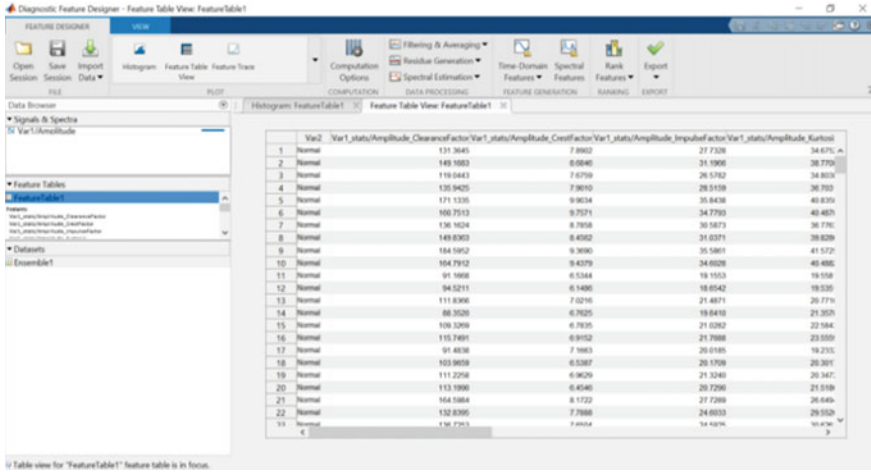
Fig. 4   Feature table

2.5   Classifier Model

The classifier model is built by using MATLAB classification learner app. The extracted features are exported to the classifier model. Here, the classification is done by using naive bayes and decision tree classification. The classifier is tested and validated using the dataset obtained from the Physionet ATM database. Here the hold out validation is given as 20% and then the dataset is trained.

3   Result and Discussion

3.1   ECG Signal Acquisition

The ECG Signal acquired through the sensor is plotted as shown in Fig. 5 using the serial port, the system gets the analog value and plots that value using MATLAB.

3.2   Pan Tompkins Algorithm

The raw ECG signal after processed through the Pan Tompkins algorithm is found as shown in Fig. 6 and the QRS peaks are detected as shown in Fig. 7.
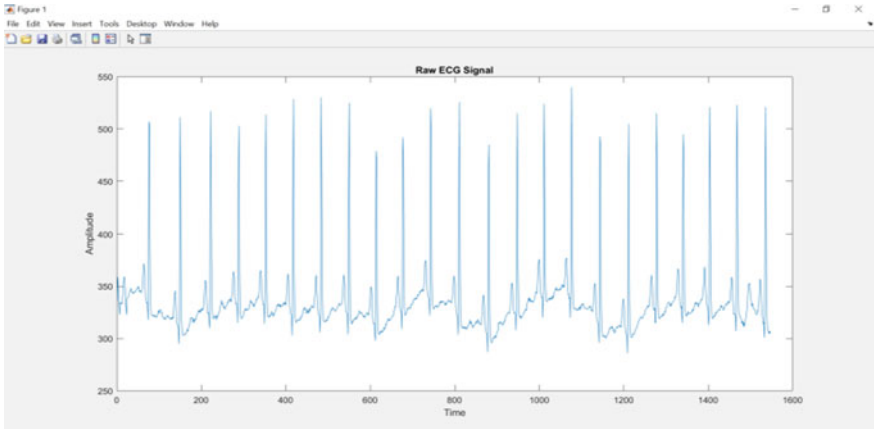
Fig. 5 Raw ECG signal
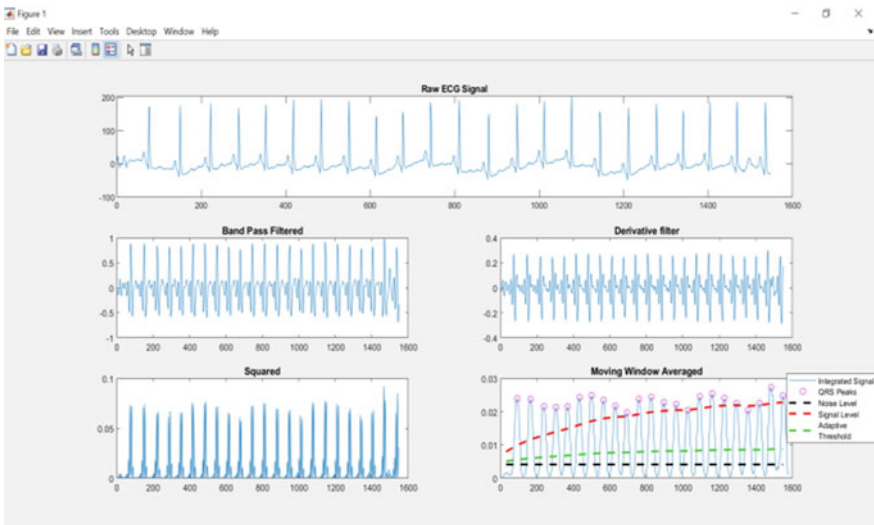
Fig. 6 Pan Tompkins algorithm implementation

3.3 S-T Segment Detection

The PQRST points detected are in the ECG Signal are shown in Fig. 8 and S-T segment plotting from the detected PQRST points are shown in Fig. 9.

The deviation in the normal and abnormal ECG are shown in Fig. 10 which serves as deciding factor for the Myocardial Infarction. The abnormal ECG may have either elevated or depressed S-T segment.
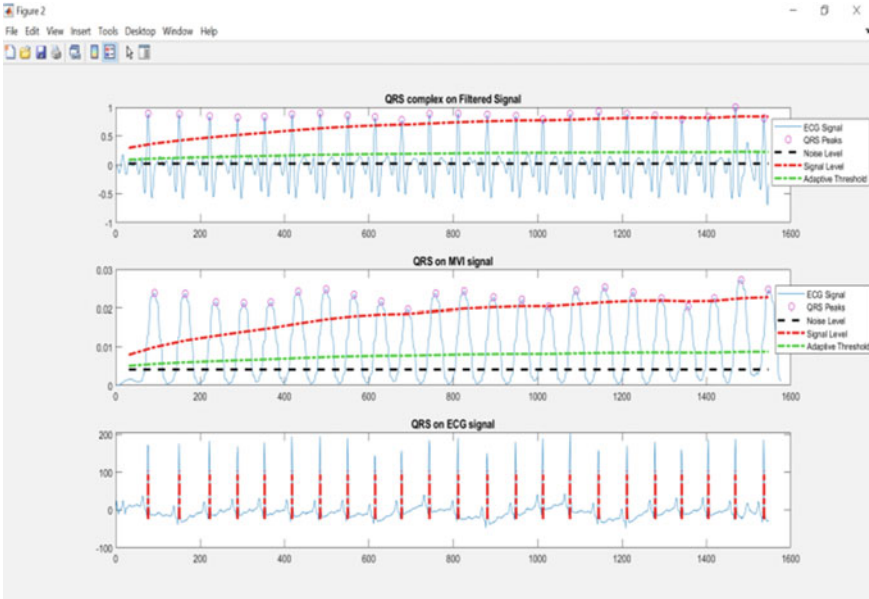
Fig. 7 QRS detection
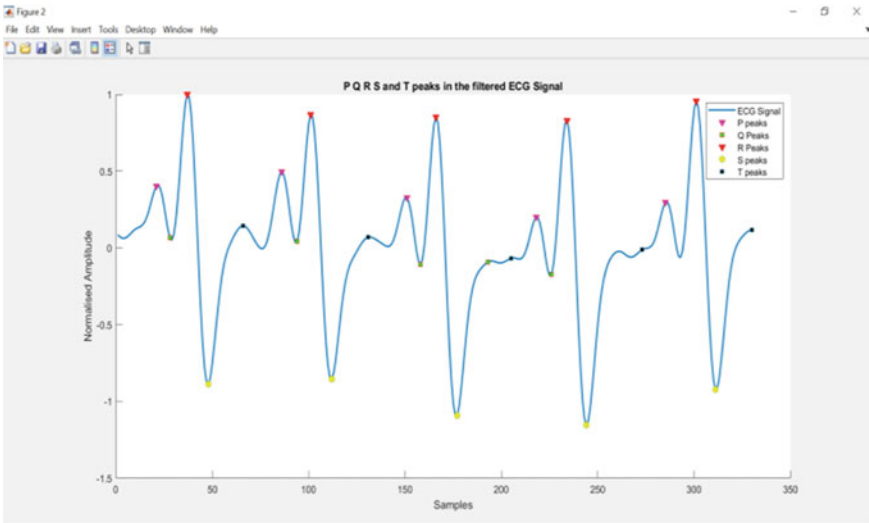
Fig. 8 PQRST signal
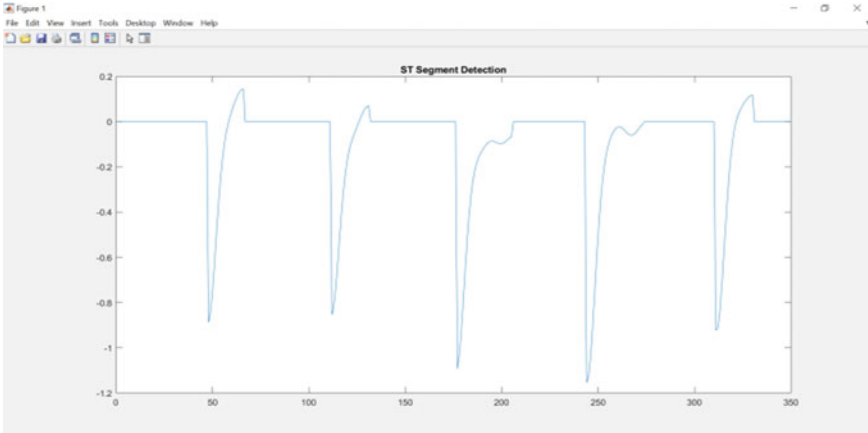
Fig. 9 ST Segment detection from filtered ECG signal
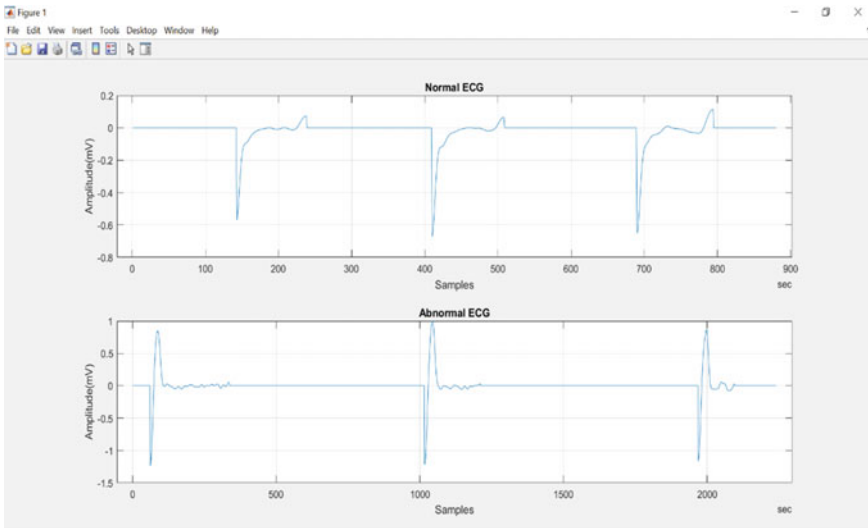
Fig. 10 Normal and abnormal ECG signal ST segment

3.4 Classifier Model

The features extracted from the data sets are given to the Classification Learner App and the scatter plot obtained for the Naïve Bayes and Decision Tree are shown in Figs. 11 and 12. From which the accuracy for Naïve Bayes obtained as 93.9% and Decision Tree as 98.5%

The confusion matrix for the both classifiers are as shown in Fig. 13.
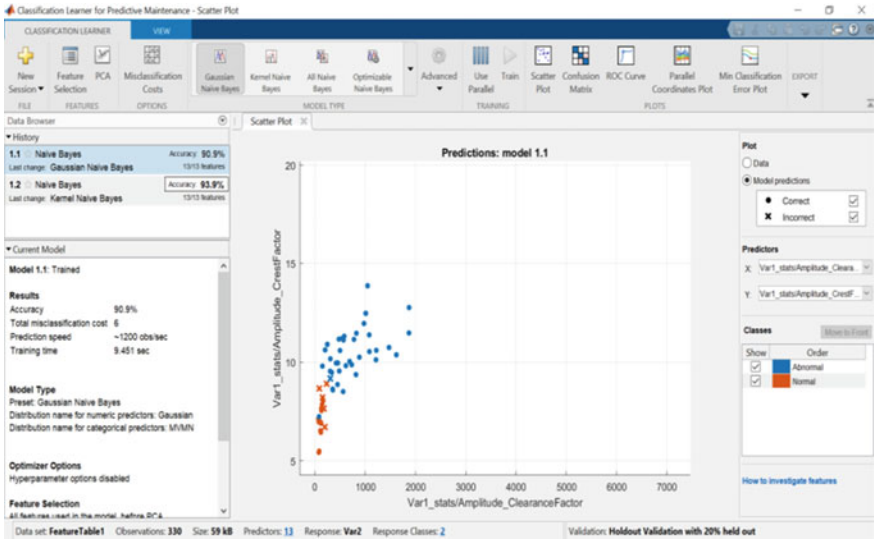
Fig. 11 Scatter plot for Naïve Bayes
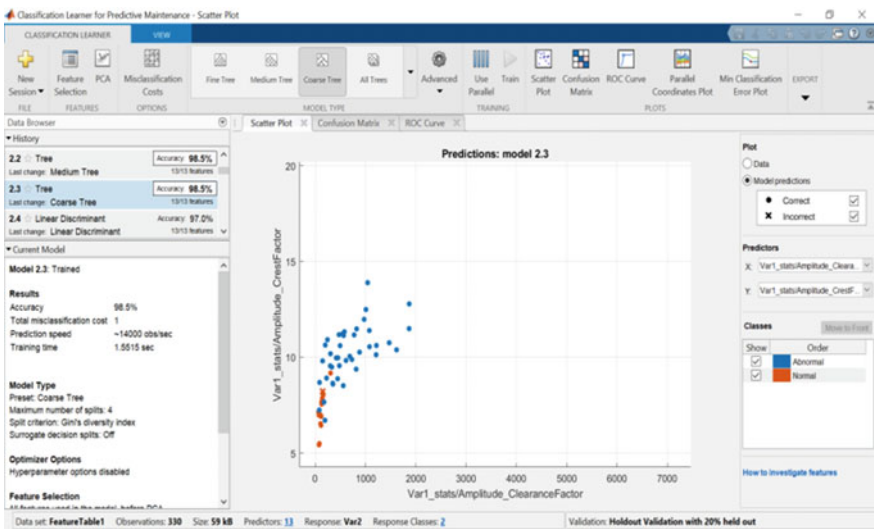
Fig. 12 Scatter plot for decision tree

4 Conclusion and Future Work

Myocardial Infarction needs an accurate and timely treatment for patients. So, in machine learning, Naïve bayes and Decision Tree were used and compared, in which

a) Naive Bayes   b) Decision Tree
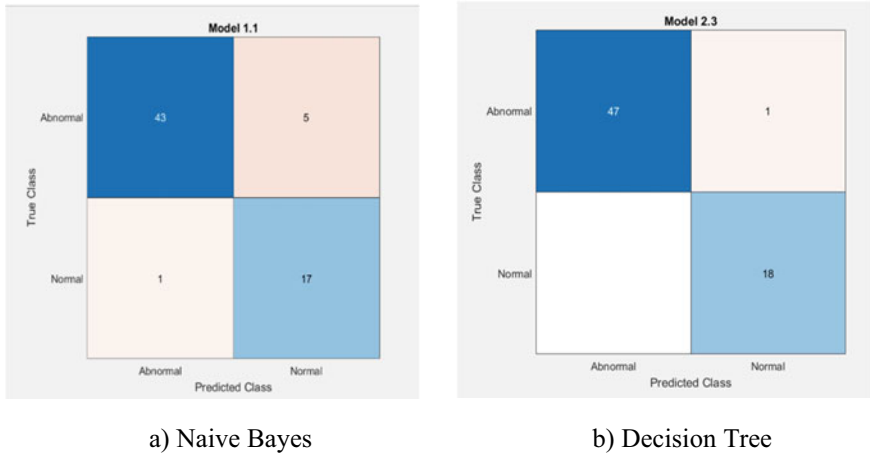
Fig. 13 Confusion matrix

the Decision Tree gave higher accuracy of 98.5%. In near future, we are trying to improve the accuracy level and planning to give an additional support through emergency assistance and transmit real-time data using IoT.

References

1. Salehi F et al (2017) The importance of electrocardiography parameters in healthy Iranian children. ARYA Atherosclerosis 13(3):159–160
2. World Health Organization (2017, May 17) Fact sheets of 'Cardiovascular diseases (CVDs).' Retrieved from https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
3. Thygesen K, Alpert JS, Jaffe AS, Chaitman BR, Bax JJ, Morrow DA et al (2019) Fourth universal definition of myocardial infarction. Eur Heart J 40(3):237–269
4. Yan Y et al (2019) The primary use of artificial intelligence in cardiovascular diseases: what kind of potential role does artificial intelligence play in future medicine?' J Geriatric Cardiol JGC 16(8):585–591
5. Celin S, Vasanth K (2018) ECG signal classification using various machine learning techniques. J Med Syst 42:241
6. Pandey SK, Sodum VR, Janghel RR, Raj A (2020) ECG arrhythmia detection with machine learning algorithms. In: Raju K, Senkerik R, Lanka S, Rajagopal V (eds) Data engineering and communication technology. Advances in intelligent systems and computing. vol 1079. Springer, Singapore
7. Sekhar BB, Likhitha V, Narendra I, Harika G (2019) Prediction and detection of heart attack using machine learning and internet of things. Int J Eng Adv Technol (IJEAT) 8(4). ISSN: 2249–8958
8. Uddin S, Khan A, Hossain M et al (2019) Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 19:281
9. Purnendu SP (2017) Machine learning and IoT for prediction and detection of stress. In: 17th international conference on computational science and its applications (ICCSA) 2017 IEEE

10. Subashini A, Sai Ramesh L, Raghuraman G (2019) Identification and classification of heart beat by analyzing ECG signal using Naive Bayes. In: 2019 Third international conference on inventive systems and control (ICISC). pp 691–694
11. Verma AR, Gupta B, Bhandari C (2020) A comparative study of ECG beats variability classification based on different machine learning algorithms. Augment Hum Res 5:16
12. Mudasir MK, Syed IA (2016) Prediction of heart disease using decision tree a data mining technique. IJCSN 5(6)
13. Chamuleau SA, van den Brink RB, Kloek JJ, Broekhuis E, de Beaumont EM, Koster RW (2005) Complicaties van een niet herkend hartinfarct; 'Complications of an unrecognized myocardial infarction'. Ned Tijdschr Geneeskd. 2005 Nov 19; 149(47):2593–9. Dutch. Erratum In: Ned Tijdschr Geneeskd. 2006 Dec 30; 150(52):2898. PMID: 16355569
14. Kshirsagar P (2020) ECG signal analysis and prediction of heart attack with the help of optimized neural network. ACJ 9(4):497–506
15. Priyan MK, Usha DG (2018) A novel three-tier internet of things architecture with machine learning algorithm for early detection of heart diseases. Comput Electri Eng 65:222–235
16. Pan J, Tompkins WJ (1985) A real-time QRS detection algorithm. IEEE Trans Biomed Eng BME-32 (3):230–223
17. Mali B, Zulj S, Magjarevic R, Miklavcic D, Jarm T (2014) Matlab-based tool for ECG and HRV analysis. Biomed. Signal Process Control 10:108–116

Design and Development of a Real Time Mouse for Human Computer Interaction

U. Vinothkumar (iD)**, D. Kavya, M. Abinesh, M. Kavin, and P. Aravind**

Abstract   In this paper there are two parts, Transmitter and Receiver. The transmitter side is designed with Peripheral interface controller (PIC), Bluetooth, limit switches and Micro electro mechanical systems (MEMS). In the receiver side there are Bluetooth and Personal Computer with Visual Basic 6.0. In the transmitter, the signals sensed by the accelerometer and limit switches, then it is given input to the microcontroller. The Analog to Digital Converter in the microcontroller converts these sensor input to digital code and fed to Bluetooth module via Universal asynchronous receiver/transmitter (UART) which triggers the module to transmit the signals to the computer. At the computer, Bluetooth captures these signals. Thus, according to the movement of hand, cursor is moved and according to limit switch press, functions analogous to clicking the button is carried out in Personal Computer with the help of Visual Basic 6.0.

Keywords MEMS · PIC · UART · Visual Basic 6.0 · ADC · Bluetooth

1   Introduction

1.1   Overview

Nowadays the human computer interaction is a need for survival. Most of industries, military, medical and domestic application require this type of interaction. Mainly interactions such as speech and gestures have been used for communication. But speech recognition lacks accuracy and probable to make errors [1] . So we require a wireless real time user friendly device which can avoid such incidents or a system that is accessible to gestures. In this paper therefore we go with the hand glove gesture recognizing system, which can perform all the function that a real time mouse does like dragging, selecting, scrolling the pointer to 3600 etc. A wearable

U. Vinothkumar (✉) · D. Kavya · M. Abinesh · M. Kavin · P. Aravind
Dr.N.G.P. Institute of Technology, Coimbatore, Tamil Nadu, India
e-mail: vinothkumar@drngpit.ac.in

wireless electronic glove that transforms hand and finger movement to real time data for different applications. This paper replaces the standard computer mouse for controlling a computer mainly used for power point presentations etc. Project uses Microchip's microcontroller IC named Peripheral Interface Controller PIC16F877A and Microchip's 'Integrated Development Environment', MPLAB, to simulate and assemble the written code. Why we call such system user friendly? Because these are some methods that we require to implement in a world where physical inabilities of humans are a cause of concern the normal condition like backache, half paralyzed, deaf and dumb etc. So we call them user friendly.

1.2 Objectives

- The objective of this paper is to design the wearable, real time and user friendly device for human computer interaction.
- To mimic all the gestures performed in the real world wearing the data glove in real-time with high accuracy.

1.3 The Scope of the Paper

The scope of the paper in future deals with the gaming sector which shall be further more improvised for virtual environment applications. Another field where this can be used is medical field, where paralyzed persons can be installed with this equipment for their communications.

This paper replaces the standard computer mouse for controlling a computer mainly used for power point presentations through wireless connection.

2 Existing Methods

In 2004 a social anthropologists 'Edward T Hall' claimed that over 60% of human interaction that we today is nonverbal in nature i.e. most of the communication can be possible by gestures [2]. In this paper we use gestures which function independently on speech i.e. we use autonomous gestures for communication. For a accurate interaction the clarity of gestures is also a case of concern so we must select components which can be user friendly can be accessible and reduce cost. Considering the case where magnetic induction coils are used as sensors [3]. Here we require separate sensors for each of the fingers respectively to improve degree of freedom but looking for economic consideration we could go for better. Therefore we go for TRI-axis MEMS sensor which have all the ability that magnetic induction coils and have reduced cost [4].

2.1 Gesture Recognition

Here we use autonomous gestures for communication purpose where motion of hand is selected as co-ordinate axis x, y, z axis respectively. Considering other kind of gesture recognition system example as for robotic interaction, where gestures are identified using the rule based approach and combination of three skin like regions at a particular image frame for satisfy a particular defined condition respectively [5].

2.2 Tri Axis Control

The tri axis control can be obtained by using MEMS sensors. It gives opportunity for free arm rotation. For each x, y, z axis program is written and it is being executed by PIC using VB software. The tri axis control overcomes the disadvantages of magnetic induction coil as it require less capital cost.

3 Methodology

Methodology section will served as a 'technical blue print' for this project to who has an interest to upgrade or modify this project in the future. In this section, it is important to know that all technical aspect regarding to this project will be included. For this section, it will divided into two parts which consist of software and hardware. For software sub-section, need MP-lab programming to program the programmable integrated circuit and for interfacing the PIC module with pc visual basic 6.0 had been used. In hardware sub-section, it will more on designing and developing the circuit using strip board, fit in and solder the components on it and communicate with microcontroller with the laptop through RS232 interface and Bluetooth. The functional block diagram of Hand Data Glove: A New Generation Real-Time Mouse for Human–Computer Interaction is shown in Fig. 1.

3.1 Data Glove

Here we uses a glove-type device which could detect hand position and movement. In this approach user require to wear a glove like device, which uses sensors that can sense the movements of hand and pass the information to the computer. These approaches can easily provide exact coordinates of palm's location and orientation, and hand configurations. The main advantage of these approach are high accuracy and fast reaction speed.

Fig. 1 Top and bottom view of hand data glove

3.2 Bluetooth Module

The Bluetooth module are used for wireless transmission of data from wearable hand glove device to personal computers. Bluetooth used here have a range of 10 m with an available frequency range of 2.4 GHz. As per requirement we can set variable distance by using amplifiers up to 100 m and so on. The Bluetooth is selected so that it is low cost, low power, wireless mesh network standard (Fig. 2).

Fig. 2 Bluetooth module
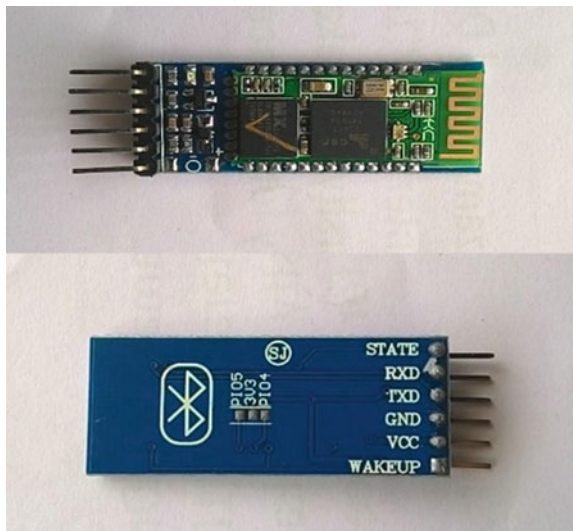
Fig. 3 MEMS accelerometer
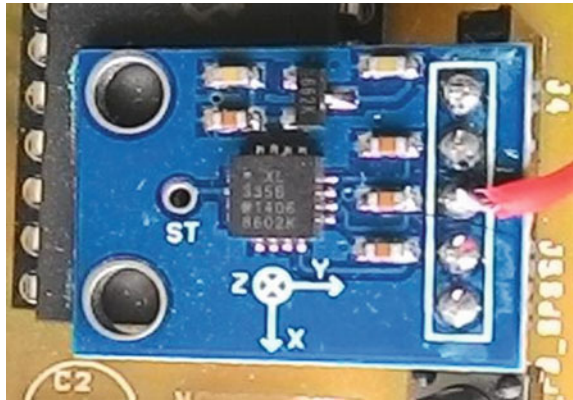
3.3 Limit Switches

The limit switches used here to obtain different function done by mouse. Each switches act for functions such as left button, right button and center click. In visual basic 6.0 each of the functions are described as special character respectively. i.e. for left button- dollar ($),right button-hash tag (#), center click-asterisk (*) characters are used respectively. For each function when we click on limit switch the following symbols are executed by PIC16F877A and the corresponding movements are generated by windows application.

3.4 MEMS Accelerometer

The MEMS accelerometer (Fig. 3) is placed above the wearable data glove. The accelerometer measures acceleration forces. These forces may be static, like the constant force of gravity, or they could be dynamic—caused by moving or vibrating the accelerometer. Here we use tri—axis MEMS which recognize the movement of arm in the form of respective co-ordinates of x, y, z respectively. Here the moment of palm is recognized.

For each positions the function generated in visual basic 6.0 i.e. for x, y, z different coding is written through visual basic 6.0 and is been executed in PIC16F877A.

3.5 Proposed Method

See Fig. 4.

Transmitter side:
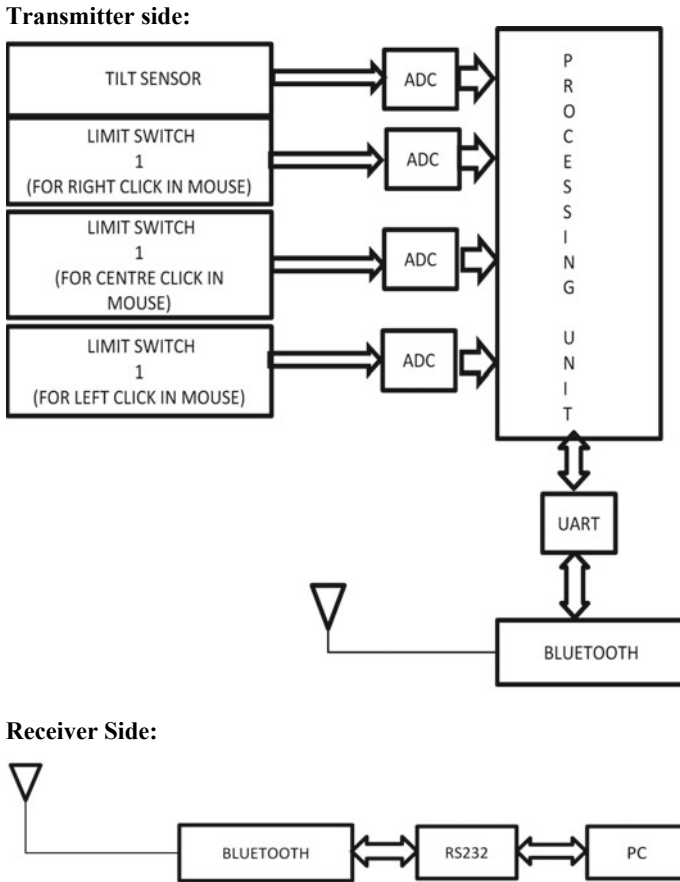
Fig. 4 Functional block diagram

4 Results and Discussion

Thus the Hand data glove is able to perform all functions of optical mouse with more degree of freedom (Fig. 5).

5 Conclusion

The hand data glove can now mimic all the gestures performed by a real time mouse. The added advantage given by degree of freedom of hand data glove is more than that of mouse. The Bluetooth act as an interfacing device between the two hardware sections. The detection of palm movement is sensed by using tri-axis accelerometer

Fig. 5 Real time mouse interaction with computer
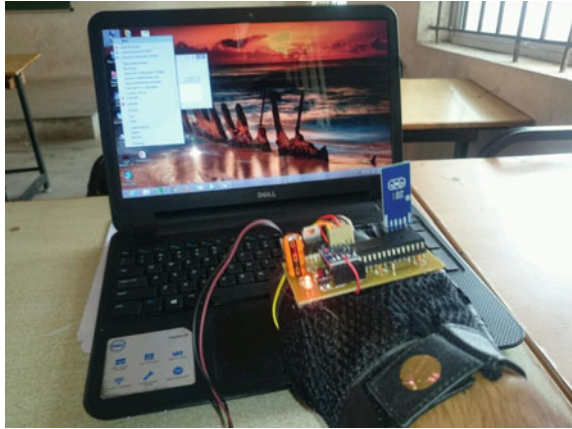
MEMS. The results also indicates left, Right and Center click operations effectively done by using accelerometer MEMS and limit switches.

6 Future Scope

The scope of the project in future deals with the gaming sector which shall be further more improvised for virtual environment applications. Another field where this can be used is medical field, where paralyzed persons can be installed with this equipment for their communications.

This project replaces the standard computer mouse for controlling a computer mainly used for power point presentations through wireless connection.

References

1. Piyush K, Jyoti V, Shitala P (2012) Hand data glove: a wearable real-time device for human computer interaction. Int J Adv Sci Technol 43
2. Shiratori T, Hodgins JK (2008) Accelerometer-based user interfaces for the control of a physically simulated character. ACM Trans Graph 27(5):1–9
3. Chin-Shyurng F, Herman S (2005) Development of a data glove with reducing sensors based on magnetic induction. IEEE Trans Indus Electron 52(2)
4. PRIME Faraday Partnership" An Introduction to MEMS" January (2002) Wolfson School of Mechanical and Manufacturing Engineering Loughborough University, Loughborough. ISBN 1–84402–020–7. http://www.primetechnologywatch.org.uk.
5. Pranathi N, Ahmed SM (2013) Tri-Axis motion detection using MEMS for unwired mouse navigation system in the future generation machines. Int J Adv Res Comput Commun Eng 2(9)

Design and Implementation of Low Cost Energy Meter Using MIT App Inventor

P. Gajendran, S. N. Deepa, and N. Rajasingam

Abstract Rapid development of low-cost sensors and smart sensors in this twentieth century makes the electronic based industries to get into the plenty of new solutions in all the sectors. In this line software tools also support in nice a way to implement the task with less effort. This proposed work going to enable such kind of an industrial need application, namely low cost smart energy measurement is the popular area in many accepts like Data control, management, saving and so on. Producing the one unit of electricity is equal to saving half unit is the common thumb rule in the electricity sector. Such a wonderful area utilizes all the advanced technologies to implement its requirement.

Keywords Low cost energy measurement · Sensors and smart devices · Open source softwares

1 Introduction

An open source hardware and software platform triggers the smart devices implementation with less effort. Every designer now-days tasting the open source tools to implement their needs with short span of time [1]. There are two main reasons behind that one is cost of overall implementation and second one is very important

P. Gajendran
Karpagam College of Engineering, Coimbatore, India

S. N. Deepa
Anna University Regional Campus, Coimbatore, India

N. Rajasingam (✉)
Dr. N. G. P. Institute of Technology, Coimbatore, India
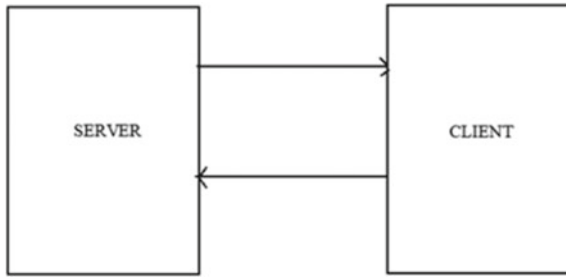
Fig. 1 Block diagram of the proposed work

one in this past decade this method accelerate the development work in the electronic industries that is online community support. This one accelerate everyone simulation work into real hardware implementation work [2]. This paper going to focus such tools like Arduino and MIT APP inventor these two open source tools enable our ideas to function in the real world environment.

A block diagram shown in Fig. 1 contains only two components in it. That the Bi-directional communications between the server and client. In conventional energy measurement topologies are defined as sensing and monitoring unit. Now these two terms are slightly modified into server and client model [3]. Server going to plays the vital role in thus proposed work. The functions are listed as follows.

- Sensing unit
- Data Manipulation
- Communication between Server and Client (Fig. 2)

Now the sever portion is clearly illustrated in the above block diagram. The heart of the server unit is Arduino UNO, powerful hardware implementation. In this decade its growth is tremendous in the area of embedded systems. Arduino community provides hardware and software support to implement any low-cost embedded solutions.

Other major components are P.T. (Potential Transformer), C.T. (Current Transformer), ZCD (Zero Crossing Detector) and Bluetooth module. Using P.T. and C.T. the important quantities required for the energy measurement in sensed and fed into
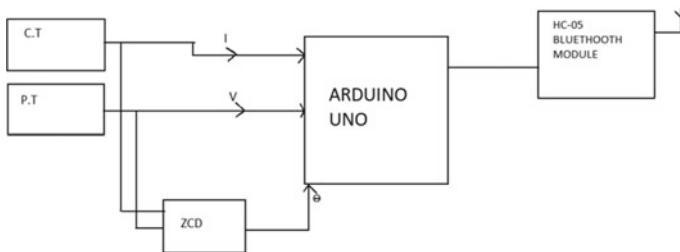
Fig. 2 Block diagram of server unit

Fig. 3 Block diagram of client unit
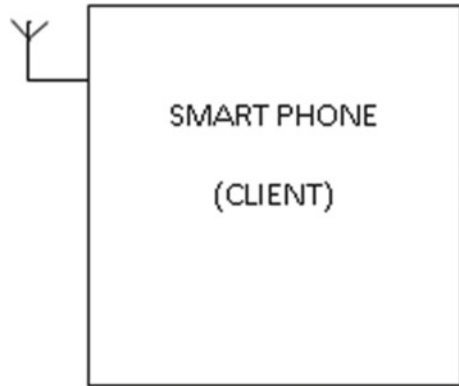
the microcontroller unit (Arduino UNO). Then the same voltage and current signal are tapped from the output of the P.T. and C.T. and gives as the input to zero crossing detectors [4, 5].

By receiving these signal ZCD provides the output as theta, that is phase angle difference between V and I. To compute P (real power), Q (reactive power), Power factor and energy. Finally, the computed signals were transmitted via Bluetooth module (Fig. 3).

Client Unit is simply a smart phone using Bluetooth protocol the message packets transmitted from the Server unit is received in the smart phone. Client Unit implementation is going to carried out with the help of MIT App Inventor. Again, this one is for the open source cloud-based software support for the App Inventor. The android app development is the one role in the client design. The Implementation is clearly illustrated in the Fig. 4. This app has access security feature; to access the particular "App Fingerprint Authentication" is added as the gateway to enter into their application. Once the authentication is verified, it allows the user to get into the network. In this way, server and client talks to each other. The calculated parameters are then transmitted from server unit to receiver. Then the same will be displayed in the dashboard window in the mobile App.

2 Implementation of the Proposed Work

2.1 Hardware Selection

Hardware units are only present in the server client. They are,

1. P.T.
2. C.T.
3. ZCD
4. Bluetooth Module
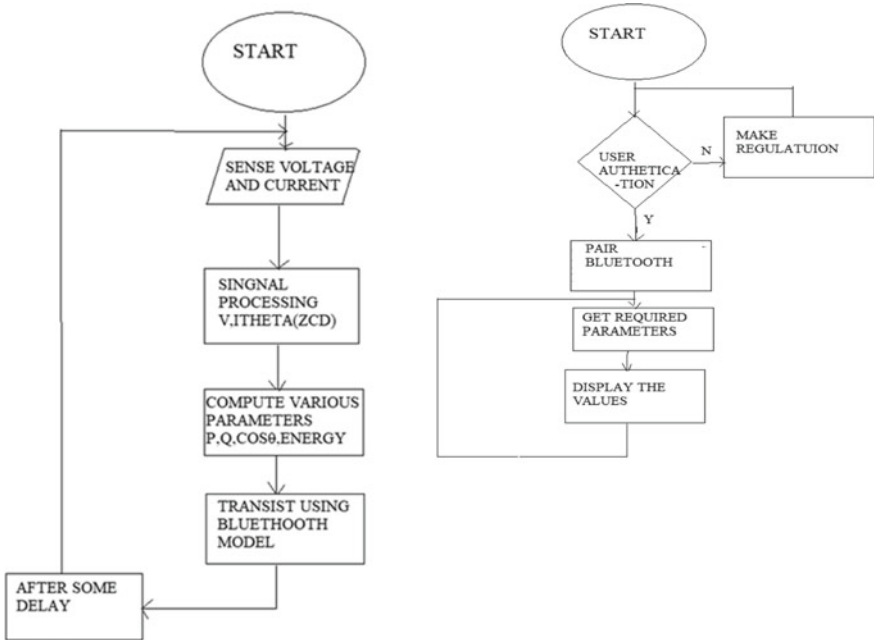
Fig. 4 Flow charts of server unit and client unit function

2.2 App Implementation

MIT App inventor tool is used to develop the android App required for this work. It has two stages one is screen development and another one is the code blocks. There are three screens used in this android app. First and second screen are used for user verification using finger print sensor. Then third screen has main dashboard window there user can take the readings of the energy meter.

Figure 5 clearly shows the components used in the screen-1 window. Here using inbuilt fingerprint sensor in the smart phone is used to read the user input. Also the Fig. 6 shows the code blocks required for finger print authentication.

Fig. 5 Finger print gateway
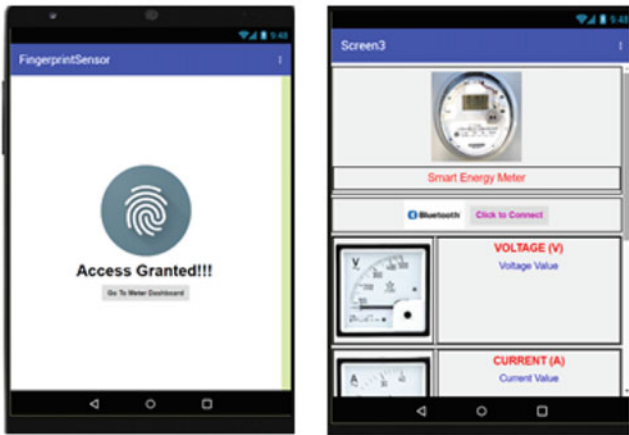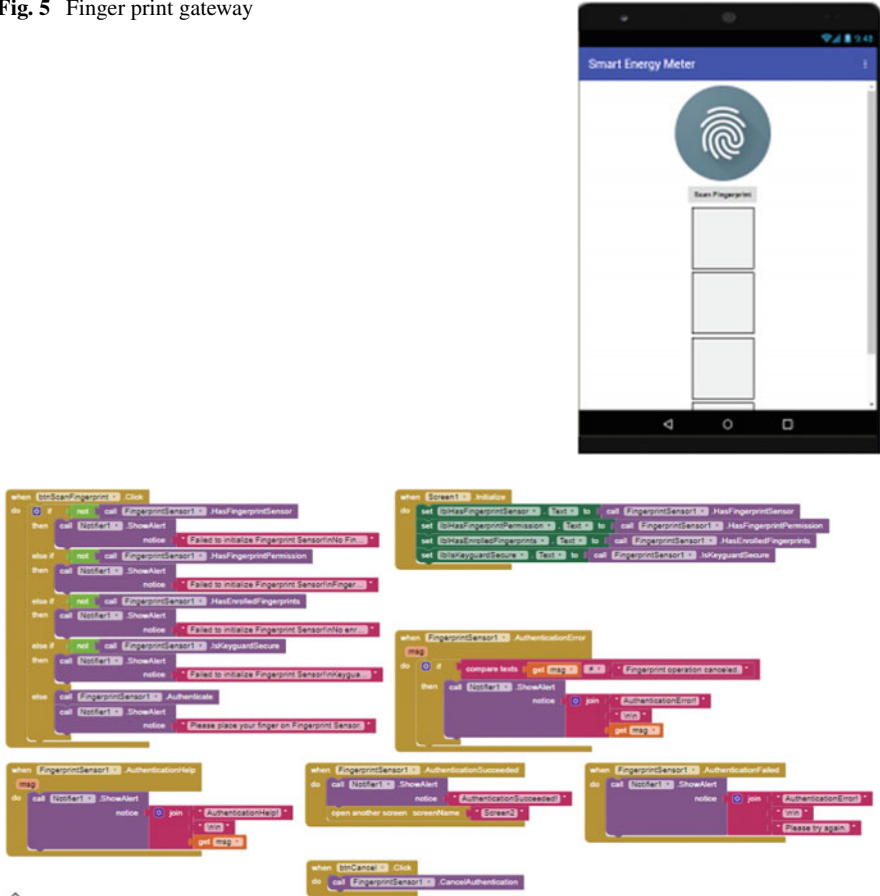
Fig. 6 Code blocks used in the screen-1 and Screen-2, Screen-3 view
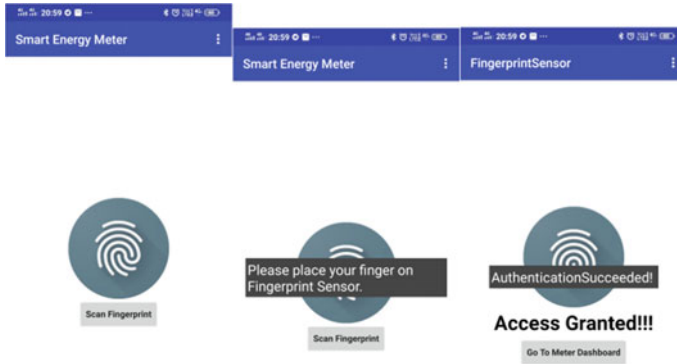
Fig. 7 Outputs from Screen-1 and Screen-2 window

As shown in Fig. 6 Screen-2 will allow entering the next level once finger print authentication is success. In screen-2 authentication confirmation is displayed then provides the entry to reach the main dashboard window that is screen-3. Here in screen-3 Bluetooth connectivity option is given to the user to pair with the server Bluetooth module. Once it's done it will start receive the transmitted parameters from the server to display it.

3   Output from MIT App Windows

At first download the App and install it in android based smart phone. One important condition the phone must have fingerprint sensor. Then open the App the first screen will appear as shown in the Fig. 7. Here the user needs to click the Scan Fingerprint button followed by scan the finger. Once it's done, it allows the user to enter in to the second screen. There it will give Authentication Succeeded message. Then user need to click the Go To Meter Dashboard button, immediately screen three will appear as in the Fig. 8.

Finally the screen-3 window will shows the actual dashboard window of the low cost energy meter. Here using Click to connect button user can pair the mobile Bluetooth with server Bluetooth module that is HC-05. Once its paired successfully, the dashboard will shows the various the electrical quantities.

4   Conclusion

In this way the proposed work is designed and implemented using Arduino and MIT App inverter tools. This one serves the requested functions as mentioned in the initial chapters. At present this work focuses on only single phase AC domestic loads. In
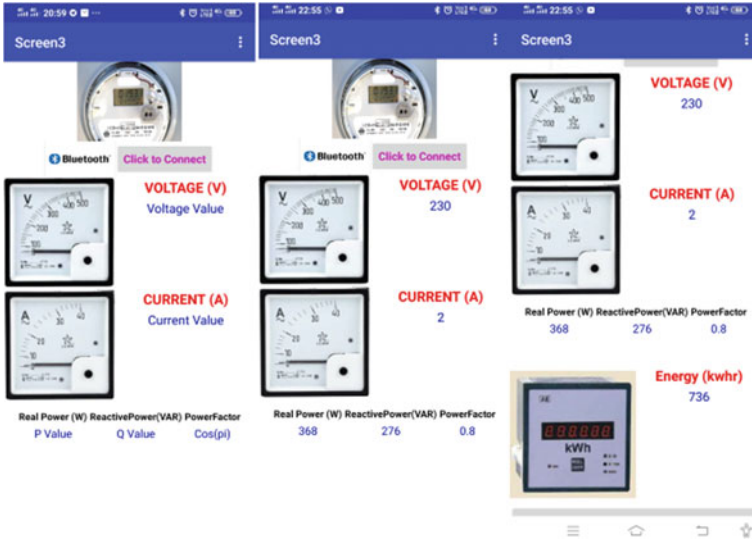
Fig. 8 Outputs from Screen-3 window

future work the same need to be extended to three phase and DC measurements. Also online cloud storage needs to implement to get in to the internet world.

References

1. Morello R, De Capua C, Fulco G, Mukhopadhyay SC (2017) A smart power meter to monitor energy flow in smart grids: the role of advanced sensing and IoT in the electric grid of the future. IEEE Sens J 17(23):7828–7837
2. Barkana BZ, McDonough W (2019) AP computer science principles: designing the hourly app in MIT app inventor. In: 2019 IEEE long island systems, applications and technology conference (LISAT) 2019 May 3, IEEE, pp 1–6
3. Pokress SC, Veiga JJ (2013) MIT app inventor: enabling personal mobile computing. arXiv preprint arXiv:1310.2830
4. Albu MM, Sănduleac M, Stănescu C (2016) Syncretic use of smart meters for power quality monitoring in emerging networks. IEEE Trans Smart Grid 8(1):485–492
5. Burunkaya M, Pars T (2017) A smart meter design and implementation using ZigBee based wireless sensor network in smart grid. In: 2017 4th international conference on electrical and electronic engineering (ICEEE) 2017 Apr 8, IEEE, pp 158–162

Detection of Sybil Attack in Internet of Things

C. Vinothini, J. Priya, P. Anitha, and T. S. Reshmi

Abstract The technology so called Internet of Things (IoT) has been knocking around discussion circles since few years back. Internet of Things, the technology behind the term is actually a simple idea, allowing the physical or dumb devices to get communicate with one another via the internet. Although the Internet of Things appears like it is just a concept at present, it will get to most of our lives very finely and ever soon. **Also** various industries will be affected distinctly. Though massive applications are there in IoT, data are getting transferred digitally in this modern world every day which paves the way for hackers to steal our private information. Hence the major grief is about digital security in todays' existence. Enormous research in this security area is in progress which includes privacy threats, vulnerabilities and secured communication etc,. Among these issues the crucial aspect is security and privacy. IoT network architecture get varies with respect to different applications which are vulnerable to threats of different types used by the hackers to make use of the IoT network. Sybil Attack is one of them which affect the communication module of the whole network. In this paper, a speculation of detection and mitigation of the Sybil attack has proposed. In contrast to other existing schemes, this proposed conception will work better in each and every situation with less time consumption.

Keywords Sybil attack · Prevention · Security · Privacy · Internet of things

1 Introduction

Internet of Things (IoT) is one of the evolving technologies employed in enormous applications in real time includes medical, agricultural, disaster management, environmental monitoring, and goes on. IoT network comprises of number of actuators

C. Vinothini (✉) · P. Anitha
Department of Computer Science and Engineering, Dr.N.G.P Institute of Technology, Coimbatore, India

J. Priya · T. S. Reshmi
Department of Information Technology, Bannari Amman Institute of Technology, Sathy, India
e-mail: reshmits@bitsathy.ac.in

and sensor nodes to sense and monitor the environmental or physical conditions. The characteristics of sensors are used to determine the performance in IoT network, natch. Few characteristics are: i. Communication failures, if anyone of the node in the network get fails while exchanging of information with others, which must be informed without any retardation. ii. Utilization of sensors, the way the sensors should produce the maximum performance. iii. Heterogeneity of sensor nodes means the sensors of distinct varieties should work in a cooperative manner. iv. Scalability, hundreds and thousands of sensors are getting employed in a network therefore the IoT should be designed highly scalable.

In spite of many more advantages, the conjugations of IoT devices present a pivotal challenge to industries in terms of intense security risks emerging from unmonitored and unsecured devices connected to the IoT network. As applications of IoT network grow continuously, data are getting transferred digitally in this modern world every day which paves the way for hackers to steal our private information. Therefore, it is understood that how IoT security is important in terms of security strategy and protection schemes that significantly protect from the chances of security attacks and even cyber-attacks on IoT devices that are interconnected to the network built for specific applications and functionalities. Without strong and healthy security, any connected physical device is vulnerable to getting compromised, breached and controlled by an attacker.

Presently, though, there is not much cause for distress. Whereas very few people are using IoT devices in a full-fledged manner, so there won't be as many attackers now to get access for them. This above-said statement is not to declare that this will be the situation always. Soon, IoT will become a major part of our life in the near future, and as days pass, it is essential to be monitored to the maximum extent and protected even more carefully.

One of the crucial challenges faced with the evolving IoT applications is privacy and security, and majority of the IoT smart devices are not structured to deal with security and privacy issues [1]. Hence the increase in number of connected devices, the chances for malicious attacks and other security threats is unavoidable.

Therefore the major grief is about digital security in todays' existence. Here comes the remaining section of this paper, Sect. 2 briefs about Sybil attack and its causes. Related works are done relevant to the proposed work given in Sect. 3. Whereas the Proposed work and Conclusion explained in Sects. 4 and 5 are respectively.

2 Sybil Attack

Physical and Network attacks, Software attacks and Encryption attacks are few attacks that are vulnerable to Internet of Things. It includes DoS (Denial of Service) attack, Node tampering, Node injection or Malicious code, Physical damage, etc., Under physical and network attacks, Sybil Attack is the one [2]. Earlier, the peer-to-peer network category experiences this type of attacks in the beginning in which

a peer or node in the network directs multiple identities effectively simultaneously and erodes the authority and power in legitimate systems.

The foremost intent of this Sybil attack is to obtain a greater influence in the network to carry out illicit (w.r.to the instructions and protocols set to the network) steps in the system. A single commodity (a host) has the potential to create and direct multiple identities (like user credentials, IP address details). At the spectators from outside, the multiple fake identities created because of the Sybil attack appear to be the real and unique identities.

The hope and belief in most IoT networks are presumed in a tacit fashion. This implicit assumption of hope and belief can be ill-treated by assailants to interrupt the network and falsify the reputations of trusted devices. Spoofing devices' identities or forging new identities allow assailants to impersonate them as legitimate devices. Devices with these sorts of illicit and forged identities are termed as Sybils [3].

As fundamental to Sybil attacks, harmful nodes try to interrupt and spoil the network operations by creating several fake and manipulated identities. The Sybil attack is a huge ruinous attack against the IoT network where numberless real identities with fabricated identities are used to get an illicit entry into the IoT network [4–6].

3   Related Works

As above-noted point, for the first time, the Sybil attack was identified in the peer-to-peer network [4]. In a research work related to Sybil attack, the author framed taxonomy for different formations of Sybil attack which targets any of the networks to interrupt or damage that classifies into i. Direct and Indirect Communication ii. Fabricated and Stolen Identities. iii. Simultaneity [1]. In all of these classifications, this Sybil attack intrudes in the network catches a legitimate node and makes it into an illegal node or a legal node by "malicious node" and gives identity to itself and then releasing several forged IDs or fabricated IDs from other true nodes.

The fake or fabricated Identities will look similar to the legitimate node identities but that is forged for non-existing nodes. This process will happen in the network is getting affected by the Sybil attack. As a result of the Sybil attack, the legal nodes in the network will think they have many legitimate peers [7, 8]. Once after misleading the legitimate node into believing malicious nodes as their neighbor nodes this attack breaks the condition of mapping between the identity and host of nodes [9].

Numerous works has been done for detecting and mitigating Sybil attacks in the network. Few methodologies and schemes discussed here for reference. They are:

Message Authentication and Passing, Random Password Comparison (RPS), Channel-based detection, Compare and Match Approach and still the progress goes on.

Though these many techniques are there in existing, some limitations also to be considered where the attack gets compromised.

For Example in the RPS method there is no repair mechanism for route if the route failure is the case. In Channel- Based Detection and RADS UWB methods focuses only on stationary networks and the indirect Sybil attacks are get detected whereas in the approach of Compare and Match has to know about complete information about all the existing nodes else it fails. And in the RFID method the communication overhead is high [10].

In a further Sybil attack detection mechanism proposed in [3] relies on the Ultra-Wide Band (UWB) ranging capabilities in the physical layer and the distance between each node and its neighbors are monitored periodically by every node which is a time-consuming process and also the mobility of the nodes were not considered [3]. Also, come across an algorithm named USAS (Unpredictable Software-based Solution) for detecting compromised nodes. Here the administers attest few selected nodes randomly to reduce the computation time of checksum, therefore there is a chance to miss nodes that are far from the control node for attesting [4]. Another work proposed in [5] for WSN architecture. The lightweight proposed method similar to message authentication and passing algorithm employs watchdog nodes that overhear Hello Packets exchanges between nodes. Then in the form of state diagram every watchdog nodes produce detected information partially and generates the final Sybil nodes list by using a detection rule defined already.

By using other personal nodes identities the Sybil activity can be created is well-known to everyone and now almost 70% of research methods in existing for Sybil attack detection are deals with Message authentication and passing method [3] that requires some modification for time consumption and cost-effectiveness.

A secure RPL routing has been proposed in this paper by detecting and mitigating Sybil attack the Rank attack. Based on the parameters used for the detection of attackers in the IoT network, the trust values of direct and indirect for all parent and child nodes were estimated [1].

One other author described the protocols named LEACH and LEACH-E. The author states that the energy consumption is high when the cluster node and the BS are getting communicated while comparing non-cluster nodes in WSN networks. Therefore it is understood that the number of clusters-heads is directly proportional to the energy consumption which leads shorten the lifetime of the corresponding network. Hence, number of head nodes to be considered [4]. Here this concept can be related to IoT network in terms of the number of sensors and actuators connected to a host.

The light-weight identity certificate method is an authentication-based approach to defeat Sybil attacks that use Merkle hash trees and a one-way key chain scheme. Since Merkle hash tree is being used, this method consumes more memory storage for information storage [5].

Interestingly another idea in [6] to mitigate Sybil attack in VPNETs also defined and proposed a novel scheme called Local Sybil Resistance (LSR) for effective defending against Sybil attack in privacy-preserving VPNETs. Here, in LSR scheme, the sensitive information of all vehicles including identity and location privacy can be well protected.

A trust-based security design for opportunistic IoT network called T_CAFE framed by the author in [7] that computes the direct trust of a node based on parameters such as correct packet ratio (CPR), amiability (Am), Frequency and forwarding ratio (FR) and ER detects and mitigates Sybil attack and also provides the additional benefit of isolating the attackers from the packet forwarding procedure.

4   Proposed Method

Our proposed algorithm is based on the authentication scheme for detecting and mitigating Sybils and this work also uses the LEACH protocol [9]. This proposed algorithm supports any kind of physical system. Here the network type chosen is IoT as a physical system, sensor, and actuator nodes are get registered to the network and generate their unique identities, then the credentials are shared to a particular host that controls a network of devices in an application. Consider a high storage actuator as a head node of the sensors which are connected to it. While sending and transferring data to the host system via the actuator, the host will be using cryptography hash function for the security purpose.

With reference to the KANGAROO CHF [8] scheme, the algorithm has been proposed for the significant IoT network.

According to the proposed algorithm, If the sensor nodes and actuator nodes are authenticated successfully once after the registration and generation of unique Identities through the KANGAROO CHF scheme then the nodes are normal and considered to be a legitimate and real node otherwise it must be a Sybil node.

Few parameters are considered here to calculate and by the result of it, the Sybil attackers can be detected and mitigated [9]. The parameters are Threshold and Calculated Throughput of the receiver similarly the packet delivery ratio and total energy consumed by the IoT network.

Once the Sybil attack is detected in the network then immediately the respective Sybils should be mitigated from the IoT network. To mitigate it, first step is to remove the identified malicious node by using any one of the existing algorithm like RSSI. Let us assume a threshold values of throughput and energy are TDTh, TDEn respectively for the IoT network. Calculate the total energy consumed by the whole network at the deployment stage. Similarly calculate the Packet delivery ratio and Throughput of the IoT network in the same manner. And once after designed IoT application has been deployed for a certain need, Calculate again the parameters throughput, packet delivery ratio and the energy consumption as said above. Now to identify and to detect the Sybil attack, the calculated values of threshold, packet delivery ratio and energy consumption at two different stages are to be compared.

Based on the difference in the values of these parameters calculated above determines the presence of Sybil attack and that will be shown as a major threat to the entire system.

Thus, our proposed algorithm detects the malicious node and mitigates the same from the IoT network. In our IoT network the Sybil attack attacks the communication

layer or network layer of the IoT system. The procedure has been detail explained in the below given algorithm.

Pseudo-code: Sybil Attack Detection

Prerequisite: Threshold and Calculated value of Throughput and Energy of N nodes.

    Step 1: Frame a physical system or a network with N nodes.
    Step 2: Based on clustering scheme, determine the Cluster Head.
    Step 3: Generate unique Identity for n nodes using LEACH protocol.
    Step 4: Connect Cluster Head and Base Station of the network.
    Step 5: If calculated value of
       Throughput is less than the threshold value of Throughput and the calculated value of Energy is less than the threshold value of Energy then there is no Sybil in the network.
    Step 6: If Step 5 fails, then Sybil is there in the network
    Step 7: Apply mitigation algorithms like RSSI to remove Sybil node from the network.

5   Conclusion

Internet of Things (IoT) is one of the evolving technologies employed in enormous applications in real time includes medical, agricultural, disaster management, environmental monitoring, and goes on.

The major grief is about digital security in todays' existence. Though many techniques are there in existing, some limitations also to be considered where the attack gets compromised. Hence our proposed algorithm based on the authentication scheme, detects and mitigates Sybil and protects the IoT network from the Sybil attack to the greatest extent. The throughput, energy and delay are calculated results in better values when Sybil Attack is mitigated. In future digital signature based or Blockchain based scheme may be proposed to detect and mitigate Sybil Attack in Internet of Things.

References

1. Tandon F (2019) Trust-based enhanced secure routing against rank and sybil attacks in IoT. In: 2019 twelfth international conference on contemporary computing (IC3), pp 1–7
2. Priya J, Gunasekaran M (2019) Security-aware and privacy-sensitive of internet of things (IoT): a review. In: 5th International conference on advanced computing and communication systems (ICACCS), pp 225–230

3. Patel ST (2017) A review: sybil attack detection techniques in WSN. In: 4th international conference on electronics and communication systems (ICECS), pp 184–188
4. Suriya U, Kumar R, Vayanaperumal R (2015) Detecting and preventing sybil attacks in wireless sensor networks using message authentication and passing method. Scientif World J 1–8
5. Amuthavalli R, Bhuvaneswaran RS (2015) Genetic algorithm enabled prevention of sybil attacks for LEACH-E. Mod Appl Sci 9(9):41–49. https://doi.org/10.5539/mas.v9n9p41
6. Zhang Q, Wang P, Reeves DS, Ning P, Carolina N (2005) Defending against Sybil attacks in sensor networks. In: 25th IEEE international conference on distributed computing systems workshops (ICDCSW'05), pp 1–7
7. Lin X, Member S (2013) LSR : mitigating zero-day Sybil vulnerability in privacy-preserving vehicular peer-to-peer networks. IEEE J Selected Areas Commun/Supplement. 31(9):1–10
8. Kandhoul N, Dhurandher SK, Woungang I (2019) T _ CAFE : a trust based security approach for opportunistic IoT. IET Commun 13:3463–3471. https://doi.org/10.1049/iet-com.2019.0657
9. Bertoni G, Daemen J, Peeters M, Van Assche G, Van Keer R (2018) In: KangarooTwelve : fast hashing based on Keccak- p, Applied cryptography and network security book, pp 1–23
10. Lim J, Yu H, Gil J (2017) Detecting sybil attacks in cloud computing environments based on fail-stop signature, symmetry. pp 1–12. https://doi.org/10.3390/sym9030035

Diagnosis of Brain Tumor Using ANN with Spatial Fuzzy Clustering and Active Level Set Contour

Hafiza Akter Munira and Md Saiful Islam

Abstract In medical science, image segmentation is a crucial and demanding factor. It has massive usage in the diagnosis of brain tumors. Magnetic resonance imaging (MRI) scans are used to detect the unnatural tissues of the brain efficiently which is always a powerful tool. The brain tumor identification and segmentation process are very steady and long delayed operation, because of the natural physical composition of human body. This paper deals with the classification of brain disease from MRI neuro-cell images. For the image segmentation and detection of brain tumor, a further combination of efficient method is proposed in this paper which uses inbuilt fuzzy C-means clustering which is further rectified by spatial clustering. Along with this segmentation, the proposed technique follows anisotropic diffusion filter and histogram of images for image enhancement. To meet the goal of separating tumor from normal brain cells, along with spatial fuzzy C-means, active level set contour is used for level-based image segmentation. The extraction of features from brain tumor is done by using DWT. With these features, the Artificial Neural Network classifier is trained. The accuracy of the proposed method is 96.7% on the patients 699 MRI data images. The result of this proposed method has been evaluated based on accuracy, sensitivity and specificity which is better than many existing methods.

Keywords MRI brain images · Spatial Fuzzy-C means clustering · Active level set contour · Anisotropic diffusion filter · Histogram of image · Artificial neural network

1 Introduction

With the rapid growth of technological advancement in image processing, the improvement of medical science is also noticeable. Different techniques are now commonly used, such as X-beam, ultrasonography, Computed Tomography (CT) and Magnetic Resonance Image (MRI). Amongst them, MRI is widely used for

H. A. Munira · M. S. Islam (✉)
Chittagong University of Engineering and Technology, Chattogram, Bangladesh

brain tumor diagnosis. The identification of tumor as a benign or non-benign is also very crucial. This can be classified as benign and malignance.

In this modern era, automated image processing is playing a crucial part in various applications. To decrease the image noise, pre-processing is used which is the most basic procedure of image analysis. The RGB to grayscale conversion is performed in conjunction with image resizing [1]. Through the median filter, the elimination of noise from the image data is performed. The anisotropic diffusion filter is then used for reducing the contrast between consecutive pixels. Image enhancement is done using histogram of images [2]. The enhanced image is then delivered to the spatial fuzzy C-means clustering [3, 4]. The extraction of features is carried out using DWT [5]. An important segmentation method called the active contour level set is used in this proposed approach to monitor the boundaries of image. To improve the level set active contour segmentation, there are many hybrid systems of fuzzy clustering [6, 7]. Actually, fuzzy clustering based on image intensity is used by such algorithms.

In this paper, fuzzy clustering incorporates with spatial information which eliminates the need of morphological operation. The rest of the paper is structured in following way: Sect. 2 provides a proper discussion of this proposed method; Sect. 3 provides the result analysis. Finally, in Sect. 4 the conclusion of this proposed method has been described.

2   Proposed Tumor Classification Methodology

The proposed methodology is implemented through the following steps, in which each step contains the elaborate description of each process. The entire process of the proposed implementation is shown in Fig. 1.
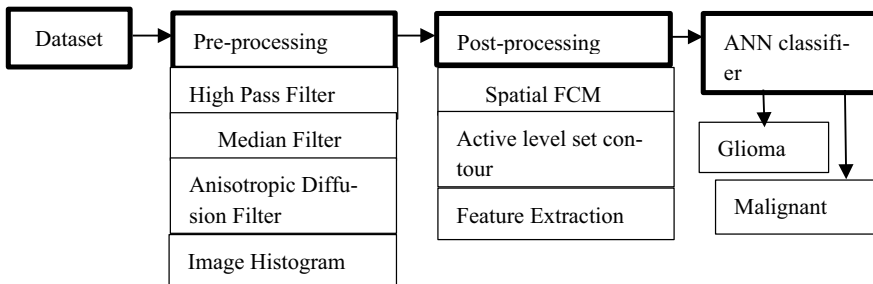
Fig. 1 Block diagram to show the entire process of the implementation

2.1 Dataset

The required datasets for MRI images have been obtain from https://github.com/sartajbhuvaji/brain-tumor-classification-dataset. Totally 699 MRI images of patients comprised of glioma and meningioma tumor had been undertaken.

2.2 Pre-Processing

In order to get more assurance and simplicity in distinguishing the tumor [8, 9], the motivation behind these methods is fundamentally to improve the image quality.

High Pass Filtering. It is used for edge detection and noise highlighting if any. The algorithm steps for this pr-processing is given as: hp_fil = $[-1\ 2\ -1;\ 0\ 0\ 0;\ 1\ -2\ 1]$;

Median Filter. Median Filter is being utilized in the subsequent stage to eliminate any undesirable clamor whenever recognized by the HPF.

Anisotropic Diffusion Filter. Anisotropic filter blurs the images without removing any edge. This filter can be expressed as,

$$\frac{\partial I}{\partial t} = \text{div}\ (c(x, y, t), \nabla I) = \nabla c \cdot \nabla I \tag{1}$$

where, $c\ (x, y, t)$ denotes the coefficient of diffusion. Through this the rate of diffusion is managed and is selected to preserve the image's edges. The diffusion co-efficient is suggested as,

$$c(||\nabla I||) = \exp\left(-\frac{||\nabla I||}{K}\right)^2 \text{ or } c(||\nabla I||) = \frac{1}{1 + \left(\frac{||\nabla I||}{K}\right)^2} \tag{2}$$

Figure 2 represents a filtered image using anisotropic diffusion filter which shows that the image edges are not blur due to filtering.

Histogram of Images. The image was enhanced using histogram of images [9]. In Fig. 3, the enhanced image is shown.

Fig. 2 Pre-processed filter image
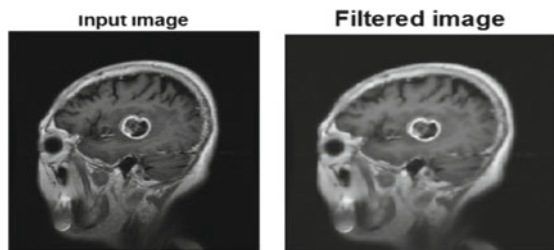
Fig. 3 Pre-processed
enhanced image
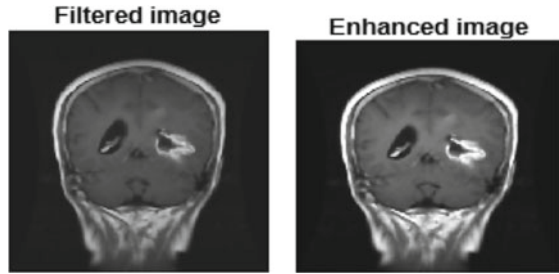
2.3 Post-Processing

The improved image from pre-processing then reaches the post-processing phases.

Spatial Fuzzy C-means. In medical problems, fuzzy C-means clustering is widely used because of its popularity [10, 11]. On the other hand, a membership function u_{mn} is used to denote the proportion of integration of nth object to the mth cluster in fuzzy C-means. The cost function of FCM is given by:

$$J = \sum_{n=1}^{N} \sum_{m=1}^{C} \mu_{mn}^{l} \llbracket i_n - V_m \rrbracket^2 \tag{3}$$

where, l (>1) is a constant through which the obtained segmentation fuzziness is observed. The following constraints are responsible for membership function:

$$\sum_{m=1}^{C} \mu_{mn} = 1; \ 0 \le u_{mn} \le 1; \ \sum_{n=}^{N} \mu_{mn} > 0 \tag{4}$$

where, μ_{mn} denotes the proportion of integration of nth object to the mth cluster in FCM. The absence of spatial data [4, 5, 9] is the biggest issues of conventional FCM algorithms for segmentation of image. The output of FCM segmentation will decrease due to the image noise and artifacts. That's why, spatial FCM algorithm is used. A spatial FCM algorithm was suggested by Chuang [4] which can be defined by:

$$\mu_{mn}^{'} = \frac{\mu_{mn}^{p} h_{mn}^{q}}{\sum_{k=1}^{C} \mu_{kn}^{p} h_{kn}^{q}} \tag{5}$$

where, the respective controlling parameters are p and q. The spatial information is carried by h_{mn}, can be expressed as:

$$h_{mn} = \sum_{k \in N_n} \mu_{nk} \tag{6}$$

where, the image pixel is centered at a local window given by N_n.

Active level set contour. Effective contrast boundaries for an image segmentation is used in active level set contour [12, 13]. With the help of active contour level set, image segmentation is a popular process [14]. Using the zero-level tracking $\Gamma\ t$, the evolution of active contours can be given by:

$$\varphi(t, x, y) \text{ less than } 0 : \text{ when, value of } (x, y) \text{ is placed in } \Gamma\ t$$
$$\varphi(t, x, y) \text{ equal to } 0 : \text{ when, value of} (x, y) \text{ remains at } \Gamma\ t.$$
$$\varphi(t, x, y) \text{ greater than } 0 : \text{ when, value of} (x, y) \text{located outside } \Gamma\ t.$$

Feature Extraction. For feature extraction, the proposed approach follows discrete wavelet transform (DWT) for statistical feature extraction. From the MRI input images, the following features were extracted: contrast, mean, energy, variance, correlation, skewness, kurtosis, inverse directional moment, standard deviation and entropy.

ANN Classifier. ANN classification is carried out after extraction of the necessary features. The number of iterations of the network classifier is 1000. The initial weights of the network are randomized. The number of hidden layers of this network is 20 and training function is Resilient backpropagation (trainrp).

3 Results Analysis

The implementation of both spatial FCM algorithms and active level set contour has been done in MATLAB 2020a. Through this proposed approach, initial segmentation of images is possible with level set active contour and the accuracy of the trained network on the 105 tested image is 96.7%. The network is trained on 489 MRI images and validation and testing image are 105.

In Fig. 4, the segmented image after using spatial FCM (a, c) and active level set contour image with edges (b, d) are shown.

In Fig. 5, the confusion matrix of the training is shown and Table 1 represents the performance of the network based on sensitivity, accuracy and specificity. The comparison of this proposed method with others is shown in Table 2.
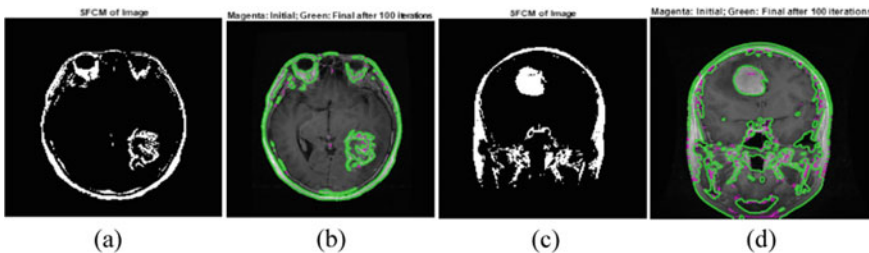
Fig. 4 Spatial FCM segmentation (**a**, **c**) and active level set contour after 100 iterations (**b**, **d**)

Training Confusion Matrix

	1	2	
1	**305** 62.4%	**9** 1.8%	97.1% 2.9%
2	**11** 2.2%	**164** 33.5%	93.7% 6.3%
	96.5% 3.5%	94.8% 5.2%	**95.9%** **4.1%**

Output Class / Target Class

Validation Confusion Matrix

	1	2	
1	**72** 68.6%	**1** 1.0%	98.6% 1.4%
2	**0** 0.0%	**32** 30.5%	100% 0.0%
	100% 0.0%	97.0% 3.0%	**99.0%** **1.0%**

Output Class / Target Class

Test Confusion Matrix

	1	2	
1	**70** 66.7%	**2** 1.9%	97.2% 2.8%
2	**0** 0.0%	**33** 31.4%	100% 0.0%
	100% 0.0%	94.3% 5.7%	**98.1%** **1.9%**

Output Class / Target Class

All Confusion Matrix

	1	2	
1	**447** 63.9%	**12** 1.7%	97.4% 2.6%
2	**11** 1.6%	**229** 32.8%	95.4% 4.6%
	97.6% 2.4%	95.0% 5.0%	**96.7%** **3.3%**

Output Class / Target Class

Fig. 5 Confusion matrix of training

Table 1 Classification performance

ANN classifier result		
Sensitivity	Specificity	Accuracy
97.38%	95.42%	96.7%

Table 2 Comparison of this proposed approach with existing ones

Method	Accuracy (%)	Result
Region based segmentation + watershed segmentation + ANN [8]	63	It cannot detect the edges smoothly, sometimes the high-density tumor tissues edge detection becomes difficult
K-means segmentation + Gabor filter + ANN	88	Not only the affected tissue is segmented, also the normal tissue is segmented with this segmentation
Fuzzy C-means segmentation [3]	92	Due to the noise and artifacts, it is not strong enough for medical image segmentation
Proposed method	96.7	It gives initial segmentation of image with proper edge detection

Thus, the improved spatial FCM segmentation is esteemed for trying to merge intensity and spatial information altogether. This algorithm esteemed less noise and it is perfect for initial level set contour in medical image processing.

4　Conclusion

For brain tumor detection, an automated image segmentation is proposed in this paper using enhanced spatial fuzzy C-means with active level set contour. The proposed approach contains the pre-processing steps where the images are being ready for further segmentation using noise removal median filter, anisotropic diffusion filter and histogram of images. Then, the images are sent to post-processing steps where the spatial fuzzy C-means is used with active level set contour for better detection of boundaries of abnormal tissues. The statistical extracted features are trained using ANN where the accuracy is found 96.7% on test data which is far better than many existing methods.

References

1. Loges Wari T, Karnan M (2010) In: International conference on signal acquisition and processing, an enhanced implementation of brain tumor detection using segmentation based on soft computing
2. Hemanth DJ et al (2009) Effective fuzzy clustering algorithm for abnormal MR brain image segmentation. In: Advance computing conference 2009. IACC 2009, IEEE international. pp 609–614
3. Suganya R, Shanthi R (2012) Fuzzy C-Means algorithm—a review. IJSRP 2(11. ISSN 2250–3153
4. Aslam A, Khan E, Beg MMS (2015) Improved edge detection algorithm for brain tumor segmentation. In: Second international symposium on computer vision and the internet (VisionNet'15), Elsevier

5. Lakra A, Dubey RB (2015) A comparative analysis of MRI brain tumor segmentation technique. Int J Comput Appl 125:5–14 (0975–8887)
6. Sudharani K, Sarma TC, Prasad KS (2015) Advanced morphologicaltechniqueforautomat-icbraintumordetectionandevaluationof statistical parameters. In: International conference on emerging trends in engineering, science and technology (ICETEST)
7. Dhage PM, Phegade MR, Shah SK (2015) Watershed segmentation brain tumor detection. In: International conference on pervasive computing (ICPC)
8. Kumbhar U, Patil V, Rudrakshi S (2013) Enhancement of medical images using image processing in MATLAB. Int J En Res Technol 2(4):2359–2364
9. Chuang KS, Hzeng HL, Chen S, Wu J, Chen TJ (2006) Fuzzy c-means clustering algorithms with spatial information for image segmentation. Comput Med Imaging Graph 30:9–15
10. Cai W, Chen S, Zhang D (2007) fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. Pattern Recogn 40:825–838
11. Osher S, Fedkiw R (2003) Level set methods and dynamic implicit surfaces. Springer, New York
12. Chan TF, Vese LA (2001) Active contours without edges. IEEE Trans Image Process 10:266–277
13. Varuna shree N, Kumar TNR (2018) Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network. In: Springer corporate information
14. Havaei M et al (2017) Brain tumor segmentation with deep neural networks. Med Image Anal 35:18–31

Current Overview of Machine Learning Application for Predicting Steam Huff and Puff Injection Production Performance

Mohammad Galang Merdeka, Syahrir Ridha, Berihun Mamo Negash, and Suhaib Umer Ilyas

Abstract Thermal Enhanced Oil Recovery (EOR) is one of the main contributors to EOR worldwide production. Steam huff and puff injection, one of its methods, is a technique in which steam is injected in a cyclical manner alternating with oil production. Reservoir simulation is considered as the most reliable solution to evaluate the reservoir performance and designing an optimized production scheme. However, it still remains time-consuming and expensive. Applying machine learning to build a predictive proxy model is a suitable solution to deal with the issue. Presently, there have been a limited number of studies covering the topic of proxy model development to estimate production performance for this injection method. This study provides a review of the machine learning implementations for estimating steam huff and puff injection production performance, starting with an introductory explanation about the method, followed by the currently deployed machine learning models along with the challenges and future prospects.

1 Introduction

As most of the giant oil fields are all currently at the mature stage, Enhanced Oil Recovery (EOR) is considered a viable solution to recover the remaining oil after the primary and secondary recovery stages. It involves the injection of a fluid or fluids into a reservoir intending to supply the energy needed to displace oil. Furthermore, the injected fluids interact with the reservoir rock and fluid to alter the physical properties and create an advantageous condition to improve oil recovery [1].

M. G. Merdeka (✉) · S. Ridha · B. M. Negash
Petroleum Engineering Department, Universiti Teknologi PETRONAS, Seri Iskandar, 32610 Perak, Darul Ridzuan, Malaysia

S. Ridha · S. U. Ilyas
Institute of Hydrocarbon Recovery, Universiti Teknologi PETRONAS, Seri Iskandar, 32610 Perak, Darul Ridzuan, Malaysia

Thermal EOR is one of the main EOR methods applied worldwide operating on the mechanism involving reduction of oil viscosity by the injection of thermal energy. As reported by the International Energy Agency, it accounts for more than 40% of total EOR production in 2015 [2]. The target oil is a high-density and high-viscosity oil. Heavy oils are such example for the target, and several reports on the thermal EOR implementations showed that almost all of the reservoirs are heavy oil reservoirs.

In evaluating the performance of a reservoir and designing an optimized production scheme, reservoir simulation is conducted beforehand. It is considered as the most reliable solution. However, the whole process of reservoir simulation still remains time-consuming and expensive due to the complexity of a reservoir system and the difficulty of building a reliable full reservoir model. Building predictive proxy model using machine learning algorithms is considered a suitable solution to deal with the issue, where proxy models will estimate the reservoir simulation results using lighter mathematical models [3]. In reservoir engineering, the proxy model has been mainly used for sensitivity analysis, risk analysis, history matching, production forecasting and production optimization [4]. The advantages of the proxy model over reservoir simulation are the computational speed and significantly economical solution, at a tradeoff of slightly lower accuracy.

This paper will provide a review of the research works related to machine learning applications in predicting steam huff and puff injection method. It is aimed to summarize the findings and highlight the advantages and disadvantages of applying machine learning approaches in this research area. The discussion will start with an introductory explanation of steam huff and puff method, including the how-to and past field successes. Then, the currently deployed machine learning models for this method will be explored. Finally, the challenges and future prospects to utilize machine learning coupled with current reservoir engineering technologies will be discussed.

2 Steam Huff and Puff Injection Method

Steam huff and puff injection, also known as cyclic steam stimulation, is a method in which steam is cyclically injected into the well. After injection, the well is shut-in, allowing the steam to "soak" into the reservoir. The high temperature of the steam will reduce oil viscosity near the steam-oil interface. This is called the soaking period. The well is opened again after the soaking period, producing oil at a higher rate. After some time, the heat dissipates, and the near-well oil viscosity returns to its original value. The production declines and after reaching an economic limitation rate, steam is reinjected into the well, starting a new injection cycle. This whole process of injection, soaking and production is called the huff and puff method and illustrated in Fig. 1. Several screening criteria of this method had been developed in the past [5–7]. In general, a typical steam huff and puff process is well suited for heavy oil reservoirs with high porosity and permeability, shallow depths and high oil saturation.
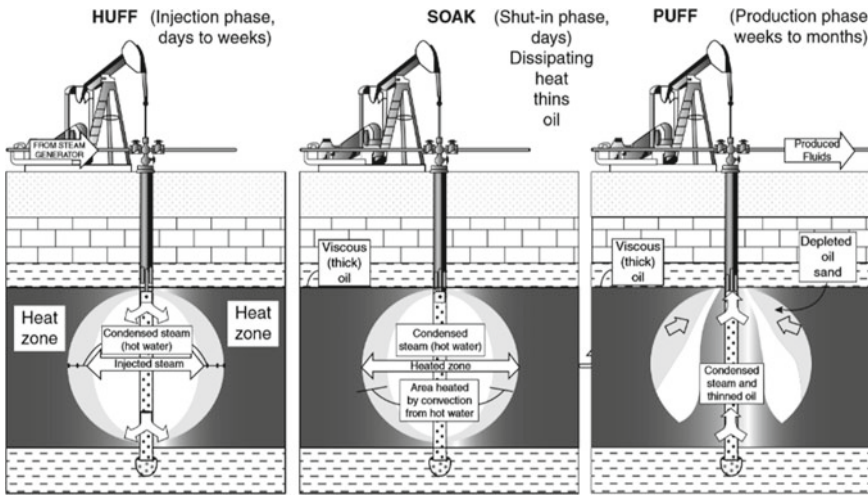
Fig. 1 Steam huff and puff injection schematic by Green and Willhite [1]

Even though the recovery factor is still smaller in comparison to the other steam injection methods, this injection method is still highly attractive due to small capital investment and quick payout. It is one of the most implemented thermal EOR methods with numerous reported commercial success in the past, for examples: Cold Lake Field in Alberta, Canada [8, 9], Tia Juana Field in Venezuela [10, 11], Midway Sunset Field in California [12], Liaohe Field in China [13] and Duri Field in Indonesia [14–16]. Furthermore, it is often used as a precursor before conducting a full field scale steam injection. To predict the performance of steam huff and puff injection, several semi-analytical models had been developed in the past [12, 17–20]. However, the drawbacks of those models are the limitations imposed by the assumption of the underlying physical equation and the need to have a reasonable history matching before using the models for prediction.

3 Machine Learning Applications in Predicting Steam Huff and Puff Injection Production Performance

3.1 Introduction

Machine learning (ML) is currently gaining a lot of attention to the recent breakthrough in data analysis. It is a tool that gives computers the ability to learn without being explicitly programmed. In oil and gas, especially reservoir engineering, most of the ML problem types were supervised learning problems, e.g. assisted history matching and forecasting and optimizing production [4, 21]. Those problems can be

challenged by applying machine learning algorithms to develop a proxy model, which is a mathematically or statistically defined function that approximates a real system or its simulation. Based on the objectives, the developed model can be utilized for specific field case or presented as a general or universal model [22]. For specific case model, the obtained field data are used to train the model and solve problems related to that field. Conversely, the universal model is used to solve a category of problems regardless of the variation of reservoir properties. It is trained using synthetic data obtained from reservoir simulation results by running numerous simulation experiment cases with varied inputs.

3.2 Existing Proxy Models for Steam Huff and Puff Injection

Although there were many research works related to steam huff and puff performance prediction, a very limited number related to the use of machine learning to develop proxy model was found. In predicting steam huff and puff performance, the most used machine learning algorithm was the Artificial Neural Network (ANN).

ANN is a model that emulates the biological neural system. It consists of nodes which are similar to neurons in the human brain. A node receives signals from adjacent nodes and processes them to produce an output. A generalized ANN model structure is presented in Fig. 2.

It consists of an input layer, an output layer, and one or multiple hidden layers in between. Each neuron from a layer is connected to the neurons from other nearby layers with a connection called weight, which represents an influence of corresponding input on the connected neuron. Then, to transfer the contained information, an activation (or transfer) function is used. This is called forward propagation and it is conducted for every connection in the network. After defining the model, the

Fig. 2 Artificial neural network structure by Arpaci [23]
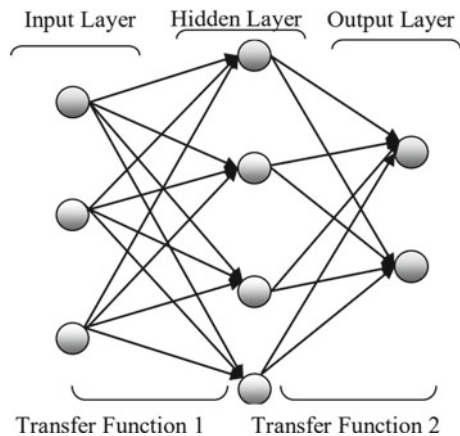
backpropagation algorithm is used to readjust the weight and minimize the error between the output and the database.

Several existing researches on ANN proxy models for steam huff and puff injection are presented below. They were presented as a general model and some are the continuation and improvements from previous works over the years. ANN was deployed to predict the production performance based on input parameters such as reservoir properties and operating conditions.

Arpaci developed ANN models for steam huff and puff injection using horizontal wells in naturally fractured reservoir [23]. The inputs of the model are reservoir rock and fluid properties, operation design parameters and fracture design parameters. The outputs are the number of cycles, cycle duration, oil flowrate, and cumulative oil production. Synthetic data were generated using the help of a thermal reservoir simulator. In addition, inverse models were developed to predict operation parameters and fracture design parameters, using performance indicators such as oil production and production period as additional inputs. The errors of each built models are less than 5%. However, most of the reservoir parameters were assumed constant, such as oil density, relative permeability, anisotropy and capillary pressures.

Sun and Ertekin developed a data-driven proxy model for steam-assisted gravity drainage (SAGD) and steam huff and puff injection in naturally fractured reservoir [24]. There are two kinds ANN models that were developed: a supervisory model that predicted the number of cycles and a sub-model that predicted the reservoir performance. It was set that the cycle automatically switches when the oil production rate hits a certain economic limit, thus the total number of cycles in the end of simulation period is not certain for every simulation case. The inputs will first visit the supervisory model to determine the number of cycles, then the sub-model will be predicting the reservoir performance, using both inputs from the database and the result of the supervisory model. The developed models show a good match with the database obtained from the numerical simulation. In addition to the ANN models, a Graphical User Interface was developed to facilitate easier use of the models to the end-users. Although this study had covered the disadvantage of less reservoir parameters from the previous study, the effect of the heterogeneity, anisotropy and capillary pressures were still neglected. The other limitations came from the fluid model assumption where the fluid composition was dead oil and water, and the number of predicted cycles in the sub-model were limited to 10 cycles.

Ersahin and Ertekin continued the previous study by Sun and Ertekin and developed ANN models for steam huff and puff injection in naturally fractured reservoir [25]. In this study, there were two main performance indicators that became the output of the model, which are cumulative production profile and viscosity contours. The simulations were performed to understand the heat-exchange mechanisms in the stimulation zone around the wellbore. Improvements were also carried out to the input parameters, e.g., the relative permeability table was generated using correlations specifically designed for sandstone rocks, as most of the thermal EOR projects were performed in sandstone rocks [26]. The results were good for predicting the outputs, with average error for cumulative production profile and viscosity contours were 3.64% and 5.32 respectively.

3.3 Challenges and Future Prospective

A limited number of research works encourage further explorations on this topic. The challenges would be coming from the difficulties of building a reliable model itself and finding the effective use of machine learning alongside conventional reservoir engineering practices.

Current models were able to be developed due to the huge amount of available data. Finding the data to train the model would be difficult if there are less implementation of the methods themselves, especially in the future, where new innovations on steam huff and puff injection method will emerge. The other challenge is the limitation of the model. Proxy model is basically an interpolation model. Hence, it can only be utilized when all the values of input parameters fall within the range of the training data. Having more data with larger intervals of inputs will help in developing a better general model.

Although proxy model provides a faster and cheaper computational solution, the weakness of the technology should not be ignored. Conventional reservoir simulation is still the most reliable tool in reservoir engineering and proxy model should not be used as a replacement. Instead, it should be used as an assisting tool to achieve reservoir engineering analysis more effectively. For example, as a starting guide in creating a field or well development plan.

The future prospects of this research topic lie on the technological innovations of steam huff and puff methods itself, as well as the advancements of machine learning and its subset, deep learning. Thermal EOR is projected to be still one of the highest contributors of EOR production in 2040, second only to miscible EOR [2]. There have been several studies and field trials of the new modifications to steam huff and puff injection, such as including additives to steam, conducting injection in different well configurations and combining cyclic injection with fracturing [27]. Therefore, one can expect more innovations in this method and their field implementations to gather more data for modelling purposes. On the other side, improvements in machine learning may highly contribute to determine the best approach to develop a proxy model for performance prediction. For instance, choosing the best training algorithm or determining the best adjustments for hyperparameters of the proxy model.

4 Conclusion

This study provided an overview of the steam huff and puff injection methods and the application of machine learning to predict the production performance. Machine learning helps to improve the reservoir simulation approaches in predicting the production performance by developing a proxy model which is highly efficient in reducing the required time and cost, at a tradeoff of slightly lower accuracy. However, it should be noted that the proxy model is not a replacement for reservoir simulation, but rather as an assisting tool that effectively helps in reservoir engineering practice.

For future prospective, the technological innovations of steam huff and puff method and the advancements of machine learning are expected to explore further towards more viable solutions.

References

1. Green DW, Willhite GP (1998) Enhanced oil recovery, 1st edn. Society of Petroleum Engineers
2. McGlade C, Sondak G, Han M (2018) Whatever happened to enhanced oil recovery?. International energy agency (2018). https://www.iea.org/commentaries/whatever-happened-to-enhanced-oil-recovery. Last Accessed 28 Oct 2020
3. Negash BM, Tufa LD, Ramasamy M et al (2017) System identification based proxy model of a reservoir under water injection. Model Simul Eng 7645470
4. Zubarev DI (2009) Pros and cons of applying proxy-models as a substitute for full reservoir simulations. In: SPE annual technical conference and exhibition, New Orleans, Lousiana, USA
5. Taber JJ, Martin FD, Seright RS (1997) EOR screening criteria revisited—part 1: introduction to screening criteria and enhanced recovery field projects. SPE Reserv Eng J 12(3):189–197
6. Farouq Ali SM (1974) Current status of steam injection as a heavy oil recovery method. JPT 13(1):54–68
7. Sheng J (2013) Enhanced oil recovery field case studies. Gulf Professional Publishing
8. Buckles RS (1979) Steam stimulation heavy oil recovery at cold lake, Alberta. In: SPE California regional meeting, 18–20 April, Ventura, California
9. Beattie CI, Boberg TC, McNab GS (1991) Reservoir stimulation in the cold lake oil sands. SPE Reservoir Eng
10. De Haan HJ, Van Lookeren J (1969) Early results of the first large-scale steam soak project in the Tia Juana. J Petrol Technol 21
11. Puig F, Schenk L (1984) Analysis of the performance of the M-6 area of the Tia Juana Field, Venezuela, under primary, steam-soak, and steamdrive conditions. In: Society of petroleum engineers AIME, SPE, vol 1. pp 263–276
12. Jones J (1977) Cyclic steam reservoir model for viscous oil, pressure depleted, gravity drainage reservoirs. In: SPE California regional meeting California
13. Liu W-Z (1997) Steam injection technology to produce heavy oils. Petroleum Industry Press Beijing, China
14. Gael BT, Gross SJ, McNaboe GJ (1995) Development planning and reservoir management in the Duri steam flood. In: SPE western regional meeting, 8–10 March, Bakersfield, California
15. Pearce JC, Megginson EA (1991) Current status of the duri steamflood project Sumatra, Indonesia. In: SPE International thermal operations symposium, 7–8 February, Bakersfield, California
16. Bae WS, Masduki A, Permadi AK, Abdurrahman M (2017) EOR in Indonesia: past, present, and future. Int J Oil and Gas Coal Technol 16(3):250
17. Boberg TC, Lantz RB (1966) Calculation of the production rate of a thermally stimulated well. J Petrol Technol 18(12):1613–1623. Society of Petroleum Engineers
18. Gontijo JE, Aziz K (1984) A simple analytical model for simulating heavy oil recovery by cyclic steam in pressure-depleted reservoirs. In: SPE annual technical conference and exhibition, 16–19 September, Houston, Texas
19. Sylvester NB, Chen HL (1988) Improved cyclic steam stimulation model for pressure-depleted reservoirs. Soc Pet Eng AIME, SPE, pp 213–226
20. Gozde S, Chhina HS, Best DA (1989) Analytical cyclic steam stimulation model for heavy oil reservoirs. In: SPE California regional meeting, 5–7 April, Bakersfield, California
21. Jaber AK et al (2019) A review of proxy modelling applications in numerical reservoir simulation. Arabian J Geosci. Saudi Society for Geosciences

22. Ertekin T, Sun Q (2019) Artificial intelligence applications in reservoir engineering: a status check. Energies. MDPI
23. Arpaci B (2014) Development of an artificial neural network for cyclic steam stimulation method in naturally fractured reservoirs. Master Thesis, The Pennsylvania State University
24. Sun Q, Ertekin T (2015) The development of artificial-neural-network-based universal proxies to study steam assisted gravity drainage (SAGD) and cyclic steam stimulation (CSS) processes. In: SPE Western regional meeting, 27–30 April, Garden Grove, California, USA
25. Ersahin A, Ertekin T (2019) Artificial neural network modeling of cyclic steam injection process in naturally fractured reservoirs. SPE Reservoir Evaluation and Engineering. SPE
26. Alvarado V, Manrique E (2010) Enhanced oil recovery: an update review. Energies
27. Alvarez J, Han S (2013) Current overview of cyclic steam injection process. J Petroleum Sci Res 2(3)

Enhanced Cluster Head Based Data Gathering (ECHGS) Technique in IoT Based Smart Irrigation System

K. Padmanaban, A. M. SenthilKumar, A. K. Velmurugan, and E. S. Madhan

Abstract Internet of Things (IOT) is an emerging technique which is used in several applications like home automation, office automation, hospital monitoring, network device monitoring, industrial related problems and smart agricultural monitoring system. In IoT environment, sensors are dynamically connected with the network to monitoring and routing the data to the server. Energy utilization and maximizing throughput are important issues in the Internet of Things (IOT). Energy plays a vital role in increasing network lifetime. Throughput improves the network performance and it is calculated based on average successful data delivered at a time instance. Various models such as cluster tree, clustering and Tree Cluster based Data Gathering Scheme (TCDGS) are used for handling maximizing energy and throughput. In this paper, a novel technique "Enhanced Cluster Head based Data gathering" (ECHGS) is presented to focus the energy and delay in the Internet of Things (IOT) based smart irrigation system for agriculture. In this proposed technique, the cluster head for the group of sensor nodes is selected based on the workload and energy. The node which has high residual energy and less workload is selected as a cluster head to collect and deliver the data to the server node effectively. The proposed technique results are evaluated and compared to the existing techniques in terms of residual energy and throughput. The residual energy is saved by 4.54% higher than the Tree Cluster based Data Gathering Scheme (TCDGS) and the throughput is maximized by 11.57% higher than the Tree Cluster based Data Gathering Scheme (TCDGS).

K. Padmanaban (✉) · A. K. Velmurugan
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India

A. M. SenthilKumar
Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R & D Institute of Science & Technology, Chennai, Tamil Nadu, India

E. S. Madhan
Department of Computational Intelligence, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India

1   Introduction

Different physical devices such as washing machines, refrigerators, air conditioners and television are connected to a heterogeneous network called the Internet of Things (IoT). Different devices connected to an IoT have different topology and battery functionality makes a complex network [1]. In IoT, the wireless sensor is an important part. Actuators of the sensors are used to sense the data and send the information to the server [2]. This data can be analyzed based on the problem specified in the application. Various sensors located inland are used to transmit the data to the server with the help of work station. Energy and lifetime are important aspects of the sensors [3].

Energy plays an important role in the sensors which are used in IoT based applications. Residual energy can increase the lifetime of the network. Residual energy is calculated based on the energy remains in the sensor nodes after transmitting the sensed data and information to the servers. Any node's failure due to less energy can lead the entire network to become inactive. In the IoT based applications, another major factor to be considered is a throughput. All the sensed data which are transmitted from the sensor nodes must reach the server with less time. If any data delivered with a high failure rate to the servers cause the poor performance of the entire IoT application [4].

Various techniques are used to focus the residual energy and throughput in the IoT applications during the data collection, data sensing and data transmitting phases. Cluster tree, Clustering and Tree Cluster based Data Gathering Scheme (TCDGS) are some of the existing techniques used to focus the energy and throughput. In all these techniques, sensors are clustered together into the group to minimize the coverage area to save the energy and minimize the delay.

The major contribution of this paper is discussed below,

1. The proposed Enhanced Cluster Head based Data gathering (ECHGS) technique improves the residual energy level.
2. The proposed technique maximizes the throughput during the data transmission between the sensors and the server.
3. The Proposed technique results are validated and compared with the existing techniques Cluster Independent Data gathering Technique (CIDT) and TCDGS.

The rest of the paper is organized as Sect. 2 details the literature review of the existing techniques. Section 3 shows the proposed technique. Section 4 includes the experimental setup and results. Section 5 concludes the work.

2 Literature

Wang et al. [5] proposed a path planning algorithm to reduce the mobile nodes count. In this technique, an efficient algorithm is designed to reduce the nodes on the path by path optimization. The results of the path planning algorithm are outperformed than the existing technique CSWEEP in terms of travel length and number of nodes. Lee et al. [6] presented a model based on Neural Network with input and output data as working hours, temperature, weather, humidity and electrical consumption for IoT applications. Model is verified with test data and performance is evaluated. Energy consumption is optimized in this work. Mean squatted error method is applied to the obtained data and simulation is done. Ke et al. [7] introduced a novel method called Joint Optimization method for Data Offloading, Renewable energy-aware, and Bandwidth allocation for IoT devices based on deep Reinforcement Learning (JODRBRL). This method minimizes the total cost which includes data delay cost, bandwidth cost and energy consumption cost by adaptive learning from the dynamic IoT environment. The JODRBRL results are better than the existing techniques Double DQN (DDQN), Dueling DQN and greedy policy.

Guo et al. [8] proposed a comprehensive energy estimation model for IoT networks. This proposed model analyzes the insufficient random-access procedure and utilize the power ramping and preamble picking for optimizing energy utilization. Also, Multi-Agent Reinforcement Learning (MARL) algorithm based on Win-or-Learn-Fast Policy Hill-Climbing (WoLF-PHC) is introduced in this work to decrease the algorithm complexity. The results are evaluated and compared with the random picking method. The proposed WoLF-PHC method results outperformed the random picking method results. Padmanaban et al. [4] presented a method called Tree Cluster based Data Gathering Scheme (TCDGS) in wireless sensor networks. In this work, nodes are clustered and form a tree-like structure. Cluster head node is identified for each cluster to maximize the residual energy. Data gathering is done by the data gathering node in the network and data is sensed from the cluster head instead of sensing all the nodes in the cluster and transfer to the base station. This particular module increases the life time of the network. The TCDGS results are better than the existing method of CIDT results.

All these existing works [9–17] have not considered residual energy and time delay in the IoT based agriculture smart irrigation system. In this proposed work, the residual energy and average time delay are considered. The residual energy is maximized for all the nodes and average time delay is minimized when compared to the existing methods CIDT and TCDGS.

3   The Proposed IoT Model for Agriculture Smart Irrigation System

The proposed architecture for the smart irrigation system in agriculture is shown in Fig. 1. The agriculture lands are shown in the diagram with circles. The clusters are formed based on the agriculture lands. Entire agriculture area is divided into lands. Each land among the total agricultural area is considered as a cluster. There are many sensors placed in each cluster. The sensors are placed in each agricultural land to collect the water level information from the land. In this paper, the Enhanced Cluster Head based Data gathering" (ECHGS) technique is used to collect the data from sensors effectively by proposing a new module to identify the cluster head. In each cluster of sensors, the cluster head is identified by conducting the poll based on the load and energy parameters of sensor.

The poll is conducted by substation periodically and the sensor node which has maximum energy and minimum load is considered as cluster head. The elected cluster head is responsible for collecting water-related data and send it to the substation. The substation forwards the data which is received from the cluster head to the base station. The substations are placed in all the agriculture land. The base station which sends the data that is received from the substation to the data server. The data server stores the water data and calculates the irrigation level for each agriculture lands. This processed information about water irrigation is given as output to the users who are accessing the data server to know the level of water irrigation to their lands. Each sensor's load and energy values are varying in every time instance. Current cluster head's load and energy level is changed after transmission of data. To maintain the efficiency of the IoT model after every data transmission from sensors, the proposed algorithm suggests an idea to identify the alternate cluster head which has more energy and less load is to be elected as cluster head by the substation.
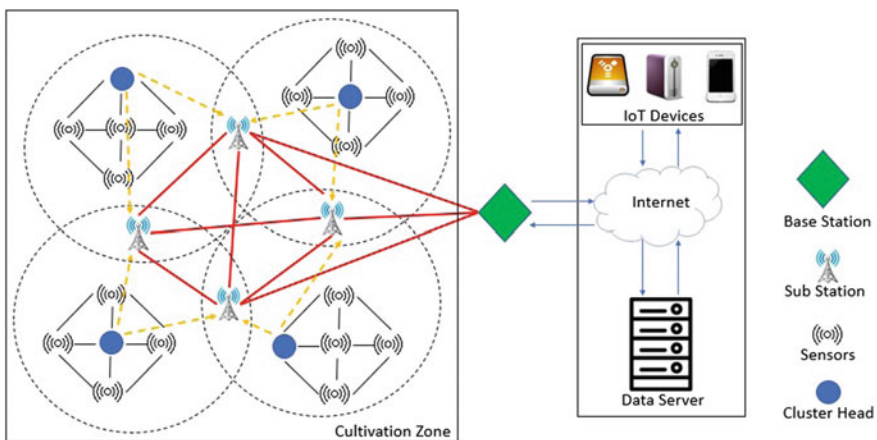
Fig. 1 Architecture diagram

The energy in a sensor is calculated based on transmitting and receiving the data in Eqs. (1) and (2). The formula used to calculate the energy is given as,

$$Tx_E(m, l) = E * m + \lambda * m * l^2 \tag{1}$$

$$Rx_E(m) = E * m \tag{2}$$

where, l is the length between two sensors, E is the consumption of energy factor denoting the power per bit during transmission and receiving the data and λ is the amplifier coefficient to send a single bit.

The proposed flowchart of ECHGS method is shown in the Fig. 2. Initially the clusters are formed based on the land area among the total agriculture. Sensors are placed equally in all the clusters. Cluster head is selected in each cluster based on the high energy and less load values.

Data gathering is done by the cluster head and send to the substation. Substation forwards the data to the data server for analyze the level of irrigation to each cluster. These analytical reports can be given as output to the users when they use the IoT based smart irrigation system.

4 Results

The proposed method results are analyzed using Network Simulator version-2 (NS2). The nodes in the simulation environment are distributed in the communication network. The channel type used in this simulation is a wireless channel, simulation time considered is 100 s, the number of nodes considered are 50, MAC type of the protocol is 802.11, traffic model used is CBR, transmission range is 250 m, network interface type used is wirelessPhy and mobility model used is random way model. In this simulation environment, the proposed method ECHGS results are evaluated in terms of residual energy and throughput. The proposed ECHGS method results are outperformed the results of TCDGS and CIDT.

4.1 Residual Energy (RE)

The proposed model in the IoT environment aims to increase the network lifetime by maximizes the throughput. The Residual Energy (RE) is calculated as the sum of remaining energy available in all the nodes as shown in Eq. (3) Residual Energy in each node is the amount of energy remaining in a sensor node at the current moment of time.

Fig. 2 ECHGS flow diagram
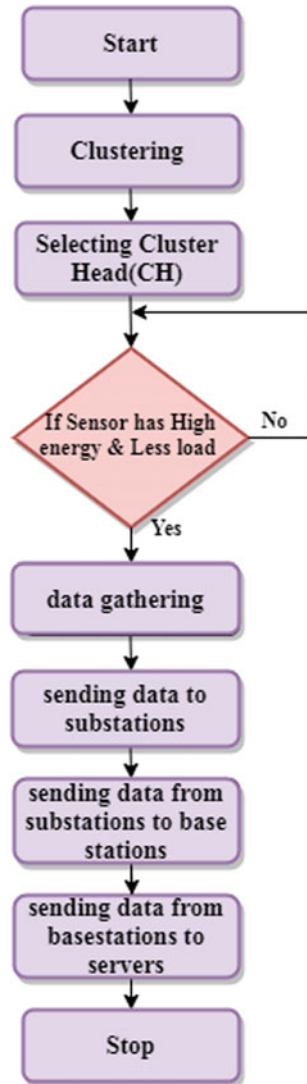
$$RE = \sum_{i=1}^{M} RE_i$$

(3)

Residual Energy is the remaining energy stored in the node. Figure 3 shows the residual energy comparison of proposed system ECHGS and existing methods TCDGS and CIDT. The proposed method ECHGS residual energy is higher than the existing methods of TCDGS and CIDT. Residual is maximized gradually from the
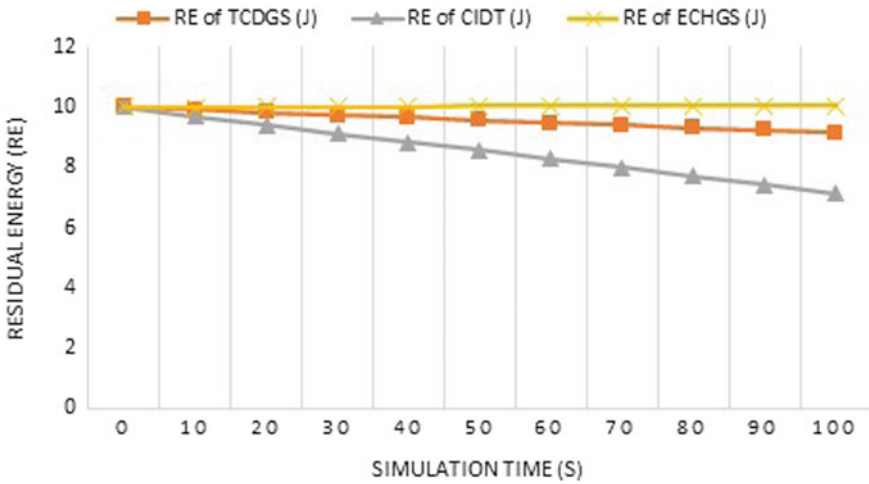
Fig. 3 Residual energy

simulation time 0 s to 100 s. The proposed ECHGS method residual energy is higher by 4.54 and 14.59% when compare to TCDGS and CIDT methods.

4.2 Throughput (Th)

The proposed model in IoT environment aims to increase the performance by maximizes the throughput. Throughput is the total successful data delivered to the server in the IoT environment. It is calculated Eq. (4) as the total successful data transferred (RP) in a unit of time (T).

$$Th = \sum_{j=1}^{N} RPj/T \qquad (4)$$

Figure 4 shows the throughput of the proposed method. The throughput is evaluated for the proposed system ECHGS and compared with the existing methods TCDGS and CIDT. Throughput of the ECHGS method is higher than the existing methods TCDGS and CIDT. Throughput is increased gradually when the simulation time 0 s to 100 s. The proposed ECHGS method throughput is higher by 11.57 and 50.00% when compare to TCDGS and CIDT methods.
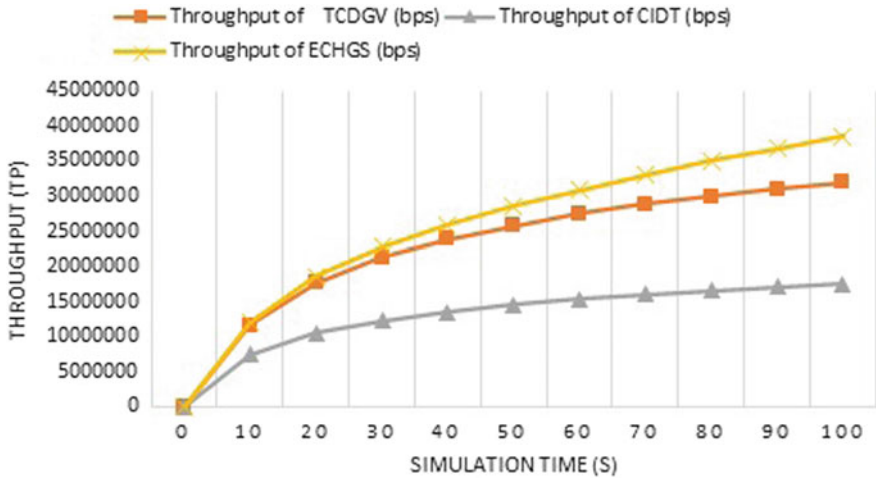
Fig. 4 Throughput

5 Conclusion

In this paper, the IoT based smart irrigation system using ECHGS method is discussed to address the solutions to increase the network lifetime and improve the network performance. The sensor nodes are clustered and a cluster head for each node is selected based on the essential parameters such as residual energy and load. The effective way of selecting a cluster head in each cluster leads to the better results from sensor node data collection step to data transfer up to the server. The proposed ECHGS method results are evaluated and compared with the existing methods of TCDGS and CIDT in terms of residual energy and throughput. The proposed ECHGS method results are better than the existing methods. In future, IoT based smart irrigation system can be implemented in real-time with other parameters packet delivery ratio and packet loss ratio.

References

1. Rapate GS, Naveen NC (2018) Energy and routing efficiency in IoT: proposal for combined approach. In: 2018 International conference on electrical, electronics, communication, computer, and optimization techniques (ICEECCOT), Msyuru, India, pp 451–454
2. Atzori L, Iera A, Morabito G (2010) The internet of things: a survey. Comput Netw 54(15):2787–2805
3. Deshpande Niranjan R, Vadane Pandurang M, Sangle Sagar D, Dighe MS (2016) A IOT-based modern healthcare system using body sensor network (BSN). Int J Innov Res Comput Commun Eng 4(11):19540–19546
4. Padmanaban K, Jagadeesh Kannan R (2016) Tree cluster based data gathering scheme (TCDGS) in wireless sensor networks. Int J Comput Technol Appl 9(61):2809–2818

5. Wang C, Ma H (2011) Data collection in wireless sensor networks by utilizing multiple mobile nodes. In: 2011 seventh international conference on mobile ad-hoc and sensor networks, Beijing, pp 83–90
6. Lee S, Cha J, Kim KS (2019) Data gathering and application to building energy optimization with sensitivity analysis for IoT applications. In: 2019 International SoC design conference (ISOCC), Jeju, Korea (South), pp 184–185
7. Ke H, Wang J, Wang H, Ge Y (2019) Joint optimization of data offloading and resource allocation with renewable energy aware for IoT devices: a deep reinforcement learning approach. In: IEEE Access vol 7. pp 179349–179363
8. Guo Y, Xiang M (2019) Multi-agent reinforcement learning based energy efficiency optimization in NB-IoT networks. In: 2019 IEEE globecom workshops (GC Wkshps), Waikoloa, HI, USA, pp 1–6
9. Appala Raju V, Sri Harsha V, Bhanu Deepthi N, Prasanth N (2018) Zonal stable election protocol for heterogeneous wireless sensor networks. Int J Eng Technol (UAE) 7:725–728
10. Dhage MR, Vemuru S (2018) A effective cross layer multi-hop routing protocol for heterogeneous wireless sensor network. Indonesian J Electri Eng Comput Sci 10(2):664–671. https://doi.org/10.11591/ijeecs.v10.i2.pp664-671
11. Goutham Chand K, Sidhendra M, Hussain MA (2018) Soil nutrient measurement in paddy farming using IoT. Int J Eng Technol (UAE) 7:356–358
12. Gupta P, Satyanarayan KVV, Shah DD (2018) Development and testing of message scheduling middleware algorithm with SOA for message traffic control in IoT environment. Int J Intell Eng Syst 11(5):301–313. https://doi.org/10.22266/IJIES2018.1031.28
13. Rao MV, Rama Krishna TV, Ganduri R, Roohi A (2018) An effective energy management system for smart office cubicles using IoT. J Adv Res Dynam Control Syst 10(2 Special Issue):338–347
14. Krishna MNV, Harsha NS, Kasula VDK, Swain G (2017) Optimization of energy aware path routing protocol in wireless sensor networks. Int J Electri Comput Eng 7(3):1268–1277. https://doi.org/10.11591/ijece.v7i3.pp1268-1277
15. Rajakumar R, Amudhavel J, Dhavachelvan P, Vengattaraman T (2017) GWO-LPWSN: grey wolf optimization algorithm for node localization problem in wireless sensor networks. J Comput Netw Commun. https://doi.org/10.1155/2017/7348141
16. Gupta P, Shah DD, Satyanarayana KVV (2016) An IoT framework for addressing parents concerns about safety of school going children. Int J Electri Comput Eng 6(6):3052–3059. https://doi.org/10.11591/ijece.v6i6.10448
17. Rao KR, Kumar TR, Venkatnaryana C (2016) Selection of anchor nodes in time of arrival for localization in wireless sensor networks. https://doi.org/10.1007/978-81-322-2671-0_5

Estimation of State of Charge Using EKF and SVR in Li-Ion Battery

P. Sannihith Reddy and R. Shanmughasundaram

Abstract Precise State of Charge (SoC) estimation is an essential requirement in electric vehicle energy management and also helps in the optimal control system design. But it is not a direct and easy task to be done. Batteries, which have a distinct nonlinear behavior because of various internal and external conditions and being complex electrochemical devices makes SoC estimation a very difficult task. Therefore, a wide range of methods is currently in place for estimating the battery SoC in real-time. This paper proposes a hybrid model which uses Extended Kalman Filter (EKF) and Support Vector Regression (SVR) to estimate the SoC. Firstly, an EKF algorithm that is based on a first-order Li-ion battery model is developed and SoC of the model is estimated. The estimated SoC value is then used as an input parameter to the SVR algorithm which acts as closed feedback to the model and helps in accurate prediction of SoC.

Keywords State of charge · EKF · SVR and Li-ion battery

1 Introduction

As air pollution due to fossil-fueled vehicle emissions has been raising at an alarming rate, the automotive sector has been looking forward to finding other efficient ways to power the engines. Electric Vehicles (EVs) seems to be a great solution to address this problem. According to the International Energy Agency (IEA) in 2018 electric car sales has been increased by 68% globally which is almost double the year-on-year average sales growth. This shows the impact of EVs in the transportation sector. However, the expert agrees on the need for the improvement in battery technology which is a major bottleneck for EVs.

Proper operation of EVs depends on fuel cells (battery pack) and their management systems. So for ideal energy transformation, stable and reliable power batteries are very important [1]. SoC is considered as an important parameter in the Battery

P. Sannihith Reddy (✉) · R. Shanmughasundaram
Department of Electrical and Electronics Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

Management System (BMS). The BMS monitors SoC and provides information about power level of the battery. The inefficient control of Li-ion battery charging process can affect the battery life and might also damage the battery. The SoC can be defines as the level of charge measured in comparison to its total capacity.

Out of the many available methods to estimate the SoC, one simple and common way is to build linear relation between battery open circuit voltage (OCV) and it's SoC. Despite, its simplicity this method is not preferred for dynamic applications. Because OCV takes long time (few hours) to settle down after a charge or discharge cycle [2]. The other notable methods include Ampere-hour, coulomb-counting, modified coulomb-counting method etc., that are used for estimating the SoC. However, these methods suffer from accumulation error with increasing time. Few advanced techniques like neural network method require huge training data and many complex calculations are involved in the process [3].

The model proposed in this paper makes use of EKF and SVR for estimating the SoC. EKF is one of the finest estimation methods available for a non-linear (but differentiable) state space model of a physical system. In EKF, SoC is estimated by comparing model outputs with the measurements acquired from the actual system. The physical system is modelled in a way that Gaussian white noises affect state and output equations [4]. The support vector regression model developed for estimating state of charge is made from the utilization of the right battery data to train and selection of suitable kernel function. The data set used in this model was collected from the CALCE battery group [5].

2   Battery Model

A mathematical model of the battery is required to build battery internal state variables and external quantitative relation to electrical characteristics. Then using external variables like terminal voltage, current and temperature internal state variables like SoC can be calculated. In this paper, the Thevenin model is chosen over its ability to accurately simulate the Li-ion battery dynamic characteristics.

In Fig. 1, the following parameters are defined as:

R_0 = Battery Internal Resistance.
C_1 = Double Layer Capacitance.
R_1 = Electrochemical Polarization Resistance.
C_2 = Capacity of the Battery.
V_1 = Voltage across RC branch.
V_2 (SoC) = Open Circuit Voltage (Function of SOC).
V_t = Terminal Voltage.
$I(t)$ = Battery current.
Applying Kirchhoff's voltage law to the model

$$Vt = V_2 - V_1 - R0I \qquad (1)$$
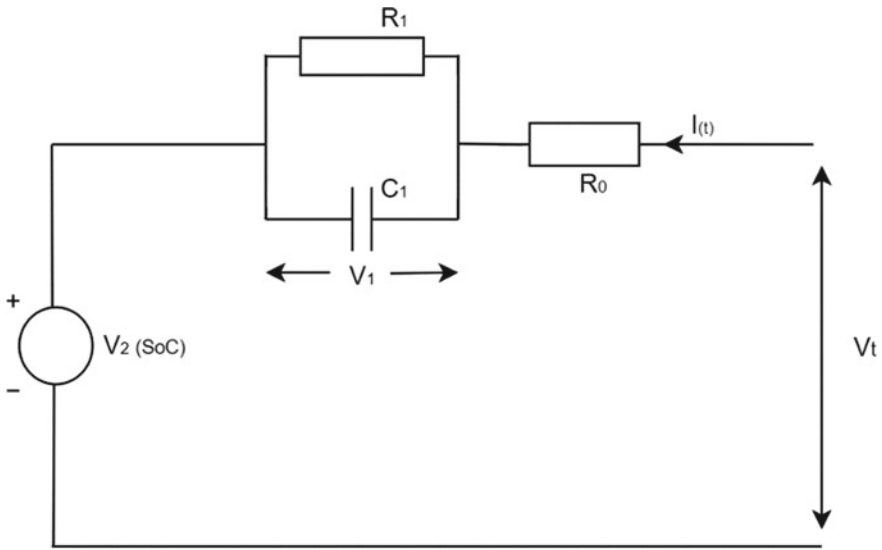
Fig. 1 Thevenin battery model

Substituting V2

$$Vt = fn(SOC) - V_1 - R0I \tag{2}$$

From the above equations, the following system equations can be derived:

$$\begin{bmatrix} SoC \\ V_t \end{bmatrix} = \begin{bmatrix} \frac{I}{kC_2} \\ -\frac{I}{R_1 C_1} V_t + \frac{1}{C_1} I \end{bmatrix} \tag{3}$$

$$V_2 = k\, SoC + V_t + I\, R_0 \tag{4}$$

Equation (3) represents state equation and Eq. (4) represents observation equation of the battery model. The discretization to get continuous equation and calculation of SoC using EKF are explained next part.

3 EKF Implementation

EKF algorithm implementation can be divided into two steps [6]:

1. Prediction step: This step involves anticipating the system future state by using present state.
2. Updating step: In this step, the value of present state estimate is updated on the basis of system output value.

From Eqs. (3) and (4), state equation of the system can be expressed as:

$$x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T, \quad x_1(t) = SoC, x_2 = V_t \tag{5}$$

Consider system input $u(t) = I$ and output $y(t) = V_t$, then battery model will be represented as below:

$$\dot{x} = f(x, u) + w \tag{6}$$

$$\dot{y} = g(x, u) + v \tag{7}$$

where w = process noise and v = measurement noise.
In Eq. (6) and (7)

$$f(x, u) = \begin{bmatrix} \frac{u}{kC_2} \\ -\frac{I}{R_1 C_1} x_2 + \frac{1}{C_1} u \end{bmatrix} \tag{8}$$

$$g(x, u) = kx_1 + x_2 + uR_0 \tag{9}$$

To linearize the model, Taylor series expansion is needed on Eq. (6) and (7). After linearization:

$$\dot{x} = A_t x + B_t u + w \tag{10}$$

$$\dot{y} = C_t x + D_t u + v \tag{11}$$

In Eqs. (10) and (11)

$$A_t = \frac{\partial f(x, u)}{\partial x} \tag{12}$$

$$B_t = \frac{\partial f(x, u)}{\partial u} \tag{13}$$

$$C_t = \frac{\partial g(x, u)}{\partial x} \tag{14}$$

$$D_t = \frac{\partial g(x, u)}{\partial u} \tag{15}$$

The discretized model is expressed as follows:

$$x_{k+1} = A_k x_k + B_k u_k \tag{16}$$

$$y_k = C_k x_k + D_k u_k \tag{17}$$

EKF algorithm calculation steps are given as:

- Initially set value of $k = 0$
- Initial condition is set to $x*k$
- P is error covariance matrix
- Q is noise covariance matrix

Update state estimation:

$$\hat{x}_k = A x_{k-1} + B u_k \tag{18}$$

Update error covariance:

$$P_k = A P_{k-1} A^T + Q \tag{19}$$

Kalman gain matrix:

$$L_k = P_k C_k^T \left(C_k P_k C_k^T \right)^{-1} \tag{20}$$

Update the estimate:

$$\hat{x}_k = \hat{x}_k + L_k \left[y_k - \left(C_k \hat{x}_k + D_k u_k \right) \right] \tag{21}$$

Update error covariance:

$$P_k = (I - L_k C_k) P_{k-1} \tag{22}$$

Using the above formulae, EKF calculates the system state. The main step of the EKF is to adjust the kalman gain according to error magnitude by comparing the predicted value with the measured and this gain is used to compute the further values. The bigger the error, greater the gain and greater the correction rate of estimated value; lesser the error, smaller the gain, and lesser the correction rate of estimated value. This provides a lucid reason as to why the EKF method is good in real-time computation performance [7].

4 Simulation

The simulation of EKF algorithm has been carried out in Matlab Simulink based on the battery model. The capacity of Li-ion battery used in this model is 2000mAh and cut-off voltage is 4.2 V.

Simulation parameters are:
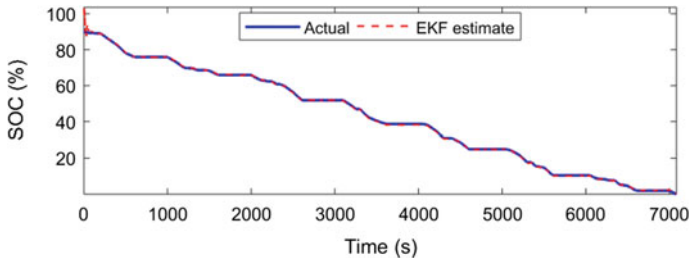
Fig. 2 SoC estimated value and actual value
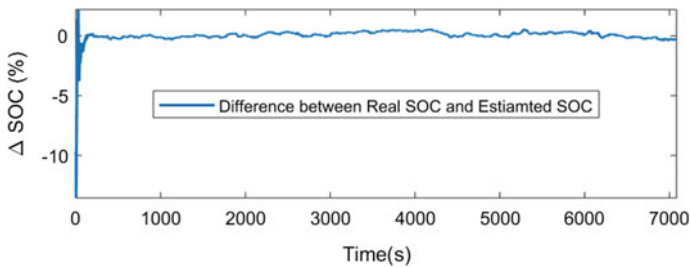
Fig. 3 Error in estimation

- Initial State: $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$

- Initial Covariance: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

- $Q = 0.1$
- $R = 100$

To validate the result of extended kalman filter algorithm, the simulation results were compared with the experimental data collected from CALCE battery group.

Figure 2 shows the estimated SoC values by EKF to the actual values. Error in the estimation can be observed from Fig. 3.

5   SVR Implementation

Usually, Support Vector Machine (SVM) is used as a classification algorithm. SVM classifies the data points by shifting data to a high-dimensional space. In a situation where linear separation of data is not possible, the data is transformed to enable drawing a hyperplane between various classes. Then, features of new data is used to predict the class to which a new data point belongs to. SVR uses similar principles as SVM, with only a few changes. In SVR, since the output is a real number predicting

the information is not a direct task. However, the main idea remains same: to minimize error, identify the hyperplane, which maximizes the margin, keeping in mind that part of the error is tolerated [8, 9].

The accuracy of the SVR model majorly depends on the proper data set, kernel function and model parameters. The model parameters should be chosen wisely or else there is a chance that the model may not behave as it is intended to. Also, select the input features that has considerable effect on the required target parameter. This helps to avoid unnecessary complexity and improves the accuracy [10].

Here in this paper SVR is implemented using python sklearn library. The data required to the model is collected from CALCE battery data group. The kernel function used is Radial Bias Function (RBF). Because, the optimal predictive outcome is better for a nonlinear kernel compared to a linear one. The model parameter C, which is the penalty parameter used to control the trade-off between training error and a testing error. This helps in generalizing the classifier to unconsidered data. C is set to 0.1 in this model. The value of C is selected by comparing the accuracy with the various C values.

Figure 4 shows the variation in model prediction accuracy for various C values. It is found that at $C = 0.1$ model is getting saturated. So C is set to 0.1 to avoid further complexity by increasing the C value. The other model parameter gamma is set to 1, gamma determines how quickly boundary dissipation happens; larger values decrease the effect of any individual support vector. Figure 5 shows how variation in c and gamma values effect the classification model.

The idea is to develop a model that has the ability to estimate SoC accurately in real-time. It is known fact that EKF is highly dependent on the battery model. But creating a battery model which can address all real-time scenarios is not an easy task. So this might affect the EKF model estimation accuracy in actual situation. To
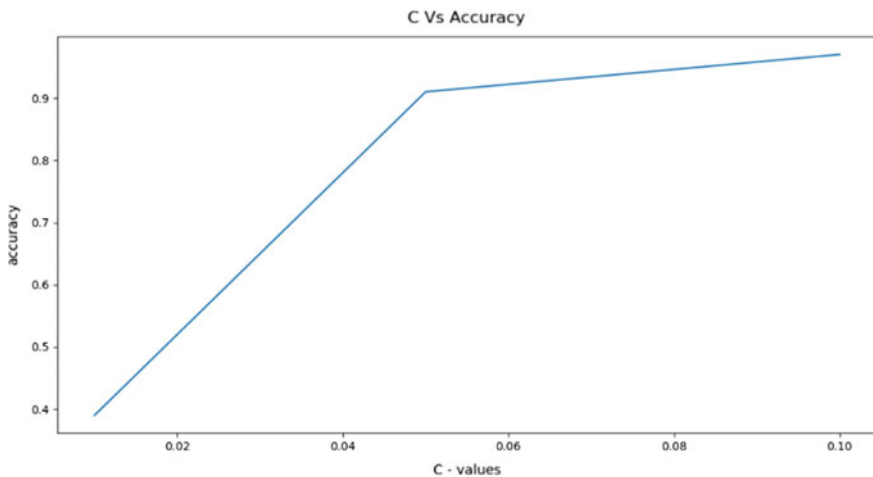
Fig. 4 C versus prediction accuracy
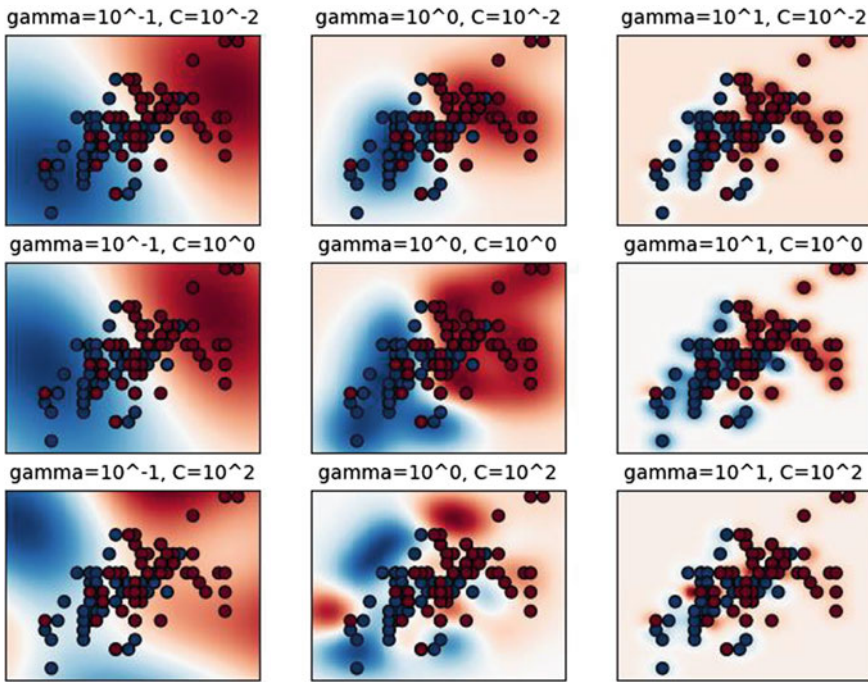
Fig. 5 SVR parameters (c & gamma) effect [11]

address this and improve the prediction accuracy, a kind of closed feedback SVR model is developed by giving SoC estimated by the EKF algorithm as one of the input parameter along with voltage, current and discharge capacity to the SVR model. This acts as indirect closed feedback loop and help the model to correct the errors in EKF model estimation, which results in better accuracy.

Figure 6 shows the variation in SoC values predicted by SVR algorithm vs actual SoC values. The SVR model prediction accuracy is found to be 99%, which is 4% more than EKF. The increase in accuracy may seem to be small but the model is capable of handling real-time scenarios in a better way when compared to EKF model alone.

6 Conclusion

A hybrid model for estimating the SoC based on EKF and SVR algorithm has been proposed in this paper. Initially required battery data set for the model has been obtained from CALCE battery group, and then a mathematical model of the battery with the same specification as of data set collected is developed. Then using the model
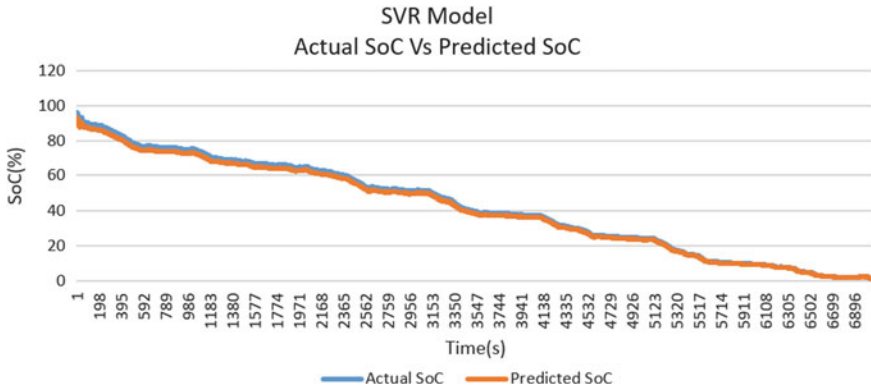
Fig. 6 SoC prediction using SVR

system state equation and observation equation are obtained. Now using EKF algorithm equations iteratively system state has been calculated and updated constantly. Finally, an accurate estimation value is achieved from EKF. Now this estimated SoC value along with other battery parameters is used to develop the SVR model, which resulted in accurate SoC prediction.

References

1. Ding-Xuan Y, Yan-xia G (2003) SOC Estimation of Lithium-Ion battery based on kalman filter algorithm. Appl Mech Mater 347–350
2. Ciortea F, Rusu C, Nemes M, Gatea C (2017) Extended Kalman Filter for state-of-charge estimation in electric vehicles battery packs. In: 2017 International conference on optimization of electrical and electronic equipment and 2017 Intl aegean conference on electrical machines and power electronics, Brasov, pp 611–616
3. Muthumanikandan S, Shanmughasundaram R (2016) Estimation of state of charge of lithium ion battery using artificial neural networks. Int J Control Theory Appl 9(10):4331–4338
4. Taborelli C, Onori S (2014) State of charge estimation using extended Kalman filters for battery management system. In: IEEE international electric vehicle conference (IEVC), Florence, pp 1–8
5. https://web.calce.umd.edu/batteries/data.html
6. Krishnakumar A, Shanmughasundaram R (2018) Simplified SOC estimation by EKF in Li-Ion cell. JARDC 3:616–622
7. Kitagawa M, Shibuya H, Takehara J (1992) Development of battery state of charge indicator for electric vehicles. In: Proceedings of the 11th international electric vehicle symposium
8. Saimurugan M, Ramachandran KI, Sugumaran V, Sakthivel NR (2011) Multi component fault diagnosis of rotational mechanical system based on decision tree and support vector machine. Expert Syst Appl 38:3819–3826
9. Sabareesh GR, Sugumaran V, Ramachandran KI (2006) Fault diagnosis of a taper roller bearing through histogram features and proximal support vector machines. In: IEEE international conference on signal and image processing, December, B V Bhoomaraddi college of Engineering and Technology, Hubli

10. Haq IN, Saputra RH, Edison F, Kurniadi D, Leksono E, Yuliarto B (2015) State of charge (SoC) estimation of LiFePO4 battery module using support vector regression. In: Proceedings of the joint international conference on electric vehicular technology and industrial, mechanical, electrical and chemical engineering, Surakarta, pp 16–21
11. https://scikitlearn.org/stable/_images/sphx_glr_plot_rbf_parameters_001.png

Evolving Spiking Neural Network for Prediction Problems in Healthcare System

Tasbiha Ibad, Said Jadid Abdulkadir, and Norshakirah Binti Ab Aziz

Abstract This paper highlights the role of evolving spiking neural networks (an enhanced version of SNN) for predicting medical diagnosis. This article aims to focus on regression problems under a supervised learning strategy. In this paper, we have trained and tested eSNN on benchmarking datasets. Among the three datasets, one is the ICU Dataset which helps in predicting the recovery ration of patients who stayed in ICU. Another dataset is Plasma_Retinol which predicts the risk of cancer-related to certain carotenoids. Dataset pharynx is a part of a study conducted in the USA to determine the success rate of two radiation types. The selected datasets are those which were previously used for BioMedical Engineering related tasks. Later the evaluation was conducted using Regression Metrics. From experiment results, it is concluded that eSNN with standard parameters without optimization performed well but there is still space available for improvement to achieve the highest possible prediction scores.

Keywords eSNN · Regression · Optimization · Prediction

1 Introduction

Artificial neural networks can be appraised as one of the vigorous classifiers. Because of their ability to handle noisy data skillfully [1–4]. ANN solves problems that are established on approved algorithmic techniques [5–7]. SNNs (Spiking Neural Networks), the 3rd generation of ANNs, plays a significant role in the processing of biological information [8]. The initial two prototypes of SNN are conspicuous which are recognized as the Threshold and Conductance model [9]. There is one derived model of SNN which is known as Evolving Spiking Neural Network (eSNN), which was first presented by [10]. eSNN is the one that evolves through learning. For encoding techniques, eSNN utilizes Population encoding. As for the training, the

T. Ibad (✉) · S. J. Abdulkadir · N. B. A. Aziz
Department of Computer and Information Sciences, Universiti Teknologi PETRONAS Seri Iskandar, Perak, Malaysia
e-mail: tasbiha_18003084@utp.edu.my

one-pass learning method has been chosen and it fits well. The target of the training is to establish an output neuron repository with class labels [11]. The objective of eSNN is to serve as a fast-learning process that is suitable for both offline and online tasks. eSNN is widely applied in classification tasks, such as face recognition [12] ecological problems [13], taste recognition [14] and has accomplished better results than conventional methods. But for regression tasks, the contribution of eSNN is not highlighted. Therefore, this study aims to provide a brief review of eSNN performance in Regression Problems.

2 Literature Review

In the literature study of eSNN, there is a wide range of its application for classification tasks. Such as the Integration of eSNN with the Harmony search algorithm [15]. Firefly Algorithm was used to optimized eSNN for classification problems [9]. Multi-Objective Differential Evolution algorithm is combined with eSNN in [16], to achieve the best accuracy results for classification. Differential evolution (DE) was integrated with eSNN in [19] to optimize it for Classification Problems. A new hybrid K-DESNN for clustering issues was established by integrating D.E and K-means with evolving spiking neural networks (K-means eSNN) in [17]. eSNN was optimized with the Harmony Search Algorithm (HSA) in [18]. To tackle the issue related to hyperparameter sensitivity of eSNN, an optimizer was needed by eSNN for parameter optimization. Therefore, this article proposed the combination of HSA as an optimizer and ESNN as a classifier for the optimization of parameters. There were several attempts to solve the issue of manual parameter tuning. To compare the results of the optimized eSNN model there is a standard eSNN with parameter values: C: 0.9, Sim: 0.1, and Mod: 0.75 introduced in [19]. As stated earlier, mostly optimized eSNN frameworks are for classification tasks. Therefore, this study aims to evaluate the standard eSNN for Regression Problems. For this paper, we have trained eSNN on popular datasets from literature, for Regression Problems.

3   Methodology

3.1   Evolving Spiking Neural Network

Algorithm 1: eSNN Training Algorithm

Require : m_l, s_l, c_l for a class label $l \in L$
1.   initialize neuron repository $R_l = \{\}$
2.   for all samples $X^{(i)}$ belonging to class l do
3. $w_j^{(i)} \leftarrow (m_l)^{order(j)}, \ \forall j \mid j$ pre-synaptic neuron of i
4. $u_{max}^{(i)} \leftarrow \sum_j w_j^{(i)} (m_l)^{order(j)}$
5. $v^{(i)} \leftarrow c_l \, u_{max}^{(i)}$
6.   if min (d $(w^{(i)}, w^{(k)})) < \ s_l$, $w^{(k)} \in \ R_l$ then
7. $w^{(k)} \leftarrow$ merge $w^{(i)}$ and $w^{(k)}$
8. $v^{(k)} \leftarrow$ merge $v^{(i)}$ and $v^{(k)}$
9.   else
10. $R_l \leftarrow R_l \cup \{w^{(i)}\}$
11.   end if
12.   end for

The training begins with the initialization of 3 eSNN parameters in the interval [0, 1]: C (proportion factor) and Sim (similarity value), and Mod (modulation factor). Each sample i that is from a similar class l is encoded into numerous pre-synaptic input neurons named as j. The weight wj is calculated using Mod_i and $order_j$. Where, Mod_i represents modulation factor from Thorpe neural model. The $order_j$ defines the rank of an emitted spike from neuron j. The output neuron similarity is estimated according to the Euclidean distance between the neuron weight vectors. If a particular neuron is believed to be very similar to others, it will combine with the truly similar one. Thus, the merging procedure involves the calculation of average weight vectors along with the threshold value. N represents the figure of samples formerly used to upgrade output neuron k.

3.2   Datasets

Benchmarking datasets from the OpenMl.org dataset repository, are used to train and test eSNN see Table 1. The chosen datasets are related to health sciences. And were collected to be used for a specific study based on regression analysis. The ICU dataset contains information of 200 survived patients after staying in an adult intensive care unit (ICU). Plasma Retinol, datafile comprises of 315 observations on 14 variables.

Table 1 Datasets description

Datasets	Number of instances	Number of attributes
ICU	200	21
Plasma Retinol	315	14
Pharynx	195	12

The pharynx data is a small portion of a large medical trial conducted in the United States for patients with cancer (squamous carcinoma).

3.3 Validation Metrics

The validation of eSNN conducted using the following famous regression metrics.

- **R-Squared**: It is a statistical degree of indicating how well the regression predictions estimated the real data values. It is also called the Coefficient of Determination
- **Mean Absolute Error:** It calculates the average amount of the errors in a set of calculations, without focusing their direction. Concisely, it is the average absolute difference between X and Y.
- **Root Mean Square Error**: It is the measure to tell how the residuals (i.e., prediction errors) are spread out over the regression line. It is the standard deviation of the residuals.

4 Result and Discussion

This section demonstrates the findings of the analysis based on prediction accuracy and other measures. Additionally, the experiments are carried out 10 times for both pieces of training as well as testing on the whole data sets. From a total number of samples, 80% of the data is the training dataset and the rest is used for testing. The experiment was conducted on ICU, Plasma-Retinol, and Pharynx Datasets from the field of Medical and Health Sciences. The results have been evaluated based on their prediction performance. According to the results presented in Table 2 the average R-square value is negative for Plasma_Retinol. It is conceivable to get a negative R-square for calculations that do not comprise of the constant period. In the case of

Table 2 Average prediction rate of eSNN

Datasets	R^2	MAE	RMSE
ICU	0.32	0.49	1.13
Pharynx	0.5	96	186
Plasma_Retinol	-0.80	116	132

negative results, R-square cannot be interpreted as the square of a correlation [20]. For ICU and Pharynx datasets R-square values are positive but very low. In general, the higher the R-squared, the better the model fits the data. Statistically, Mean Absolute Error (MAE) is one of the several metrics for reviewing and evaluating the excellence of a machine learning model. The unit of MAE value is the same as the target variable. The average MAE value for ICU 0.49, for Plasma_Retinol it is 96, and for Pharynx data it is 116. This value indicates that the prediction error for ICU is very low. But for the rest of the two datasets, error values are high which means eSNN did not predict accurately. RMSE tells us how focused the data is across the line of finest fit. If the RMSE will be 0, it implies that the points lie on the regression line (and therefore there are no inaccuracies). The unit of RMSE value is also the same as the target variable. But the higher the value the greater the prediction error, which indicates the poor performance of the model. According to average values from Table 2, it can be seen that the RMSE value for Plasma_Retinol is 132, and for Pharynx data it is 186 which describes that eSNN did not predict well for these two datasets. As the rate of error is high. But for the ICU dataset, the error rate of eSNN prediction performance is low i.e., 1.13. This shows good eSNN performance. The overall results of eSNN prediction performance is a combination of good and bad fit. In this study, eSNN was with standard parameters, so it could be hypothesized that eSNN can improve its performance with its optimized parameters (Fig. 1).

5 Conclusion and Future Directions

In this study, we have tested standard eSNN for regression problems. The model is trained on widely used datasets from OpenML.org. During the literature analysis, it was found that eSNN was widely used and tested for classification problems. But regression problems are neglected except for a few stdies. So, this study aims to test eSNN for the regression task. Here we used eSNN with standard parameter values. As stated, earlier eSNN is sensitive to its parameters. Therefore, the selection of parameter values is a crucial task. Here eSNN with standard values of parameters is trained and tested. According to the results eSNN performed not so well for regression datasets. Hence there is wide room for improvement for the same problems. Future work could be the optimization of eSNN with several well-known Optimization Algorithms. So that eSNN would achieve the best possible results in solving regression problems.
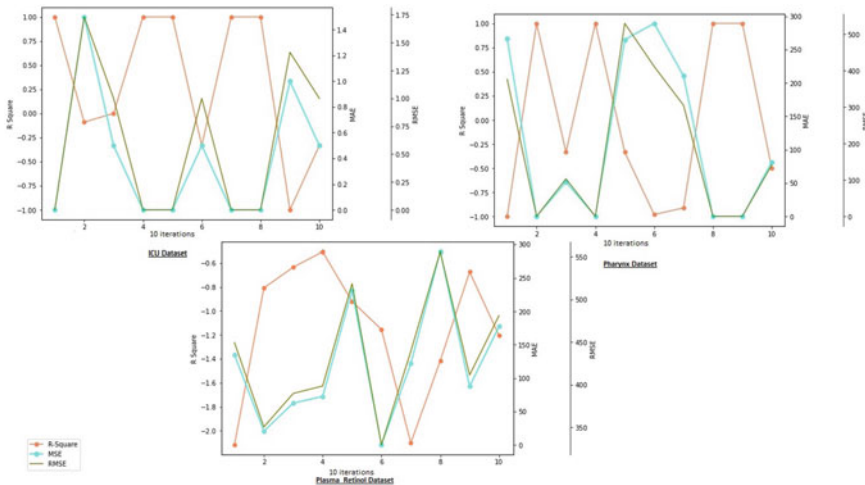
Fig. 1 Graphical representation of prediction score at every 10 iterations

References

1. Morris A, Josifovski L, Bourlard H, Cooke M, Green P (2000) A neural network for classification with incomplete data: application to robust ASR. In: Proceedings ICSLP 2000 (No. CONF)
2. Abdulkadir SJ, Suet-Peng Y, Foong OM (2013) Variants of particle swarm optimization in enhancing artificial neural networks. Austral J Basic Appl Sci
3. Abdulkadir SJ, Yong S, Zakaria N (2016) Hybrid neural network model for metocean data analysis. J Inform Math Sci 8(4):245–251
4. Abdulkadir SJ, Yong S-P, Marimuthu M, Lai F-W (2014) Hybridization of ensemble Kalman filter and non-linear autoregressive neural network for financial forecasting. Mining Intell Knowled Explor 72–81
5. Abdulkadir SJ, Yong SP (2014) Empirical analysis of parallel NARX recurrent network for long-term chaotic financial forecasting. In: 2014 International conference on computer and information sciences ICCOINS 2014—a conference of world engineering science and technology congress ESTCON 2014—proceedings
6. Abdulkadir SJ, Yong SP (2015) Scaled UKF–NARX hybrid model for multi-step-ahead forecasting of chaotic time series data. Soft Comput 19(12):3479–3496
7. Abdulkadir SJ, Yong S-P (2013) Unscented Kalman filter for noisy multivariate financial time-series data. In: International workshop on multi-disciplinary trends in artificial intelligence
8. Ghosh-Dastidar S, Adeli H (2009) Third generation neural networks: spiking neural networks. In: Advances in computational intelligence. pp 167–178
9. Roslan F, Hamed HN, Isa MA (2017) The enhancement of evolving spiking neural network with firefly algorithm. J Telecommun Electron Comput Eng

10. The ECOS framework and the ECO learning method for evolving connectionist systems. J Adv Computat Intell 2(6):195–202
11. Lobo JL, Laña I, Del Ser J, Bilbao MN, Kasabov N (2018) Evolving spiking neural networks for online learning over drifting data streams. Neural Netw 1(108):1–9
12. Wysoski SG, Benuskova L, Kasabov N (2008) Fast and adaptive network of spiking neurons for multi-view visual pattern recognition. Neurocomputing 71(13–15):2563–2575
13. Schliebs S, Platel MD, Worner S, Kasabov N (2009) Quantum-inspired feature, and parameter optimization of evolving spiking neural networks with a case study from ecological modeling. In: 2009 international joint conference on neural networks 2009 Jun 14 IEEE
14. Soltic S, Wysoski SG, Kasabov NK (2008) Evolving spiking neural networks for taste recognition. In: 2008 IEEE international joint conference on neural networks (IEEE World Congress on Computational Intelligence) 2008 Jun 1. IEEE, pp 2091–2097
15. Yusuf ZM, Hamed HN, Yusuf LM, Isa MA (2017) Evolving spiking neural network (ESNN) and harmony search algorithm (HSA) for parameter optimization. In: 2017 6th international conference on electrical engineering and informatics (ICEEI) 2017 Nov 25. IEEE, pp 1–6
16. Saleh AY, Shamsuddin SM, Hamed HN (2015) Multi-objective differential evolution of evolving spiking neural networks for classification problems. In: IFIP international conference on artificial intelligence applications and innovations 2015 Sep 14. Springer, Cham
17. Saleh AY, Hameed HN, Najib M, Salleh M (2014) A novel hybrid algorithm of differential evolution with evolving spiking neural network for pre-synaptic neurons optimization. Int J Adv Soft Comput Appl 6(1):1–6
18. Saleh AY, Hamed HN, Shamsuddin SM, Ibrahim AO (2017) A new hybrid k-means evolving spiking neural network model based on differential evolution. In International conference of reliable information and communication technology 2017 Apr 23. pp 571–583
19. Saleh AY, Shamsuddin SM, Hamed HN (2016) A memetic harmony search algorithm based on multi-objective differential evolution of evolving spiking neural networks. Int J Swarm Intel Evol Comput 5(130):2
20. Kitchenham BA, Pickard LM, MacDonell SG, Shepperd MJ (2001) What accuracy statistics measure. In: IEEE proceedings-software. Jun 1. vol 148(3). pp 81–5. Menard S. Applied logistic regression analysis. Sage, 2002

Prototyping Pro-Active Wearable Gadget for the Surveillance of Coal Miners in Pakistan

Rahul Kumar, Ramani Kannan, Jayasankari Ganasan,
Ghulam E Mustafa Abro, Abdul Sattar, Vipin Kumar, and Nirbhay Mathur

Abstract After an advent of internet of things (IoT), the entire dynamics of this world has been changed so rapidly. Being a student of science and technology, it is very much essential for us to think about the potential alternatives for social issues. In this regard, one may notice that there are thousands of people who die annually in the coal mines of Pakistan and rescue team cannot help in any way. This is one of the most emerging issues right now in other developing countries too where extraction of minerals from in depth mines is going on. This paper identifies this issue and suggests the prototyping of pro-active IoT-enabled wearable gadget for the smart surveillance of coal miners within the mines of Pakistan. This wearable IoT device not only monitor the hazardous gas, pulse rate of miner, humidity and temperature conditions within and exact altitude where miner is working currently but it will also provide a pro-active control to miners to generate any sort of query they face while working within the dark and depth mines using error palette. This entire data will be received at a specific Internet protocol (IP) using internet broker. This will also help the rescue officers to decide their rescue flow accordingly.

Keywords Coal miners · Internet of things (IoT) · Internet protocol · Proactive approach · Surveillance

R. Kumar · R. Kannan · N. Mathur
Department of Electrical and Electronics Engineering, Universiti Teknologi PETRONAS, 32610 Perak, Malaysia

J. Ganasan (✉) · A. Sattar
Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, 32610 Perak, Malaysia
e-mail: Jayasankari_20000248@utp.edu.my

G. E. Mustafa Abro
Department of Electrical and Electronics Engineering, Hamdard University, Karachi 74600, Pakistan

V. Kumar
Department of Civil and Environmental Engineering, Universiti Teknologi PETRONAS, Perak, Malaysia

1   Introduction

In the era, where internet of things (IoT) has evolved the dynamics of every part and corner of the world, there are still some areas where the utilization of IoT-enabled systems is still at the stage of infancy. One of such areas is the coal mining industry where various Alliances of support workers and coal miners witnessed the loss of their loved ones. In the list of recent incidents, one may see the tragedy of Sanjdi coal mine of Baluchistan, Pakistan where more than 10 people died, and 8 people were in critical situation because of the leakage of some poisonous gas.

So far, one may see the serious concerns of International Industrial Union for coal miners and specifically the focus of Pakistan central mines labor federation (PCMLF) to resolve such issues and they both are in favor of IoT enabled systems i.e. wearable gadgets. The proposed IoT-enabled pro-active system in this paper cannot be implemented within coal mine due to constraints of communication but if managed on a wireless sensor nodes or network than thousands of the lives may be secured easily. Even in the case of any uncertainty i.e. collapse of a mine, the rescue team can perform their duties more efficiently without wasting unnecessary time.

The suggested system in this manuscript proposes an IoT-enabled system, that comprises of several sensors as discussed in the section of methodology later and this make this pro-active gadget as one of its kinds. There are most of the cases where these rescue professionals start their operation at a very unknown place where there is very less success probability. By utilizing the proposed system, these professionals can save maximum coal miners because this system will not only be used for surveillance purpose unlike in [1]. This gadget is pro-active in nature which means now the coal miners can intimate their location to base station if they are fortunately alive and as per altitude the rescue team initiate the operation.

2   Literature Review

This section brief some of the latest strategies that were previously used in the active sensorization and detection of physical parameters. These same techniques will be discussed in general to produce a statement that these sensors can be interfaced with any sort of micro-controller to design a pro-active wearable gadget. This pro-active wearable gadget will not only sense the physical parameters within the local mines of Pakistan to update the base station [1] but this will also enable the miners to generate a query too. The brief study of literature can be witnessed under this section that can acquire the exact whereabouts of a miner and let us know about his surroundings easily. Moreover, previously people used such techniques only for monitoring and surveillance. One can see the ubiquitous and pervasive technology for data acquisition [2]. This approach found to be suitable because of less computational capability but less interaction between the system and user. Some of the researchers have proposed the single dimension audio signal communication but this will increase

the requirement of storage device. These techniques if utilized properly then the rescue team will not dig at unknown place and instead of this, they will know exactly the spot from where they can rescue the miner [3]. Researchers of today's era are still engaged in proposing the wearable and easy to carry devices, but all are limited to surveillance [4].

After studying several research contributions, it has been noted that previously proposed techniques i.e. An alert system design based on android applications [5] were either limited to just monitoring or to generate the alert only. The one of the major shortcomings of such strategies where they were entirely unable to share the exact location of miner who is in trouble within the debris.

One may raise the questions regarding the accuracy of using such sensors; thus, it is to share that these same sensors have been proposed in the design of various sensitive incubators where exact data of babies are communicated using wireless communication [6]. As a researcher, one may see various people have contributed various alternatives like this to resolve many critical as well as sensitive issues [7]. One can see the use of such sensors in the design of sewer-bot that is designed to clear the drain line and senses the temperature, humidity, and leakage of any hazardous gas [8] (Table 1).

The leakage of hazardous gases results in various tragedies thus, there have been many sensors and ways to identify the leakage such that average slope-based multiplication [9]. In the catalogue of sensors that may detect the leakage of such hazardous and explosive gases, one may find number of sensors that have made our work

Table 1 Summary of all previously proposed techniques

Sr.	Proposed Technique	Cons	Refs.
1.	Smart wearable jacket design	Miners cannot generate any query in case of emergency or uncertainty	[1]
2.	Pervasive computing strategies	The algorithm is complex and less flexible	[2]
3.	2-D signal communication	Needs more storage and there is more distortion	[3]
4.	Smart wrist watch	Mobile communication cannot done be within mines	[4]
5.	Android application based systems	Every miner cannot afford the smart mobile phones	[5]
6.	Mean square and average slope multiplicative algorithms	The process time is huge and in our scenario it misfits because of process delays	[6]
7.	Magnetic low frequency techniques	These low magnetic frequencies causes cancer	[10]
8.	MEMS systems	Still there is no full range of sensors that are required for our case	[11]
9.	Intelligent robotic applications	The mapping constraints of mines as every environment is different	[12]

easier to detect the leakage of such explosive or poisonous gases i.e. methane [10]. The main constraint for utilizing such sensors is to establish the communication network within the underground environment. This type of communication is already established using low frequency magnetic fields and their immediate transmission approach [11]. While studying the literature related to such utilization of sensors, one may come across with MEMS based sensors and devices that are very robust when it comes to use them within mines or underground environment but this require the digital communication followed by the standard IEEE 802.15.14 protocol [12]. This IEEE standard is proposed to communicate at low rate within a wireless area network. Almost in all sort of research contributions so far, researchers are somehow more tilted towards monitoring of an unknown area and generating the alerts with the detection of some physical parameters. Some researchers have also proposed some of the robots for this task i.e. the famous wolverine version 2 which was introduced for the first time in the year of 2001 by mine health and safety administration (MHSA), United States of America [13]. This is not the first time that robots are introduced in such critical applications, they have been used before such that ground hog, Gemini scout and numbat proposed for the similar issues. One may see the Table 1, that highlights the summary of such techniques and systems proposed for the similar cases.

The mentioned Table 1 presents a brief comparison among the proposed techniques. One may see most of the proposed strategies are based generating either the alert or just for monitoring purpose. In addition to this, majority of the techniques cannot communicate their global positioning satellite (GPS) location, temperature, humidity, exact depth and lastly, the status of miner whether he is alive or not.

3   Methodology

In this section, the hardware implementation for the pro-active wearable gadget is proposed. This manuscript presents some of the efficient sensors to be embedded with the system to work as per objectives as highlighted in the below mentioned Table 2.

In this entire pro-active wearable gadget, the user will first turn on the system by pressing the single pole single through switch (SPST). After just turning it on, the sensors will start gathering the results of all physical parameters and send it to Arduino board.

Once Arduino Mega board will receive it, this will further be processed it to send it to static internet protocol (IP) using ESP8266 Wi-Fi module. One can see the entire data i.e. temperature, humidity, pulse rate, altitude, and exact location on a single IP-oriented screen. The role of error palette is to provide a pro-active access to miner. This is because if anything happens in front of him for which he needs to update the base station then this smart pro-active wearable gadget may provide him this facility to communicate easily using this palette. This error palette comprises of 01 switch and 01 Led dully connected with the same main system. This switch

Table 2 Proposed component list with their purpose

Sr.	Proposed sensor	Purpose
1.	DHT 11s	For detection of runtime temperature and humidity values
2.	MQ-2 sensor	To detect the leakage of any hazardous gas within mine
3.	Pulse sensor	To acquire the pulse of miner working within the mines
4.	GPS module	Proposed to know the exact longitudinal and latitudinal values of miner
5.	BMP 180 sensor	To acquire the altitude of miner at what depth he is working
6.	ESP8266 Wi-Fi module	To link the sensor values with internet of things
7.	Arduino mega board	To process the entire sensed values
8.	Error palette	To send the queries from miner's side to base station
9.	Single pole single through switch SPST	To turn on the system

is there for generating an emergency query within the mine and to indicate for help from base station. One may see the entire block diagram as mentioned in Fig. 1. The entire prototype as shown in the Fig. 1 can be initiated in acrylic sheet [1] or may be mounted on any cloth fabric or one may opt for a safety helmet and embed this system with in it. In this manuscript the proposed system is not mounted on anything to show that it can be adjusted into any form. The entire schematic diagram is shown in Fig. 2. Whereas the hardware prototype is illustrated in Fig. 3.

4 Results and Discussion

The device has been tested for the results through different ways. In order to check out the change in altitude and location, this device had been moved to different height and positions and the entire results were updated accordingly on serial monitor of Arduino integrated development environment shown in Fig. 4. Moreover, for checking the pulse values, the sensor has been given an extended wire so that the miner may place it on of his finger. The entire data has been sent to IP and one can see the display as illustrated in Fig. 5.

In above figures, the altitude is shown in feet one may convert it into meters too. The minus (−) sign shows the depth at which the miner is working.

After putting the values on LatLong.net, we come across the exact location of the webinar as shown in Fig. 6 given below:

Whereas the user can also utilize the error pallet and may generate the query. After generating the query, the base station may send the help or rescue team to help
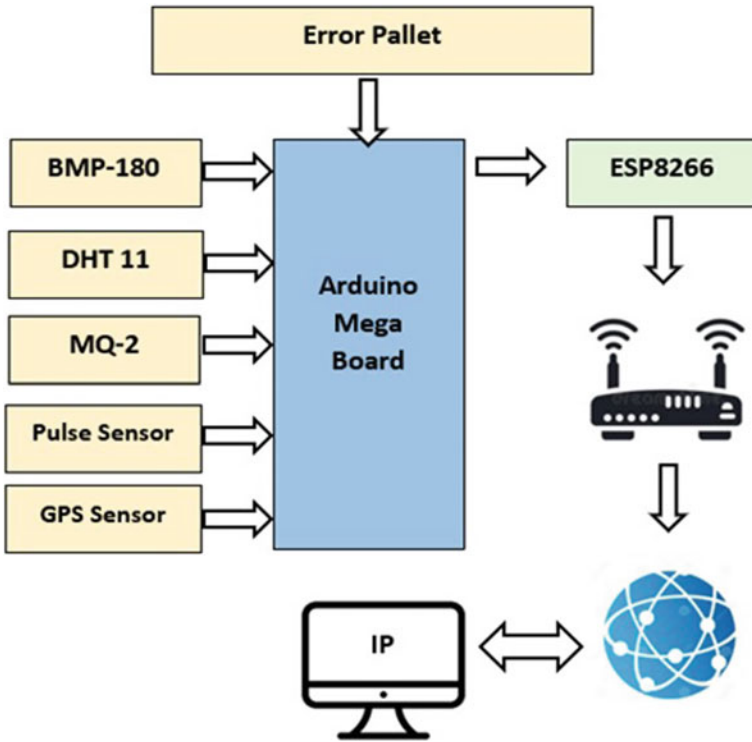
Fig. 1 Block diagram of proposed pro-active wearable gadget
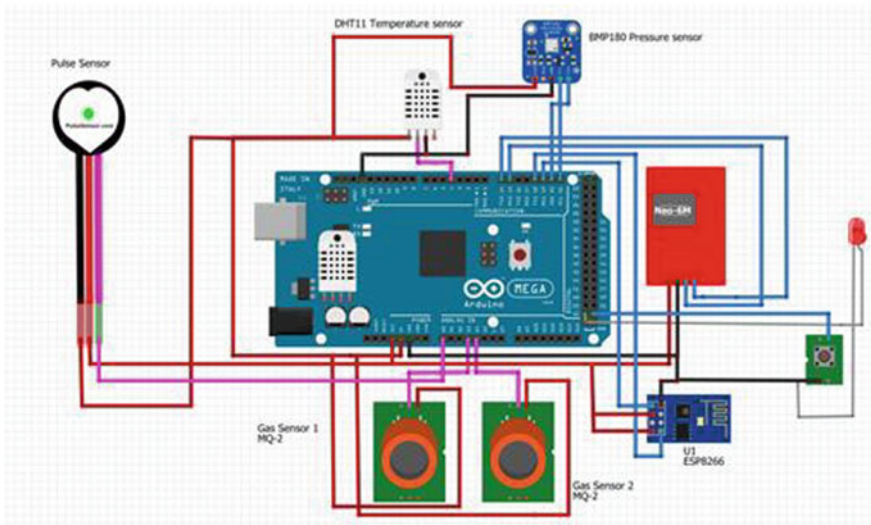
Fig. 2 Schematic diagram of proposed pro-active wearable gadget
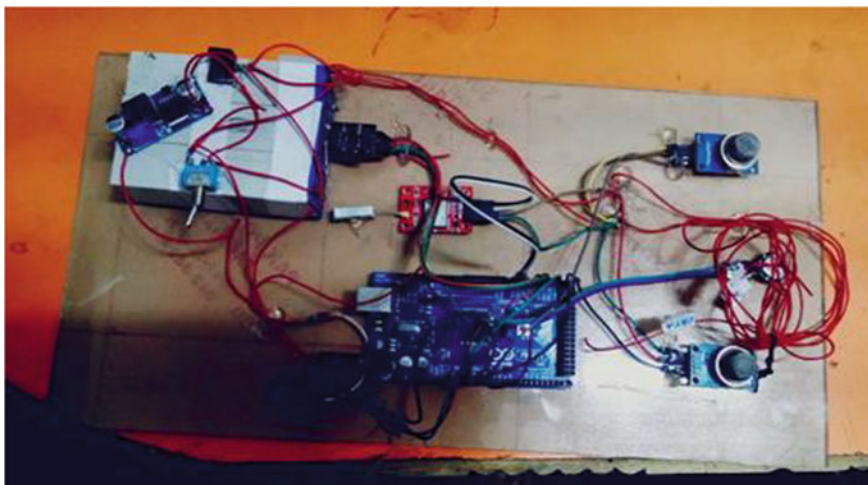
Fig. 3 Proposed pro-active wearable gadget
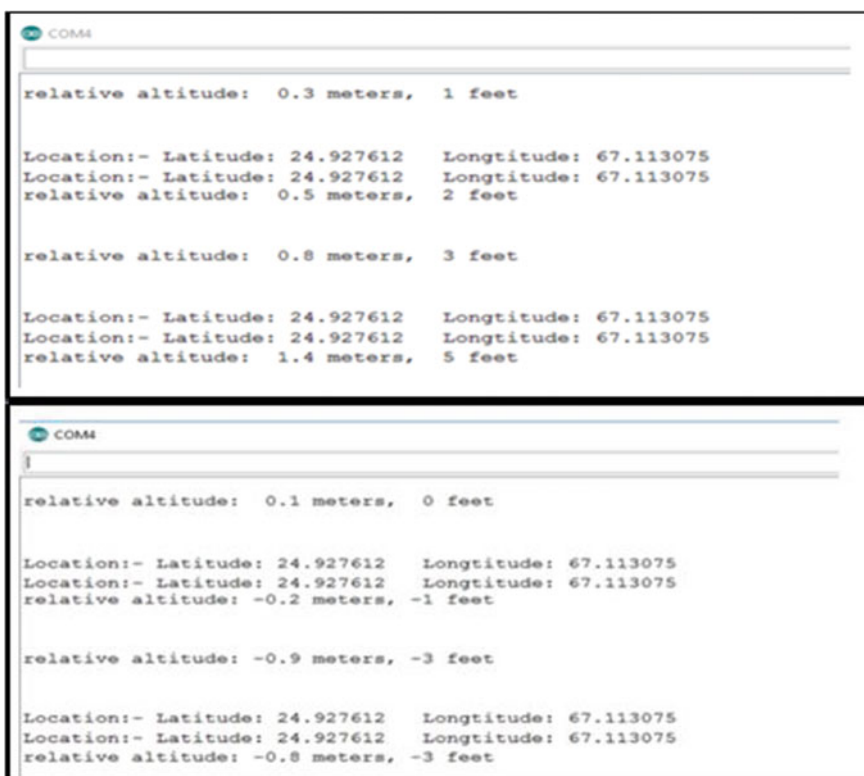
Fig. 4 All physical parameters and their display on serial monitor

Fig. 5 All physical parameters and their display on IP

Fig. 6 Display of location on LatLong.net via values acquired from device

Fig. 7 Display of an emergency query generated by error palette

the miner within the mines. The query will be also be generated on IP display unit as shown below in Fig. 7.

5 Conclusion

This manuscript provides a non-conventional strategy in the form of a pro-active device that aware almost every physical parameter within the mines that can affect the life of coal miners. In addition to this, provide an easy way to monitor the environment remotely from base station. Unlike other strategies, this prototype of pro-active wearable gadget provide a facility to miners as well to generate the emergency query so that if sensors are failed in recording any sort of uncertainty even then the rescue team and essential help may be obtained within short span of time.

The manuscript presents a system design that along with the detection of several parameters such that miner's pulse rate, availability of hazardous or poisonous gas, updating exact altitude and GPS positions make this subject one of its kind. The only thing due to which it is much more efficient and reliable is the pro-active facility provided within through which now the miner can also communicate the base station to ask for any sort of help in emergency or uncertain conditions. In future, the authors are trying to work out on data visualization using a dashboard to monitor the live moment for the coalminers.

References

1. Abro GEM, Shoaib AS, Safeeullah S, Ghulam A, Kun- dan K, Fiaz A (2018) Prototyping IOT based smart wearable jacket design for securing the life of coal miners. In: 2018 International conference on computing, electronics and communications engineering (iCCECE), IEEE, pp 134–137

2. Lukowicz P, Baker MG, Paradiso J (2010) Guest Editors' introduction: hostile environments. IEEE Pervasive Comput 9:13–15

3. Kwon GH, Smith-Jackson TL, Bostian CW (2011) Sociocognitive aspects of interoperability: understanding communication task environments among different organizations. ACM Trans Comput-Human Interac 18:1–21

4. Cernea D, Mora S, Perez A, Ebert A, Kerren A, Divitini M, Gil de La Iglesia D, Otero N (2012) Tangible and wearable user interfaces for supporting collaboration among emergency workers. In: Herskovic V, Hoppe HU, Jansen M, Ziegler J (eds) Collaboration and technology. Springer, Berlin, Heidelberg, pp 192–199

5. Divitini M, Farshchian BA, Floch J, Mathisen BM, Mora S, Vilarinho T (2012) Smart jacket as a collaborative tangible user interface in crisis management. In: Proceedings of the workshop on ambient intelligence for crisis management

6. Chen W, Nguyen ST, Coops R, Oetomo SB, Feijs L (2009) Wireless transmission design for health monitoring at neonatal intensive care units. In: Applied sciences in biomedical and communication technologies, ISABEL 2009. 2nd international symposium on, IEEE, pp 1–6

7. Özmen A, Tekce F, Ebeǧlu MA, Taşaltin C, Öztürk ZZ (2012) Finding the composition of gas mixtures by a phthalocyanine-coated QCM sensor array and an artificial neural network. Sensors Actuators B Chem 115(1):450–454

8. Markham A, Trigoni N (2012) Magneto-inductive networked rescue system (MINERS): taking sensor networks underground. In: Proceedings of the 11th international con ference on information processing in sensor networks, ACM, pp 317–328

9. Tanmoy M, Das PS, Mukherjee M (2011) Rescue and protection system for underground mine workers based on ZigBee. Int J Adv Comput Eng Architect 1:101–06

10. Reddy AH, Kalyan B, Murthy CSN (2015) Mine rescue robot system–a review. Procedia Earth Planet Sci 11:457–462

11. https://www.cdc.gov/niosh/mining/content/emergencymanagementan-dresponse/commtracking/advcommtrackingtutorial3.html

12. https://pdfs.semanticscholar.org/d60c/0fd4782ab327ff61625f0863cba0e20bf67f.pdf

13. Jing L, Yingchun S, Wenhui C, Xufang B (2012) Application research on gas detection with artificial olfactory system. Sensors Actuators 3:233–254

Design of a Self-Tuning PID Controller for a Temperature Control System Using Fuzzy Logic

Md. Tauhidul Islam, Ariful Islam, Rahul Kumar, Ghulam E Mustafa Abro, Sourav Majumdar, and Vipin Kumar Oad

Abstract Temperature plays a significant role in industrial realm, especially in metallurgy and chemical industries. As maintaining optimal conditions in manufacturing can make a huge difference to its profitability and efficiency. Without providing a proper temperature control, the entire production lines can fall apart. Every product needs to be manufactured in an individual and specific conditions. If the temperature is too high or too low then it can affect the output product, potentially it may ruin completely and leads to wasting of raw materials. To maintain the stability of temperature requiring in any situation a temperature control system is needed. Typically classical conventional PID controllers are being used but these controllers have some lack of efficiency in controlling. These controllers exhibit higher overshoot, long rise time and settling time. Self- tuning Fuzzy PID controllers are more efficient than those classical PID control- lers. In this paper we proposed a design of Fuzzy Logic based self-tuning PID controller to reduce overshoot and conquer long rise time and settling time. This design provides more suitable and effective control to the temperature control system in industries.

Md. Tauhidul Islam
Department of Electronics and Telecommunication Engineering, Chittagong University of Engineering and Technology, Chittagong, Bangladesh

A. Islam
Department of Fundamental and Applied Sciences, Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Perak Darul Ridzuan, Malaysia

R. Kumar (✉)
Department of Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Perak, Malaysia

G. E. Mustafa Abro
Department of Electrical and Electronic Engineering, Hamdard University, Karachi, Pakistan

S. Majumdar
Department of Electronics and Communication Engineering, Khulna University of Engineering and Technology, Khulna, Bangladesh

V. K. Oad
Department of Civil and Environmental Engineering, UTP, Perak, Malaysia

1 Introduction

Temperature control systems are being used in a wide variety of industries to manage
various operations or processing of manufacture. A temperature controller controls
temperature so that the process value matches the set point, but the response differs
due to the characteristics of the controlled object and the controlling method of the
temperature controller.

Various strategies can be exerted to the temperature control system, like PI, PD,
PID [1, 2], Artificial Intelligence (AI) [3], Fuzzy Logic [3], Genetic Algorithm (GA)
[1, 4], Fuzzy Self-tuning PID [3] etc. Recently, some strategies based on PID control-
ling and tuning method have been proposed in order to ameliorate system perfor-
mances. Gani et al. have proposed an ameliorated design using Genetic Algorithm
(GA) on optimal PID tuning of temperature system [1]. Salsbury proposed a feed
forward control method as a replacement of conventional PI feedback control [5].
Some has also done with self-tuning fuzzy PID controllers [3]. However those clas-
sical conventional PID controllers are less effective and some with modern strategies
like GA or AI, there is no doubt about their effectiveness but these strategies are too
hard to implement and costly too. Among these Self-tuning Fuzzy PID controller's
algorithm is the easiest strategy to implement and it has better effectiveness [4,
6]. This paper predominantly focuses on designing a new self-tuning Fuzzy PID
controller to maintain required temperature.

The proposed design of self-tuning Fuzzy PID controller is mainly based on Fuzzy
Interference System (FIS). The procedure of fuzzy interference system includes three
phases: (i) Fuzzification: In this phase, crisp inputs are transposed into degrees of
membership. The degree of membership is discerned by plugging crisp inputs into
the membership function affiliated to the fuzzy set. (ii) Rule evaluation: In this phase,
each and every fuzzy rule is imposed with a strength value. The strength is discerned
by the degrees of memberships of the crisp inputs in the fuzzy sets of preceding
portion of the fuzzy rules. (iii) Defuzzification: In this phase the fuzzy outputs are
transposed into crisp outputs [7, 8] (Fig. 1).

2 Model of the System

In this paper we are modeling a temperature control system of an industry.
Considering a first order transfer function for the controlling system:

$$G(s) = \frac{K}{Ts + 1} e^{-\tau s} \tag{1}$$
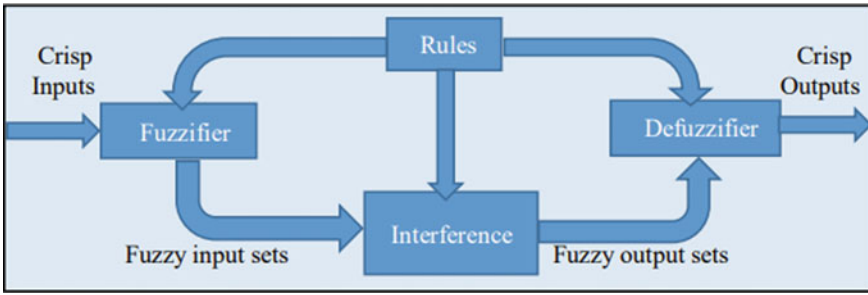
Fig. 1 Block diagram of fuzzy interference system

where, $K = 1$ is defined as the static gain and $\tau = 2.8$ as pure lag time, $T = 3$ as time-constant of the controlled object [9, 10].

2.1 Conventional PID

PID controller is vastly used in every controlling system. The first letter of the name of three controlling terms make up unitedly PID. It is actually a total combination of proportional, integral and derivative terms of controlling system. Here PID controller is used for controlling initially for the system. Equation of conventional PID can be expressed as follows:

$$u(t) = K_P e(t) + K_i \int e(t) + K_d \frac{de(t)}{dt} \qquad (2)$$

where, K_P is proportional gain, K_i is integral gain, K_d is derivative gain and e is error present in the controller. These are the tuning parameters of the controller (Fig. 2) [11, 12].
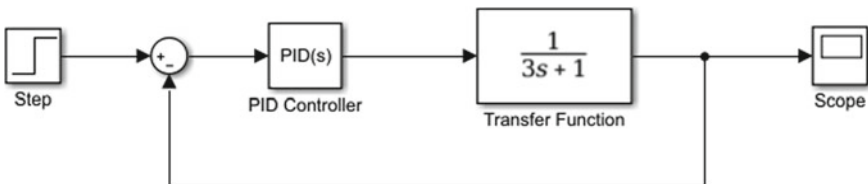
Fig. 2 Block diagram of conventional PID
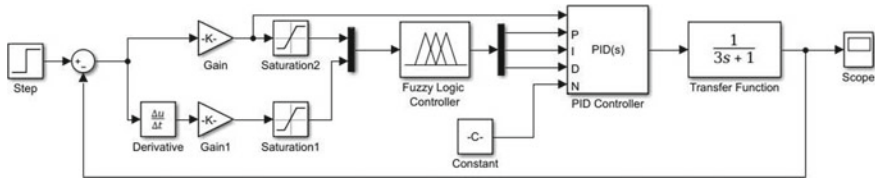
Fig. 3 Block diagram of self-tuning -fuzzy PID

2.2 Self-Tuning Fuzzy PID

The block diagram of the proposed system is shown in Fig. 3. The system is designed in MATLAB through Simulink. Firstly, a Fuzzy Interference System is designed by defining membership functions, universe of discourse and fuzzy rules in MATLAB through Fuzzy Logic Toolbox. Then this FIS system has been implemented in the de- signed block system in Simulink.

In this paper Mamdani method of Fuzzy Interference system (FIS) is used to process the fuzzy system. In MATLAB, there is a tool named Fuzzy logic toolbox which is used to create membership functions, universe of discourse and fuzzy rules. Here, in this interference system, the crisp inputs are system error (e) and rate of system error changing (ec) and the crisp outputs are proportional gain (K_P) integral gain (K_i) and derivative gain (K_d). The range for the system error (e) is taken as [−1 1], rate of change of error (ec) is [−1 1] and range for the output is considered as proportional gain (K_P) at [1 5], integral gain (K_i) at [0 1] and derivative gain (K_d) at [0 1.5]. These ranges are obtained from the observations of conventional PID controller since the conventional tuning gives as K_P 1.285, K_i as 0.27 and K_d as 0.9 (Fig. 4) and (Table 1).

3 Result and Analysis

The proposed design has been simulated in MATLAB through Simulink. Figure 5 shows the step response of both the conventional PID controller and proposed PID controller. The step response is determined at the step input temperature of 500 °C.

The comparison of performance and robustness between these two controller is shown in Table 2.

The analysis of the simulation result and comparison table exposes that the proposed self-tuning controller overshoots 0.63% with peak value of 503.16 whereas the conventional PID controller overshoots 13.82% with peak value of 569.13. The proposed method reduced the overshooting problem as well as providing better dynamic performance and robustness. So, It is clear that the proposed self-tuning Fuzzy PID controller for temperature control system provides better efficiency in controlling.
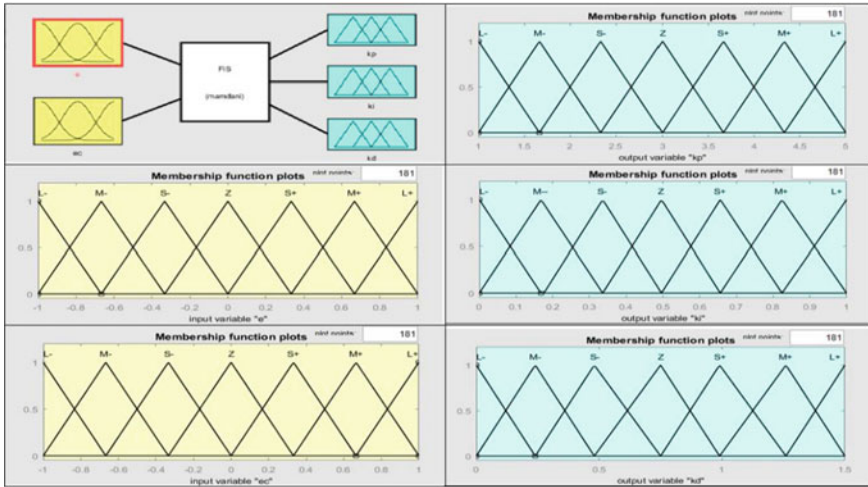
Fig. 4 FIS system and input–output membership functions

Table 1 Fuzzy rules

e		ec						
		L−	M−	S−	Z	S+	M+	L+
L−	Kp	L+	L+	M+	M+	S+	S+	Z
	Ki	S+	S+	Z	Z	Z	L+	Z
	Kd	L−	L−	L−	M−	M−	Z	Z
M−	Kp	L+	L+	M+	M+	S+	Z	S−
	Ki	S−	S−	S−	S−	Z	Z	M−
	Kd	L−	L−	M−	M−	S−	Z	Z−
S−	Kp	M+	M+	M+	S+	Z	S−	M−
	Ki	L−	L−	M−	S−	Z	S+	M+
	Kd	M−	M−	S−	S−	Z	S+	S+
Z	Kp	M	S+	S+	Z	S−	M−	M−
	Ki	L−	M−	M−	S−	Z	S+	M+
	Kd	M−	M−	S−	Z	S+	S+	M+
S+	Kp	S+	S+	Z	S−	S−	M−	M−
	Ki	L−	M−	S−	S−	Z	S+	S+
	Kd	S−	S−	Z	S+	S+	M+	M+
M+	Kp	S+	Z	S−	M−	M−	M−	L−
	Ki	M−	S−	S−	S−	Z	S+	S+
	Kd	Z	Z	S+	M+	M+	L	L+
L+	Kp	Z	S−	S−	M−	M−	L−	L−
	Ki	S+	Z	Z	Z	Z	L+	L+
	Kd	Z	S+	S+	M+	L+	L+	L+

Where (L−) = Negatively large, (M−) = Negatively Medium, (S−) = Negatively Small, Z=Zero, (S+) =Positively Small, (M+) = Positively Medium, (L+) = Positively Large. And these are the linguistic variables.
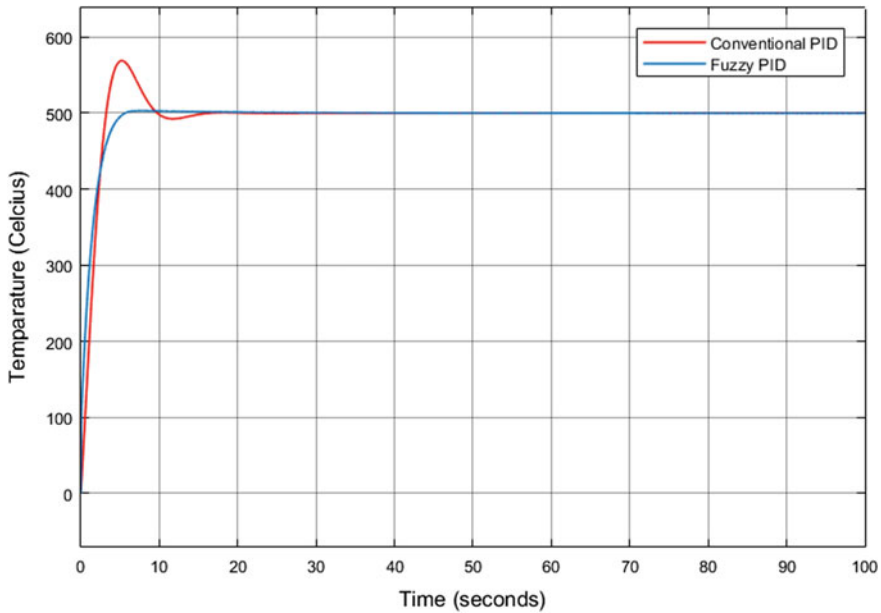
Fig. 5 Response graph after simulation

Table 2 Comparison table of performance and robustness

Parameter	Conventional PID	Fuzzy PID
Rise time	2.43 s	2.99 s
Settling time	8.87 s	4.56 s
Overshoot	13.82%	0.63%
Peak	569.13	503.16
Peak-Time	5.21 s	7.66 s

4 Conclusion

In this paper a self-tuning fuzzy PID controller has been designed for controlling temperature system for industrial management. Fuzzy interference system (FIS) has also been described elaborately in this paper. After simulating the system in MATLAB, the results expose that this self-tuning fuzzy PID has almost zero overshoot with a better dynamic performance than conventional PID. Analyzing the proposed algorithm and results of this paper, it is concluded that fuzzy PID is more effective and it can be replaced with conventional PID. This system can be ameliorated by farther modification in the designing algorithm like membership functions, rules and gains which impacts the controlling performance by extenuating both the overshoot and dynamic response.

References

1. Gani MM, Islam MS, Ullah MA (2019) Optimal PID tuning for controlling the temperature of electric furnace by genetic algorithm. SN Appl Sci 1:880
2. Anwar AZ, Jabeen B (2019) Prototyping non-holonomic hovercraft for path planning and obstacle avoidance. Sir Syed Univers Res J Eng Technol 9(1)
3. Baogang H, Hao Y (2001) Review of fuzzy PID control techniques and some important issues. Automatic Sinica 27(4)
4. Rehman NU, Kumar K (2018) Implementation of an autonomous path planning and obstacle avoidance UGV using SLAM. In: 2018 International conference on engineering and emerging technologies (ICEET), IEEE, pp 1–5
5. Salsbury TI (1998) A temperature controller for VAV air-handing units based on simplified physical models. HVAC and R Res
6. Abro GEM, Jabeen B, Ajodhia KK, Rauf A, Noman A (2019) Designing smart Sewerbot for the identification of sewer defects and blockages
7. Hassan MMM (2013) Current studies on intrusion detection system, genetic algorithm and fuzzy logic. Int J Distrib Parallel Syst (IJDPS) 4(2)
8. Mugisha JC, Munyazikwiye B, Karimi HR (2015) Design of temperature control system using conventional PID and intelligent fuzzy logic controller. In: Proceedings of 2015 international conference on fuzzy theory and its applications (iFUZZY) The Evergreen Resort Hotel (Jiaosi), Yilan, Taiwan, November 18–20
9. Abro GEM, Bazgha J, Manan A (2019) Stabilization of Non-ho- lonomic 03 dof hovercraft using robust RST control design. Sukkur IBA J Emerg Technol 2(1): 45–50
10. Diwakar TK, Vivek Y, Raut KH (2014) PID tuning rules for first order plus time delay system. Int J Innov Res Electri Electron Instrument Control Eng 2(1)
11. Wu X, Wang X, He G (2020) A fuzzy self-tuning temperature PID control algorithms for 3D bio-printing temperature control system. In: Chinese control and decision conference (CCDC) by IEEE. ISSN: 1948–9447
12. Jiang W, Jiang X (2012) Design of an intelligent temperature control system based on the fuzzy self-tuning PID. In: International symposium on safety science and engineering in China, 2012 (ISSSE-2012), Procedia Engineering, vol 43. pp 307–311

Design of a Low-Cost High-Speed and Large-Capacity Data Logger Using MicroSD Card

Muhammad Sajjad, Mohd Zuki Yusoff, and Junaid Ahmad

Abstract Data logging is an important process for evaluating the functionality and performance of many electronic systems such as internet of things (IoT). Most of the mixed-signal systems digitize the signals from the sensors to generate input signals and then perform some digital signal processing on these input signals to generate output signals. Data logging of these input and output signals helps to evaluate the performance of the systems. Many of these systems contain an FPGA for digital signal processing. The FPGA can also be used to facilitate data logging. This paper presents a data logging and data extraction system which is a high capacity, low cost and ensures data integrity over gigabytes of the data. The system has been evaluated by logging data in a system, having four inputs and one output signal each having a sampling rate of 1 kHz, for five hours. Data integrity has been observed for each sample of the data.

Keywords MicroSD · VerilogHDL · SPI interface · FPGA · State machine · MATLAB · Data logging

1 Introduction

Sensors are an integral part of electronic systems. Systems such as IoT can have a lot of sensors. They sense some physical quantities and convert them into electronic signals. After digitization, these signals are passed to signal processing modules as input signals. These modules perform digital signal processing on the input signals and generate some output signals. Most of the electronics systems contain

M. Sajjad (✉) · M. Z. Yusoff
Universiti Teknologi PETRONAS, Perak, Malaysia
e-mail: muhammad_19001746@utp.edu.my

M. Z. Yusoff
e-mail: mzuki_yusoff@utp.edu.my

J. Ahmad
Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Swabi, Pakistan

programmable logic devices like FPGA as a signal processing module. Data logging of the input and the output signals is necessary for evaluating the performance of the systems and verifying the functionality of the signal processing techniques in real-time. In case of reliable data logging, the inputs can be passed to verified high-level software routine to generate new signals. These signals can be compared to the recorded output signals for verifying the functionality of the signal processing modules.

One way for data logging is writing data to NOR-flash, NAND-flash, or EEPROM which are normally present in the system electronics [1]. But these memories are small in capacity and can hardly store a few megabytes of data [2, 3]. So these devices are not suitable for data logging in high throughput systems. They also require erasing before writing which reduces their data writing speed and add another task for the data logging controller. Normally these devices are mounted on PCBs by the soldering process and cannot be separated from the electronic systems. Because of this, the electronics systems have to be moved in the lab to recover logged data for offline analysis. Another way is using commercial off-the-shelf (COTS) data recorders [4]. These data recorders can store data in gigabytes and are interfaced with RS-232/RS-422 ports [5]. They record data to file system of the attached USB flash or microSD card. Main drawback of these COTS recorders is that they take electrical power and some space and hence are not a suitable choice for critical missions.

The proposed data logging system consists of a microSD card connected to four general-purpose digital pins (for implementing SPI) of a FPGA-based main processing board. Raw data from the sensors and the processed data are written on sectors of the microSD card. The digital design architecture ensures recording of each sample, handling of contention arising from simultaneous read and write to the buffers and matching of cyclic redundancy checks (CRC) during write operations to the microSD card. Most of the existing data loggers use parity bit for error detection which only applies to byte level errors in contrast to CRC which applies to the whole data of the sectors. The data logging system has been tested for five hours duration and no missing or corrupted data have been observed.

The paper is organized as follows. Section 2 discusses SD card Interfacing. Section 3 presents implementation detail, followed by the Data and Information Retrieval in Sect. 4. Results and Interpretation is presented in 5. Finally, the conclusion is presented in Sect. 6.

2 SD Card Module

The Secure Digital card is widely used for data logging systems. It has flash memory and a microcontroller embedded in it. The flash memory control operations such as writing, reading, and erasing are performed inside the SD card. The data transfer from the host system to an SD card is in the form of data blocks which are in units of 512 bytes. Two modes can be used for data communication between a host controller and an SD card i.e. SD mode and SPI mode. By default, the SD card is

in SD mode. Both modes have their pros and cons. SPI communication protocol is a bit simpler as compared to its default mode because the SD card can be attached to any host controller using some GPIO pins. Therefore, this mode is most suited for cost-effective embedded systems. Moreover, the SD card requires two different clock frequencies for its initialization mode (100–400 kHz), and data transfer (up to 25 MHz) mode as specified in the SD card product manual whereas the commercially available loggers record data from a RS422/RS485/RS232 which can hardly manage data transfer up to 430600 Hz. Keeping in view all the challenges of the SD card, a state machine has been designed to cater to the data archival on the microSD card. After powering up the system, configure SPI clock frequency between 100–400 kHz and put MOSI and CS pins to the high state. Send 74–80 clock pulses to the serial clock pin of the SD card to make it ready for communication. Put CS to low and send CMD0 to reset the card. The card will respond with 0×01 when it will be in the idle state. After the idle state is achieved, send CMD8 with proper argument and CRC before the start of the initialization process. The proposed data logging system architecture comprises of a control unit, a microSD card module, two multiplexers, and two block RAMs.

2.1　Control Unit

A Control unit is like a brain for the data logging system. A state machine is designed to control operations. The flow chart for the state machine is shown in Fig. 1. At first, the control unit will get the SD card ready flag from the SD card module to start its operation. After the SD card is ready, the control unit will cater to the data flag from sensors. After getting the data flag, the Control Unit will provide an 8-bit address to a Buffer and store the sensors data on Block RAM. When block RAM is full, the control unit will generate a RAM full flag to the SD card module to store the respective RAM data on the sector of the SD card. The control unit also stops the data writing process by providing an End data transmission signal to the SD card module.

2.2　Block RAM

Single port Block Rams are implemented using IP cores. Each Block RAM has a size of 500 bytes, having a 9-bit address, 8-bit data-in, 8-bit data-out, and its Read–write operation is controlled by the control unit. The reason for its size of 500 bytes size is that the sector length of the SD card is 512 bytes and the data packet after each 1 ms is 25 bytes. The reason for using two block RAMs is that when the first RAM is full and is being dumped on the SD card, the incoming data from sensors is stored on the second RAM to avoid any data loss. Write address as it will take data from sensors, and then write it on RAM. Whereas the address from the SD card is named as the
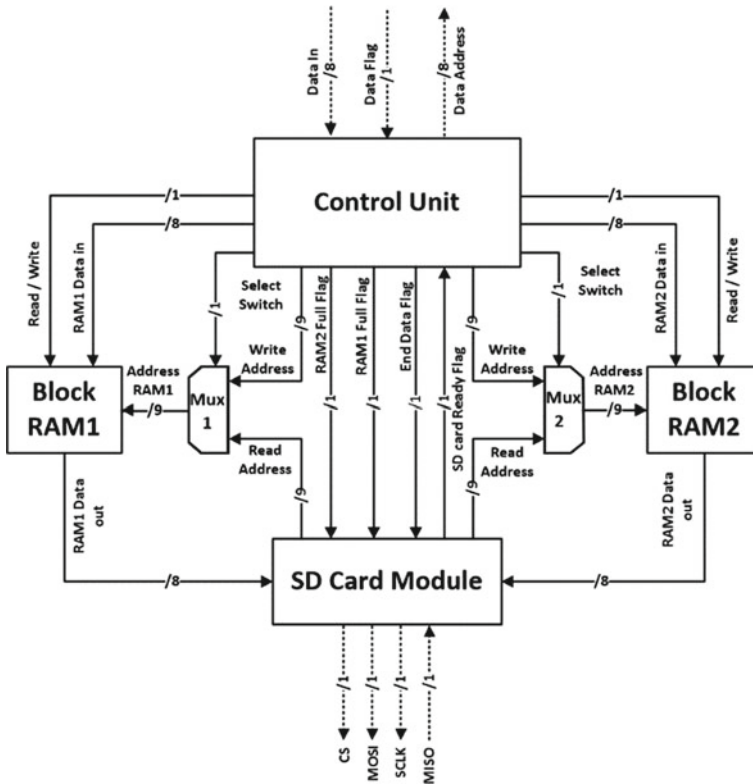
Fig. 1 Architecture of the data logging system

Read address as it will read the stored data from RAM. The initialization process of the card will start when CMD1 is received by the card. This initialization process will end when the response is changed from 0×01 to 0×00. After initialization data block length is set in units of 512 bytes by sending CMD16 to the card. CMD25 is then sent to the card to configure it for multiple block writing. After the command is accepted by the card, the host can send data packets to the card.

3 Implementation

Data logging on a PC using serial cables is not possible in remote locations or for the case of a drone flight. Acquiring real-time data in such a scenario is a tedious task. The proposed data logging system is Implemented using four sensors connected to the 4 channels of a 24-bit ADC, which is interfaced with an FPGA using SPI protocol as shown in Fig. 2. The data from sensors is stored in an internal buffer of FPGA. When the buffer is full, it generates a trigger to inform the control unit about
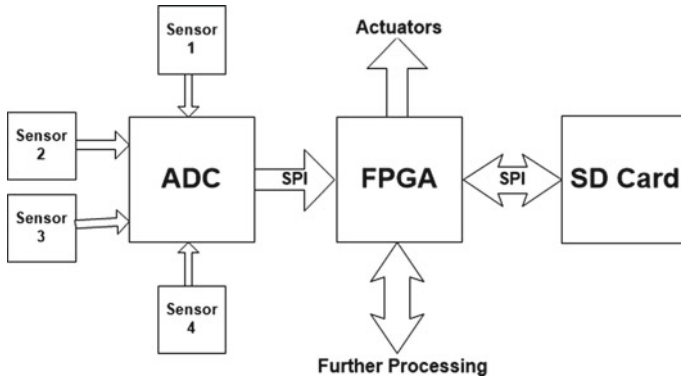
Fig. 2 Experimental setup of the system

data availability. The reason for using a buffer is that its access time is very small as compared to external memories. Data is also available to actuators to perform necessary actions and is then stored on the SD card for offline analysis. This data from sensors can also be made available for further processing.

S1(n), S2(n), S3(n), and S4(n) represents the sensor outputs of first, second, third and fourth sensor respectively (shown in Fig. 2), so, S(n) is the output of algorithm which is running on these sensors' outputs. The algorithm in this proposed system is just an integration of all sensors. Hence, every sample of sensors data is used in this integrator algorithm ensuring data integrity. The equation for this integration-based algorithm is (1).

$$S(n) = S(n - 1) + S1(n) + S2(n) + S3(n) + S4(n) \tag{1}$$

As ADC is a 24 bit, so each sensor's data is a 24-bit data. After applying a scale factor on the sensors data, it becomes a 32-bit data i.e. 4Bytes. So, there are 16 Bytes of sensors data, and 4 bytes of last iteration summation result hence, making it a data of 20 Bytes. This data is then framed by a start byte, 2 bytes of checksum, and 2 bytes of stop bytes to make it a data packet of 25 Bytes. After 1 ms, a data trigger is generated that store these 25 bytes on Block RAM. Block RAM is 500 Bytes, so after 20 ms Ram is full and is then ready to dump its data on the sector of the SD card. The last 12 locations of each sector of the SD card are filled with 0xFF. Sensors data is written in LSB to MSB format. Clock frequency for initialization and data transfer mode has been set to 384.615 kHz in its native mode so it takes approximately 1.33 ms to write 512 bytes on a sector. After every 20 ms, a sector is written on the SD card. So, in 5 h duration, it writes on 900,000 sectors of an SD card. Each sector is of 512 Bytes, so, the total data is 450 MB.

4 Data and Information Retrieval

For missions involving high data rates for a long duration, data logging in a single file is not a feasible solution because the software used for post-processing or offline analysis crashes due to limited memory. To avoid this situation, some computer programs like hex editors can be used. With the help of MicroSD Card Reader, data written on the sectors of the microSD card can be viewed and manipulated. A data file can be created by specifying addresses for both the start sector and the end sector. In this way, multiple files, having a manageable size for post-processing software, can be generated through a hex editor. These files contain sector addresses and bytes of data written by data logging setup. So, we cannot use them without further processing. To make the files usable for analysis, these files can be further processed by specially designed software routines on Integrated Development Environment (IDE) like Microsoft Visual Studio or LabVIEW. Such routines can process these files on the byte level. So, tasks like checksum matching, removing header or footer, and concatenating relevant bytes to make a float, double, or integer type data values can be performed easily.

5 Results and Interpretation

For example, Fig. 3 shows a comparison of offline and online (or real-time) performance of the algorithm and hence the electronic system. Figure 3 also shows outputs
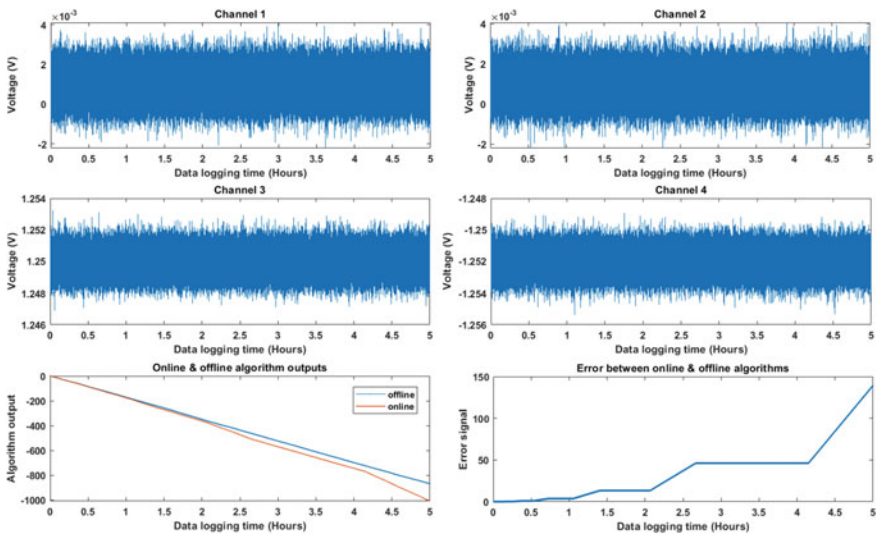
Fig. 3 Online and offline algorithm comparison

of the same algorithm running in MATLAB (offline mode) and the FPGA (online or real-time mode). This algorithm has been fed with the same data from the sensor and ideally the graphs in Fig. 3 should track each other. But this type of mismatch does not imply any data logging or algorithm error. This kind of deterministic error is due to different type of truncation modes in double-precision floating-point computations in MATLAB and FPGA. Since the algorithm consists of integration, which sums up all the errors arising in the whole mission, so the error tends to increase with time.

6   Conclusion

This paper has presented a data logging setup for recording digitized input and output signals in an electronic. The presented method ensures that every sample of the signals is intact even for the missions with data size reaching to gigabytes. By using the proposed data logging setup, large-capacity data can be recorded with full integrity at a very low cost. This can help to speed up the evaluation and development of IoT.

References

1. Kim GJ, Baek SC, Lee HS, Lee HD, Joe MJ (2006) LGeDBMS: a small DBMS for embedded system with flash memory. In: VLDB, CiteSeer, pp 1255–1258
2. Lee SW, Moon B (2007) Design of flash-based DBMS: an in-page logging approach. In: Proceedings of the 2007 ACM SIGMOD international conference on management of data, pp 55–66
3. Kim J, Kim JM, Noh SH, Min SL, Cho Y (2002) A space-efficient flash translation layer for compactflash systems. IEEE Trans Consum Electron 48(2):366–375
4. De-Leon HL, Quiros RE (2005) Self-contained flight data recorder with wireless data retrieval. United States Patent 6,898,492
5. Yamawaki Y (1995) Traffic accident data recorder and traffic accident reproduction system. United States Patent 5,446,659

Machine Learning in Healthcare: Current Trends and the Future

Usman Ahmad Usmani and Jafreezal Jaafar

Abstract Today, an abundance of electronically stored medical image data and DL algorithms can be used to recognize and detect patterns and anomalies in this kind of dataset. Computers and algorithms can interpret the imaging data as a very qualified radiologist can see irregular skin, lesions, tumours and brain bleeds. Consequently, the use of AI/ML tools/platforms to help radiologists is poised to grow exponentially. This approach addresses a vital issue in the healthcare sector as well-trained radiologists are challenging to come by worldwide. These professional professionals are, in most cases, under tremendous pressure due to the influx of digital medical data. We analyze and address the current state of A.I. applications in healthcare. A.I. can be applied to various healthcare data forms (structured and unstructured). Popular A.I. techniques include machine learning for structured data such as classic support vectors and neural networks, modern in-depth learning unstructured data natural language processing.

Keywords Healthcare trends · Machine learning tools · ML disease detection · Robotic surgery

1 Introduction

Identifying diseases rare or hard to diagnose also depends on detecting what are known as' edge cases. As this type of ML system relies on large datasets of raw pictures (and different transformations) of these conditions, they are more precise than humans for this type of detection [1]. They should improve automation productivity and intelligent decision making in primary/tertiary and public health care. This would have the most significant effect on A.I. instruments since the quality of life will increase for thousands of people worldwide [2].

U. A. Usmani (✉) · J. Jaafar
Universiti Teknologi Petronas, Seri Iskandar, 32610 Perak, UTP, Malaysia
e-mail: usman_19001067@utp.edu.my

The Microsoft Project Inner Eye, which uses ML methods for segmentation and detecting tumours with 3D images, is an excellent test case. It can assist in the pre-case preparation, navigation and effective tumour contours for radiation preparation. M.R.I.s and other sophisticated imaging systems are fitted with ML algorithms, increasingly used for early cancer detection. ML tools often offer significant value by increasing the surgeon's display with information such as cancer's location during robotic procedures and other image-led procedures. Therefore, the use of AI/ML software/networks to help radiologists should be exponentially expanded. Unfortunately, these details are messy and unstructured too. In particular, patient data can not be used for basic computational and analytical modelling instead of regular business transaction data [3]. The need for the hour is a powerful and agile AI-enabled platform, which can link and analyze a diverse mix of data types (i.e. blood pathology, genomics, radiological images, medical history) [4]. These systems should also dig extensively and uncover the secret trends through the analyses [5]. Furthermore, doctors and other health practitioners should understand and understand their findings to achieve their success with great confidence and clearness. These bills are appropriate for the interpretable A.I. and distributed ML systems and can fulfil these systems' requirements within a short period. Surgery robots can provide human surgeons with unique assistance, enhance their vision and manoeuvrability during operations and make precise and minimally invasive incisions [6]. The optimal stitch geometry and wound cause less pain. There are inspiring opportunities to apply AI/ML to such robots in digital operations.

1.1 The Motivation for ML in Healthcare

The following article provides a detailed summary of this. Software-centred robot communication with the aid of massively distributed computing. Insights and guidance based on the history of Surgery (personal and machine) and findings (favourable or not). A.I. has built space for augmented reality guidance and guidance in real-time. Telemedicine and remote surgery prospects for reasonably easy operations. Daily citizens in the United States have had a long and bitter debate about actual health care costs and complexities. A.I. and similar methods powered by data are unique to solve some of the root causes of the problems—long waiting, concerns of unreasonable charges, a long, overly complicated appointment process, and lack of access to the right physician [7]. The best thing about this kind of A.I. application is that data security is not a big issue: a daunting and challenging healthcare system problem [8]. The organizational challenge often contains no sensitive patient data concerning illness, diagnosis or medicine. Still, it consists of data concerning funding, money, marketing or human resources concerns as is the case for any modern organization [9, 10]. These same problems plague conventional organizations for several decades, and AI/ML techniques are now part of a solution. This is because massive databases and smart search algorithms, which are a vital part of A.I. systems, are excellent for specific patterns or optimization problems. Hospitals and public health agencies

should also use advanced AI/ML technologies and strategies in their organizational aspects every day. Such programmes' core purpose should be to build AI-supported networks to improve healthcare outcomes for the most extensive section of ordinary citizens. The ultimate aim of current processes in conventional organizations is to maximize benefit. Robust A.I. methods for managing medical operations need to differentiate themselves from traditional systems by mixing empathy with income. Increasingly, important names in the pharmaceutical industry are selecting A.I. and ML technologies to resolve the challenging issue of useful drug testing [11].

1.2 Artificial Intelligence Techniques for Detection of Diseases

This article offers several famous examples—like Sanofi, Genentech, Pfizer. These case studies cover all sorts of clinical fields—metabolism, cancer therapies, immunooncology. Beyond the traditional long-term method, A.I. techniques are being used more and more to speed up the fundamental processes of early selection and mechanism discovery. The Berg biotechnology company uses its A.I. method for study in its research. For start-ups, the evaluation of multi-channel data, including research papers, patents, clinical trials and patients records, may include state-of-the-art inferential approaches, Markov chain models, enhanced learning and natural language (N.L.P.s). The key objectives are to identify patterns and establish high-dimensional representations stored in the cloud and used during drug discovery [12, 13]. Also, A.I. instruments can help shape why and under what conditions diseases are more likely to arise. Physicians can direct and plan to intervene (personalized) even before they begin to display symptoms. The aim is extraordinarily complicated and challenging. It is to identify exact treatment options for an individual based on an ongoing personal medical history, decisions about behaviours, genetic data and pathological tests. The ongoing COVID-19 crisis has shown the value of performing hundreds of simultaneous development and therapeutic trials of vaccines. Traditional statistical modelling approaches are almost impossible to use, designed for small experiments to collect data and distinguish patterns from all these various sources, often delivering high uncertainty levels. For such a planetary scale, A.I. techniques must bear the challenge of solving. A wide range of exciting and forward-looking approaches to AI/ML and platforms have been investigated in the field of health care. Sujets from radiology assistants to intelligent medical management, personalized medicine and digital public health monitoring have been explored [14–16]. Data protection issues and legal frameworks are still recognized as barriers to the complete adoption of these systems. The type of data that third parties can access and use legally (for example, the owners of the A.I. and ML software, physical devices or platforms) may be quite challenging to decide. As technologists and IA/ML practitioners, it is essential to have a bright future in which the AI-algorithms support ordinary citizens' trillions to improve their primary health and wellbeing.

2 ML Potential Applications in Healthcare

With the rapid population growth, it seems impossible to record and evaluate large patient information quantities. Machine learning gives us the ability to automatically identify and process this data, making the healthcare system more complex and resilient. Machine education in healthcare combines into a common thread two types of fields: informatics and medical science [17–19]. Machine learning methodology gives medical research development and analyses complex medical data for further study.

Many researchers are working to develop new dimensions and features in this area. Google recently developed an algorithm of learning machines to identify cancerous tumour mammograms. Stanford also has a deep learning algorithm to assess cancer of the skin [20, 21]. There are annual conferences, such as Machine Learning for Healthcare, which seek to explore new automated medical technologies for better service [22, 23]. Machine education aims to increase the machine's prosperity, efficiency and reliability. However, the machine learning method in a health system is the doctor's brain and experience what a patient requires a human touch and concentration [24, 25]. This can not be replaced by machine learning or any other technology. An automated computer can provide better service. Here are the top 10 machine learning frameworks for healthcare.

2.1 ML-Based Diagnosis for Heart Diseases

The heart is one of the primary organs of our body. We also have several heart disorders like Coronary Artery Disease (CAD), CHD, etc. (Fig. 1).
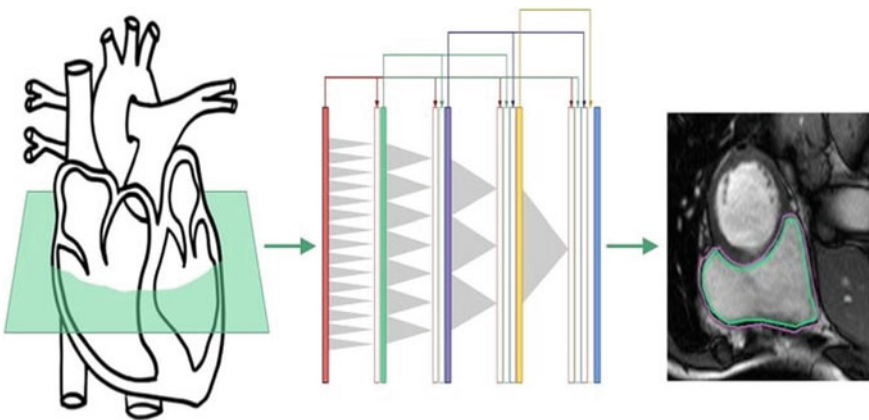
Fig. 1 Diagnosis of heart disease

Many researchers work on algorithms for the diagnosis of cardiovascular disease. It's a really hot problem of research in the world. One of the main advantages of machine learning in health care is an integrated cardiovascular diagnostics method [26–28]. Researchers use various tracked machine learning algorithms as a teaching algorithm, such as the Support Vector Machine (SVM) or Naive Bayes. The Dataset for U.C.I. Heart Disease can be used for training or assessment. For data analysis, the WEKA data mining tool can be used. Alternatively, cardiac disease's diagnostic method can be established using the Artificial Neural Network (ANN) approach. One of the most common and dangerous diseases is diabetes. This is also a significant cause of other extreme disorders and deaths. This condition can affect our different body parts, including the heart and kidney nerves.

2.2 ML-Based Prediction of Diabetes

A machine learning approach aims to predict diabetes and early-stage patients. Random forest, K.N.N., Decision Tree and Naive Bayes can be used to construct a prediction system as a classification algorithm. Naive Bayes beats other accurate algorithms (Fig. 2).

Because of its excellent efficiency and less time for estimation, the diabetes dataset can be downloaded from here. It comprises 768 data points, each with nine functions [29, 30]. The liver is the second most crucial internal organ in our body. It plays a significant role in metabolism. Many liver diseases such as cirrhosis, chronic hepatitis, liver cancer can be targeted (Fig. 3).
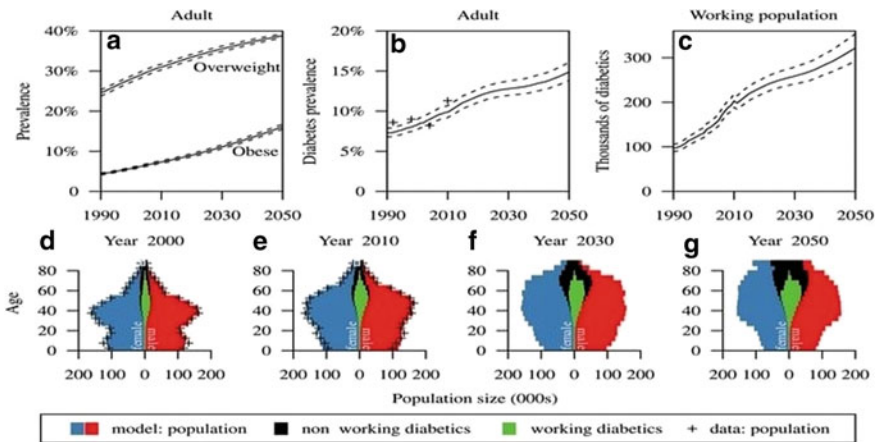
Fig. 2 Prediction of diabetes
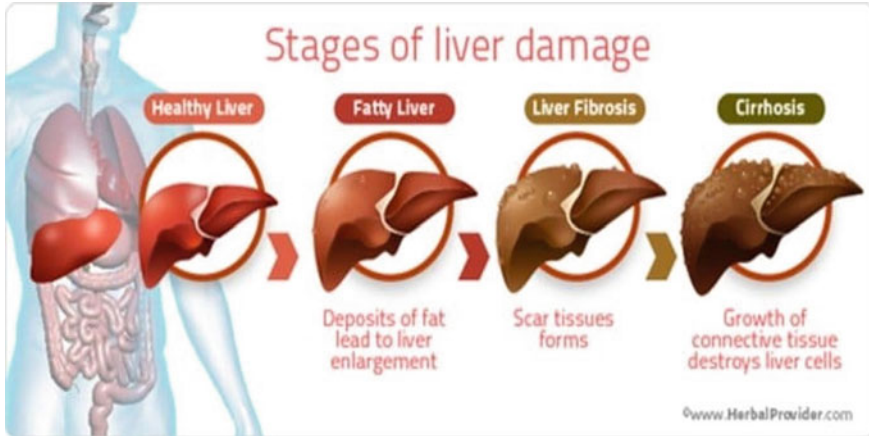
Fig. 3 Liver disease prediction using ML techniques

Recent concepts have been used significantly in machine learning and data mining to prevent liver disease. Disease with extensive medical records is challenging to predict.

However, researchers use machine learning principles like grouping, clustering and many more to solve this problem. Dataset (I.L.P.D.) can be used for hepatic disease prediction. This dataset comprises ten variables or liver diseases. A classifier can be used with the Vector Machine Support (SVM). You can use M.A.T.L.A.B. to create a prediction system for liver disease.

2.3 ML-Based Robotic Surgery

One of the primary qualifications for healthcare training is robotic Surgery. This application will soon be a promising field. This application can be divided into four subcategories, including automatic suturing, surgical capacity assessment, surgical materials, and surgical modelling. Automation of sutures minimizes machine life and fatigue (Fig. 4).

The Raven Surgical Robot, for example. Researchers attempt to use a machine learning approach to test the surgeon's performance in the least invasive robot procedure. University of California researchers, San Diego (UCSD), explore machine learning technology to improve surgical robotics. Since robots can not function efficiently in neurosurgery, the manual operational workflow requires time and can not automatically provide input. The use of machine learning is a benchmark application of machine learning in medicine in drug discovery [31]. The Microsoft Project Hanover aims to implement precise medicine learning machine technologies. Many Organizations are currently using drug discovery machine learning technology. Benevolent AI, for example. It aims to use artificial intelligence (A.I.) to

Fig. 4 Robotic surgery using ML

discover drugs. Machine learning has many advantages in this field, including accelerating the process and reducing the failure rate. Machine learning also optimizes the development and expense of drug discovery (Fig. 5).
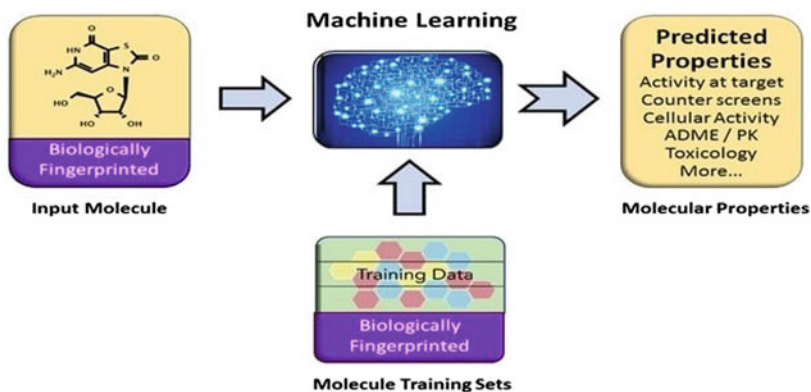
Fig. 5 Prediction and detection of cancer using ML

2.4 Machine Learning in Radiology

An intelligent electronic health record system can boost computer training, including classification documents and optical character recognition. This programme aims to create a plan to sort patient requests via email or convert a manual recording system into an automated system. This application seeks to develop a secure and easily accessible framework. Electronic health records' rapid growth has increased patient healthcare information storage and advancement. For example, duplicating data removes data errors. It could be used as a classifier or artificial neural network (ANN) to create an electronic health recorder managed learning algorithm, like the Support Vector Machine (SVM). Aidoc provides radiologists with machine learning software to speed up detection. The purpose is to interpret the medical picture to recognize the body's abnormalities understandably. The supervised algorithm for machine learning is used mainly in this field. Machine learning technology is used for medical image segmentation. Segmentation is the process by which structures are represented in an image. The graph cutting method is used primarily for image segmentation. For analysis, standard Radiology Text File Language Processing is used. The use of computer radiology can also improve patient care. The clinical trial may be a series of studies requiring responses for biomedical or pharmaceutical efficiency and safety. This research aims to reflect on the ongoing progress of treatment (Fig. 6).

It takes a lot of time for this clinical trial. In this area, machine learning has an enormous impact. An ML system can provide full service and real-time monitoring. The advantage of using machine learning technology is that it can be remotely controlled in clinical trials and studies. Machine learning also provides patients with a healthy clinical environment. The clinical trial's efficiency can be improved with supervised computer education in healthcare.
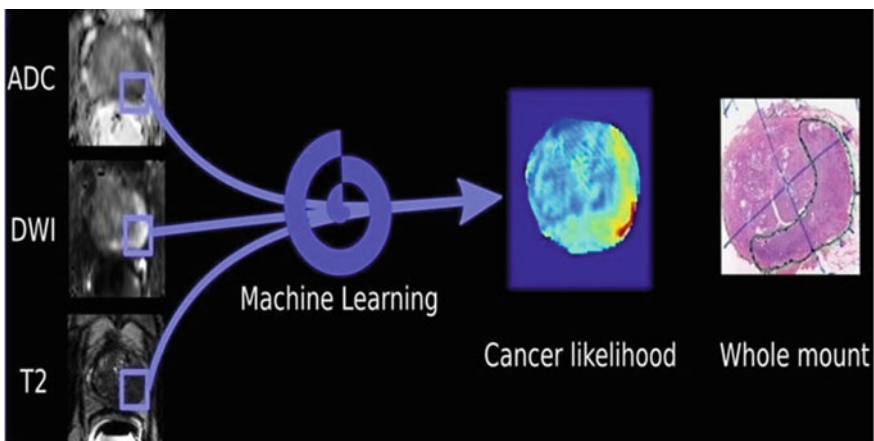
Fig. 6 Machine learning in radiology

Regarding machine learning performance, more knowledge almost always provides better outcomes, and the health sector relies on multiple details. McKinsey aims to deliver as much as $100B a year for big data and machine learning in pharmaceutical and medicine focused on improved decision making, optimized creativity, research/clinical testing and emerging technology for doctors, consumers, insurers and regulators.

What is the source of this fundamental knowledge? If we look at named data points, we will see research and growth, physicians, nurses, patients, caregivers, etc. The range of various sources (at present) forms part of the synchronization and use of this information to enhance the health system and treatments. The biggest challenge at the intersection of machine learning and healthcare today is finding ways to collect and use various kinds of data to help analyze, prevent and manage people. Burgundy ML applications in pharmaceuticals and medicine offer hope for a future in which data synchronicity, research and ingenuity are a daily fact. At Emerj, we research A.I.'s effect on the pharmaceutical industry as part of our A.I. Landscape Opportunity service. Global drug companies have used A.I. Incentive Environments to assess where A.I. works in their business, and what applications drive its value. Disease disorder identifying and diagnosing is a crucial element of medical ML science. In 2015, more than 800 cancer medicines and vaccines were investigated, according to a survey conducted by Pharmaceutical Research and Manufacturers of America (Fig. 7).

In an interview with Bloomberg Science, research scientist of the Knight Institute Jeff Tyner pointed out that it was also a struggle to find ways to use all the findings while it was interesting. "The concept of a biotechnologist working with computer scientists and informatics is so interesting," Tyner said. It is no wonder that great players were the first to jump on the car, especially in areas of significant need, such as detecting and treating cancer. I.B.M. Watson Health announced I.B.M. Watson
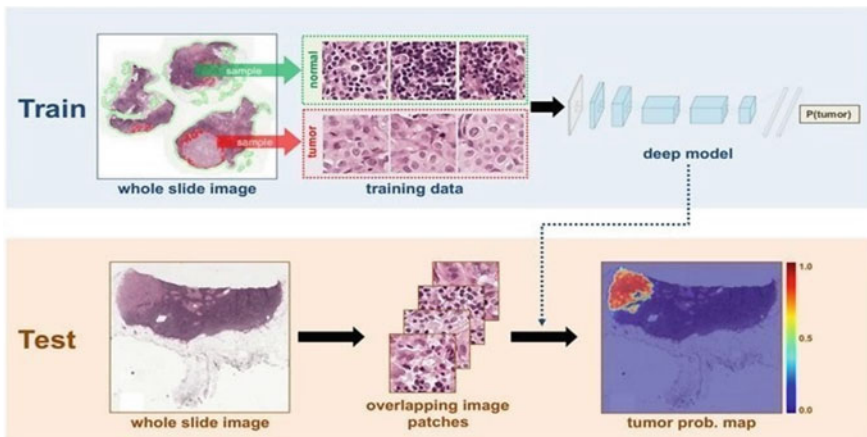
Fig. 7 Cancer detection using ML

Genomics in October 2016, a joint project with Quest Diagnostics to support accurate medicines using cognitive computing, genomic tumours, and a biopharmaceutical company located in Berg. Current research programmes include intravenous tumour dosing experiments and prostate cancer identification and management. In Google, the Deep Mind Health disclosed several cooperation with the United Kingdom last year, including Moorfields Eye Hospital in London, where macular degeneration technology is developed in older eyes. Currently, supervised learning allows clinicians to choose from selective diagnoses or predict patient risk from symptoms and genetic data, for example. I.B.M. Watson Oncology is a pioneering medical reform and history institution to expand treatment possibilities. This personalized treatment approach has significant consequences for patients to boost their wellbeing and minimize overall healthcare costs. For instance, reductions in health care costs will fluctuate and decrease as more people stick to medication and treatment plans.

Behavioural adaptation is also a significant driver in the prevention system, addressed in December in an interview with Emerj by Catalia Health's Cory Kidd. And there are many start-ups with varying levels of success in cancer diagnosis, prevention and care (for example). Somatix—a B2B2C software platform provider with an ML database analysis that "recognizes mouth-tohand movement," especially when smoking stops. SkinVision is the first to build the software for 'SkinVision' which was DermCheck for 'Just, Just C.E. Cert.' The first time this app was developed, DermCheck sent images to dermatologists (people rather than machines) via phone to return to a personalized treatment plan. In this area, the frontier appears to be precision medicine with the detection of 'multifactorial' mechanisms and alternate therapeutic routes. This study is unaccompanied, often mainly limited to detecting data patterns without prediction (recently included in supervised learning). Machine learning has many practical applications that can help develop and conduct clinical trials. Advanced predictive analytics can identify clinical problems and thereby contribute to lower, quicker, and cheaper testing, for example, for social media, doctor visits and genetic information for target groups. According to McKinsey, several more ML applications for improved clinical trial performance have been developed, including determining sample sizes for better quality, addressing and adjusting patient recruitment discrepancies, and using electronic health records to remove information errors (e.g. double entry). DeepMind Wellness, Google, with University College London Hospital, collaborated on machine-learning algorithms to identify different levels of safe and cancerous radiation. DeepMind and U.C.L.H. work to distinguish healthy and cancerous tissue variations on ML applications. More knowledge is covered in our radiology machine learning applications. Classification of documents (sorting patient requests by email, for example) with the aid of vector machines and optical character recognition (transforming a digitized character of cursive or another manuscript). The MATLAB ML and the handwritten character recognition network of the Google Cloud Vision API provide only two examples of creativity.

Intellectual and electronic health records are being developed over the next decade by the M.I.T. Clinical Machine Learning Community. Diagnostics, clinical decisions and personal care guidelines, including integrated ML/AI, where supported.

The M.I.T. research site notes that 'the need to build robust machine learning algorithms will learn from smaller data on the labelling, interpretation and generalization of natural languages in medical and institutional environments.' ML and A.I. technology will monitor and predict all outbreak ovarian diseases through satellite-based information, historical Web-based information and sociological information. Today's opioid crisis is a clear example of A.I. Vector support. Artificial neural networks were used to predict malaria outbreaks based on temperature, mid-month precipitation, total positive cases, and other data points. Epidemic severity predictions are urgently needed, especially in third-world countries lacking in medical facilities, education, and healthcare access. ProMED-mail is a software that monitors and detects emerging diseases in real-time.

Significant issues still have to be addressed when integrating ML technology in pharmaceuticals and medicinal products: data processing is one of the most pressing problems today. Patient information is also sensitive and not easily accessible. It seems safe to believe that most of the population is concerned with sharing information rather than privacy concerns. It is noteworthy to note, in a survey carried out in the United Kingdom by the Welcome Foundation in March 2016, that only 17% of the respondents never decided to distribute their anonymized information to third parties other than study.

The more open algorithms should comply with the strict drug development regulations; the causes of machine findings must be seen and understood. It is essential in the pharmaceutical industry to attract data science talent and develop a deep pipeline of Skills. The breakdown of data silos and the promotion of a "data-centered perspective" across industries is critical in helping the industry to transform their thinking into long-term progressive changes and to find value. Pharmaceutical firms have refused to update or fund research programmes in history unless there is a considerable and immediate monetary benefit. The relocation of already chaotic electronic records through networks will be a first step towards creating tailor-made care solutions.

3   Natural Language Processing in Healthcare

Natural Language Processing provides exciting opportunities for swimming through and using enormous data currently unused to improve results, reduce costs and increase care efficiency. We found the N.L.P. drivers in healthcare in the first part of this two-part series. For the increasing amount of data already produced in siLos and daily systems, the A.I. sector appears necessary. The paper discusses the factors which contribute to the growth and implementation in the health care sector of natural language processing, the potential benefits of performance and the future of artificial intelligence and machine learning in the healthcare industry. It is a synthetic intelligence division which mainly focuses on perception and human data or speech.

The technology involves a range of sub disciplines, including language questions, natural language creation and language processing. Firstly, the technology employed in health care is double: analyzing individuals' speech and extracting significance.

The two leading roles of technology in machine learning and N.L.P. in healthcare are the most important. Over the years, N L.P. has advanced with validated R.O.I. recognition, enabling clinics to transcribe notes with useful E.H.R. information using natural language treatments in healthcare. Front-end speech recognition eliminates medical records' role rather than cure, while back-end technology works to identify and correct transcription errors before moving them to human evidence. While the market is almost overwhelmed by language recognition technology, many companies disrupt the field with deep learning algorithms in mining applications that give more significant opportunities. Clinical recording developments—Healthcare computer training has resolved clinical recording and free doctors from a complicated E.H.R. manual system. This was possible due to the dictation from speech to text and the wording of the data entry capturing structured data while processing. As machine learning progresses in healthcare, essential knowledge can be gathered from other new sources, and P.H.M. and V.B.C. initiatives can be powered by analytics. Study of data mining—The integration of data mining into healthcare systems allows organizations to reduce decision-making subjectivity and provide valuable medical knowledge. Data mining can be a cyclic technology until every HCO has started developing a successful business plan for better patient care. Computer-assisted coding—NLP-driven C.A.C. promises to prove coder accuracy im-. Computer-aided code collects details and information on therapy to capture and optimize any code. The research showed that current suppliers in the production of C.A.C. solutions on the market must redirect their answers to a value-driven paradigm's demands and ensure they function as expected. Many health I.T. programmes have a regulatory burden because metrics such as ejection fractions are not stored as separate values. Health systems have to determine if an expulsion fraction (Fig. 8).

Health systems have to determine if an expulsion fraction is reported in a note and save any value in a format which can be used for accurate reporting on the organizational automated analytical reporting platform. New N.L.P. cases, such as clinical study matching, N.L.P. and machine education, have an immediate impact. Some companies are now trying to solve N.L.P. engine problems to suit the test. N.L.P. simplifies and turns the test match process into a seamless one with innovations. Prior approval—A report showed that the charge for doctors' last approval needs is growing. These standards improve overall activities and interfere with the delivery of treatment. The question as to whether payers earned and approved reimbursement was not rendered after a while as a result of language processing. I.B.M. Watson and Anthem are now working on a payer N.L.P. module for quick prior authorization assessment. Medical decision-making encourages physicians' correct option for natural languages production and engineering. Some healthcare areas require extra monitoring, such as medical errors. Risk Adjustment and Hierarchical Category Category Coding, a risk adaptation model, was initially developed to predict patients' potential costs. N.L.P. is also used for the control and evaluation of symptoms by physicians. H.C.C. coding is increasingly prevalent in value-based payment models. H.C.C. uses ICD-10 coding to assign each patient risk ratings. The natural language processing will allocate a risk factor to patients and estimate healthcare costs.
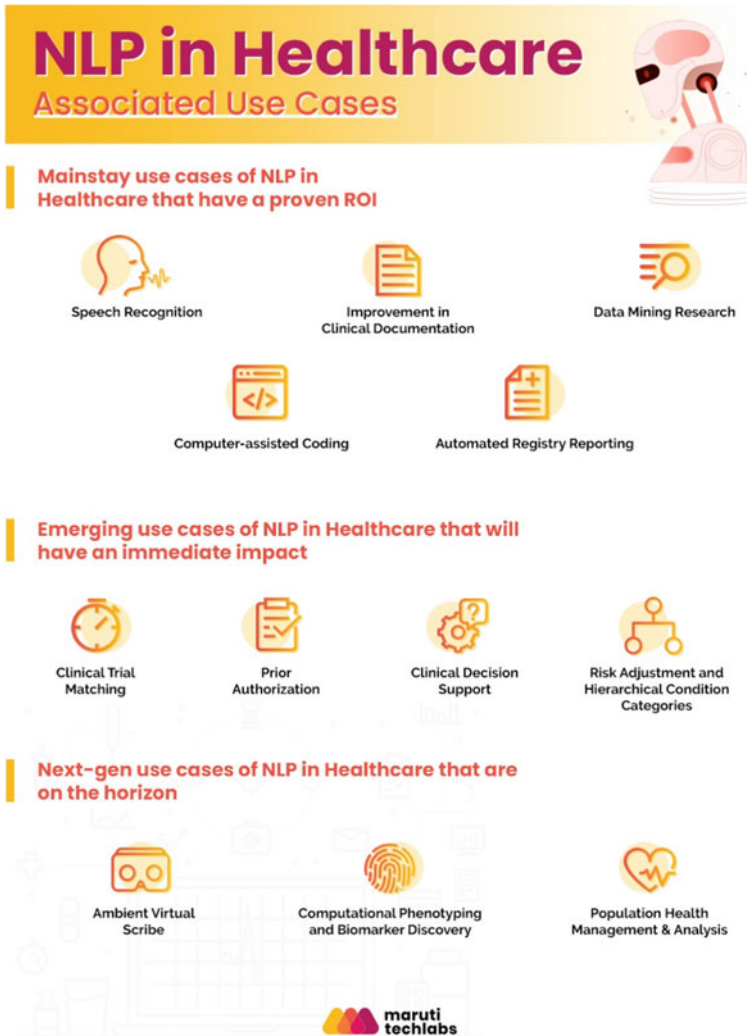
Fig. 8 N.L.P. in healthcare

Environmental Virtual Scribe—Language software must substitute human scribes full of clinical documentation. N.L.P. and Artificial Health Intelligence will turn clinical documentation into a wholly changed game in that scenario. The N.L.P. can also enable physicists to teste patients in machine phenotyping and biomarker development. Compared to the specialist's expertise, N.L.P. recognizes phenotypes through the patient's present conditions. N.L.P. may also be used to determine speech patterns and diagnoses for neurocognitive injuries such as Alzheimer's and dementia. An application for the N.L.P. will eventually identify a subset of ethnic or racial groups in population surveillance to monitor and map health differences. There is no

granularity for defining significant socio-cultural disparities and population tracking in the current administrative databases.

However, N.L.P. offers a strong basis for further research and development. The N.L.P. will use health systems to improve its managing treatment and solutions approach. Machine-learning firms can boost healthcare physicians' workflows and patient experiences. Natural language processing solutions in the field of healthcare are complete: enhancing patient experience with the doctor and E.H.R. Language recognition systems may help close the distance between complicated medical issues and the perception of the wellbeing of patients. N.L.P. can be an ideal way to deal with E.H.R. Many clinicians take N.L.P. as an alternative form of typing and writing notes, including growing patient health knowledge. Although patients have access to their health information through an E.H.R. system, the specifics are difficult for most people to understand. Just a fraction of the patients can choose their medical history. This changes as machine learning is implemented in healthcare. The N.L.P. tools help to measure and improve treatment outcomes while improving healthcare quality. Value-based pay will allow medical facilities to assess doctors' efficiency and recognize care inequalities. N.L.P. algorithms can help HCOs detect potential treatment errors. The N.L.P. algorithms can gather useful data from extensive databases to classify sensitive patients and provide healthcare practitioners with the best methods for coping with challenging patients. For predictive healthcare research, identifying high-risk patients and the development of the diagnostic process can be carried out using Pervasive Analytics and natural speech processing and predictional healthcare. For emergency rooms, complete data is essential.

For example, delays in diagnosing Kawasaki disease cause significant complications if they are in some way, overlooked or violated. As scientific evidence shows, an NLP-based algorithm has classified risk patients with Kawasaki's disease with 93.6% sensitivities and 77.5% accuracy than the clinician's manual test. Many French researchers are working on another NLP-based patient algorithm to map, diagnose and prevent HAI. The N.L.P. generated unstructured data to classify early indications and clinicians as such. Further tests have also been carried out to analyze and predict heart failure-treated patients' likelihood automatically. Natural language processing was used to assess a patient's risk of hospital admission and death for the past 30 days and provide free text updates for the last 24 hours. At the end of the successful experiment, the algorithm was better than expected, and the overall positive predictive performance of the model reached 97.45%. The advantages of using N.L.P. will undoubtedly be used to evaluate and forecast these conditions in patients by applying a wide range of algorithms. The promotion of superior care decisions and patient health outcomes includes natural language processing and other ML approaches.

4   Conclusion and Discussion

If your child has lost its wisdom teeth, individual doses of Vicodin are more likely to be given. They may have Bactrim for an infection in the urinary tract (U.T.I.).

Few patients, hopefully, will ever receive the same dose of any drug in the not too distant future. Few people use the same medication if we know enough about the patient's biology and history. The promise of personalized medicine is a world in which guidelines are customized to everyone's wellbeing and disease remedies, their medical history, genetic features, past ailments, nutrition, levels of stress and much more. However, it might eventually expand to minor conditions (i.e. to give somebody a smaller dose of bacteria for U.T.I., or an unparalleled version of bacteria formulated for the Hooman Hakami from Medtronic, a Medtronic and I.B.M. diabetes video) Medtronic would like to see insulin control pumps function independently, monitor blood glucose levels and, if appropriate, administer insulin w While Western Medicine has maintained its primary focus on the treatment and improvement of diseases, proactive treatment and intervention are essential. This application is being advanced by the first wave of IoT devices (particular FitBit). Machine training may be used to track the worker's performance or tension and seek beneficial improvements in risk groups (not just to re-live or cure symptoms after reversals). Ethical concerns regarding "increasing" human physical and (particularly) mental capacity are severe and probably will become more critical as the technology of improvement in the next 15 years.

On the other hand, any human expert could easily defeat the narrower computer vision application (if the model has been adequately trained). Federal "red tape" or HIPAA also can make sports more "goliath" for the medical profession than "david." New social networks will catch teenagers up and beat Snapchat and Facebook because of their virality, marketing and user interface. You only need a dozen engineers at the right time, just like Instagram, but it is unlikely that a dozen engineers have the industry, although they have earned ten million dollars. This labyrinth can demand more money, contacts and know-how than any small Silicon Valley company can collect. It seems like I.B.M. or Medtronic can have a healthy lead in medical innovation because of these factors alone. In recent years, ML has contributed to several areas, including the vision and processing of natural languages. Many existing training data (e.g. images in vision or sentences in natural language) can benefit from more complex models. We are now on the verge of a substantial change in HE. Through applying ML to electronic health data, including genomic data, more and more accessible, epidemiologists in health care can better understand the underlying risk for acquiring infectious diseases and transmission mechanisms and develop specific strategies for reducing HAI. It is important to note that, while powerful, ML can not recognize relationships that do not exist in the data.

Furthermore, ML does not substitute for regular statistical or random screening. Instead, ML can be used to extend the existing HE toolbox. Interdisciplinary teams would have the most significant influence on the future and work together to interpret the data.

References

1. Lipton RB, Scher AI, Steiner TJ, Bigal ME, Kolodner K, Liberman JN, Stewart WF (2003) Patterns of health care utilization for migraine in England and the United States. Neurology 60(3):441–448
2. Islam MN, Inan TT, Rafi S, Akter SS, Sarker IH, Islam AKM (2020) A survey on the use of A.I. and ML for fighting the COVID-19 pandemic. arXiv preprint arXiv:2008.07449
3. Lu C, Strout J, Gaudreau R, Wright B, Marcus FBDC, Buch V, Andriole K (2020) An overview and case study of the clinical A.I. model development life cycle for healthcare systems. arXiv preprint arXiv:2003.07678
4. Kaur J, Mann KS (2017) AI-based healthcare platform for real-time, predictive and prescriptive analytics using reactive programming. J Phys: Conf Ser 933:012010
5. Tripathi G, Ahad MA, Paiva S (2020) S2HS-A blockchain-based approach for the smart healthcare system. Healthcare 8:100391
6. Yoon JE, Suh CJ (2019) Research trend analysis by using text-mining techniques on the convergence studies of A.I. and healthcare technologies. J Inf Technol Services 18(2):123–141
7. Gil-Lacruz M, Gracia-Pérez ML, Gil-Lacruz AI (2019) Learning by doing and training satisfaction: an evaluation by health care professionals. Int J Environ Res Public Health 16(8):1397
8. Das S, Sanyal MK, Application of A.I. and soft computing in healthcare: a review and speculation 8:21
9. Clifford GD (2020) The future A.I. in healthcare: a tsunami of false alarms or a product of experts? arXiv preprint arXiv:2007.10502
10. Troncoso EL (2020) The greatest challenge to using AI/ML for primary health care: mindset or datasets? Front Artif Intell 3:53
11. Lysaght T, Lim HY, Xafis V, Ngiam KY (2019) AI-assisted decision-making in healthcare. Asian Bioethics Rev 11(3):299–314
12. Drysdale E, Dolatabadi E, Chivers C, Liu V, Saria S, Sendak M, Wiens J, Brudno M, Hoyt A, Mazwi M (2019) Implementing A.I. in healthcare
13. Johnson SL (2019) A.I., machine learning, and ethics in health care. J Legal Med 39(4):427–441
14. Stanfill MH, Marc DT (2019) Health information management: implications of artificial intelligence on healthcare data and information management. Yearb Med Inform 28(1):56
15. Pawar U, O'Shea D, Rea S, O'Reilly R (2020) Explainable A.I. in healthcare. In: 2020 international conference on cyber situational awareness, data analytics and assessment (Cy- Bersa). pp 1–2
16. Panesar A (2019) Machine learning and A.I. for healthcare. Springer
17. Farroha J (2019) Security analysis and recommendations for A.I./ML-enabled automated cyber medical systems. Big Data: Learn Anal Appl 10989:109890
18. Adadi A, Berrada M (2020) Explainable A.I. for healthcare: from black box to interpretable models
19. Reddy S, Allan S, Coghlan S, Cooper P (2020) A governance model for the application of A.I. in health care. J Am Med Inf Assoc 27(3):491–497
20. Hernandez-Boussard T, Bozkurt S, Ioannidis J, Shah NH (2020) MINIMAL (MINimum Information for Medical A.I. Reporting): Developing reporting standards for artificial intelligence in health care. J Am Med Inf Assoc
21. Hunter P (2019) The advent of A.I. and deep learning in diagnostics and imaging: machine learning systems have the potential to improve diagnostics in healthcare and imaging systems in research. EMBO Rep 20(7):e48559
22. Greco L, Percannella G, Ritrovato P, Tortorella F, Vento M (2020) Trends in IoT based solutions for health care: moving A.I. to the Edge. Pattern Recognit Lett
23. Henriksen A, Bechmann A (2020) Building truths in A.I.: making predictive algorithms doable in healthcare. Inf Commun Soc 23(6):802–816
24. Halminen O, Tenhunen H, Heliste A, Seppälä T (2019) Factors affecting venture funding of healthcare A.I. companies. ICIMTH 268–271

25. Terry N (2019) Of regulating healthcare A.I. and robots. Available at SSRN 3321379
26. Srivastava SK, Singh SK, Suri JS (2020) State-of-the-art methods in healthcare text classification system: A.I. paradigm. Front Biosci (Landmark edition) 25:646–672
27. Morley J, Machado C, Burr C, Cowls J, Taddeo M, Florida L (2019) The debate on the ethics of A.I. in health care: a reconstruction and critical review. Available at SSRN 3486518
28. Tan Y, Jin B, Yue X, Chen Y, Vincentelli AS (2020) exploiting uncertainties from ensemble learners to improve decision-making in healthcare A.I. arXiv preprint arXiv:2007.06063
29. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y, Dong Q, Shen H, Wang Y (2017) Artificial intelligence in healthcare: past, present and future. Stroke Vascular Neurology 2(4):230–243
30. Ellahham S, Ellahham N, Simsekler MCE (2020) Application of artificial intelligence in the health care safety context: opportunities and challenges. Am J Med Qual 35(4):341–348
31. Fritchman K, Saminathan K, Dowsley R, Hughes T, Cock MD, Nascimento A, Tere- Desai A (2018) Privacy-preserving scoring of tree ensembles: a novel framework for A.I. in healthcare. In: 2018 IEEE international conference on big data (Big Data), pp 2413–2422

Predictive Maintenance for a Turbofan Engine Using Data Mining

Ismaila Mahmud, Idris Ismail, and Zuhairi Baharudin

Abstract Airplane safety remains one of the crucial areas that must have a robust maintenance strategy due to its impact in the transportation of human beings and goods. Predictive maintenance is a vital means of ensuring complex system such as turbofan engines in airplane are being used safely and optimally. The advent of information and communication technologies provide ways to collect useful data for maintenance strategies and decision making. The acquired data are unstructured and may contained incomplete information. Data mining transform the data to become meaningful and useful for machine learning application. In this paper, data mining techniques for predictive maintenance are presented, with machine learning algorithms applied to predict the maintenance conditions of a turbofan engine and the results are compared. The results show that support vector machine has a slightly better accuracy than the other methods.

Keywords Predictive maintenance · Machine learning · Turbofan engine · Data mining

1 Introduction

Maintenance is a process that ensure tools, machine (simple or complex) are kept in a good working condition and ensuring that it is repaired when the need arises as a result of breakdown within the possible time. Thus, maintenance has been around since the existence of machineries. It has also evolved with the changes that accompany new innovations in machine systems. Traditionally, maintenance tends to be reactive-respond to failures in equipment or devices. And as a result, when failure occurs, it becomes very catastrophe. Human life and properties have been lost due

I. Mahmud (✉) · I. Ismail · Z. Baharudin
Department of Electrical/Electronic Engineering, Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia
e-mail: ismaila_19001025@utp.edu.my

I. Mahmud
Department of Electrical Engineering, Ahmadu Bello University, Zaria, Nigeria

to insufficient maintenance and safety strategies in the aviation industry. A well planned maintenance procedure could decrease the rate of failure, ensure system availability, reduce the cost of maintenance and increase system reliability [1, 2]. The complexity of airline systems makes conventional maintenance methods not to meet the requisite maintenance requirements for the aviation industry. With the advent of smart systems, gives room to have some high level machine flexibility, availability and high quality of services and products [3]. The advancement brings about new form of challenges which is the complexity of systems and processes. However, digitization in major sectors and couple with advancement in information and communication technologies (ICT), brings new prospect for maintenance strategies [4]. Maintenance measures could be predicted through the analysis of real time data and recorded data. Predictive maintenance means having smart way of handling maintenance from the usual reactive done before [5]. The lack of structured data pose as a challenge in implementation of predictive maintenance strategies [6, 7] but the use of Data mining allows unstructured data to become useful.

Prognostics have been used in the area of maintenance for quite sometimes. Prognostic and health management (PHM) is a means where failure rates of machine or equipment are estimated which enables monitoring its overall effect on the entire system and provides ground for decision making of maintenance actions to be taken or otherwise [8]. PHM could be achieved through the following methods; model-based or data-driven methods. In the model-based approach, it has the ability to incorporate physical knowledge of the system. However, there are usually difficulties in finding the mathematical representation of complex systems. Unlike the model-based, the data-driven method has the ability to utilize high dimensionality data to perform prediction. Still, a dataset that has a run-to-failure in it, to be used for modelling are quite not available. This is as result of most data does not have fault evolution due to the nature of design of the system and managerial demands until failure occur. Moreover, time and cost could hind obtaining actual system fault evolution. Propriety issue is another factor that could make access to available data hard [9]. All these constitute challenges and limits wide research in this area. However, an alternative has been initiated through the use of run-to-failure data simulation and thus, reinforced research in the field. Data-driven methods build the prediction model through the use of machine learning algorithms. Examples of classical learning algorithms include approaches based on artificial neural networks (ANNs), k-means clustering, logistic regression, Decision Trees (DTs) and support vector machines (SVMs) [10]. In addition, data-driven algorithms sometimes make more accurate predictions than physics-based models because mathematical models cannot accommodate enough complexity. This paper investigates data mining and machine learning application on simulated run-to-failure data from NASA turbofan engine dataset. Four machine learning algorithms such as decision tree, support vector machine, random forest and gradient boosted tree were applied, and their accuracies compared. The organization of the paper is as follows; Sect. 2 covers the related work. In Sect. 3, data miming is being described with its features and the machine learning algorithms used in the paper are highlighted. Case study and results are presented in Sect. 4 and Sect. 5 respectively. Finally, the conclusion is in Sect. 6.

2 Related Work

There have been some research works reported in the literature on predictive maintenance. Researchers have proposed novel methodologies to produced high accuracy in predictive maintenance. In Behera et al. [11], the authors proposed a data-driven prognostic technique for predictive maintenance based on tree ensemble algorithms. They used NASA turbofan engine dataset and found out that gradient boosted trees performed better than random forest. Alberto et al. [12] proposed a data-driven predictive maintenance process that perform data acquisition and processing for a real machine process. The data was obtained from computer numerical control (CNC) turning centre with two tool posts. Machine learning models were used in assessing the remaining useful life of the machining tool. The concept shows that predictive maintenance approach could be used in areal machine process where preventive maintenance remain the feasible maintenance strategy adopted. Accorsi et al. [13] developed a data mining framework for condition-based maintenance with different machine learning classifiers. The models were applied to real-world case study, the fault prediction accuracy was quantified and compared. The results show that random forest having slightly better accuracy than other methods. In Bukhsh et al. [14], the authors propose three tree-based classification methods for predictive maintenance of railway switching systems. The models were able to trigger the activity type and predict the maintenance need of the railway switches. Lee et al. [15] applied predictive maintenance to two machine tool system elements using artificial intelligence algorithms the two machine tool systems consist of spindle and cutting tools. The component's conditions were obtained from flank wear and bearing's remaining useful life for cutting tool and spindle respectively. Artificial neural network and support vector machine were used in classifying the tool's condition. Okoh et al. [16] proposed a statistical techniques-based data-driven prognostic approach for predictive maintenance. The model was conducted on a synthetic data validated by industrial experts. In this paper, data mining based predictive maintenance is used for turbofan engine with machine learning algorithms to predict the engine conditions.

3 Data Mining

Data mining provide means where hidden trends, patterns, and associations on datasets are discovered. It involves many methods such as statistics and machine learning to solve wide range of the problem at hand [17]. In a typical classification problem, neural network uses gradient descent training to solve it, while a decision tree uses greedy search algorithms to solve it. The main purpose of data mining is model creation [18]. A framework to execute Data Mining similar to standard one like Cross Industry Standard Process for Data Mining (CRISP-DM) [19] with modification to suit our classification model case is developed. The steps in the framework includes; Data Preparation, Feature Selection and Prediction model.

3.1 Data Preparation

Row data are often not useful due to having presence of noise or too weak and in some instance very large. Such data have to be process and reduced to a form, it could be useful and easily be applied on predictive model. Some of the actions taking to improve the signal quality include data quality assessment, operating condition operation discovery and data normalization. The tools used in data preparation include filtering, amplification, data compression, data validation and de-noising.

3.2 Feature Selection

The data goes to another form of process to make it more suitable for use. At this stage, features are extracted and selected. The aim is to extract indictors or features with fault or patterns and also having minimal subset features from problem domain while maintaining original feature structure with high accuracy. The purpose of feature is to select relevant and informative features effectively.

3.3 Prediction Models

Decision tree, support vector machine, random forest and gradient boost tree are the machine learning algorithms adopted for the prediction models in this paper. These are supervised methods which build its models through establishing links between input variables and target variables.

Decision Tree (DT). A decision tree is a machine learning algorithm that built its classification model using a tree structure flowchart. Decision tree algorithm is constructed with three set of nodes; root node, internal node and the leaf which represents the class label. Among the algorithms used in developing decision tree are C4.5 [20], classification and regression tree (CART) [21]. These algorithms are implemented base on divide and conquer strategy where the training set is being recursively partition.

Support Vector Machine (SVM). SVM is a statistical learning concept use to perform classification through obtaining optimal hyperplane between two classes. The hyperplane is the linear separator and in case of nonlinear classes, a kernel function is used [22]. Kernel is a mathematical function that transformed nonlinear data to be classified by a linear hyperplane and is given by

$$K\left(x, x'\right) = \exp\left(-\gamma \left|\left|x - x'\right|\right|^2\right) \qquad (1)$$

where $\|x - x'\|$ is the Euclidean distance between two data points x and x'. γ is the control parameter that maintain error trade-off between variance and bias in the model. Cost function is used in training SVM for error minimization

$$\frac{1}{2}\|w\|^2 + C \sum_{k=1}^{n} \xi_k \tag{2}$$

perform under these constraints

$$y_k \left(w^T x_k + b \right) \geq 1 - \xi_k \ \text{ and } \ \xi_k \geq 0 \tag{3}$$

where w is weight vectors, C is the penalty parameter, ξ_k is the slack parameters used in handling non-separable data, T is matrix transpose, x_k is the training data points, b is a constant and y_k is the class label of the training data.

Random Forest (RF). Random forest is one of the ensembles of decision tree. In random forest, N subsets of training data are generated from the original dataset using bootstrap sampling approach. Decision tree T_i is constructed from the training of the subset. Random forest is built from using N—decision trees [14]. The prediction of random forest is the average predictions of the trees given as;

$$f(x) = \frac{1}{N} \sum_{i=1}^{N} T_i(x) \tag{4}$$

Gradient Boost Tree (GBT). Gradient boosted tree is also an ensemble of decision tree that produces weak tree classifiers and it use boosting approach to build the base model. In the boosting approach, multiple models are built through resampling of the training instance that previously could not be estimated and by having some loss function (sum of squared errors or cross entropy) minimized [14]. If the gradient overall ensemble function is $F_m(x)$ and it generates a new model $F_{m+1}(x)$ for the base-learner $h(x)$. For a given squared error loss function, the algorithm is express as:

$$\rho_m = argmin_p \sum_{i=1}^{N} \left[\zeta \left(y_i, F_{m-1}(x_i) + \rho h(x_i) \right) \right] \tag{5}$$

$$F_m(x) = F_{m-1}(x) + \rho_m h(x) \tag{6}$$

where N is the total number of training instances, ζ is the loss function and ρ is the gradient step-size.

4 Case Study

The prognostic and health management dataset used in this paper is from NASA prognostic research center. It is a simulation of turbofan engine under several continuous conditions. The data was captured using C-MAPSS (Commercial Modular Aero Propulsion System Simulation) software. The turbofan engine architectural structure with its primary components is shown in Fig. 1.

Air goes into the engine through the fan and low pressure compressor. Thereafter, the air passes through high pressure compressor that mixed it with fuel and ignited by a combustor. This results in lifting up discharge air velocity by the combust fuel in the high pressure compressor (HPC) that controls the low pressure turbine (LPT) and high pressure turbine (HPT). A 9000 lbs thrust class engine with various sensors installed on it was used for the data readings collection. Several operating conditions were used in gathering the engine information such as temperature variation (within the range of -51–39 °C) and the engine's altitudes from sea level to 40,000 ft [11]. Table 1 provides some details of the four subsets of testing and training datasets in the C-MAPSS dataset.

There is k-by-26 matrix-order stored in each subset where 26 is the number of input attributes and k is the row count of the dataset. The three operational settings have their combine data represented in every tuple of the dataset. The dataset is put such that it start from engine number, followed by the cycle number, columns 3–5 are the values of the operational settings and the remaining are the 21 different sensor reading values attached on the aircraft engine [23]. It is considered that in each tuple,
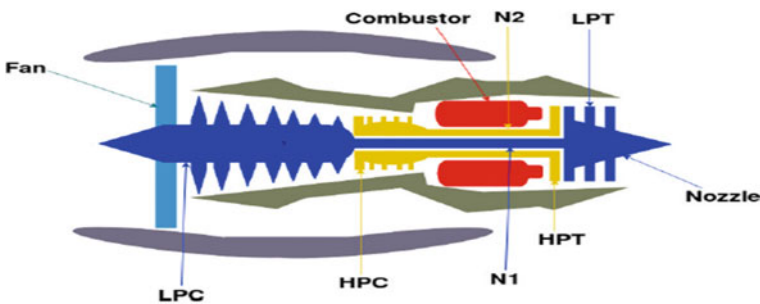
Fig. 1 Architecture of turbofan engine

Table 1 Details of C-MAPPS dataset

Datasets	FD001	FD002	FD003	FD004
Training trajectory	100	260	100	249
Test trajectory	100	259	100	248
Operating conditions	1	6	1	6
Fault condition	1	1	2	2

the training and testing subset hold the total life-cycle time of the engine and the last entry is the condition where the engine is on the brink of failure or declared unhealthy [24, 25]. The major purpose of RUL estimation is to identify the condition when the engine might need inspection or replacement. The RUL attribute targets could be define as normal, warning and critical depending on the particular value in the RUL attribute.

5 Results and Discussion

In this section, the FD001 dataset is used in building and testing the predictive maintenance model for the turbofan engine. The four machine learning algorithms; decision tree, support vector machine, random forest and gradient boosted tree are trained to monitor and predict the conditions of turbofan engine.

Confusion matrix is used in evaluating the performance of the models' prediction ability. In the matrix, each row and column show the true and predicted values of the model and can reveal which classifier is confused. A true positives are positive cases correctly classified as positive class and true negatives are negative cases correctly classified as negative class. Conversely, a false positives are negative cases incorrectly classified as positive class and false negatives are positive cases incorrectly classified as negative class. True positive rate is the proportion of correctly classified positives while false negative rate is the proportion of misclassified positive cases (error rate per class).

The input dataset for the models is extracted from FD001 dataset of NASA prognostic and health management dataset. It is 20626×26 dataset with 26 input features. 21 features represent sensor signals employ in monitoring the conditions of the turbofan engine. The first 5 attributes consisting of; engine numbers (100), number of cycles (time) and three operating conditions. In the pre-processing of the dataset, 14 features are selected from the 21 sensor signals as being meaningful for the prediction models. Principal component analysis (PCA) is applied to the selected features to reduce the dimensionality of the features [15]. PCA is a mathematical process of mapping multi-dimensional dataset to lower dimension and maintaining its original information. The mapped dataset is used as the input for the models. The prediction conditions are labelled as; critical, warning and normal which corresponds to the remaining useful life of the dataset. Table 2 shows the accuracy obtained by each model which measures the overall performance of the models. The results are

Table 2 Accuracy of the predictive maintenance by the models

Model	Accuracy (%)
Decision tree (DT)	83.8
Support vector machine (SVM)	88.6
Random forest (RF)	86.0
Gradient boosted tree (GBT)	83.6
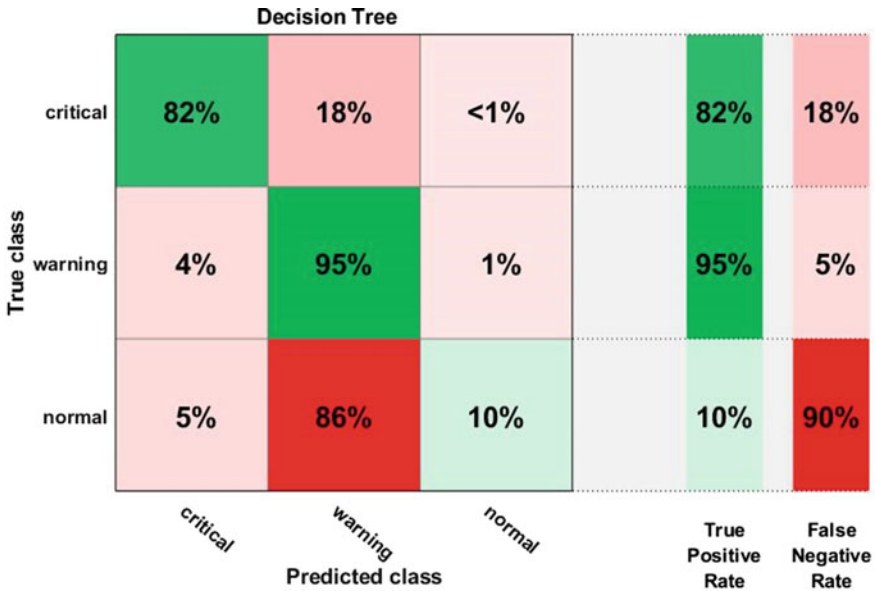
Fig. 2 Confusion matrix for decision tree

obtained using hold-out validation, by using 75% for training and the remaining 25% for testing.

SVM had the better result with 88.6% accuracy while GBT had the least result of 83.6%. Figures 2, 3, 4 and 5 show the confusion matrices of the models.

From the Fig. 3, it shows that SVM correctly predict critical condition 87%, warning condition 96% and normal condition 38%. The SVM maintained better prediction among the models. Though, RF perform relative similar with the SVM in all the indicators as shown in Fig. 4

6   Conclusion

In this paper, machine learning algorithms through data mining were employed in the development of predictive maintenance of turbofan engine. Four models; decision tree, random forest, gradient boosted tree and support vector machine were trained. Three engine's conditions (normal, warning and critical) represent the remaining useful life of system. The results show that SVM had better overall prediction accuracy than the other three decision tree-based models. In the future, we intend to enhance the performance of the decision tree-based model by having it hybridize with a metaheuristic algorithm.
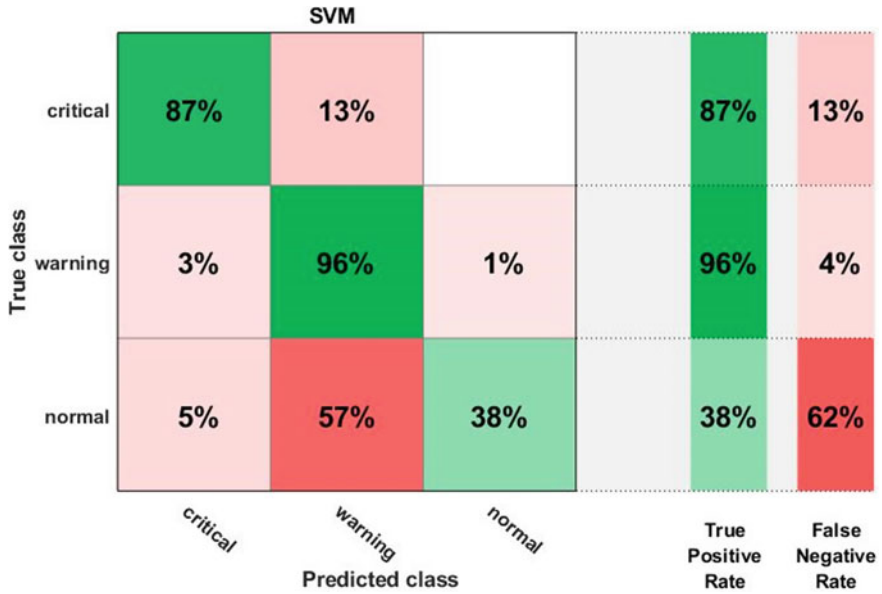
Fig. 3 Confusion matrix for support vector machine
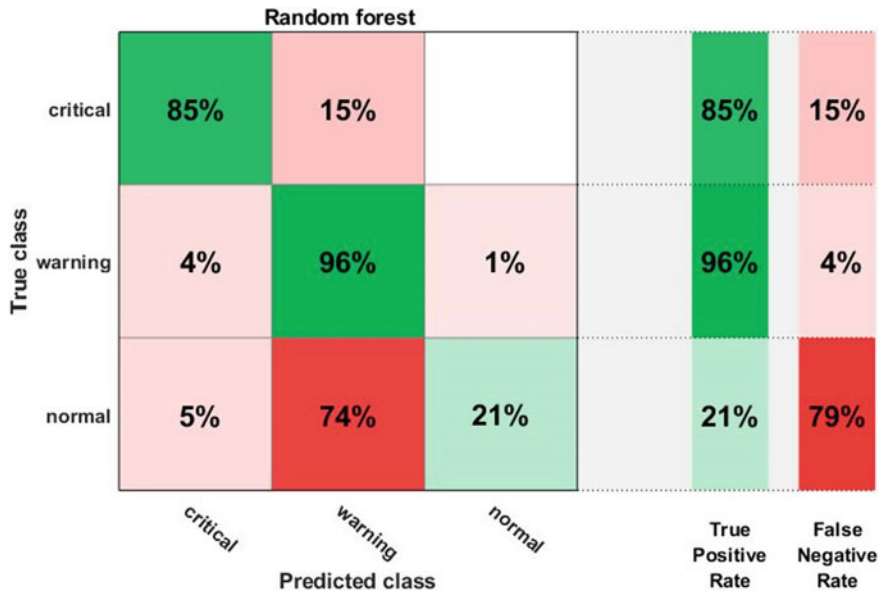
Fig. 4 Confusion matrix for random forest
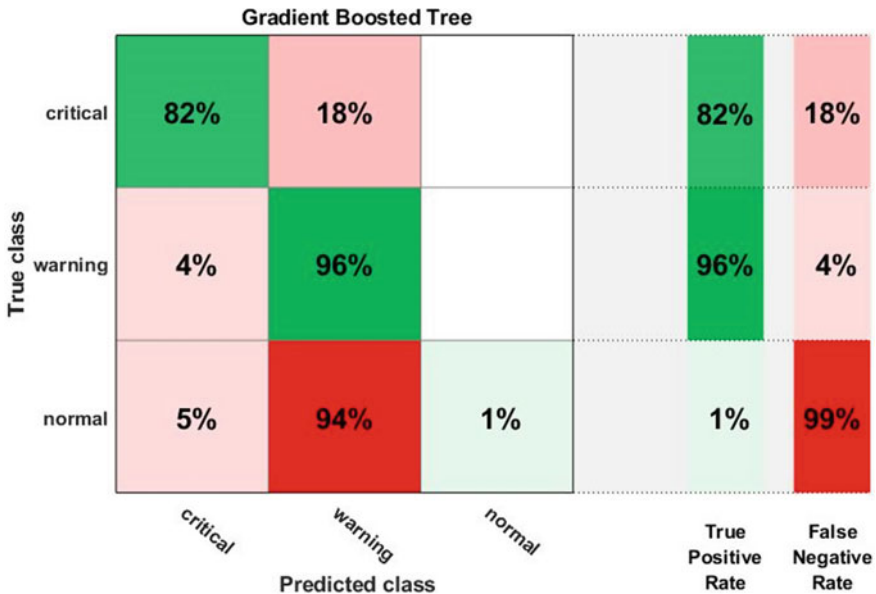
Fig. 5 Confusion matrix for gradient boosted tree

References

1. Li Z, Wang K, He Y (2016) Industry 4.0-potentials for predictive maintenance. Adv Econ, Bus Manag Res
2. Li Z, Wang Y, Wang K-S (2017) Intelligent predictive maintenance for fault diagnosis and prognosis in machine centers: industry 4.0 scenario. Adv Manuf 5(4):377–387
3. Fitouri C, Fnaiech N, Varnier C, Fnaiech F, Zerhouni NJIP (2016) A decison-making approach for job shop scheduling with job depending degradation and predictive maintenance. FAC-PapersOnLine. 49(12):1490–1495
4. El Maraghy W, El Maraghy H, Tomiyama T, Monostori L (2012) Complexity in engineering design and manufacturing. CIRP Ann 61(2):793–814
5. Khoshafian S, Rostetter C (2015) Digital prescriptive maintenance. In: Internet of things, process of everything, BPM everywhere, pp 1–20
6. Colledani M et al (2014) Design and management of manufacturing systems for production quality. CIRP Ann 63(2):773–796
7. Roy R, Stark R, Tracht K, Takata S, Mori M (2016) Continuous maintenance and the future–foundations and technological challenges. CIRP Ann 65(2):667–688
8. Lee J, Wu F, Zhao W, Ghaffari M, Liao L, Siegel D (2014) Prognostics and health management design for rotary machinery systems—reviews, methodology and applications. Mech Syst Signal Process 42(1–2):314–334

9. Saxena A, Goebel K, Simon D, Eklund N (2008) Damage propagation modeling for aircraft engine run-to-failure simulation. In: 2008 international conference on prognostics and health management. IEEE, pp 1–9
10. Kotsiantis SB, Zaharakis I, Pintelas P (2007) Supervised machine learning: a review of classification techniques. Emerg Artif Intell Appl Comput Eng 160:3–24
11. Behera S, Choubey A, Kanani CS, Patel YS, Misra R, Sillitti A (2019) Ensemble trees learning based improved predictive maintenance using IIoT for turbofan engines. In: Proceedings of the 34th ACM/SIGAPP symposium on applied computing, pp 842–850
12. Jimenez-Cortadi A, Irigoien I, Boto F, Sierra B, Rodriguez G (2020) Predictive maintenance on the machining process and machine tool. Appl Sci 10(1):224
13. Accorsi R, Manzini R, Pascarella P, Patella M, Sassi S (2017) Data mining and machine learning for condition-based maintenance. Procedia Manuf 11:1153–1161
14. Bukhsh ZA, Saeed A, Stipanovic I, Doree AG (2019) Predictive maintenance using tree-based classification techniques: a case of railway switches. Transp Res Part C: Emerg Technol 101:35–54
15. Lee WJ, Wu H, Yun H, Kim H, Jun MB, Sutherland JW (2019) Predictive maintenance of machine tool systems using artificial intelligence techniques applied to machine condition data. Procedia CIRP 80:506–511
16. Okoh C, Roy R, Mehnen J (2017) Predictive maintenance modelling for through-life engineering services. Procedia CIRP 59:196–201
17. Kovacs K, Ansari F, Geisert C, Uhlmann E, Glawar R, Sihn W (2019) A process model for enhancing digital assistance in knowledge-based maintenance. In: Machine learning for cyber physical systems. Springer, pp 87–96
18. Rokach L, Maimon OZ (2008) Data mining with decision trees: theory and applications. World scientific
19. Shearer C (2000) The CRISP-DM model: the new blueprint for data mining. J Data Warehousing 5(4):13–22
20. Quinlan J (1993) C4. 5: programs for machine learning Morgan Kaufmann San Francisco. CA, USA
21. Breiman L, Friedman J, Stone CJ, Olshen RA (1984) Classification and regression trees. CRC press
22. Tong MT (2019) Using machine learning to predict core sizes of high-efficiency turbofan engines. J Eng Gas Turbines Power 141(11)
23. Babu GS, Zhao P, Li X-L (2016) Deep convolutional neural network-based regression approach for estimation of remaining useful life. In: International conference on database systems for advanced applications. Springer, pp 214–228
24. Peel L (2008) Data driven prognostics using a Kalman filter ensemble of neural network models. In: 2008 international conference on prognostics and health management. IEEE, pp 1–6
25. Wang T, Yu J, Siegel D, Lee J (2008) A similarity-based prognostics approach for remaining useful life estimation of engineered systems. In: 2008 international conference on prognostics and health management. IEEE, pp 1–6

Cyber Security Breaches and the Long-Run Effect on Firms' Market Value: A Conceptual Framework

Syed Emad Azhar Ali and Fong-Woon Lai

Abstract Incidents of cyber security breaches are on the rise and pose a potential threat to businesses. Especially for publicly traded firms as they could create a long-lasting influence on their financial performance and, thus, the market value (MV) of a firm. Following the footsteps of the efficient market hypothesis, previous studies have examined only the short-run impact on MV ensuing to security breach announcements. Therefore, this study aims to conceptualize the impact of security breaches on MV as manifested by long-run abnormal returns those firms. The study is expected to provide a meaningful insight to investors and managers on the long-run interconnection between cyber security breaches and firms' market value.

Keywords Efficient market hypotheses · Firms' market value · Long-run abnormal returns · Security breaches

1 Introduction

The Global Risk Reports by World Economic Forum (WEF) [1, 2] perceived a higher risk for cyber-attacks with very high likelihood and impact following the enormous risks associated with natural disasters. The increasing dependency on IoT and IR4.0 is making the business operations more vulnerable for incidents of cyber-attacks such as spyware, malware, Denial of Service (DOS) attacks, phishing, ransomware, and other vulnerabilities [3–5].

Measuring and analyzing the economic impact of security breaches has been a point of interest from the scholars' perspective. The effect of a security breach on the trust and confidence of business stakeholders, including the investors in the stock market [3, 6, 7]. A firms' MV is one indicator to assess the change in the confidence of investors. Therefore, various studies have studied the effect of security breaches on the MV of a firm using theory for Efficient Markets [6, 8–14]. In most

S. E. A. Ali (✉) · F.-W. Lai
Department of Management and Humanities, Universiti Teknologi PETRONAS, Darul Ridzuan, 32610 Perak, Malaysia
e-mail: syed_17007896@utp.edu.my

of these studies, the negative impact was witnessed on MV manifested by negative abnormal returns. These studies were underpinned by the 'Semi-Strong Efficient Market Hypothesis (EMH) [15, 16], according to which stock prices adjust quickly to all new information. Accordingly, the short-run examination of a security breach event has been performed in these studies by analyzing the stock price behavior (i.e., from few days before an event to few days after an event), exhibiting an immediate analysis of MV. While the short-run analysis helps get the market's quick reaction to an event announcement, it is imperative to explore the long-run analysis to estimate a more realistic economic impact of a breach announcement. Studies addressing this problem are unclear whether an event of security breach will affect the long-run MV of firms.

1.1 Long-Run Impact on MV

According to the Global Cost of Data Breach Report [17], the time to identify and contain a data breach has been continuously increasing from 257 days in 2017 to 280 days in 2020, with estimated response cost ranging around $1 million for each firm. The announcements such as security breaches indicate that the firms are reluctant to disclose complete details of a breach on their first announcement. In recent times, we have witnessed events of security breaches where the details concerning a breach are disclosed months after the first breach announcement. For instance, in July 2019, Equifax was penalized with US$ 700 million by the Federal Trade Commission (FTC) and Consumer Financial Protection Bureau, after an enormous data breach in 2017. From the investors' perspective, they will probably have a close watch on subsequent disclosures and announcements by the firm concerning a security breach announcement that can influence their investment decision making. Hence, the impact of a security event, especially concerning MV, cannot be judged by analyzing the announcement effect only using the traditional event study methodology, as examined by [7, 18]. Especially in cases where the announcement and economic impact vary.

Considering the above arguments, probably, the abnormal effect of a security breach announcement by a listed firm can have a meaningful impression on future cash flows, the required rate of return, financial distress, and credit rating. All these concerns will be manifested on MV in the long run as well. Overall, it can be hypothesized that the security breach announcement might lead to abnormality in the long-run indicators of MV in the stock market, such as long-run abnormal returns. To the best of our knowledge, the long-run impact on MV after a security breach has not been addressed in the literature. An unfavorable abnormality might manifest that influence in its long-run stock returns.

2 Literature Review and Hypothesis Development

An influential research group has discussed various issues relating to the management of security risks, such as security investments [19, 20], institutional stimulus on innovation, and security [21]. Another line of research focuses on the market implications of disclosures linked to the security of information [67, 68] and security breaches [11, 22–28] by having an underpinning of 'efficient market theory' and the methodology of an event study. Among all event studies on security, the current paper is compatible with those scholarly works which have inspected the impact of the security breach on the stock market. Among all event studies on security, the current paper is compatible with those scholarly works which have inspected the impact of the security breach on the stock market [18, 29–31]. It can be elucidated from most of the studies that the security breach will have a significant unfavorable influence on the overall confidence of stock investors as signaled by negative abnormal returns within 1 or 2 days of the breach announcement. All these systematic research works have based only on the short-run impacts of breaches while only anecdotally exploring the long-run implications of security breaches.

2.1 Studies on Long-Run Impact:

The review of existing works reveals that the implications of events for the firm's long run prospects, especially the firms' MV have become a common concern for researchers. Long-run effects have been mostly explored with the BHAR model, 1–3 years after the incident. Within the information management avenue, the BHAR model was used to address the long-run value of the company about the impact of Capability Maturity Model (CMM) [32], Enterprise Resource Planning (ERP), Supply Chain Management (SCM), and Customer Relationship Management (CRM) [33, 34]. As far as we know, the impact of a security breach on the company's long-run MV has not yet been researched in the avenue of security. This research is the first to conceptualize the effects of the security breach on the firms' MV in the long-run.

Hypothesis Development and Conceptual Framework.

Long-run Abnormal Returns in the period before the breach announcement. Despite breached firms' intention to delay the announcement, there could be signs by which security breaches can somewhat be anticipated. For example, the firm's webpage might become inaccessible for their customers and other stakeholders, a very sluggish internet browsing, customers might face access issues on the firm's website, pointless popping up of messages at the time of customer's log-in, abrupt changes to system passwords or accounts, browser warning of errors. Moreover, the concerned users might face infection alarms through antivirus toolkits installed in their systems. Thus, these signals of system malfunction can negatively influence the trust and confidence of customers. Likewise, the firms' MV in the stock market might be influenced, and

**Long-run Market
Value of Firms**

Security Breach
Announcement

H_1 → Abnormal returns before
announcement

H_2 → Abnormal returns after
announcement
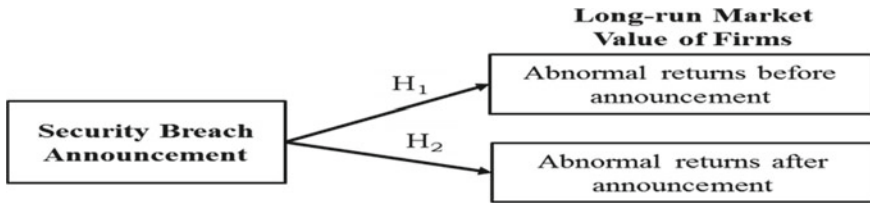
Fig. 1 A conceptual framework for analyzing the impact of security breaches on long-run MV

they might have allocated a likelihood of a security breach for the concerned firm. It is also probable that the financial impact of a security breach might have already incorporated by investors in their stock returns even before the actual security breach announcement. Hence, it can be hypothesized that:

H1: In the period before the announcement, abnormal returns of security breached firms will be negative.

Long-run Abnormal Returns in the period after breach announcement. To fully evaluate the financial influence of security breaches, it is also vital to analyze stock price behavior in the post-announce era. Studies in recent times have also reported statistically unusual long-run stock price reaction ensuing to the announcement of an event. Examples include repurchase offers, spin-offs, initiations and omissions on dividends,open market repurchases, stock splits, seasoned equity offerings, shareholder fights, and initial public bidding (see [35] all such studies). Likewise, security breaches can also impact firm financials even after the announcement as consumers and suppliers respond to security breaches. Since these impacts must be part of quantifying the overall economic impact of security breaches, stock returns should be measured over extended timelines ensuing an announcement. Accordingly, there are equal possibilities of positive and negative returns in the post-announcement period. Our second hypothesis is:

H1b: In the period after the announcement, abnormal returns of security breached firms will be negative.

The conceptual framework of this study is exhibited in Fig. 1. The announcement of a security breach is an exogenous construct. At the same time, the endogenous construct is demonstrated by long-run MV.

3 Methodology

The methods and estimation techniques we have proposed in measuring the long-run abnormal returns are different from those usually used in event studies when the short-run impact of events is analyzed on the stock market. As the methodology of event study frequently provides skewed approximations of both the ultimate economic influence and the test statistics [36, 37]. Our proposed methodology is based on robust

and more accurate methods, which were newly established and used frequently in studies [34, 38–42].

3.1 Assessing the Long-Run Abnormal Returns

The literature argues as to how to measure long-run abnormal returns [35, 36]. The first problem is the relevant variables to be controlled for calculating abnormal long-run returns. Previous studies on long-run stock valuation have been primarily controlled for a company's systemic risk (or beta). According to literature, the size, market-to-book ratio, and previous performance are imperative predictors of stock returns [43–45]. The present consensus, therefore, appears to be that abnormal returns have to be determined after controlling size, market-to-book ratio, and previous performance [38].

Buy-and-hold abnormal returns (BHARS) using one-to-one match samples:

Our proposed approach is one-to-one matching in which each sample firm is compared to the appropriate control firm having similar size, market-to-book ratio, and prior performance. Then it is recommended to create two individual samples of one-to-one:

i.   Choose a firm nearest in size to the sample firm (size-matched).
ii. Choose a firm nearest to the sample firm in terms of its previous performance from the sample firm (performance-matched).

The abnormal return is the gap between the sample firm's BHARs and its matched control firm, ensuring consistency with the null hypothesis. Accordingly, both parametric and non-parametric tests can be used for statistical inferences. This paper provides an estimate of abnormal buy-and-hold-abnormal-returns (BHARs) by daily return statistics. The raw returns and the return of the sample company are compounded first through the frame for BHARs measurements. The abnormal return is the disparity between the sample firm's cumulative returns and its benchmark (Fig. 2). BHARs are determined as:
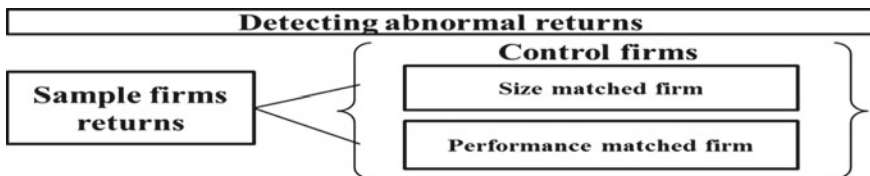
Fig. 2 One-to-one matched sampling methodology [36]
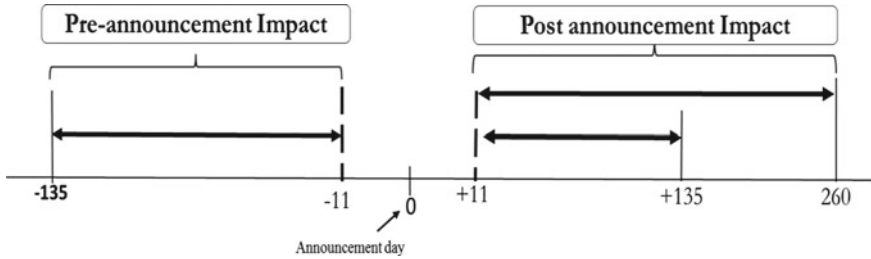
Fig. 3 Mapping of calendar time for assessment of MV

$$BHAR_i = \prod_{t=1}^{T}(1 + R_{it}) - \prod_{t}^{T}(1 + R_{bt}) \qquad (1)$$

$$\overline{BHAR_{it}} = \frac{1}{N}\sum_{i=1}^{N} BHAR_i \qquad (2)$$

3.2 Mapping of Security Events for Assessment of MV

To collect findings over time, we conceptualize to map the calendar to date of the occurrence for each entity in our sample. The day of the announcement is day 0, the next day for trade is day one, and the day before the announcement is day-1, etc. It is advised to forecast abnormal returns for at least 18 months, beginning six months prior announcements through two periods of six months each after the security breach notification. It consists of 250 trading days each year.

Furthermore, it contemplated deducting from both sides a 2-week duration (10 trading days) while measuring abnormal changes in stock returns. Moreover, security breaches may, in the period before the disclosure, have already impacted stock prices, and thus, the MV. Therefore, before the actual announcement, abnormal changes may occur (Fig. 3). In the following three periods of six months each, we degree buyand-hold abnormal returns and standard deviation from daily stock returns.

- *Six months pre-announcement: trading days -135 to -11,*
- *Six months post-announcement: trading days 11–135,*
- *Twelve months post-announcement: trading days 136–260.*

4 Conclusion

Findings in future studies based on the proposed framework might emphasize the significance of security and determine that it cannot be disregarded. Events of security

breach do not occur unwittingly, and over time, users have a lower level of tolerance for such events. Also, the news or media attention exposes the presence of flaws in the firm's security, rendering the firm more likely to be the victim of such assaults in the future. Taking SONY as an example, this reveals that SONY's loss of MV of 293 million dollars a year after the incident and total loss of 937 million two years after the event would result from a security breach event. It implies that investors would reevaluate the worth of the company in the stock exchange as they face tangible and intangible risks from security breach incidents.

References

1. Collins A (2019) The global risks report 2019. World economic forum, Geneva, January 15 2019. Available: https://www.weforum.org/reports/the-global-risks-report-2019
2. Collins A (2018) The global risks report 2018. World Economic Forum, Geneva, January 17 2018. Available: https://www.weforum.org/reports/the-global-risks-report-2018
3. Smith KT, Jones A, Johnson L, Smith LM (2019) Examination of cybercrime and its effects on corporate stock value. J Inf Commun Ethics Soc 17(1):42–60
4. Jansen J, Junger M, Montoya L, Hartel P (2013) Offenders in a digitized society. In: Cybercrime and the Police. Eleven International Publishing, Hague, ND, pp 45–59
5. Lukonga I (2018) Fintech, inclusive growth and cyber risks: focus on the MENAP and CCA regions. In: IMF working papers, vol 18. Available: https://doi.org/10.5089/978148437490 0.001
6. Cavusoglu H, Mishra B, Raghunathan S (2004) The effect of internet security breach announcements on market value: capital market reactions for breached firms and internet security developers. Int J Electron Commer 9(1):70–104. https://doi.org/10.1080/10864415.2004.110 44320
7. Bose I, Leung ACM (2014) Do phishing alerts impact global corporations? A firm value analysis. Decis Support Syst 64(August):67–78
8. Ettredge M, Richardson VJ (2002) Assessing the risk in e-commerce. In: Presented at the annual Hawaii international conference on system sciences. Hawaii, US
9. Campbell K, Gordon LA, Loeb MP, Zhou L (2003) The economic cost of publicly announced information security breaches: empirical evidence from the stock market. J Comput Secur 11(3):431–448. https://doi.org/10.3233/JCS-2003-11308
10. Hovav A, D'Arcy J (2004) The impact of virus attack announcements on the market value of firms. Inf Syst Secur 13(3):32–40. https://doi.org/10.1201/1086/44530.13.3.20040701/830 67.5
11. Kannan K, Rees J, Sridhar S (2007) Market reactions to information security breach announcements: an empirical analysis. Int J Electron Commer 12(1):69–91. https://doi.org/10.2753/JEC 1086-4415120103
12. Malhotra A, Kubowicz Malhotra C (2011) Evaluating customer information breaches as service failures: an event study approach. J Service Res 14(1):44–59
13. Sinanaj G, Muntermann J (2013) Assessing corporate reputational damage of data breaches: an empirical analysis. In: Presented at the 26th Bled E-Conference—E-Innovations challenges and impacts for individuals, organizations and society. Bled, Slovenia

14. Tweneboah-Kodua S, Atsu F, Buchanan W (2018) Impact of cyberattacks on stock performance: a comparative study. Inf Comput Secur 26(5):637–652. https://doi.org/10.1108/ICS-05-2018-0060

15. Fama EF, Fisher L, Jensen MC, Roll R (1969) The adjustment of stock prices to new information. Int Econ Rev 10(1):1–21. https://doi.org/10.2307/2525569

16. Schwartz RA (1970) Efficient capital markets: a review of theory and empirical work: discussion. J Financ 25(2):421–423

17. Ponemon (2020) Cost of a data breach report, vol 2020. IBM Security, Michigan, USA. Available: https://www.ibm.com/security/digital-assets/costdata-breach-report/#/

18. Hovav A, Han JY, Kim J (2017) Market reaction to security breach announcements: evidence from South Korea. Data Base Adv Inf Syst 48(1):11–52. https://doi.org/10.1145/3051473.3051476

19. Gordon LA, Loeb MP (2006) Budgeting process for information security expenditures. Commun ACM 49(1):121–125

20. Gordon LA, Loeb MP (2002) The economics of information security investment. ACM Trans Inf Syst Secur 5(4):438–457

21. Hsu C, Lee JN, Straub DW (2012) Institutional influences on information systems security innovations. Inf Syst Res 23(3–2):918–939. https://doi.org/10.1287/isre.1110.0393

22. Lowry PB, Cao J, Everard A (2011) Privacy concerns versus desire for interpersonal awareness in driving the use of self-disclosure technologies: the case of instant messaging in two cultures. J Manag Inf Syst 27(4):163–200. https://doi.org/10.2753/MIS0742-1222270406

23. Hovav A, D'Arcy J (2005) Capital market reaction to defective IT products: the case of computer viruses. Comput Secur 24(August):409–424

24. Deane JK, Goldberg DM, Rakes TR, Rees LP (2019) The effect of information security certification announcements on the market value of the firm. Inf Technol Manage 20(3):107–121. https://doi.org/10.1007/s10799-018-00297-3

25. Wang T, Kannan KN, Ulmer JR (2013) The association between the disclosure and the realization of information security risk factors. Inf Syst Res 24(2):201–218. https://doi.org/10.1287/isre.1120.0437

26. Gordon LA, Loeb MP, Sohail T (2010) Market value of voluntary disclosures concerning information security. MIS Q 34(3):567–594. https://doi.org/10.2307/25750692

27. Goldstein J, Chernobai A, Benaroch M (2011) An event study analysis of the economic impact of IT operational risk and its subcategories. J Assoc Inf Syst 12(9):606–631. https://doi.org/10.17705/1jais.00275

28. Cao J, Calderon T, Chandra A, Wang L (2010) Analyzing late SEC filings for differential impacts of IS and accounting issues. Int J Account Inf Syst 11(3):189–207. https://doi.org/10.1016/j.accinf.2010.07.010

29. Rosati P, Deeney P, Cummins M, Van der Werff L, Lynn T (2019) Social media and stock price reaction to data breach announcements: evidence from US listed companies. Res Int Bus Financ 47(September):458–469

30. Jeong CY, Lee SYT, Lim JH (2019) Information security breaches and IT security investments: impacts on competitors. Inf Manag 56(5):681–695. https://doi.org/10.1016/j.im.2018.11.003

31. Johnson MS, Kang MJ, Lawson T (2017) Stock price reaction to data breaches. J Financ Issues 16(2):1–12

32. Filbeck G, Swinarski M, Zhao X (2013) Shareholder reaction to firm investments in the capability maturity model: an event study. Eur J Inf Syst 22(2):170–190

33. Dehning B, Richardson VJ, Zmud RW (2007) The financial performance effects of IT-based supply chain management systems in manufacturing firms. J Oper Manag 25(4):806–824

34. Hendricks KB, Singhal VR, Stratman JK (2007) The impact of enterprise systems on corporate performance: a study of ERP, SCM and CRM system implementations. J Oper Manag 25(1):65–82

35. Fama EF (1998) Market efficiency, long-term returns, and behavioral finance. J Financ Econ 49(3):283–306

36. Barber BM, Lyon JD (1997) Detecting long-run abnormal stock returns: the empirical power and specification of test statistics. J Financ Econ 43(3):341–372. https://doi.org/10.1016/S0304-405X(96)00890-2
37. Kothari S, Warner JB (1997) Measuring long-horizon security price performance. J Financ Econ 43(3):301–339
38. Lyon JD, Barber BM, Tsai CL (1999) Improved methods for tests of long-run abnormal stock returns. J Financ 54(1):165–201. https://doi.org/10.1111/0022-1082.00101
39. Lee I, Loughran T (1998) Performance following convertible bond issuance. J Corp Finan 4(2):185–207
40. Hendricks KB, Singhal VR (2014) The effect of demand–supply mismatches on firm risk. Prod Oper Manag 23(12):2137–2151
41. Hendricks KB, Singhal VR (2005) An empirical analysis of the effect of supply chain disruptions on long-run stock price performance and equity risk of the firm. Prod Oper Manag 14(1):35–52
42. Ali SEA, Khurram S (2017) Impact of demographic and health factors on GDP growth of South Asian Countries. Int J Acad Res Bus Soc Sci 7(3):2222–6990
43. Carhart MM (1997) On persistence in mutual fund performance. J Financ 52(1):57–82
44. Fama EF, French KR (1996) Multifactor explanations of asset pricing anomalies. J Financ 51(1):55–84
45. Jegadeesh N, Titman S (1993) Returns to buying winners and selling losers: implications for stock market efficiency. J Financ 48(1):65–91
46. Ali SEA, Lai F-W, Dominic PDD, Brown N, Lowry PB, Ali RF (2021) Stock market reactions to favorable and unfavorable information security events: A systematic literature review. Comput Secur 110(November):102451
47. Ali SEA, Lai F-W, Hassan R (2020) Socio-economic factors on sector-wide systematic risk of information security breaches: Conceptual framework. Paper presented at the 9th International Economics and Business Management Conference, Melaka, Malaysia

Towards Early Distribution of Container-Based Microservices in Cloud Computing Environment

Abdul Saboor, Ahmad Kamil Mahmood, Mohd Fadzil Hassan, Syed Nasir Mehmood Shah, Muhammad Aadil Siddiqui, Saeed Ahmed Magsi, and Muhammad Junaid

Abstract Microservices gained the popularity for use in scalable cloud application. The application based on microservices uses intensive network communication to call other microservices or to pass on the messages to hundreds of microservices running inside the cloud/edge nodes. This study addressed the early distribution of container-based microservices and proposed two distribution strategies named as Random Distribution and Design Pattern Distribution. In random distribution approach the microservices are assigned arbitrarily to the available data centers. While in design pattern distribution the microservices are grouped together on the basis of behavioral design patterns which identifies common communication patterns among objects. The proposed solution was tested using custom built simulation environment and the results showed that the early distribution of microservices according to the design pattern of the application resulted in significant reduction of network calls to the microservices hosted at other network nodes or data centers.

Keywords Cloud/edge computing · Microservices · Container · Optimization

A. Saboor (✉) · A. K. Mahmood
High Performance Cloud Computing Center (HPC), Universiti Teknologi Petronas, 32610 Sri Iskandar, Malaysia
e-mail: abdul_19001745@utp.edu.my

A. K. Mahmood
e-mail: kamilmh@utp.edu.my

M. F. Hassan
Institute of Autonomous System for Autonomous Facilities, Universiti Teknologi Petronas, 32610 Sri Iskandar, Malaysia
e-mail: mfadzil_hassan@utp.edu.my

S. N. M. Shah
KICSIT, Institute of Space Technology (IST), Islamabad, Pakistan

M. A. Siddiqui · S. A. Magsi · M. Junaid
Universiti Teknologi Petronas (UTP), 32610 Sri Iskandar, Malaysia

1 Introduction

In the last decade cloud computing has gained interest of the industries and research institutes. The Cloud Computing enables ubiquitous computing and provides the convenient on-demand access to the shared pool of configurable resources, such as storage, network, server, applications and services [1]. The cloud data centers mainly relies on the concept of virtualization to provide access to the shared resources and cloud services [2]. For the last few years there had been a shift towards the containerization. The concept of containers creates lightweight and flexible environment by allowing applications to share operating system. Recent advancement in container technology have encouraged organization and researchers to use the containers in the cloud environment. The adoption of containers helped in realization of microservices at cloud level. Microservices offer number of advantages such as scalability, reliability, re-usability, fast response time, reduced cost, and many more [3, 4]. Microservices also comes with number of challenges. The microservices adoption challenges include security vulnerabilities [5, 6], finding the right size and number of services [3].

As microservices are independent units and often needs to interact with other microservices. The large number of microservices and their interaction thus increases the communication overhead. One way to provide reliable and fast communication is by decreasing the communication among physical machine hosted at multiple data centers. This study suggests different ways to distribute the microservices on physical machines so that the microservices have to use lesser communication network. The microservices are ordered/managed in such a way that they mostly interact with the microservices available at same physical machine and do not have to interact most of time with the microservices hosted on other physical machines.

2 Literature Review

In cloud computing one of the most critical factor is to provide quality of service (QoS) within the defined constraints. Traditionally, heuristic scheduling algorithms had been used to provide QoS by minimizing the processing time and schedule tasks efficiently [7], but the design shift of cloud services from monolithic applications to loosely coupled micro-services brought in new challenges such as increased makespan, heavy network utilization, and increased cost. These challenges are addressed by the researchers time to time. According to Zahang [8] container placements on VM is an issue. The VM placement and container placement should not be addressed individually, therefore Zahang proposed a solution named *Container-VM-PM*. The solution resulted in optimization of the placement of new containers and minimized the number of physical machines and resource wastage.

In a study by Lin [9], the container-based microservice scheduling was done based on multi-objective optimization method using an ant colony optimization algorithm.

The solution took into account the utilization of computing and memory resources of physical machines However, the study considered microservice scheduling in single container only, and did not considered the producer consumer relationship of the microservices. Similarly, Lv et al. [10] proposed the container distribution strategies to address the problem of heavy communication workloads among containers. For container placement they proposed the worst fit decreasing algorithm, and for container assignment proposed a two stage Sweep and Search algorithm. The experimental results showed the reduction in communication overhead among the containers.

Zhou et al [11] proposed the optimal placement schema for containers in the cloud environment. They introduced one-shot algorithm which work as placement schema for the container cluster, and an online algorithm which breaks down the online decision making into on-spot decision depending on resource price.

Recently another study by Sampaio et al [12] proposed an adaption mechanism, named as REMaP (Run-time Microservices Placement). In this model the automatic run-time placement of microservices is based upon usage history of resources and mciroservices affinity.

The literature review showed one common concern that heavy network communication and latency is one of the major issue in container based cloud environment. Especially the decomposition of a particular application into many small microservices results into large round trips between service calls. The possible solution could be aggregating multiple services or batch service call in a single round trip, or replacing expensive inter-process communication call with language-level function calls [13]. This study propose an early distribution of container-based microservices in cloud environment such that the microservices are grouped in such a way that it results in lower number of network calls. The solution addressed the distribution of microservices and containers at early stages of application deployment.

3 Methodology

For this paper the distribution of microservices is considered in such a way that one container hosts one and only one microservice in the cloud/edge environment. This section describes two proposed microservice distribution approaches i.e. Randomized Distribution, and Design Pattern Distribution.

3.1 Randomized Distribution

The random distribution and assignment approach does not take into account any performance parameters while assigning the containers to randomly chosen physical machine of the data centers, refer Fig. 1. The services are deployed arbitrarily
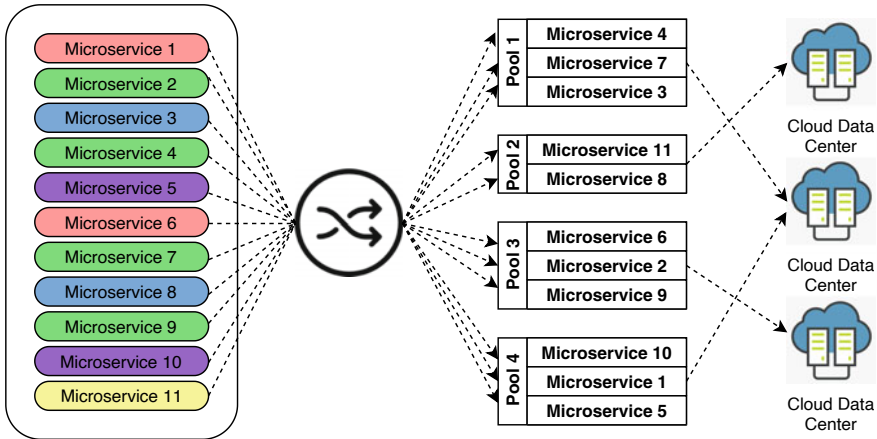
Fig. 1 Random selection and distribution of microservices to the data centers

i.e. without any consideration of performance parameters. Therefore this way of assignment can also be termed as a classical approach.

The microservices are assigned randomly and the assigned containers are distributed across random location. Thus resulting in no consideration of design pattern or microservices interaction.

The random distribution approach is shown in Fig. 1 where the randomize function selects a microservices on random basis and assign it to the random pools. Then these pools of microservices are randomly assigned to the available data centers.

In such a random distribution there is high chance that a microservice assigned to on data center is totally dependent on a microservice available at other data center. Thus, the execution of such microservices is likely to place multiple network communication calls between services for successful execution.

3.2 Design Pattern Distribution

The term *pattern* used by software designers describes the software abstraction, thus patterns used across other procedures, subroutines, and objects may also combine for more common abstraction [14]. According to Erl [15] it is set of related patterns which provides the building blocks and these building blocks are built upon the other blocks thus defining the pattern application sequence. For this study we considered the behavioral patterns as these design patterns are concerned with communication between objects.
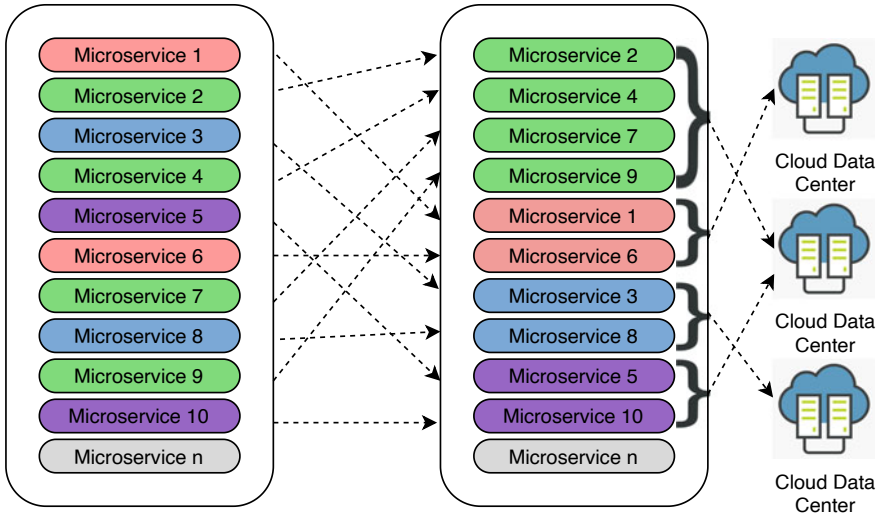
Fig. 2 Arrangement of microservices according to their interaction dependencies and assignment to the data centers

The proposed design pattern distribution schema is shown in Fig. 2, where microservices are assigned according to the relation and interaction defined in software design. If two or more microservices are dependent on each other for the execution of services, then those microservices are pooled together. Then these pools are assigned to the respective data centers.

3.3 Results and Discussion

To assess the proposed techniques, we designed the simulation program using Java to produce the microservices utilization data sets. To get the insights of the produced data sets we used an online tool called plotly [16]. Plotly provided online analytic and visualization of the data sets. The data sets were produced for one hour of microservices calls, and for the assessment of the proposed system it was considered that maximum of 1000 calls can be originated for the microservices at any given time. The anatomized simulation process however reached maximum of 729 simultaneous microservices calls at a given time. The number of service calls originated from the system are shown in Fig. 3. They are the simulated calls for microservices which will be used for testing the random and design pattern distribution. To get the test results for classical random approach, the service calls were originated for the microservices distributed randomly at different location i.e. data centers. The number of network calls placed by the microservices in random distribution are shown in Fig. 4. It clearly
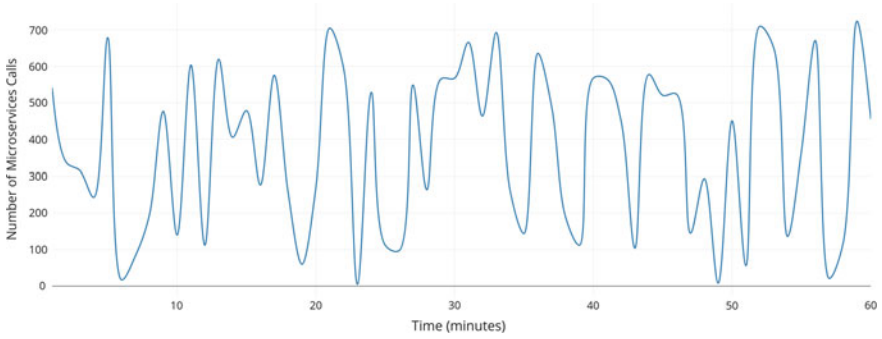
Fig. 3 Number of microservices calls originated from the system over a period of one hour
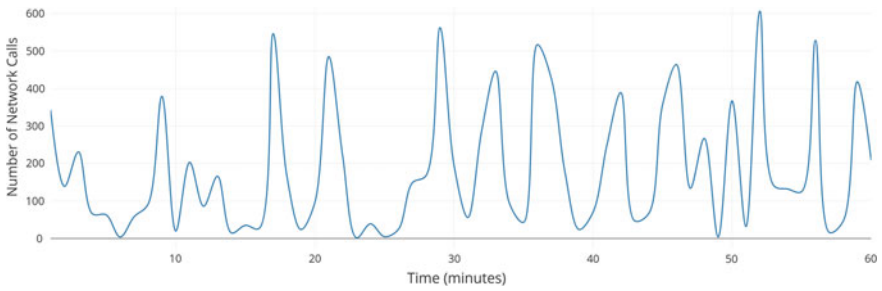
Fig. 4 Number of network calls in random distribution of microservices

shows that heavy network traffic is generated by such a distribution. On average 189 network calls were placed at a given time.

In the second simulation the design pattern distribution strategy was tested and the results showed that the number of network calls reduced significantly (refer Fig. 5). On average 84 network calls we placed at a given time, which are very low when compared to random distribution. The comparison of both distribution strategies is shown in Fig. 6.
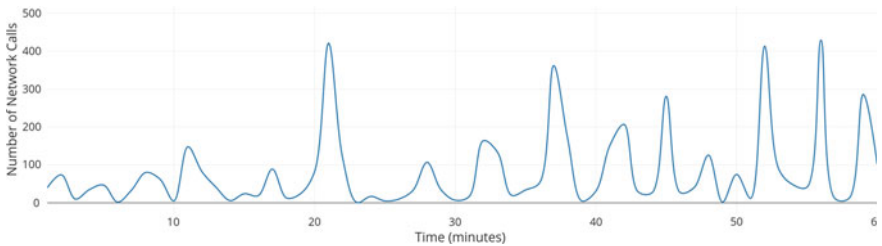
Fig. 5 Number of network calls in design pattern distribution of microservices
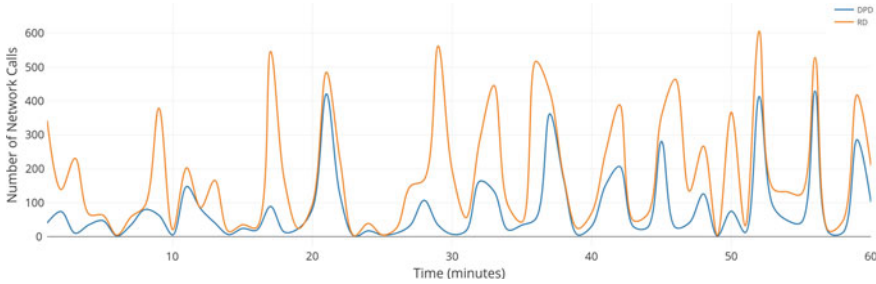
Fig. 6 Comparison of random distribution (RD) and design pattern distribution (DPD) network calls

During the multiple simulation run of the system, it was noted that for couple of time the design pattern based distribution of network calls exceeded the random distribution network calls, but it was rare. This was due to the ideal match of the microservices calls and the distribution done during the random distribution. Thus less network calls placed in such a configuration were by coincidence.

3.4 Conclusion and Future Work

The design shift of cloud services from monolithic applications to loosely coupled micro-services brought in many challenges. This study addressed the communication issues and proposed two microservices distribution strategies to reduce the network calls. The random distribution and design pattern distribution were suggested, and the simulation tests showed that the early distribution of microservices according to the design pattern of application significantly reduce the network call to other microservices.

The proposed system was tested using a custom built data sets and the system further needs to be tested using real cloud traces. In future the study will also be extended by defining other early distribution strategies such as priority based distribution and weighted mean average distribution.

References

1. Mell P, Grance T, et al (2011) The nist definition of cloud computing
2. Gong C, Liu J, Zhang Q, Chen H, Gong Z (2010) The characteristics of cloud computing. In: 2010 39th international conference on parallel processing workshops. IEEE, pp 275–279
3. Amaral M, Polo J, Carrera D, Mohomed I, Unuvar M, Steinder M (2015) Performance evaluation of microservices architectures using containers. In: 2015 IEEE 14th international symposium on network computing and applications. IEEE, pp 27–34

4. Malavalli D, Sathappan S (2015) Scalable microservice based architecture for enabling dmtf profiles. In: 2015 11th international conference on network and service management (CNSM). IEEE, pp 428–432
5. Newman S (2015) Building microservices: designing fine-grained systems. "O'Reilly Media, Inc."
6. Yu D, Jin Y, Zhang Y, Zheng X (2019) A survey on security issues in services communication of microservices-enabled fog applications. Concurrency Comput: Pract Experience 31(22):e4436
7. Juarez F, Ejarque J, Badia RM (2018) Dynamic energy-aware scheduling for parallel task-based application in cloud computing. Futur Gener Comput Syst 78:257–271
8. Zhang R, Zhong AM, Dong B, Tian F, Li R (2018) Container-vm-pm architecture: A novel architecture for docker container placement. In: International conference on cloud computing. Springer, pp 128–140
9. Lin M, Xi J, Bai W, Wu J (2019) Ant colony algorithm for multi-objective optimization of container-based microservice scheduling in cloud. IEEE Access 7:83088–83100
10. Lv L, Zhang Y, Li Y, Xu K, Wang D, Wang W, Li M, Cao X, Liang Q (2019) Communication-aware container placement and reassignment in large-scale internet data centers. IEEE J Sel Areas Commun 37(3):540–555
11. Zhou R, Li Z, Wu C (2019) An efficient online placement scheme for cloud container clusters. IEEE J Sel Areas Commun 37(5):1046–1058
12. Sampaio AR, Rubin J, Beschastnikh I, Rosa NS (2019) Improving microservicebased applications with runtime placement adaptation. J Internet Services Appl 10(1):1–30
13. Richardson C (2019) Microservices patterns: with examples in Java. Manning publications
14. Coplien JO (1998) Software design patterns: common questions and answers. In: The patterns handbook: techniques, strategies, and applications, vol 13, pp 311
15. Erl T (2008) SOA design patterns (paperback). Pearson Education
16. Sievert C, Parmer C, Hocking T, Chamberlain S, Ram K, Corvellec M, Despouy P (2017) Plotly: create interactive web graphics via 'plotly. js'. R package version 4(1):110

An AI-Based Chicken Disease Management System

Luyl-Da Quach , **Nghi Pham Quoc** , **Nhien Huynh Thi** ,
Nhan Le Thanh , **Xuan Truong Thi Thanh** , **Duc Chung Tran** ,
and Mohd Fadzil Hassan

Abstract Food stability has always received worldwide attention, especially in the development of the poultry industry. However, the poultry's diseases have caused the loss of the poultry population and direct income of the owners. In this work, a system comprising of a website and a mobile application has been developed to support disease identification in chickens; the disease spread locations can be managed and traced with maps. In addition, the system allows farmers to make contact and get support from experts through chat, voice calls, and video calls. It also has some e-commerce functions for improving sales of poultry's owners. The system uses the improved ResNet-50 model, with an accuracy of about 93.56%.

Keywords Disease identification in chickens · Identify disease spots · Application for livestock · ResNet-50

L.-D. Quach (✉) · N. P. Quoc · N. H. Thi · N. Le Thanh · X. T. T. Thanh
Software Engineering Department, FPT University, Cantho city, Vietnam

N. P. Quoc
e-mail: nghipqce140179@fpt.edu.vn

N. H. Thi
e-mail: NhienHTCE140122@fpt.edu.vn

N. Le Thanh
e-mail: NhanLTCE140044@fpt.edu.vn

X. T. T. Thanh
e-mail: XuanTTTCE140149@fpt.edu.vn

D. C. Tran
Computing Fundamental Department, FPT University, Hanoi, Vietnam
e-mail: chungtd6@fe.edu.vn

M. F. Hassan
Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Seri Iskandar, Perak, Malaysia
e-mail: mfadzil_hassan@utp.edu.my

1 Introduction

According to the World Health Organization (WHO), as of September 2020, the world has about 30 million patients infected with COVID-19, of which 940,000 deaths. This shows the rapid spread of the epidemic in recent years. Therefore, to limit the spread of disease, governments of countries have introduced policies to restrict travel, avoid crowding, isolate society, etc. However, the demand for food is continuously increasing, especially livestock and agriculture.

According to the United States Department of Agriculture statistics on July 10, 2020, despite the epidemic of COVID, the demand for food imports from China was increasing, especially chicken products. Despite the decline from the US and China, production is still about 1% higher than the world (volume reached 100 million tons) [1]. The demand for poultry products is increasing. According to the forecasts 2016–2025 of The Organization for Economic Co-operation and Development (OECD), the output of poultry meat from consumer demand in 41 countries and 12 regions could increase from 115,247 million tons (in 2006) to 131,225 million tons (in 2025) [2].

The poultry industry's development also entails many risks, not only causing economic losses but also affecting the ability to expand the scale of livestock production. According to statistics from the World organization for animal health, in the period from July 10 to July 30, there are about 39,161 birds destroyed in Africa and Asia with the H5N8 avian flu standard [3] appeared. Some new countries, such as the Philippines and South Africa. In 2014, the H5N2 disease was incubated between November 26–28 and December 2, which resulted in the destruction of over 48 million birds in a short period [4]. Also, poultry pathogens such as Escherichia coli, Salmonella, and Campylobacter jejuni can directly affect humans by consuming poultry products [5]. Therefore, developing technology for monitoring and monitoring is a very urgent matter.

In the world, there has been researched on diagnosing and detecting diseases in plants and animals. Astill [6] conducted research on the identification and detection of avian influenza virus based on new big data technology. Meanwhile, Hemalatha [7] identify chickenpox with machine learning methods such as Support Vector Machine in conjunction with Gaussian Radial Basis Function and Extreme Learning Machine, the uses in this study using grey-level matrices with the calculation of mean, standard deviation, noise, etc. for statistics. Based on the humming sound signals in chickens, Rizwan [8] utilized the extreme machine learning method and support vector machine to classify healthy and active chickens. Specifically, disease data is recorded in 20 min per day and labeled of 25 consecutive days. The results showed potential automatic sound monitoring of flock health. The above studies gave us that there have been both accuracy and feasibility in diagnosing and detecting diseases. However, the urgency of having an automatic system of identification and diagnosis of diseases in chickens using artificial intelligence applications is also implied more and more clearly. Breeders can identify the problems related to their chickens timely and promptly. According to Astill [9], some standard methods of detecting disease

in poultry, such as biosensor, wearable sensor, image analysis, surveillance robot, etc. were developed. They were unable to promote infectious diseases and limited to having to install on individual instances, though. Since then, it was essential to develop a new system based on monitoring, machine learning, and history tracking.

The identification of chicken diseases based on machine learning and deep learning methods has been carried out in many nations and territories, consisting of Vietnam. Previously, Quach [10] performed the experimental research for utilizing the deep learning model ResNet-50, and VGGNet to evaluate four diseases in chicken with 74.1% accuracy. Currently, the group has examined to integrate the ResNet-50 model, SURF characteristics, and K-means clustering method to improve the accuracy in identifying disease in chicken. The results proved that this method helped increase the model's accuracy to 93.56% and the number of features available for training [11].

In this research, a system for identifying and managing disease zones in poultry using artificial intelligence was built. The system formed two basic components, such as:

- Mobile application for smallholder farmers using the trained ResNet-50 model focuses on disease identification in poultry through images. The application also has additional functions such as helping users purchase livestock products such as feed, veterinary medicine, and vaccines from stores linked to the system. An important and indispensable function is to contact veterinarians for their advice.
- The website manages the poultry disease zone based on the user's coordinates when they identify the disease. The coordinates will be displayed on the map to help experts evaluate the disease area for timely treatment measures. Besides, the website needs an interface for experts to advise users when needed.

2 Methodology

2.1 Research Methods

The methods are applied include as follows:

- Document study method: Based on documents about models, algorithms, project implementation, and libraries.
- Experimental methods: Collect data, train models, write software, and test the system.
- Statistical method: Evaluating the accuracy of the model, assessing the model's reliability.

2.2   Research Content

Scrum is an Agile software development model used in small to medium projects [12]. By dividing the project into small phases to easily manage and adapt to changes in the development process. The stages used in the study included:

- Stage 1: Overall research planning, implementation planning, and development goals to be achieved.
- Stage 2: Analyzing the overall requirements of the project, thereby identifying the problem to be built, the level of prediction, and the system to be met, etc.
- Stage 3: Data collection and training. This stage was done using:

  - Programming language: Python.
  - Libraries: Tensorflow, Sklearn.

- Stage 4: Design and program the system based on phases 1, 2, and 3. This stage was done using:

  - Programming language: Python, Javascript.
  - Framework: Flash, ReactJS, React Native, Bootstrap4.
  - API: Mapbox, Facebook.
  - Database management: Sqlite.

3   Application

To build the system, research based on the requirements, the process of finding out management problems in practice, and building the model include the Main functional model of the system (use case model), relational database model, business model, and application development.

3.1   Disease Identification System

The study collected 492 images, divided into four disease classes: avian pox, bird-flu, Marek, and infectious laryngotracheitis. Details of the number and illustrations of the disease are shown in Table 1 and Fig. 1.

After obtaining the data, the study conducted steps aimed at data transformation, including the following steps:

- Step 1: Use the SURF feature [12] to perform characteristic extraction from diseased chickens. Each SURF feature has 64 vectors with parameters 5000 and 10,000. The image is depicted as Fig. 2 and the quantity in Table 2.
- Step 2: Research using the K-Means algorithm [13] with parameter k = 4 to filter the pixels from the obtained feature vector. As a result, the study obtained a result

Table 1 Statistics of chicken disease image number

Labels	Total species
Avian pox	138
Bird-flu	147
Marek	96
Infectious laryngotracheitis	111
Total	492

Fig. 1 Illustrative image of chicken diseases, from left to right: Bird-flu, Avian pox, Infectious laryngotracheitis, Marek
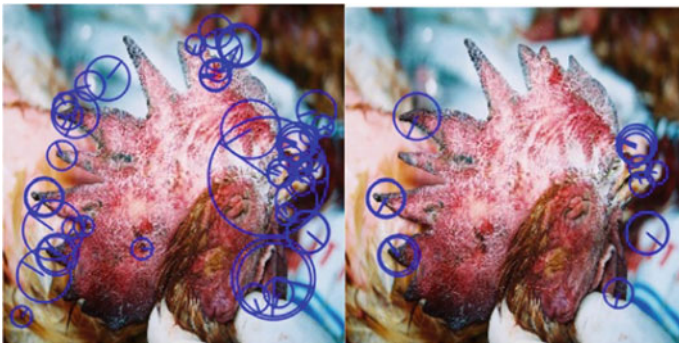
Fig. 2 Extract the SURF feature with parameters 5000 and 10,000

Table 2 Statistic number of characteristics collected from chicken disease data collection

Labels	Parameters	Number of features
Avian pox	5000	2568
Bird-flu	5000	5272
Marek	5000	2191
Infectious laryngotracheitis	10,000	1310
Total		11,341

Table 3 Set features after remaining options

Labels	Total species
Avian pox	472
Bird-flu	475
Marek	326
Infectious laryngotracheitis	120
Total	1393

Fig. 3 Examples of samples selected after step (ii), from left to right, are diseases: Bird-flu, Avian pox, Marek, Infectious laryngotracheitis

set of 1393 characteristics of the four obtained diseases described in the table. The results obtained include 1393 characteristics to train with four different diseases with the number as Table 3 and illustrations Fig. 3.

- Step 3: The research using the ResNet-50 model [14] to train the disease identification dataset. The study conducted experiments, selected parameters and input size $64 \times 64 \times 3$ with default. The reseach got the best result was 93.56%, with a learning rate of 0.01, and the worst result on our test was 90.93, with a learning rate of 1. Efficien-cy The best yield of our adjusted model on training and test sets, learning rate $= 0.01$, is shown in Fig. 4. Research results are used to illustrate the process of disease identification for the system.

3.2 Main Use Case Model of the System

The use case diagram is used to represent the relationship between the use case and the system agent. The actors in the diagram are users (people), some internal or external applications. Use cases are a major function of the system [15]. In the system of research (see Fig. 5), parts of the use case diagram include:

- Target users: Administrators, owners of livestock products shops, veterinary establishments, and smallholder farmers.
- Use cases that describe the system's main functions are disease identification in chickens, contact advice with experts from veterinary facilities, e-commerce, map zone management, and user decentralization.
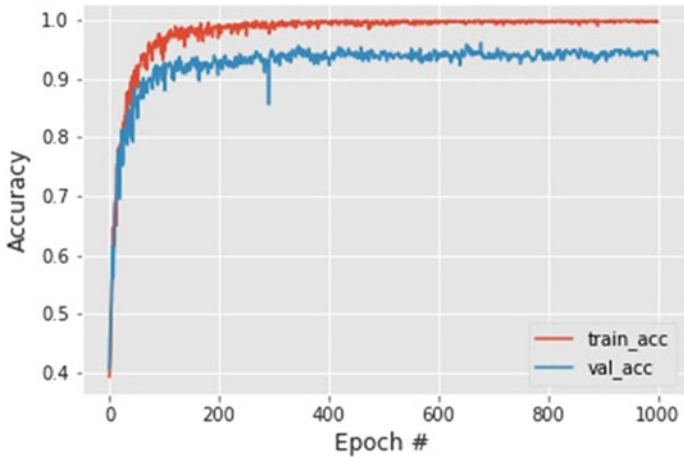
Fig. 4 Training results with learning rate $= 0.01$, epochs $= 1000$, input size $= 64 \times 64 \times 3$
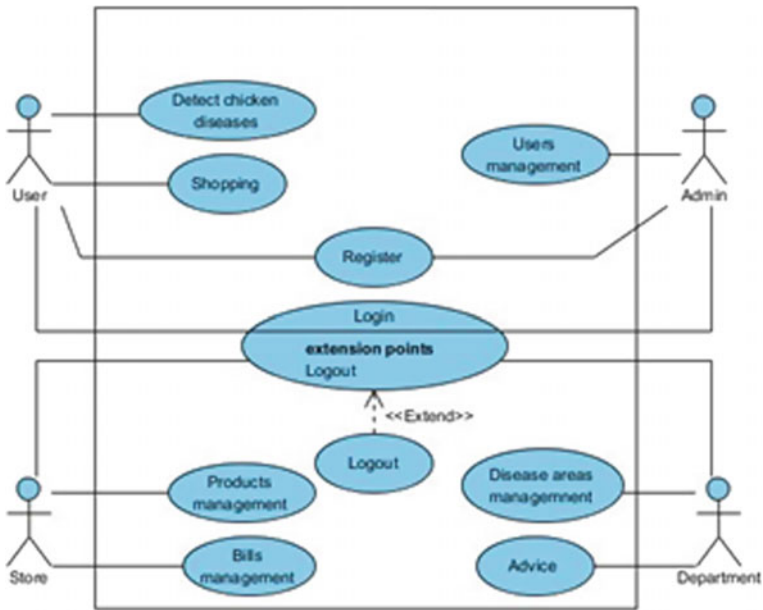
Fig. 5 Use case diagram in poultry disease management systems

3.3 Relational Database Model

The relational database model is the model that builds a digital database based on data relationships, as proposed by the E. F. Codd in 1970. A software system will

use this database to avoid data redundancy and ensure data consistency is described [17]. Therefore, research building a basic management system includes the entities. The system functions to meet the requirements set out include store management, products, users, etc.

3.4   Business Model

Based on the relational database model and use case diagram, the research team has built up the system's operation diagram with essential functions on two platforms. Dissipation includes:

1. Business Model of the Website System

    The system includes a home page to login; disease and density zoning maps; consulting contact by message, call, or video call. On the management website of the system, the research team used the streets-v9 map from Mapbox GL-JS [17] with the map center coordinates of [16] and the zoom level of 4.8 to allow the territory of Vietnam to be displayed immediately when users access and direct consulting function on the website (see Fig. 6).
2. Business Model of Mobile Application:

    The mobile phone's application includes authentication of users through registration and login, homepage, image-based diagnostics through library or camera, contact consultants such as system website, purchase, and farm management. Based on analytical models, the research team has developed a mobile application that includes functions for small poultry farmers (see Fig. 6).

Fig. 6 Video demo of the proposed system

4 Conclusion

With the available data, research has shown the feasibility of identifying the disease in chickens. With the combination of the SURF feature determination method, the K-means algorithm has helped the ResNet-50 model improve the training and recognition accuracy and avoid the limitations of the CNN model, such as saturation and overfitting 93.56% accuracy with four common diseases. The demo video of the application is shown in the QR code (Fig. 6). Through the application of the image recognition model and the deployed system, the system can develop in the following directions:

- Collecting data directly from users: The purpose is to increase data, improve model accuracy, and apply other CNN models.
- By zoning the data using the K-means algorithm, the model found that an individual chicken can suffer from many diseases. Therefore, the application of this model helps to detect any problems on the individual body.
- Improved model accuracy through a combination of image recognition and descriptive (text) data.
- Develop a farm management model according to world standards such as AsianGAP, GlobalGap, etc. from here; it is possible to query the origin of clean poultry meat to serve consumer needs.
- Develop an ecosystem to link farmers with veterinary facilities, livestock product suppliers to make e-commerce easier.
- The disease management system helps trace the origin of the disease and helps the regulator timely detect and localize the infection to avoid spreading to other areas and affecting people.

References

1. Kuberka L, Vuillemin J, Mezoughem C (2020) Livestock and poultry: world markets and trade. US Department of Agriculture Foreign Agricultural Service, Washington, DC, p 16
2. OECD (2016) Table 3.A1.4—World meat projections. In: OECD-FAO Agricultural Outlook 2016–2025. OECD Publishing, Paris. https://doi.org/10.1787/agr_outlook-2016-table102-en
3. World Organisation for animal health (2020) Highly pathogenic avian influenza (HPAI) report. No. 12 https://www.oie.int/fileadmin/Home/eng/Animal_Health_in_the_World/docs/pdf/OIE_AI_situation_report/HPAI_asof30072020.pdf. Accessed 08/2020
4. Shriner SA, Root JJ, Lutman MW, Kloft JM, VanDalen KK, Sullivan HJ et al (2016) Surveillance for highly pathogenic H5 avian influenza virus in synanthropic wildlife associated with poultry farms during an acute outbreak. Sci Rep 6:1–11. https://doi.org/10.1038/srep36237
5. Mellata M (2013) Human and avian extraintestinal pathogenic Escherichia coli: infections, zoonotic risks, and antibiotic resistance trends. Foodborne Pathog Dis 10(11):916–932
6. Astill J, Dara RA, Fraser E, Sharif S (2018) Detecting and predicting emerging disease in poultry with the implementation of new technologies and big data: a focus on avian influenza virus. Front Vet Sci 5:263. https://doi.org/10.3389/fvets.2018.00263

7. Hemalatha MS, Maheswaran R (2014) Recognition of poultry disease in real time using extreme learning machine. In: Procceedings of the internation conference on inter disciplinary research in engineering & technology 2014 [ICIDRET 2014], pp 44–50

8. Rizwan M, Carroll BT, Anderson DV, Daley W, Harbert S, Britton DF, Jackwood MW (2016) Identifying rale sounds in chickens using audio signals for early disease detection in poultry. In: 2016 IEEE global conference on signal and information processing (GlobalSIP). USA, pp 55–59. https://doi.org/10.1109/GlobalSIP.2016.7905802

9. Astill J, Dara RA, Fraser ED, Sharif S (2019) Corrigendum: detecting and predicting emerging disease in poultry with the implementation of new technologies and big data: a focus on avian influenza virus. Front Vet Sci 5:337

10. Quach L-D, Quoc NP, Tran DC, Hassan MF (2020) Identification of chicken diseases using VGGNet and ResNet models. In: EAI INISCOM 2020—6th EAI international conference on industrial networks and intelligent system, springer publishing (presented)

11. Quach L-D, Quoc ND, Thi NN, Tran DC, Hassan MF (2020) Using SURF to improve ResNet-50 model for poultry disease recognition algorithm. In: International conference on computational intelligence 2020. Malaysia

12. Bay H, Tuytelaars T, Gool L (2006) SURF: speeded up robust features. In: Leonardis A, Bischof H, Pinz A (eds) ECCV 2006, vol 3951. LNCS, pp 404–417

13. Steinley D, Brusco MJ (2007) Initializing k-means batch clustering: a critical evaluation of several techniques. J Classif 24(1):99–121

14. He K, Zhang X, Ren S, Sun J (2016) Identity mappings in deep residual networks. European conference on computer vision. Springer, Cham, pp 630–645

15. Khurana N, Chhillar RS, Chhillar U (2016) A novel technique for generation and optimization of test cases using use case, sequence. Act Diagram Genet Algorithm. JSW 11(3):242–250

16. Paredaens J, De Bra P, Gyssens M, Van Gucht D (2012) The structure of the relational database model. vol 17. Springer Science & Business Media. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=The+structure+of+the+relational+database+model&btnG=

17. Eriksson O, Rydkvist E (2019) An in-depth analysis of dynamically rendered vector-based maps with WebGL using Mapbox GL JS

Automatic Polyp Segmentation in Colonoscopy Images Using Single Network Model: SegNet

Chin Yii Eu, Tong Boon Tang, and Cheng-Kai Lu

Abstract Colorectal cancer is the third most common diagnosed cancer worldwide. Early detection and removal of adenoma during the colonoscopy examination may increase the survival probability. A novel computer-aided tool for automated polyp segmentation in colonoscopy images is described in this work. SegNet, a deep convolutional neural networks has been chosen to map low resolution features with the input resolution for automated pixel-wise semantic polyp segmentation. Publicly available databases, CVC-ClinicDB, CVC-ColonDB, and ETIS-LaribPolypDB were used to train and to test the model. The outcome demonstrated the proposed method is feasible as it attains an average of 81.78, 92.35% for mean intersection over union, and dice coefficient, respectively for testing on a combination of the aforementioned datasets.

Keywords Colorectal cancer · Computer-aided diagnosis · Convolutional neural network · Polyp segmentation

1 Introduction

Accounting all types of cancer, colorectal cancer is the third most common diagnosed cancer worldwide [1]. As the name indicates, colorectal cancer is a malignant tumor found in colon or rectum and often due to the specific type of polyp [2].

C. Y. Eu
Department of Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Malaysia

T. B. Tang
Centre of Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Malaysia

C. Y. Eu (✉) · C.-K. Lu
Smart Assistive and Rehabilitative Technology (SMART), Institute of Health and Analytics (IHA), Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Malaysia
e-mail: chengkai.lu@utp.edu.my

Currently, optical colonoscopy is the gold standard to screen colon for polyp detection and prevention of colorectal cancer. However, the polyp miss rate in colonoscopy screening is approximately 10–30% [3]. Missed and misdiagnosed polyp are due to various polyp structure [4]. Thus, computer-aided diagnosis (CAD) tool needs to be developed to help medical expert in enhancing the accuracy and promptness of medical diagnosis.

Automated polyp segmentation task can be presented in various methods which essentially divided into two domain, traditional handcrafted method and modern deep learning method. Those traditional methods often use features for polyp segmentation. For example, Sánchez-González et al. [5] used texture and color features for training classifiers. The author utilize image windowing to increase generalization and effectiveness. Active contour segmentation method was proposed by Dutta et al. [6], where the active contour framework was modified to work without edges.

Deep learning techniques have been massively initiated to the image analysis domain and outperform the old traditional ways consequently in current era. In Vázquez et al. [7], they used Fully Convolutional Network—8 (FCN-8) to up-sample the image to recover spatial resolution for per pixel prediction task. Huang et. al. [8] proposed and compared two deep neural networks, DeepLab_v3 with Long Short-Term Memory networks (LSTMs) and Segnet with LSTMs. The LSTMs are utilized for keeping localization information that could slowly vanish along the encoder networks. Another two proposed methods namely Dilated ResFCN and SE-Unet were also compared in Guo and Matuszewski [9].

In this paper, a SegNet based CAD tool was proposed for polyp segmentation. The proposed method has been validated in publicly available datasets, CVC-ClinicDB, CVC-ColonDB and ETIS-Larib. The rest of the paper presented as follows. The method used in this algorithm are demonstrated in Sect. 2, and the results obtained are further discussed in Sect. 3. Lastly, Sect. 4 concludes this paper.

2   Material and Method

2.1   Polyp Datasets

Three publicly available polyp databases namely CVC-ClinicDB, CVC-ColonDB, and ETIS-Larib polyp were used to train and evaluate the proposed method. Out of 1187 images, 70% of the databases were used to train the segmentation model network; while the rest were used to test the model. CVC-ClinicDB contains 612 standard definition still images of 384×288 resolution. CVC-ColonDB database has 379 images of 574×500 resolution. ETIS-LaribPolypDB is a polyp database that has 196 high-definition still images with a resolution of 1225×966. Each image was associated with own individual polyp masks. The ground truth segmentation masks were manually drawn and verified by expert clinicians.
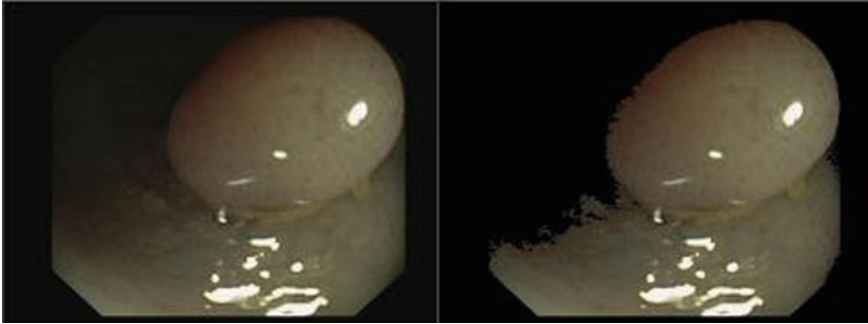
Fig. 1 Left image represents original image and right image represents pre-processed image

2.2 Image Pre-Processing and Normalization

Feature extraction through any image processing is greatly affected by the image quality itself. In this research work, thresholding was performed on image datasets as a method for pre-processing. Image pixel threshold values were set for both red channel and green channel as 50 and 15 respectively. Those pixel values that lower than pre-defined threshold values were set to zero as the non-polyp background pixel. The model network trained can be more biased to the polyp features rather than the non-polyp background features since the image was mostly contributed by non-polyp pixels. Next step is to normalize the image size to meet the input requirement of network. The image datasets have been resized to 360×480. Figure 1 shows the original image and the pre-processed image.

2.3 Proposed Model—SegNet

SegNet with VGG-19 framework architecture was used as our model. Datasets in RGB image format were pre-processed, followed by model network training and in the end test the model with testing images. The general steps involved in this automated system are presented in Fig. 2.

SegNet is an architecture that was designed to map low resolution features with the input resolution for pixel-wise semantic segmentation [10]. Encoder network of SegNet is essentially the same as typical Convolutional Neural Network (CNN).
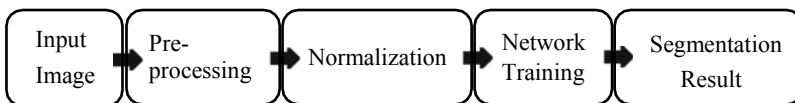
Fig. 2 Overall research methodology
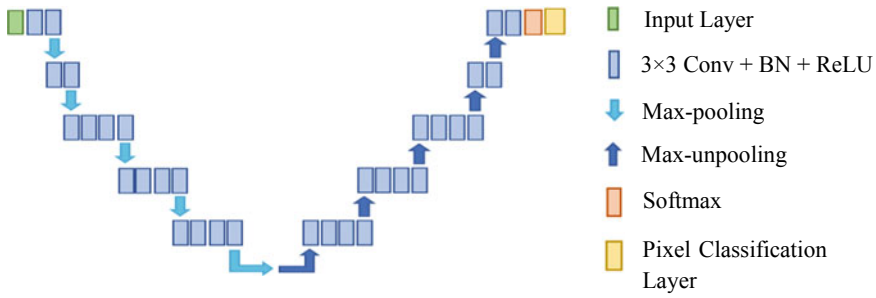
Fig. 3 Overview of model network, SegNet

Encoder network consists of input layer, convolution layer (Conv), activation layer and pooling layer. In convolution layer, a filter will be convolving around receptive field [11]. Each of these filters can be thought of as feature identifiers. The filters on the first layer convolve around the input image will be getting high values through convolution when a filter found its desired feature from the input image. The same thing will be iterated through all the location of input layer to form feature map. We included batch normalization (BN) after each convolutional layer to normalize the output, allowing for a larger learning rate that accelerate the training procedure. Trainable parameters that remain untrained previously are able to start training. The following layer is Rectified Linear Unit (ReLU), an activation layer which can limit the generated output. Pooling layer which usually comes after activation layer, is applied to decrease the spatial size of the convolved feature. By doing this, the computational power required to process the data can be greatly reduced. Moreover, it is useful for extracting dominant features which are rotational and positional invariant. There is a decoder network in SegNet which corresponded to the encoder network. It utilizes max-pooling indices from previous encoder network to up-sample the input feature map encoded from the encoder network [10]. At the final layer before SegNet last output layer, there is one layer called Softmax. This layer works to segment the polyp by activates output neurons via generating discrete probability distribution across all the output neurons with the help of cross-entropy loss calculation. Overview of the entire model network has been shown in Fig. 3.

3 Result and Discussion

In this project, pre-trained VGG-19 network was used as the framework architecture. The network has been modified into SegNet by adding deconvolution network to meet this project objectives. The training datasets were fed into the network together with ground truth images for machine learning process. Testing images will be used to evaluate the performance of the network in terms of mean Intersection over union (mIoU) and dice coefficient. The network was trained with training options set to
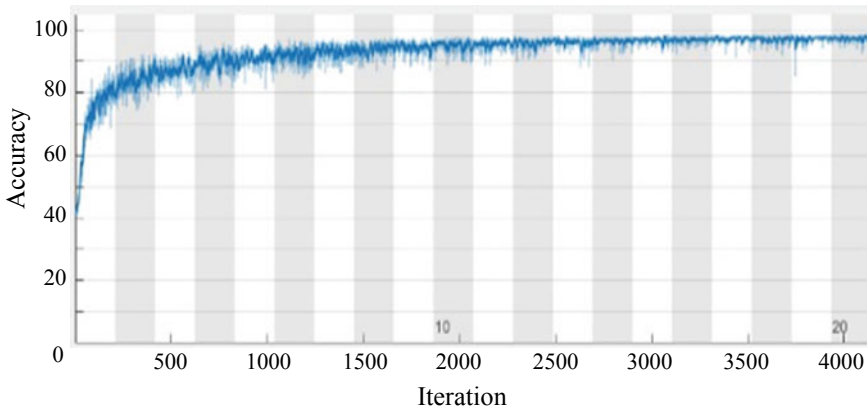
Fig. 4 SegNet training process

0.9 momentum, 0.001 learning rate, 20 epoch and mini batch size of 4. The network was trained using single CPU, Intel® Core TM i7-2600 CPU @ 3.40 GHz, 8 GB DDR3 RAM. Figure 4 shows the graph of training process. Our method results are detailed in Table 1. This method was evaluated by comparing the performance with other proposed work and the results surpass them.

According to Fig. 5a, the results show that entire polyps were successfully segmented. Some over-segmented polyps can be seen in these results prove that this model is poor in determining the edges of polyps. From Fig. 5b, the results obtained were moderate. The outcome shows more over-segmented polyp regions than the good results from Fig. 5a. The model did missed segment some true polyp region and false polyp segmentation on light-reflected region. Some poor performed outcomes also determined and presented in Fig. 5c. The model totally missed the

Table 1 Result comparison with previous research works

Research work	Mean intersection over union	Dice coefficient
Sánchez-González et al. [5]	–	87.31%
Akbari et al. [12]	–	81.00%
Nguyen and Lee [13]	–	89.60%
Huang et al. [8]	77.64%	–
Guo and Matuszewski [9]	–	82.93%
Yu et al. [14]	83.40%	89.10%
Bagheri et al. [15]	–	82.00%
Guo et al. [4]	96.00%	98.00%
Proposed Method	81.87%	92.35%
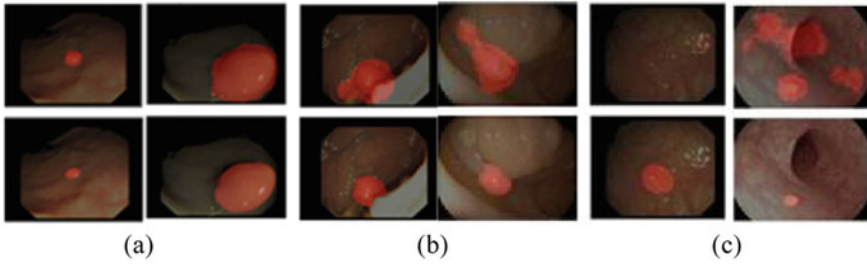
Fig. 5 First row of images show segmentation results, second row of images show ground truth. **a** Good segmentation results. **b** Moderate segmentation results. **c** Poor segmentation results

location of polyp and most of segmented non-polyp pixels were found. These poor results indicate that the model is not intelligent enough and unable to tackle more complicated structure polyp and poor condition colon.

Our proposed model obtained 81.87% mIoU, lower than [4] and slightly lower than [14] by 1.53%. Our dice coefficient reaches 92.35%, surpassed all previous works that are all under 90% except [4] which is 98.00%. Our model outperforms [14] on dice coefficient by 3.25%. Our model performed poorer than [4] due to our model contains only one single network, while [4] proposed an ensemble of three different networks into one final model. Our model is weak in classifying edges of polyps, possibly due to some of polyp features being eliminated across deeper encoder path especially convolution and max-pooling operations.

4 Conclusion

In this paper, SegNet was proposed as segmentation network. SegNet is an encoder-decoder network that map low resolution features with the input resolution for pixel-wise semantic segmentation. The proposed method were tested on CVC-ClinicDB, CVC-ColonDB, and ETIS-LaribPolypDB databases. Our algorithm achieves results of 81.78, 92.35% for mIoU and dice coefficient, respectively. There were certainly some flaws in our model indicated there are rooms to improve further. In future work, we planned to do research on color spaces that give more dominant polyps features beside RGB image. We will also try to solve the problem of useful polyp features being eliminated down the encoder path by preserving features and feed to decoder path. Development of CAD tool in polyp segmentation would surely give beneficial impact on future colorectal cancer diagnosis and management.

References

1. "Cancer" World Health Organization (2018). Available: http://www.who.int/news-room/fact-sheets/detail
2. Cao Y, Liu D, Tavanapong W, Wong J, Oh JH, de Groen PC (2006) Automatic classification of images with appendiceal orifice in colonoscopy videos. In: 2006 international conference of the IEEE engineering in medicine and biology society. Available: https://www.ncbi.nlm.nih.gov/pubmed/17946108
3. Zheng Y, et al (2018) Localisation of colorectal polyps by convolutional neural network features learnt from white light and narrow band endoscopic images of multiple databases. In: 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC). Honolulu, HI, pp 4142–4145. https://doi.org/10.1109/EMBC.2018.8513337
4. Guo X, et al (2019) Automated polyp segmentation for colonoscopy images: a method based on convolutional neural networks and ensemble learning. In: 2019 American association of physicists in medicine, vol 46, no 12. doi.org/https://doi.org/10.1002/mp.13865
5. Sánchez-González A, Garcia-Zapirain B, Sierra-Sosa D, Elmaghraby A (2018) Colon polyp segmentation using texture analysis. In: 2018 IEEE international symposium on signal processing and information technology (ISSPIT). Louisville, KY, USA, pp 579–588. https://doi.org/10.1109/ISSPIT.2018.8642748
6. Dutta S, Sasmal P, Bhuyan MK, Iwahori Y (2018) Automatic segmentation of polyps in endoscopic image using level-set formulation. In: 2018 international conference on wireless communications, signal processing and networking (WiSPNET). Chennai, pp 1–5. https://doi.org/10.1109/WiSPNET.2018.8538615
7. Vázquez D et al (2017) A benchmark for endoluminal scene segmentation of colonoscopy images. J Healthcare Eng 2017(4037190):9. https://doi.org/10.1155/2017/4037190
8. Huang C, Xiao W, Chang L, Tsai W, Liu W (2018) Automatic tissue segmentation by deep learning: from colorectal polyps in colonoscopy to abdominal organs in CT exam. In: 2018 IEEE visual communications and image processing (VCIP). Taichung, Taiwan, pp 1–4. https://doi.org/10.1109/VCIP.2018.8698645
9. Guo YB, Matuszewski BJ (2019) GIANA polyp segmentation with fully convolutional dilation neural networks. In: Proceeding of the 14th international joint conference on computer vision, imaging and computer graphics theory and applications, vol 4. Czech Republic, GIANA, pp 632–641. ISBN 978-989-758-354-4, https://doi.org/10.5220/0007698806320641
10. Badrinarayanan V, Kendall A, Cipolla R (2017) SegNet: a deep convolutional encoder–decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 39(12):2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615
11. Deshpande S (2016) A beginner's guide to understanding convolutional neural networks. Available: https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks/. Accessed 12 July 2019
12. Akbari M, et al (2018) Polyp segmentation in colonoscopy images using fully convolutional network. In: 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC). Honolulu, HI, pp 69–72. https://doi.org/10.1109/EMBC.2018.8512197
13. Nguyen N, Lee S (2019) Robust boundary segmentation in medical images using a consecutive deep encoder-decoder network. IEEE Access 7:33795–33808. https://doi.org/10.1109/ACCESS.2019.2904094
14. Yu J, Pan H, Yin Q, Bian X, Cui Q (2019) Fully convolutional dense nets for polyp segmentation in colonoscopy. In: 2019 IEEE 35th international conference on data engineering workshops (ICDEW). Macao, Macao, pp 306–311. https://doi.org/10.1109/ICDEW.2019.00010
15. Bagheri M, et al (2019) Deep neural network based polyp segmentation in colonoscopy images using a combination of color spaces. In: 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC). Berlin, Germany, pp 6742–6745. https://doi.org/10.1109/EMBC.2019.8856793

An Agile System for FiO_2 Regulation in Ventilators

S. Prabakar, J. Samson Isaac, R. Karthikeyan, K. Porkumaran, and Ramani Kannan

Abstract Respiratory failure cases are many flowing and filling the ICUs of the hospitals at any point of time. In cases of respiratory failure, either short-term or long-term ventilation is required for the patient for survival. Maintaining of O_2 saturation in patients is of prime objective in mechanical ventilation which in turn contributes for their recovery. Patients connected to the ventilator are constantly monitored for O_2 saturation. Any fall in O_2 saturation requires immediate intervention of clinician to introspect the problem and sort out the issues. One of the main reasons for low O_2 saturation is low value FiO_2 settings on the ventilator. This FiO_2 setting has to be increased or decreased manually as per monitored SpO_2 analyzer. There are always possibilities that exist, that patient on ventilators may go unnoticed or get less attention during emergencies cases entering the ICU. In this project we approached to automatically blend Air/O_2 mixer of an ICU ventilator to maintain a target SpO_2 Saturation on the patient, say 98%. Algorithm mechanism is designed in such a way, continuous FiO_2 changes will lead to achieve targeted SpO_2 level of the patient.

Keywords SpO_2 · FiO_2 · ICU ventilator · Automatic intelligence · Blender correction · Half the difference

S. Prabakar (✉)
Sona College of Technology, Salem, India

J. Samson Isaac
Biomedical Engineering, Karunya University, Coimbatore, India

R. Karthikeyan
Dr. N.G.P. Institute of Technology, Coimbatore, India

K. Porkumaran
Sri Sairam Engineering College, Chennai, India
e-mail: porkumaran@gmail.com

R. Kannan
EEE, Universiti Teknologi Petronas, Perak, Malaysia

1   Introduction

Respiratory failure cases are flowing and filling the ICUs of the hospitals every now and then. This respiratory distress cases range from road accidents to snake bite and suicide cases. All the above cases require either short-term or long-term ventilation for their survival. For the fast recovery of patients, maintaining O_2 saturation level is the prime objective of mechanical ventilation.

Normally, Alveolar Ventilation is unconsciously regulated to maintain constant arterial blood gas tensions, despite variable levels of oxygen consumption and CO_2 production. Many drugs and techniques used in anesthesia interfere with control or mechanics of ventilation, and it is the Anesthetists responsibility to ensure the adequacy of ventilation during the preoperative period. Equipment related to ventilation is consequently of great importance to the Anesthetist and Anesthetic Technician. Correct use of the equipment relies on a good understanding of basic respiratory physiology as well as how the individual ventilator operates.

Patients connected to the ventilator are constantly monitored for O_2 saturation. Any fall in O_2 saturation requires immediate intervention of clinician to introspect the problem and to sort out the issues. The main reason for Low O_2 Saturation is low FiO_2 settings on the Ventilator. So there are always some possibilities that exist, such as patients under ventilation are left unnoticed or given less attention that may lead to even death for patients under ventilation.

FiO_2 is Fractional Inspired Concentration of Oxygen. It is a mixer of Air and Oxygen, blend at a specific proportion, delivered to the patient through ventilator. Air/O_2 Mixer or Blender is one of the main components in any ventilator that mixes Air and O_2 at a particular proportion that is delivered to the patient to maintain his/her O_2 Saturation. The range of mixing is from 21 to 100% of Oxygen as atmospheric air which always contains 21% Oxygen by nature.

1.1   Scope and Significance of Proposed Work

The main objective is to blend Air/O_2 mixer of an ICU Ventilator automatically to maintain a target SpO_2 Saturation on the Patient, say 98%. The SpO_2 feedback of the patient, that is monitored will automatically titrate the FiO_2 mechanism on the Ventilator to achieve the target SpO_2 level. The corresponding algorithm is designed for the continuous FiO_2 changes which will lead to attain the intended SpO_2 level of the patient.

According to the algorithmic process, the continuous feedback system will lead to control the flow of ventilator's FiO_2 according to the patient's SpO_2. Automating the Blending process gives a great relief to the clinician in patient management and allows them to concentrate on other areas that need more time. This is a very safe method gives mental relief to doctors and nursing staffs and at the same time patient is also ensured of minimum O_2 Saturation for his good life.

The importance of ventilation and oxygenation have analyzed through previous work analysis. The problems of the manual setting of FiO$_2$ in ventilators which may lead to many medical accidents due to clinician's unawareness represented through many papers. These papers help to analyze about the problems faced while the patient is under ventilation and also lead to an innovative approach.

1.2 Significance of the Proposed Work

The proposed research work aims to develop an efficient hardware-software approach to continuously receive and monitor the sensor data received from fingertip based pulse oximetry which is connected to a mobile app via Bluetooth. The app continuously receives the data and checks it against a standard value. If the value starts to fall into the abnormal range, a text will be sent so that the person may take their oxygen supplement before the saturation of oxygen in the blood falls too below the normal values.

In this project we have used Reflectance pulse oximetry which is an alternative to transmissive pulse oximetry. The advantage of this method over transmissive method is that it does not require a thin section of the person's body and is therefore well suited to a universal application. Continuous monitoring pulse oximeters are typically used for long period (1 h or more) oxygen (SpO$_2$). Continuous pulse oximeters can be used for spot checks (quick readings), sleep studies and constant monitoring.

2 Proposed System Design

A complete design and approach for the automatic intelligence and feedback system are described. It includes the block diagram and algorithmic approach in depth. Figure 1 shows the input, processing and output modules of the automatic FiO$_2$ regulation in ventilators.

After the analysis of the deviation in the O$_2$ saturation, FiO$_2$ value is newly computed according to the algorithm designed. The computed FiO$_2$ value is feedback to the ventilator as an input. Depending on that new value, the SpO$_2$ value will gradually increase and gives relief to the patient. This is done repeatedly until the patient attains the normal O$_2$ saturation level.

2.1 Working Process of the Proposed System

The SpO$_2$ sensor detects the oxygen saturation from the patient. That percentage output as a feedback as well as the set FiO$_2$ from the ventilator are given as inputs to
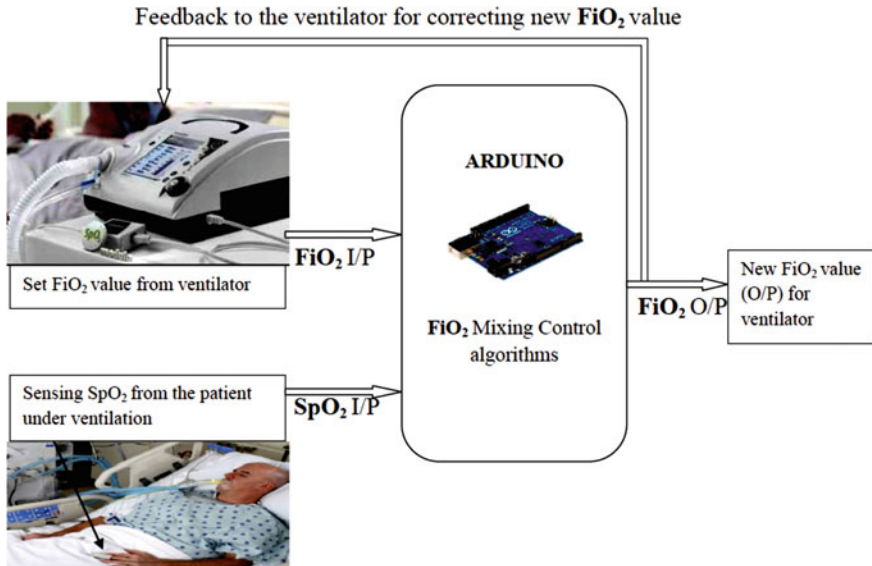
Fig. 1 Block diagram of automatic FiO$_2$ regulation in ventilators

the ARDUINO hardware via the input channels. The flow control algorithm is fed to the ARDUINO hardware to give the relevant output. The algorithm as follows:

Step 1: Set the FiO$_2$ value in ventilators and sense the SpO$_2$ level from the patient using sensors.
Step 2: Calculate the difference between the optimal and measured SpO$_2$ and determine the "Half the difference" for blender correction mode.
Step 3: Compute the new FiO$_2$ value for the ventilator to supply on patient.
Step 4: Then again sense the SpO$_2$ and repeat the same steps to attain the desired SpO$_2$ value on the patient.

According to the above algorithm a new FiO$_2$ value is given as a output from the ARDUINO hardware. That FiO$_2$ value is given as a feedback to the ventilator. Thus the SpO$_2$ value is automatically modified according to the new FiO$_2$ value.

2.2 Automatic Intelligence Algorithm for FiO$_2$ Regulation

The algorithm is developed from the SpO$_2$ deviation on the patient under ventilation in order to compute the desired FiO$_2$ automatically. Figure 2 provides the flow of FiO$_2$ computation from the obtained values.

Initially we will consider two values such as FiO$_2$ and SpO$_2$. FiO$_2$ is the fractional inspired concentration of oxygen, which is the set value in the ventilator (for say 50%) and SpO$_2$ is saturation of O$_2$%, which is measured from the patient through the sensor.

Fig. 2 Steps involved in new FiO$_2$ computation
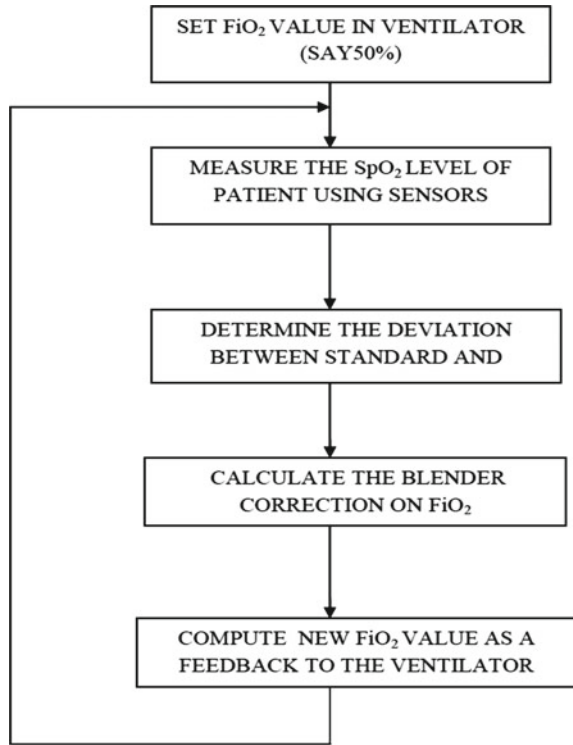
The deviation between the standard SpO$_2$ value, which is the normal level of oxygen saturation i.e. 98% and the measured SpO$_2$, which is theme assured value from the patient through sensor is determined. The determined value is the difference between the standard SpO$_2$ value and the measured SpO$_2$ value. According to the difference, the FiO$_2$ will be changed according to the at present SpO$_2$ value. A new FiO$_2$ value will be computed as a result it is given to the ventilator as a feedback input. According to that FiO$_2$ value which is newly computed, the SpO$_2$ value gets modified. So this progress is done continuously and the corrections are done gradually in loops (Table 1).

We are here taking a situation that a patient connected to a Ventilator with FiO$_2$ set value of 50% results in the patient's saturation of O$_2$ 80%. But, normal SpO$_2$ to be achieved is 98%.

Here there is a difference between normal and achieved saturation (98–80%) which is 18%. This is an indication that the FiO$_2$ set on the ventilator is in adequate and it has to be increased. This increase has to be systematic and it should tend to increase the O$_2$ saturation towards the normal level of SpO$_2$. Hence the FiO$_2$ (Air/O$_2$ mixer) is increased to a level that is "Half of the Difference" between target and achieved SpO$_2$, i.e., 9%. Now the new (Air/O$_2$ mixing) FiO$_2$setting has to be increased from 50 to (50 + 9) 59%.

Table 1 FiO2 estimation for regulation in ventilators

Set FiO$_2$ in (%)	Measured FiO$_2$ in (%)	Required SpO$_2$ in (%)	Deviation [C-B]	Blender corr. mode (changes reqd. in FiO$_2$)	New FiO$_2$ setting on vents
A	B	C	D	E	F
50	50	98	48	24	74
50	55	98	43	21.5	71.5
50	60	98	38	19	69
50	65	98	33	16.5	66.5
50	70	98	28	14	64
50	75	98	23	11.5	61.5
50	80	98	18	9	59
50	85	98	13	6.5	56.5
50	90	98	8	4	54
50	95	98	3	1.5	51.5
50	98	98	0	0	50
50	99	98	−1	−0.5	49.5
50	100	98	−2	−1	49

Let us assume that this increase in FiO$_2$ is taking the patient to a higher O$_2$ saturation of 88% from 80%. Still the Patient's SpO$_2$ is below normal by a value of 10% (98–88%), that is deficit. This monitored value gives a feedback to the Air/O$_2$ mixer to increase the FiO$_2$ to another "Half of the Difference" that is 5%.Now the new FiO$_2$ on the ventilator would be 59 + 5% i.e., 64%. This process will tend to continue till the patient achieves the targeted SpO$_2$ level of 98%.The above process is the systematic way of maintaining a patient's O$_2$ saturation through "Intelligent Feedback System".

3   Result Analysis and Discussion

This proposed approach delivered results in great relief to the clinicians as it is automatically titrating the FiO$_2$. To understand clearly, consider that the patient connected to an ICU ventilator are managed by clinicians and attended by nursing staff. We are here taking a situation that a patient is connected to a ventilator with FiO$_2$ setting of 50% that results in the patient's saturation of SpO$_2$ 80%. Normal SpO$_2$ to be achieved is 98% and so the algorithm will generate the desired computed value.

Feedback from the SpO$_2$ monitor is interfaced with the blender of the ventilator to achieve an optimum SpO$_2$ level of 98%. Any deviation in the SpO$_2$ value will trigger the blender to increase or decrease the FiO$_2$ to maintain the optimum 98%

of O$_2$ saturation. Patient saturation is automatically maintained to the optimum level throughout the time period, he/she connected to the ventilator 24 $\times$ 7. Human error or failure to re-titrate the FiO$_2$ to achieve the target SpO$_2$ is avoided. It is ensured that any point of time patient is not de-saturated. The above process is the systematic way of maintaining a patient O$_2$ saturation through "Intelligent Feedback System".

This proposed approach for designing and prototyping using Arduino hardware is reducing the complexity on the implementation of our real world project since it delivers the powerful hardware and software technology. In the software simulation the coding is done accordingly for the implementation of the algorithm for the automatic feedback system. The coding inclusive of LCD display, algorithmic calculation for automation process and stepper motor implementation for triggering the value according to saturation level measured from the patient and required delay time is given for the progress to be done systemically at required interval period according to the situation of the patient. The LCD display is connected for displaying the current SpO$_2$ and computed FiO$_2$ value for our verification.

After the software part done then the coding is checked under virtual simulation board and ensured the coding done is working as per the expectation. Then the coding is fed into the Arduino board, the Arduino board is connected to the external power supply, SpO$_2$ sensor for the saturation measurement, stepper motor for triggering of the value, relay for the switching the stepper motor and LCD display for viewing the required values. Initially the saturation is measured from the patient according to the that value, if any fall occurs then the relay board starts to trip and makes the stepper motor to rotate until it acquires an desired value and gives that value as a feedback the ventilator.

Thus this blending is done repeatedly until the required SpO$_2$ is obtained. Thus by using this technology the automatic FiO$_2$ regulation can be achieved. We need not
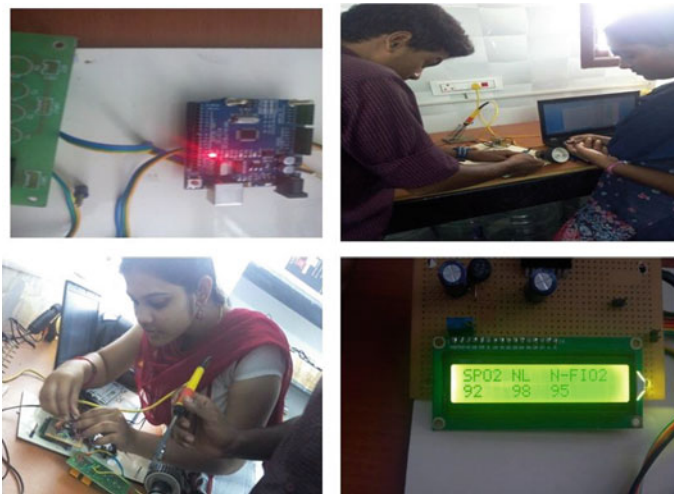
Fig. 3 Intelligent system for FiO$_2$ regulation in ventilators

require any manual adjustments of clinician and nurses. Whenever there is a fall in O_2 saturation occurs the ventilator automatically blends the FiO_2 settings according to the patient's oxygenation (Fig. 3).

Automating the Blending process gives a great relief to the clinician inpatient management. They can also over-ride this mechanism in case they want to have direct control of the patient. This mechanism allows the clinicians to concentrate on other areas that need more time. This is a very safe method gives mental relief to doctors and nursing staffs and at the same time patient is also ensured of minimum O_2 Saturation for his good life.

4   Conclusion

Thus the proposed work concludes that automating the Blending process gives a great relief to the clinician in patient management. Thus patient needs not require continuous monitoring and manual correction. They can also over-ride this mechanism in case they want to have direct control of the patient. The automatic regulation can be done not only for FiO_2 but also for all the blood gas values in different algorithmic procedures like PaO_2, $PaCO_2$ etc. Thus the system will lead to complete automatic operation on ventilator without clinicians' periodical check and also ensures the patients good care. This mechanism allows the clinicians to concentrate on other areas that need more time. This is a very safe method gives mental relief to doctors and nursing staffs and at the same time patient is also ensured of minimum O_2 Saturation for his good life.

References

1. Zhou F (2009) Component-based design of software for embedded control systems: the medical ventilator case study. In: Guan W, Sierszecki K, Angelov C (eds) Embedded software and system, ICESS'09, international conference. Mads Clausen Institute for Product Innovation, University of Southern Denmark, Soenderborg, pp 157–163
2. Ionescu CM (2009) Flow controlled artificial ventilation of a COPD patient. In: De Keyser R (ed) Control conference (ECC). System and Automation Department, Ghent University, Gent-Zwijnaarde, Belgium pp 2475–2481
3. Jafari MM (2005) Senior member, IEEE, Control systems engineering group, Tyco Healthcare Puritan Bennett R&D—"Robust Feedback Design for Proportional Assist Ventilation-System Dynamics and Problem Definition" decision and control, 2005 European Control Conference, CDC-ECC'05, 44th IEEE Conference, pp 4389–4844
4. An P (2010) High-precision ventilator with compensation algorithm and automatic regulation. In: Song JT (ed) Database technology and application (DBTA), 2nd international workshop. Schema of Electronics and Information Engineering, Ningbo University of Technology, Ningbo, China, pp 1–4
5. Tehrani FT (2007) A new decision support system for mechanical ventilation. In: Engineering in medical and biology society, 29th annual international conference of the IEEE. California State University, Fullerton, pp 3569–3572

6. Tzavaras A (2011) Locating of the required key-variables tobe employed in a ventilation management decision support system. In: Weller PR, Prinianakis G, Lahana A (eds) Engineering in medicine and biology society EMBC, Medical Instrumentation Technology Department, Technological Educational Institute (TEI) of Athens, Greece, pp 112–115
7. Yi W (2009) Fuzzy control for regulating level of pressure support ventilation. In: Zhang Q, Wang Y, Xiong J (eds) Measuring technology and mechatronics automation, ICMTMA'09, international conference, vol 2. Mechatronics and Automatic Schema, Natural University of Defense Technology, Changsha, China, pp 562–565
8. Li Y (2005) Investigation on dynamics of the flow control unit in ventilator systems and its fundamental performance limitations. In: American control conference, vol 3. VIASYS HealthCare, Yorba Linda, CA, USA, pp 2181–2186

Computer-Aided Diagnostic Tool for Classification of Colonic Polyp Assessment

Win Sheng Liew, Tong Boon Tang, and Cheng-Kai Lu

Abstract Colorectal cancer is the third most common malignancy and the fourth leading cause of cancer-related deaths worldwide. This paper presents a combination of techniques (e.g., pre-processing, transfer learning, principal component analysis, and support vector machine) to detect the polyp during colonoscopy. In particular, we carefully choose the pre-trained deep convolutional neural networks (i.e., AlexNet, GoogLeNet, ResNet-50, and VGG-19) according to their performance extracting features. A publicly available database, Kvasir, is used to train and test the detection model. The result indicates that to use ResNet-50 as a pre-trained network provides the best results among the rest. Our proposed model achieves an accuracy of 99.39%, and its sensitivity and specificity are 99.39, 99.41, and 99.38%, respectively.

Keywords Colorectal cancer · Transfer learning · Polyp · Deep convolutional neural network · Support vector machine

1 Introduction

Colorectal cancer (CRC) is one of the leading causes of cancer deaths based on the analysis of the American Cancer Society (ACS) as the number of cancer deaths was increased year by year from 1997 to 2015 [1]. The small and tiny polyp might be missed [2], which caused the chance of getting converted into a giant neoplastic polyp. Hence, the discovery of colonic polyps is essential as the polyp can be inhibited by removing it and reduce the incidence of CRC, whereas the detection of malignant polyps at an early stage may result in a curative intervention [3]. However, it is not as difficult to detect polyp as before [1] because of the elegant automated computer-aided diagnosis (CAD). Those CAD tools can detect the polyp and provide polyp location depending on the curvature of polyps. Several CAD systems have been developed for various cancer diseases such as prostate cancer, brain cancer, lung

W. S. Liew · T. B. Tang · C.-K. Lu (✉)
Department of Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Malaysia
e-mail: chengkai.lu@utp.edu.my

cancer, etc. These CAD systems obtain the images through the hardware renders and save them as digital movies using image processing software. Therefore, doctors can accurately carry out the diagnosis and decision based on the CAD system's information.

With the CAD scheme's rapid development, the detection system is continuously optimized and improved with AI techniques. AI is considered as a machine emulation with human thinking. Thus, deep learning (DL) applications in AI allow the systems to become intelligent, learning, and self-organizing [4]. Pre-processing is vital for every detection system, as every image has different sizes and features (i.e., color, density, shape, and texture) [5]. The images need to be pre-processed to reduce the image's noise and degradation; normalization is required to resize the images into the same size. Besides, the images need to undergo feature extraction and selection to carry out classification task with a different type of classifiers such as ensemble classifier [6], support vector machine (SVM) [7], Naïve Bayes classifier [8, 9], random forest [9] and K-Nearest Neighbors (KNN) [9].

2 Material and Method

In the research, a publicly available database, Kvasir, is used to train, validate, and test the detection model. All images are color colonoscopy images in JPG format with a resolution of 720×576 pixels.

A total number of 2000 images in the dataset consisted of 1000 normal cases (non-polyp) and 1000 abnormal cases (polyp). The colonoscopy images are split into 70% for training and 30% for testing. The model is developed through the MATLAB platform on a Windows 10 64-bit operating system with Intel® Core i5-2430 M CPU @ 2.40 GHz, 8 GB RAM, and NVIDIA GeForce GT 525 M GPU. For this research work, each step is illustrated in Fig. 1.
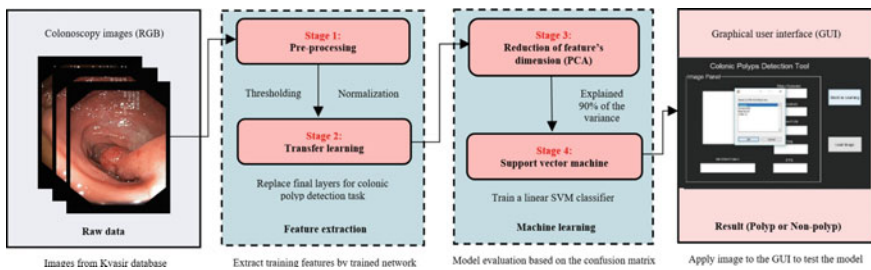
Fig. 1 Overview of the research methodology

2.1 Pre-processing

During colonoscopy screening, the images will be affected due to the visual condition in the bowel. This is because there is no ambient light in the bowel, but only the capsule on-board light as the illuminating source. Under this condition, the images captured by the camera will have some artifacts, vignettes, and illuminations [5, 10]. To enhance the polyp feature in the image, the thresholding method is used to separate the polyp from the background, which indicates the image consists of two classes of pixels (foreground and background) [11]. A color image has three channels, which are red (R), green (G), and blue (B); the threshold values of R and G channels are scaled to 0 if the values are below 50 and 15, respectively. On the other hand, the B channel is not utilized for thresholding due to pixel values of polyp and background are almost the same.

Furthermore, the images needed to resize into the same scale through normalization to feed into the neural networks for training. Nevertheless, different networks have different input sizes; AlexNet can only process the images that are 227×227 pixels, while GoogLeNet, ResNet-50, and VGG-19 process images with 224×224 pixels. The normalization equation is computed as follows:

$$y'(ResizedImage) = \frac{y - \overline{y}}{\sigma} \tag{1}$$

where y is the original pixel vector, $\overline{y}$ is the mean of the pixel vector, and σ is its standard deviation.

2.2 Transfer Learning

Deep learning (DL) is potent as it helps us a lot in complex tasks. However, it is not easy to create a DL model because it requires generous data to train the deep neural network from scratch. To resolve this issue, transfer learning (TL) is a powerful and effective solution to limited data, especially in the medical research field [12].

In addition, TL can improve the accuracy of the network while reducing its training time [13]. The original architecture of pre-trained networks such as AlexNet, GoogLeNet, ResNet, and VGG can be used and fit into own task by replacing the last three layers with the new specific task from the database. Besides, the neural networks can learn from training to progress their proficiency [14]. Nevertheless, different neural networks have various performances in terms of accuracy and training time due to the number of deeper layers and the training database's size.

2.3   Classification

From the view of machine learning, one of the classification methods is supervised learning. Support vector machine (SVM) is a classifier with supervised learning [15], and it provides remarkable performance and accuracy with less computation power. SVM divides data into two classes (normal and abnormal cases) using hyperplanes in a high dimensional feature space.

However, SVM has a higher dimensionality, which may lead to model overfitting. Therefore, principal component analysis (PCA) is utilized to reduce data dimensionality [16] to the maximum extent. To achieve dimensionality reduction, PCA's feature elimination is one of the techniques as it remains interpretability of variables [17].

3   Result and Discussion

3.1   Performance of Models

The models are evaluated based on the confusion matrix with true positive (TP), true negative (TN), false positive (FP), and false negative (FN), where TP: a polyp is detected in a frame that contains a polyp; TN: No polyp is detected in a frame without a polyp image; FP: a polyp is detected in a frame without a polyp image; FN: A polyp is missed in a frame that contains a polyp. The accuracy, sensitivity, and specificity are calculated from the confusion matrix as follow:

$$Accuracy(ACC) = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Sensitivity(TPR) = \frac{TP}{TP + FP} \tag{3}$$

$$Specificity(TNR) = \frac{TN}{TN + FP} \tag{4}$$

Table 1 Performance of each neural network

CNNs	Accuracy	Sensitivity	Specificity	AUC
AlexNet	0.9635	0.9645	0.9625	0.9935
GoogLeNet	0.9727	0.9704	0.9750	0.9964
ResNEt-50	0.9939	0.9941	0.9938	0.9988
VGG-19	0.9791	0.9645	0.9938	0.9986

The summary of results is provided in Table 1. Within these four deep neural networks, ResNet-50 is better than the rest of the networks with an accuracy of 99.39%, sensitivity of 99.41%, and specificity of 99.38%.

In Table 1, the area under receiver operating characteristics (AUROC) is measured based on the receiver operating characteristics (ROC) curve in Fig. 2. The ROC curve is a graph of true positive rate (TPR) versus false positive rate (FPR) to show a classification model's performance at all the classification thresholds. TPR is sensitivity; while FPR is calculated as follows:

$$FPR = 1 - Specificity \qquad (5)$$

Although every neural network has different computation time for data pre-processing and training, but the exact training iteration time is still depending on the hardware or the size of mini-batch set. Figure 3 shows the graph of accuracy versus training time for four neural networks. AlexNet is the fastest network, but it has the lowest accuracy. In contrast, VGG-19 takes the longest time for training and its accuracy is still lower than ResNet-50.

Due to the network complexity, the number of convolution layer directly relates to the performance of networks. ResNet-50 has the most convolution layers, and it tends to overfit the training data. As a result, it has the highest accuracy for the colonic polyp detection task and higher training time than the AlexNet and GoogLeNet. In contrast with training time, the number of learnable parameters plays an important role.
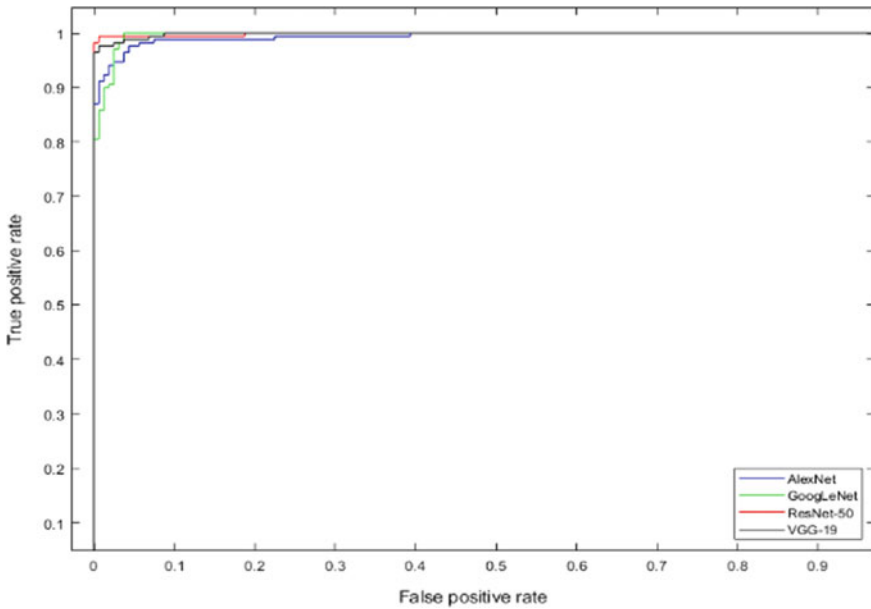
Fig. 2 ROC curve for each neural network
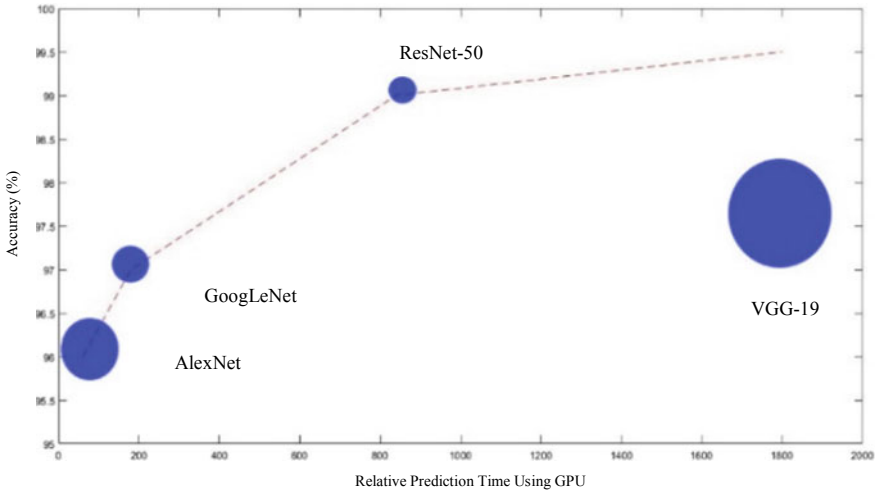
Fig. 3 Accuracy against relative training speed for each neural network

VGG-19 takes the longest time for the training because it has the biggest learnable parameters among the four networks.

Based on the discussion above, ResNet-50 is chosen for this research work. The results obtained by using the proposed method are compared with the state-of-the-art techniques. The comparison results have been made with previous works in Table 2. Our proposed method has better performance evaluation scores than other studies. In this way, the proposed method increases CAD systems' success rate with the most effective medical image detection method for the colonic polyp.

Table 2 Comparison of our results with previous studies

Name of study	Accuracy (%)	Sensitivity (%)	Specificity (%)
Bae and Yoon [6]	89.00	95.00	81.00
Xu and Suzuki [17]	91.00	95.00	90.00
Fan et. al [18]	93.70	95.40	74.70
Fiori et. al [19]	94.00	95.00	88.00
Proposed study	99.39	99.41	99.38
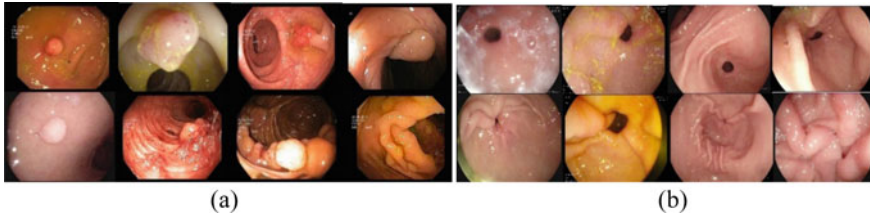
Fig. 4 Sample of colonoscopy images. **a** With polyp. **b** Without polyp

3.2 Graphical User Interface (GUI)

After the training process, the network's parameters are saved as a MAT file so that there is no training required again during the testing of the image. For the testing, a graphical user interface (GUI) is created to load the images, and it is designed as in Fig. 1. There are two buttons available in the GUI, which are machine learning and load image. First, the user requires to choose a model trained by four different CNN architectures. Once a network is chosen, all the parameters will be shown in the GUI, and it is ready to load the image for testing. Then, the classification result will be shown in the identified class column of the GUI.

3.3 Dataset

From the dataset, there are two categories of images, which are with polyp and without polyp. These 2000 colonoscopy images have different visual quality and features such as the brightness, distribution of pixel intensity, position, and polyp size. To prove the diversity of data, we have various types of images in Fig. 4, which allows the detection model to learn with multiple features.

4 Conclusion

To conclude, an automated polyp detection system was developed using pre-processing, principal component analysis, transfer learning as a feature extractor, and SVM as a classifier. The presence of polyps is detected with an accuracy of 99.39% from the Kvasir publicly accessible database. The sensitivity and specificity of polyp detection are 99.41%, and 98.75, 99.38%, respectively. The methods developed so far are therefore promising as the basis for subsequent expert human assessment.

References

1. Sareena, Mittal A, Kaur M (2016) Computer-aided-diagnosis in colorectal cancer: a survey of state of the art techniques. In: 2016 international conference on inventive computation technologies (ICICT), pp 1–6
2. Tajbakhsh N, Chi C, Gurudu SR, Liang J (2014) Automatic polyp detection from learned boundaries. In: 2014 IEEE 11th international symposium on biomedical imaging (ISBI), pp 97–100
3. Paik DS, Beaulieu CF, Rubin GD, Acar B, Jeffrey RB, Yee J et al (2004) Surface normal overlap: a computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT. IEEE Trans Med Imaging 23(6):661–675
4. Bose BK (1994) Expert system, fuzzy logic, and neural network applications in power electronics and motion control. Proc IEEE 82(8):1303–1323
5. Gueye L, Yildirim-Yayilgan S, Cheikh FA, Balasingham I (2015) Automatic detection of colonoscopic anomalies using capsule endoscopy. In: 2015 IEEE international conference on image processing (ICIP), pp 1061–1064
6. Bae S, Yoon K (2015) Polyp detection via imbalanced learning and discriminative feature learning. IEEE Trans Med Imaging 34(11):2379–2393
7. Romain O, Histace A, Silva J, Ayoub J, Granado B, Pinna A, et al (2013) Towards a multimodal wireless video capsule for detection of colonic polyps as prevention of colorectal cancer. In: 13th IEEE international conference on bioinformatics and bioengineering, pp 1–6
8. Gautam A, Singh P, Raman B, Bhadauria H (2016) Automatic classification of leukocytes using morphological features and Naïve Bayes classifier. In: 2016 IEEE region 10 conference (TENCON), pp 1023–1027
9. Devika R, Avilala SV, Subramaniyaswamy V (2019) Comparative study of classifier for chronic kidney disease prediction using Naive Bayes, KNN and random forest. In: 2019 3rd international conference on computing methodologies and communication (ICCMC), pp 679–684
10. Mamonov AV, Figueiredo IN, Figueiredo PN, Tsai YR (2014) Automated polyp detection in colon capsule endoscopy. IEEE Trans Med Imaging 33(7):1488–1502
11. Jeyavathana B, Ramasamy B, Pandian A (2016) A survey: analysis on pre-processing and segmentation techniques for medical images. Int J Res Sci Innov (IJRSI)
12. Shie C-K, Chuang C-H, Chou C-N, Wu M-H, Chang E (2015) Transfer representation learning for medical image analysis, pp 711–714
13. Felzenszwalb PF, Girshick RB, McAllester D, Ramanan D (2010) Object detection with discriminatively trained part-based models. IEEE Trans Pattern Anal Mach Intell 32(9):1627–1645
14. Ansari A, Bakar AA (2014) A comparative study of three artificial intelligence techniques: genetic algorithm, neural network, and fuzzy logic, on scheduling problem. In: 2014 4th international conference on artificial intelligence with applications in engineering and technology, pp 31–36
15. Ratheesh A, Soman P, Nair MR, Devika RG, Aneesh RP (2016) Advanced algorithm for polyp detection using depth segmentation in colon endoscopy. In: 2016 international conference on communication systems and networks (ComNet), pp 179–183
16. Ibrahim MFI, Al-Jumaily AA (2016) PCA indexing based feature learning and feature selection. In: 2016 8th cairo international biomedical engineering conference (CIBEC), pp 68–71

17. Xu J-W, Suzuki K (2014) Max-AUC feature selection in computer-aided detection of polyps in CT colonography. IEEE J Biomed Health Inform 18(2):585–593
18. Fan L, Song B, Gu X, Liang Z. Semi-supervised graph embedding-based feature extraction and adaptive kernel-based classification for computer-aided detection in CT colonography. In: 2012 IEEE nuclear science symposium and medical imaging conference record (NSS/MIC), pp 3983–3988
19. Fiori M, Musé P, Aguirre S, Sapiro G (2010) Automatic colon polyp flagging via geometric and texture features. In: 2010 Annual international conference of the IEEE engineering in medicine and biology, pp 3170–3173

Live Obstacle Detection Audio System for Visually Impaired People

P. Anitha, C. Vinothini, J. Priya, and S. Guhan

Abstract Object detection is the complex task for the visually impaired people when they are on move. This paper proposes smartphone's capability more useful for visually impaired people to overcome day to day problems with live navigation. Depth sensing camera is used to differentiate foreground objects from background objects which is used to place real world objects in different layers. As soon as the user moves to the particular layer the fore coming objects are detected and notified to the user by audio instructions. Information from the internet can be collected using the mobility services. The headphone with mic is used by the system for receiving user's voice input. Emergency tasks like sending SMS and location are sent to the subject's guardian while the subject is in danger with the help of player. This concept allows visually impaired people to walk both indoors and outdoors independently. Experimental results obtained show that the navigation system eases the user's travelling experience in complex indoor and outdoor environments. Hence, it serves as challenged people-friendly device for helping the visually impaired people to travel hassle-free and safely. The goal is to provide an effective means of navigation for the visually-impaired.

Keywords Obstacle detection · Audio navigation · Visually impaired

1 Introduction

The number of visually impaired people is on the raise over the recent years. Report of the world health organization (WHO) claims about 253 million people live with vision impairment. Among them, 36 million people are visually impaired and 217 million have moderate to severe vision impairment. Most of the visually impaired people

P. Anitha (✉) · C. Vinothini · S. Guhan
Department of Computer Science and Engineering, Dr. N.G.P Institute of Technology, Coimbatore, Tamilnadu, India

J. Priya
Department of Information Technology, Bannari Amman Institute of Technology, Sathyamangalam, Tamilnadu, India

is aged 50 or older and live in the developing countries struggling when they are alone. Every day these visually impaired people face problems in understanding and interacting with the surroundings, particularly those that are unfamiliar. It is hard for a visually impaired person to go out alone and there are not many available products that can assist them. However, research has been going on for decades for developing an effective device for visually impaired people. In this paper, a device with new features that can aid in multiple tasks while maintaining low development cost. The device can easily guide the visually impaired people and is able to give proper directions. The most valuable thing for a disabled person is gaining independence. A blind person can lead an independent life with some specifically designed adaptive things for them. There are lots of adaptive equipment that can enable a blind person to live their life independently but they are not easily available in the local shops or markets. A blind person needs to hunt and put much effort to get each equipment that can take them one step closer towards independence.

2   Related Works

Various methodologies have been proposed for the welfare of vision-impaired people using the concept of AI and ML. Here are some of them in brief.

Li et al. [1] proposed an approach pioneering effort of indoor robotic navigation, NavChair used sonar sensor to detect objects. A vector field histogram (VFH) algorithm and later Minimal VFH were proposed to provide effective sonar-based obstacle detection and avoidance. For updating a local map with detected obstacle in the form of uncertainty grids, VFH used sonar readings. It was used to calculate a polar histogram to represent the geometric information of the obstacles. At last, NavChair found a direction with minimum VFH cost and followed it to bypass obstacles. The RGB-D indoor navigation system. An emerging research focus for the robotics community to improve the mobility of blind and visually impaired people is Intelligent assistive navigation. Numerous studies have been carried out for indoor navigation on mobile devices.

Bai et al. [2] proposed a system that includes a depth camera for acquiring the depth information of the surroundings, an ultrasonic rangefinder consisting of an ultrasonic sensor, a MCU (Microprogrammed Control Unit) for measuring the obstacle distance, an embedded CPU (Central Processing Unit) board acting as main processing module, which does such operations as depth image processing, data fusion, AR rendering, guiding sound synthesis, etc., pair of AR glasses to display the visual enhancement information and an earphone to play the guiding sound.

Feng Lan, Shanghai Jiao Tong University, Shanghai, China [3] proposed a smart glass prototype system based on Intel Edison. The Intel Edison is a tiny computer offered by Intel as a development system for wearable devices. The board's main SoC is a 22 nm Intel Atom "Tangier" (Z34XX) that includes two Atom Silvermont cores running at 500 MHz and one Intel Quark core at 100 MHz (for executing RTOS Viper OS). The SoC has 1 GB RAM integrated on package. There is also 4 GB eMMC

flash on board, Wi-Fi, Bluetooth 4 and USB controllers [4]. Edison is powerful in computing but small in shape that makes it perfect to build a wearable device. Public signs in cities can act as guides for people. Following the public signs, people can find public infrastructure, such as public toilets, bus stations, subway stations, hotels and so on. So finding and recognizing the public signs outsides would help someone who is visually impaired or visually impaired gain increased independence and freedom. Therefore, they have implemented an application of public signs recognition in this developed smart glass system. This application can automatically detect, analyse and recognize all kinds of public signs around the visually impaired and visually impaired and give corresponding voice hints through wireless bone conduction headphones. With the help of this smart glass system, the visually impaired may find bus stations, subway stations, restaurants, hotels and so on. The visually impaired may even go for a trip alone.

Bourne et al. [4] proposed a project to solve instrumentation problems encountered by visually impaired employees or potential employees who are not suited to (nor do they require) complex microcomputer or synthetic speech circuitry for practical solutions. A common problem encountered by custom cabinetmakers, as well as individual homemakers, is finding the structural members within walls ("studs") onto which heavy shelves and wall hangings can be mounted. There have been many devices for finding "studs," but these are traditionally visual instruments. They have successfully modified a commercial stud finder to present auditory feedback to the visually impaired cabinetmaker and handyman". The Zircon "Stud Sensor" (also marketed by Radio Shack as the "Archer Stud finder") is a relative capacitance meter that senses hidden structures by their capacitance effect; it presents these relative indications to a sighted user via a column of LEDs. By reverse engineering, a "relative capacitance" analog signal was found which could be used to drive a voltage-controlled oscillator. Thus, with their modified instrument the visually impaired user hears a distinct rise in pitch when the capacitance increases—which happens when a stud is found. Four prototypes are being made for field trials, and several visually impaired individuals who have tested the unit have found it effective and easy to use.

Rentschler et al. [5] proposed a system to test the walker when run on a two-drum test machine. The walker was then run through the functionality test and visually inspected for cracks or damage. This approach helps the visually impaired navigate hassle free through streets.

Elaiyaraja et al. [6] proposed a system in which the prototype consists of an optical glass with head-mounted camera that captures images and offloads the computation to the database. The database is built using Microsoft Azure API and consists of details of all known persons to the visually impaired. Based on the computations, the database performs facial recognition and identification. If it is a face that already exists in the database, then the person's name and face features will be returned. If it is an unknown face, then it returns that the person is unknown to the visually impaired. With the help of identification information obtained from the unknown person, he/she can be added to the database. The Text to Speech API converts the input text into human-sounding custom speech and returned as an audio file.

Some of the implemented guide systems like smart guide proposed by Manufali [7] gives the user a prototype of an intelligent guide for the blind person is successfully designed, implemented and tested. The prototype device is facilitating the movement of blind person by warning him/her about any nearby obstacles in order to help him/her during daily activities. The guidance will be provided in the form of audio instructions through the headset and based on real-time situation for both indoor and outdoor environments. But it cannot detect what is against the user and it is also low battery storage which cannot long last for long duration.

3 Proposed Framework

3.1 Overview

The Audio navigation system proposed can be implemented in the smart phone itself. It is more battery efficient and portable. A dedicated voice assistant directs the visually impaired with help of google maps. Notifications will be given through headphones via voice commands so that it cannot confuse the user. It will guide the person during walk so one can make sure he is stable while stepping. This requires the internet connection; however, most of the processing is done offline. This system also allows the user to do basic mobile functionalities such as calling and messaging through voice commands. Figure 1 represents the overview of the flow in the system.

3.2 Proposed System

Obstacles in the visually impaired's pathway are captured with live relay using the phone's camera system shown in Fig. 2. The data is then processed and used in guiding the visually impaired via an obstacle free path. This is implemented with the help of TensorFlow powered API. Initial process of this audio navigation system is to find the obstacles before the user. It is required to guide the user how to avoid them without encountering. A depth-sensing camera with IR sensor in a mobile can provide us the data how far an object is away from the camera. Input data is processed through tensor flow and open-cv. Detected objects with its accuracy of detection will be the outcome of this module.

Depth sensing camera helps with identification of objects in layers which conveys which object is to encountered first. The camera also detects the floor using the floor level deduction which helps to identify pits and peaks. Floor detection is done by the barometer which is installed default in the phone.

Audio navigation is required to notify the person "How to move" and "What is before them". First objective of this module is filter what to be notified and what must be neglected.

Fig. 1 System overview
diagram
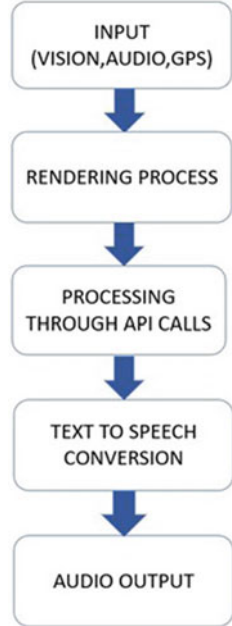
Fig. 2 Mobile camera system

The input to this module will be from the text output of camera vision. The detected objects distanced detected by depth cam, which is shown in Fig. 2, and the object itself will be formed, as a sentence through a program. It will be played as audio to the user. The text to audio conversion is done by gTTS engine service. In terms of storage, data is stored internally.

Objective of this module is to get users audio input and direct them to the places where they like to move. Audio input is processed as text by gTTS module. User's location is obtained from the device as input. The input is given to "DIRECTIONS API" to find the directions for their given place from the user's location. The reply from the API is used by the user to guide them all the way. Nearby API locates the nearby location to their location.

A visually impaired person downloads the developed application in his/her smartphone and performs the initial setup with the guardians help. The person then connects the headphones to the smars application helps to navigate to the desired destination with the help of our assistant providing audible directions. In addition to this, the assistant also provides recommendations for nearby eateries, shops, hospitals etc. In order to use these features the person has to long press the headphone button.

3.3 Applications

The main task of the application is to detect lively the obstacles like roadcrosses, pits, persons and other objects. It can also identify an object when the user captures it. Once it captures the objects it conveys the user by the voice commands. If an user needs to know the nearby locations like hospital it can be provided by asking the system. It provides the hospitals nearby. If the user needs any other locations like hotels or shops it provides them those location details within 1 km. Basic mobile functionalities such as call, message and checking battery status. It locates one to their home with live routing. It says the way to their home in steps. In an emergency it can send SMS to a friend or doctor with the current location details. It can also make call to their Friend or doctor immediately once the system hears the keywords. It allows the visually impaired navigate through cities independently. It can be used for live tracking globally with some constraints. This system will be trained for an individual user which gives them more accessibility.

4 Experimental Result

The camera analyzes and recognize objects. It then conveys this information to the visually impaired via the headphone. As an example consider the above two images. Figure 3 shows how the camera recognize the human faces along with the certainty that the object is a human. Similarly, Fig. 4 shows how camera recognize the bottle. Figure 5 shows the voice assistant interface which will converted as background process. The interface consist of five parts: Results, Response, Get_gps, Live tracking and input. In input field the user's voice is taken as input. Get_gps gives the user location. Input can be given through voice or text. Live tracking can give the obstacle detection lively when the user points at any direction. Audible reply is given by the

Fig. 3 TensorFlow vision detecting persons
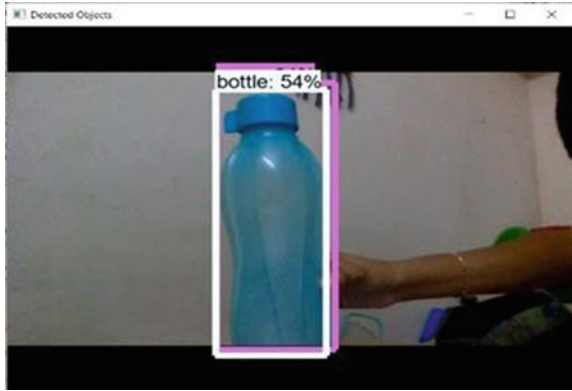
Fig. 4 TensorFlow vision detecting bottle
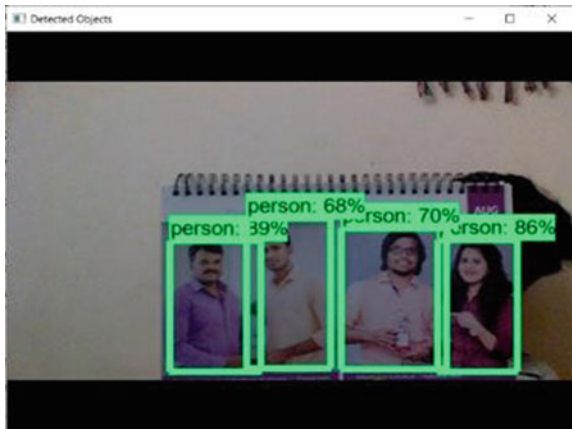
assistant and also the response field will carry the reply. The app logo within an app drawer is shown in Fig. 6.

Most of the audio comments are given by google assistant's default voice with the pitch rate and speed of 20%.It can be adjust in accordance with google account maintenance (Table 1).

On real time comparison between the existing systems and our system we came up with a prediction amount shown in the below matrix. The system detection rate of persons is most accurate as it uses object detection principle for classification of objects. The existing system 1 is down in accuracy because it uses ultra-sonic sensors. These sensors may miss the person's body when there gaps in between them. The existing system 2 uses an object detection which is less accurate to recognize a person. Our system and existing system 2's detection rate of glass objects is slightly less because of the light diffraction whereas ultrasonic sensors don't miss glass object because of light. Pits detection is completely missed by existing system 1 as it considers it as far away object. Our system considers it by using depth sensing

Fig. 5 Application interface
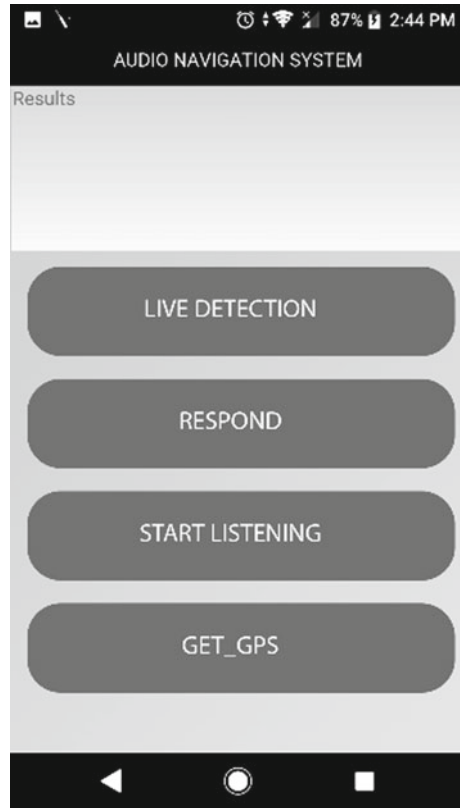
which detects floor level and varies a pit. Whereas in system 2 it misses the accuracy without using floor level detection. We use TensorFlow most accurate algorithm and online api for detecting road crosses where as other systems lacks it.

5 Conclusion

A majority of the proposed system involve the use of the 3d glass or an obstacle detection system to assist the visually impaired for navigation. The systems are often complex to operate and expensive to implement. This paper describes an approach similar to ones discussed but differs in terms of simplicity and cost. The system proposed requires basic smartphone, dedicated service with internet connectivity and a headset to achieve live navigation for visually impaired effectively. Obstacle detection and audible guidance using voice assistant gives an easier way of access.

The system developed is portable in the sense that it requires a smartphone with internet connectivity. Thus, a cost effective and simplistic solution is made possible.

Fig. 6 Application icon in the app-drawer
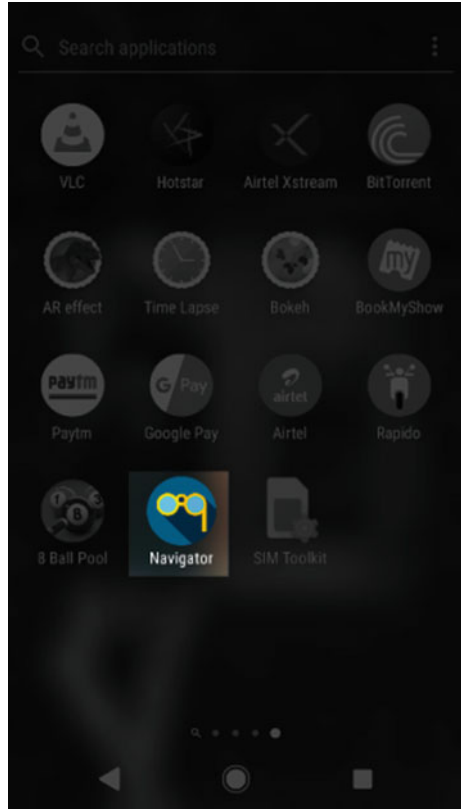
Table 1 Comparative analysis on performance of various supportive systems for visually impaired

	Obstacles	Predicted (%)		
		Smart audio system	Smart guide [7]	Smart guiding glasses [2]
Actual (%)	Persons	95	75	85
	Glass	60	80	60
	Pits	85	2	80
	Road cross	92	0	20
Average (%)	Overall	83	39.25	61.25

Requires internet connection at all times to ensure proper functionality. Problems of usage in places with limited network connectivity. Sometimes, radiation from mobile device imposes health hazards for the life.

6 Future Enhancement

Live detection is to be made more accurate and fully offline capable. This is similar to a person walking beside the visually impaired person and providing direction. This approach will be implemented with machine learning. The image processing is done using TensorFlow. The ultimate goal is to guide the visually impaired through cities effectively and hands-free.

References

1. Li B, Munoz JP, Rong X (2019) Vision-based mobile indoor assistive navigation aid for blind people. IEEE Trans Mobile Comput 18(3):702–714
2. Bai J, Lian S, Liu Z, Wang K, Liu D (2018) Smart guiding glasses for visually impaired people in indoor environment. IEEE Trans Consum Electron 63(3):258–266
3. Lan F, Zhai G, Lin W (2015) Lightweight smart glass system with audio aid for visually impaired people. In: TENCON, IEEE, region 10 conference
4. Bourne RRA, Flaxman SR, Braithwaite T, Cicinelli MV, Das A, Jonas JB, et al (2017) Vision Loss Expert Group. Magnitude, temporal trends, and projections of the global prevalence of visually impairedness and distance and near vision impairment: a systematic review and meta-analysis. Lancet Glob Health 5(9):e888–e897
5. Rentschler AJ, Cooper RA, Blasch B, Boninger ML (2003) Intelligent walkers for the elderly: performance and safety testing of VA-PAMAID robotic walker. J Rehabil Res Dev 40(5):423–432
6. Elaiyaraja, Aishwarya, Akshaya, Monikha M (2019) AJNA—a smart assistance for the visually and cognitively impaired. Int J Adv Res Electric, Electron Instrument Eng
7. Manoufali M, Aladwani A, Alseraidy S, Alabdouli A (2011) Smart guide for blind people international conference and workshop on current trends in information technology

Reducing the Aleatoric Uncertainties of Failure Prediction Using Singular Value Decomposition

Ahmad Kamal Mohd Nor, Srinivasa Rao Pedapati, and Masdi Muhammad

Abstract Uncertainty is a vital indicator in evaluating the prediction of real-world deep learning models. Without uncertainty, important decision-making process linked to safety, security and investment cannot be executed correctly by the users. While probabilistic deep learning methods have shown promise in quantifying prediction uncertainty, very few works in managing these uncertainties can be found in the literature. In failure prediction especially, it is favourable that the model produces the lowest level of prediction uncertainty, i.e., the highest level of confidence in order to initiate any degree of problem-solving mechanism. In this work, we present a technique to reduce Aleatoric uncertainty associated with noisy data using Singular Value Decomposition (SVD). Sensor data from industrial assets are presented in SVD matrix form and higher SVD modes more susceptible to contamination are eliminated to denoise the data. We compare the uncertainty level between the original dataset with the ones denoised by SVD and several known denoising methods in a Remaining Useful Life (RUL) prediction problem. Our results show that SVD denoising treatment outperforms the other denoising methods in reducing prediction uncertainty and improving prediction performance.

Keywords Uncertainty · Prognostic · PHM · RUL · CMAPSS

1 Introduction

Long term exploitation of aerospace assets is one of the goals sought by the industry. Maintenance efforts to ensure operational longevity have evolved steadily from simple corrective maintenance, to preventive measures and more recently to a more systematic framework called Prognostic and Health Management (PHM). PHM aims to optimize maintenance activities by accurately predicting failure event. It has the effect of minimizing maintenance operations, reducing workload and decreasing

A. K. M. Nor (✉) · S. R. Pedapati · M. Muhammad
Universiti Teknologi Petronas, 32610 Perak, Malaysia
e-mail: ahmad_18002773@utp.edu.my

maintenance interventions in hazardous places, resulting to significant maintenance cost reduction [1–4].

Remaining Useful Life (RUL) estimation, where the health of engineered system is predicted, is one of the core activities in PHM [5, 6]. In this domain, we witness the rise of data-driven prognostic in the last 2 decades, fuelled by its simplicity of deployment, the ever-increasing volume of data and the advancement wireless technology. Amongst all, deep learning has emerged as one of the most powerful tools due to its ability to map nonlinear relation of the system's variables.

Deep learning methods can be categorized as either being frequentist or probabilistic in nature [7, 8]. Looking into RUL prognostic angle specifically, frequentist methods predict point estimates RUL while the probabilistic ones predict the distribution or intervals of RULs [9, 10]. The latter thus has the advantage of predicting with uncertainties. In real world application, where ground truth RUL is non-existent, the only indication to trust the model's prediction is uncertainty quantification. This is especially important when manpower or investment decision need to be made following a failure prediction.

1.1 Uncertainty Reduction

Aleatoric uncertainty in deep learning occurs due to noise or stochasticity captured in the input data, which is a common situation in real world applications [11–13]. This uncertainty thus capture noise inherent to the observation. In an ideal failure prognostic model, the uncertainty should be higher when the prediction deviates from the ground truth and stays low in the case of correct prediction.

In real world sensor signal, it is hard to differentiate between the noise and the clean data. Though many denoising and smoothing algorithm have been proposed, most of them need the intervention of domain expert to choose the appropriate level of denoising and smoothing that minimizes the risk of information loss.

To this end, SVD has emerged as a formidable tool in denoising data without the intervention of domain expert [14–17]. By representing noisy data in SVD matrix, we can eliminate the higher SVD modes which are more prone to contamination than the lower ones, retaining only the clean SVD modes that form a clean signal.

In this paper, we present an uncertainty reduction technique based on SVD signal denoising. We demonstrate this ability using CMAPSS gas turbine dataset. To the best of our knowledge, this is the first paper of its kind dedicated in investigating the ability of denoising methods in reducing deep learning uncertainty. This work is important as in real world condition, having the lowest uncertainties, i.e., the highest confidence in failure prediction is the goal for an accurate decision making. This paper is organized as follows. In the next section, related publications are reviewed. In Sect. 3, we elaborate the SVD denoising methodology and performance evaluation. Then, in Sect. 4, a case study is described before we present the results and discussion of our work in Sect. 5. Finally, we give our concluding remarks in the last section.

2   Related Work

SVD has become a standard tool in denoising data in general due to its ability to eliminate singular values residing in noisy subspaces and retaining only the clean ones. In Wang et al [18] a combination of Wavelet Transforms and SVD is used to denoise image data. The wavelet transform is used to divide the image into low and high frequencies divisions. Then SVD is employed to denoise the high frequency parts to enable a clean image reconstruction. In Zhang et al. [19], a two-division recursion SVD is proven to perform best in improving contaminated signals amongst the various SVD model variants. The methods also outperform other signal decomposition techniques such as Empirical Mode Decomposition (EMD), ensemble EMD (EEMD) and Variational Mode Decomposition (VMD). SVD is deployed on matrix of image patches in He et al. [20], and by doing so, separating the signal subspace from the noise subspace. Then, SVD denoising is performed on each group of similar patches that were previously combined as training samples, producing leading singular vectors candidates that represent the corresponding patch. Singular vectors coming from noise populated samples and noise subspace are then eliminated, thus leaving the clean singular vectors that form the local basis of each patch. In Wang et al. [21], the Singular Value Difference (SV) between the noisy and clean image is exploited to denoise image in wavelet domain. SVD is applied to denoise SVs in high frequency parts of the noisy image.

Apart from image denoising, SVD also performs well with general data. In Liu et al. [22], seismic data is denoised by SVD in the frequency domain. Raw seismic data is first presented in the frequency domain by Fourier transform. Then SVD is employed to decompose the frequency spectra into eigen images, which will be then cleaned by selecting clean singular values. In Zhao et al. [23], a reweighted Singular Value Decomposition (RSVD) framework which has the advantage of improving weak features while denoising is proposed. The periodic modulation intensity, an indicator that represents the quality of Singular Components (SCs) information level, is introduced to rank the SCs. Finally, the clean signal is reconstructed based on the information level presented by the SCs.

3   Methodology

3.1   LSTM Architecture with Probabilistic Layer

500 evaluations of Bayesian hyperparameter optimization are employed to find the best hyperparameters for the proposed LSTM architecture [24]. The evaluations are done in the hyperparameter search space presented in Table 1. The Health Index (HI), corresponding to the normalized value of the RUL target for each cycle is used as the training target. The model output layer is set as a Normal distribution layer which produces simultaneously the mean and standard deviation of the HI prediction.

Table 1 Bayesian optimization search space

BayesOpt search space					
Hidden units	Dense layer size	Dropout rate	Mini batch size	Max epoch	Learning rate
10–1000	10–500	0.0–0.5	26–130	10–2000	5e-4 to 1e-3

The training thus intends to minimize the gap between the mean of the predicted HI distributions with the HI target values while at the same time predicting variable standard deviations.

In LSTM, gating functions acts as data flow regulator, preserving or forgetting important or unimportant memory, respectively. LSTM consists of Input Layer, Hidden Layer and Output Layer with a recurrent mechanism enabling the output of the Hidden Layer to be added with new input. This made LSTM particularly powerful in modelling sequential data. Below is the mathematical representation of the model.

LSTM gating equation:

$$f_t = \sigma\left(W_f.X_t + R_f.h_{t-1} + b_f\right) \tag{1}$$

$$i_t = \sigma(W_i.X_t + R_i.h_{t-1} + b_i) \tag{2}$$

$$o_t = \sigma(W_o.X_t + R_o.h_{t-1} + b_o) \tag{3}$$

$$M_t = \varphi(W_c.X_t + R_c.h_{t-1} + b_c) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * M_t \tag{5}$$

$$h_t = o_t * \emptyset(C_t) \tag{6}$$

$W* =$ Input Weight.
$R* =$ Recurrent Weight.
$b* =$ Bias.
$\sigma, \varphi, \emptyset =$ Activation Functions.
Looking at the equations above, Eq. (1)–(3) are Forget Gate, Input Gate and Output Gate equations. Equation (4) is used to calculate the candidate value that will be added to the new state "t" (5) together with the old state "t-1", regulated by the Forget Gate. Equation (6) represents the final output of LSTM. "." is matrix multiplication while "" is point to point multiplication [25].

3.2 Singular Value Decomposition (SVD)

We apply the new SVD denoising methodology as presented in [26, 27]:

1. A is the matrix of clean data expressed as:

$$A \in \mathbb{R}^{T \times D}$$

   with T, the time steps and D the number variables, i.e., column of features and with the condition that $T < D$.

2. A can be expressed in a reduced *singular value decomposition* form as:

$$A = U S V^{\mathsf{T}} = \sum_{k=1}^{T} u_k s_k v_k^{\mathsf{T}} \tag{7}$$

   With $U \in \mathbb{R}^{T \times T}$ and $V \in \mathbb{R}^{D \times T}$ as orthogonal matrix and $S \in \mathbb{R}^{T \times T}$ as a diagonal matrix. The kth SVD mode is composed of:

$$u_k \equiv U_{1:T,k}, \ (Left \ Singular \ Vector)$$

$$s_k \equiv S_{kk}, \ (Singular \ value)$$

$$v_k \equiv V_{1:D,k}, \ (Right \ Singular \ Vector)$$

3. In the same form, noisy data $\hat{A}$ can be express as:

$$\hat{A} \in \mathbb{R}^{T \times D}$$

   and E is the matrix of noise containing random noise with standard deviation, $\in$ that can be expressed as:

$$E \in \mathbb{R}^{T \times D}$$

   with their respective SVD representations and the same descriptions as mentioned before:

$$\hat{A} = \hat{U} \hat{S} \hat{V}^{\mathsf{T}} = \sum_{k=1}^{T} \hat{u}_k \hat{s}_k \hat{v}_k^{\mathsf{T}} \tag{8}$$

$$E = \overset{'}{U} \overset{'}{S} \overset{'}{V}^{\mathsf{T}} = \sum_{k=1}^{T} u_k' s_k' v_k'^{\mathsf{T}}$$

4. Writing $\hat{A}$ in function of A and E:

$$\hat{A} = A + E \tag{9}$$

3.3   Noise Filtering with SVD

Lower SVD modes in noisy data stay quite clean, thus filtering the higher effected ones could give a cleaner signal. We assume the noise to be independent, identically distributed (i.i.d.) Gaussian noise.

1. We project noisy time series data $\tilde{A}$ as SVD matrix to obtain $\tilde{u}_k$, $\tilde{s}_k$ and $\tilde{v}_k$.
2. By fitting a *Marchenko-Pastur distribution* to the tail of noisy singular values $\tilde{s}$, we can calculate the measurement error $\in$. For full calculation, please refer to [14].
3. We can then estimate the root mean square error (RMSE) of the modes with:

$$RMSE(\tilde{v}_k) \approx min\left\{ \sqrt{\frac{2}{D}}, \frac{\in}{\tilde{s}_k} \left[\frac{D-w}{D} + \frac{w}{D} \sum_{\substack{m=1 \\ m \neq k}}^{T} \frac{\tilde{\lambda}_m\left(3\tilde{\lambda}_k - \tilde{\lambda}_m\right)}{\left(\tilde{\lambda}_m - \tilde{\lambda}_k\right)} \right]^{\frac{1}{2}} \right\}$$
$$with\,\tilde{\lambda}_k \equiv \tilde{s}_k \tag{10}$$

4. For i.i.d Gaussian noise, the *spatial correlation parameter*, f and the *effective smoothing window width*, w, should be 1 [26]:

$$f = w = 1$$

5. Once RMSE of the modes obtained, we can calculate the parameter t_k, which represent the cleanness of a mode.

$$t_k = \frac{\log(RMSE(\tilde{v}_k)) - \log\left(\sqrt{\frac{2}{D}}\right)}{\log(RMSE(\tilde{v}_1)) - \log\left(\sqrt{\frac{2}{D}}\right)} \tag{11}$$

6. After that, we estimate the rank for minimum loss data reconstruction with the rank r_{min} as:

$$r_{min} \equiv maximum\,of\,k\,such\,that\,t_k > 5\% \tag{12}$$

7. We then reconstruct the clean singular values:

$$\bar{s}_k = \begin{cases} \tilde{s}_k - \left(\in {}^\prime \hat{s}_k \right)^2 & (k < k_c) \\ 0 & (k_c \leq k) \end{cases} \tag{13}$$

8. With $\in {}^\prime \hat{s}_k$ as the *Marchenko-Pastur distribution* and k_c as the minimum index k that satisfies:

$$\tilde{s}_k < \in {}^\prime \hat{s}_k \tag{14}$$

9. Finally, the clean signal can be reconstructed with r_{min} and $\bar{s}_k$ obtained previously:

$$\overline{A_r} = \sum_{k=1}^{r_{min}} \tilde{u}_k \bar{s}_k \tilde{v}_k \tag{15}$$

3.4 Performance Evaluation

The prediction uncertainty is presented by the RUL distribution's rolling standard deviation. Since the premier objective of the work is to evaluate the level and behavior of uncertainty, comparisons are made between the full feature dataset with data denoised by SVD, Simple Moving Average (MOVAVG), Exponential Weighted Moving Average (EWMA), MATLAB Empirical Mode Decomposition (EMD) and MATLAB Wavelet denoising. We use the default MATLAB settings for EMD and Wavelet.

For moving average methods (simple and exponential), we choose the denoising window size that filter noise whose characteristic best resemble a white noise:

- 0 mean residual noise value.
- Monotonic rolling standard deviation.
- Low autocorrelation.

Secondly, we compare the root mean square error (RMSE) between the mean of RUL distribution and the ground truth [28, 29].

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^{M} \left(RUL_{truth}^{(i)} - Mean_{pred}^{(i)} \right)^2} \tag{16}$$

With $RUL_{truth}^{(i)}$ as the ground truth RUL for gas turbine i, $Mean_{pred}^{(i)}$ as the mean of RUL distribution for gas turbine i and M as the total number of gas turbine.

Fig. 1 Simplified turbofan diagram
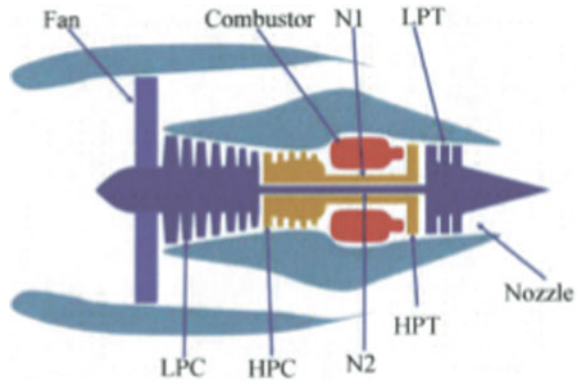
4 Case Study

4.1 CMAPSS Turbofan Dataset

Turbofan failure dataset from NASA Commercial Modular Aero Propulsion System Simulation (CMAPSS), published by Nasa Prognostic Centre (PCoE) of Ames Research Centre is used in this study [30]. The simplified diagram of the studied turbofan is shown in Fig. 1.

The chosen training and testing data consist of 100 recorded turbofan degradations. Each Training and Testing data correspond to a turbofan whose health condition deteriorate after certain cycle [30]. Each fleet training and testing data are time series sequence comprising of Time (Cycle), 3 Operating Conditions (OC) and 21 sensors measurements as in Table 2. The OC refers to different operating regimes combination of Altitude (O–42 K ft.), Throttle Resolver Angle (20–100), and Mach Number (0–0.84) [31]. High levels of noise are incorporated, and the faults are hidden by the effect of various operational conditions [31]. However, no RUL information is given in the training dataset. Only the ground truth RUL for each turbofan, are given. Data pre-processing is thus applied to select the appropriate sensors, visualize the sensor trends, and extract each cycle HI target.

4.2 Sensor Selection

Only strictly monotonic sensors are selected [32]. These sensors are useful as they best represent single trending degradation contrary to irregular and unchanged signals. 14 sensor signals, corresponding to sensors 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 17, 20, 21 are used.

Table 2 Sensors description

Sensor	Ref	Description	Unit
S1	T2	Total temperature fan inlet	^{0}R
S2	T24	Total temperature at LPC outlet	^{0}R
S3	T30	Total temperature at HPC outlet	^{0}R
S4	T50	Total temperature at LPT outlet	^{0}R
S5	P2	Pressure at fan inlet	psia
S6	P15	Total pressure in bypass-duct	psia
S7	P30	Total pressure at HPC outlet	psia
S8	Nf	Physical fan speed	rpm
S9	Nc	Physical core speed	rpm
S10	Epr	Engine pressure ratio (P50/P2)	N/A
S11	Ps30	Static pressure at HPC outlet	psia
S12	Phi	Ratio of fuel flow to Ps30	Pps/psi
S13	NRf	Corrected fan speed	rpm
S14	NRc	Corrected core speed	rpm
S15	BPR	Bypass ratio	N/A
S16	farB	Burner fuel–air ratio	N/A
S17	htBleed	Bleed enthalphy	N/A
S18	Nf_dmd	Demanded fan speed	rpm
S19	PCNfR_dmd	Demanded corrected fan speed	rpm
S20	W31	HPT coolant bleed	lbm/s
S21	W32	LPT coolant bleed	lbm/s

4.3 Piece-Wise Linear Degradation

To obtain RUL sequence labels for training, piece-wise linear degradation model is employed [33, 34]. Each fleet health is thus considered stable in the beginning until the start of deterioration which is the failure start point that initiate a linear degradation until failure.

Each fleet is associated with time series corresponding to the total operational duration of the fleet and the last cycle indicate the final instant before failure. Thus initially, we model the RUL of a fleet to start with the value of the last cycle and degrade linearly until 0 as shown in Fig. 2.
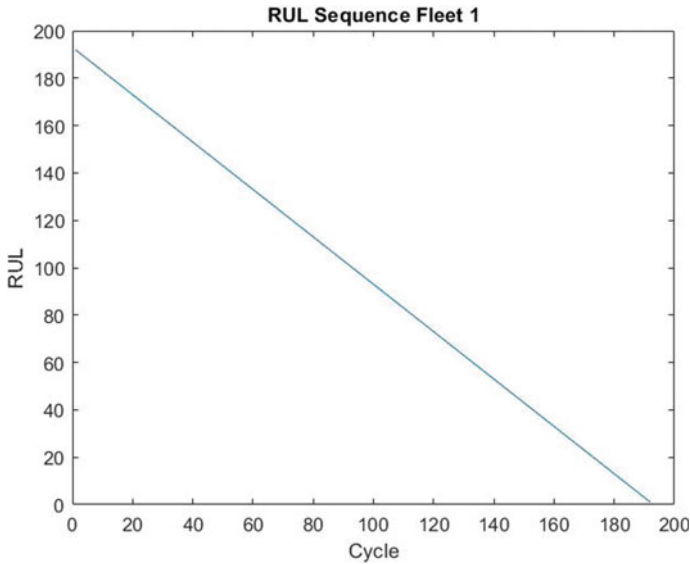
Fig. 2 Initial HI assessment for fleet 1

The failure start point in each sensor is obtained by using Cumulative Sum (CUSUM) anomaly detection technique, which returns the first index of the upper or lower cumulative sums of each sensor signal that have drifted beyond 5 standard deviations from the target mean. This indicates the initiating points of degradation [35]. Then, the mean of the failure start points for each sensor can then be calculated. From these means, we can calculate the mean of the means of the sensor's detection, which is our failure start point. Combining the linear degradation obtained earlier and the failure start point, the transformed Fleet 1 RUL sequence is presented in Fig. 3.

4.4 Data Sequence Length and RUL Values

We show the sequence length plot of the training data in Fig. 4. From this plot, we derive the RUL plot as shown in Fig. 5 following steps mentioned in Sect. 73.3.

In Fig. 6 we show the sequence length of testing and data its associated expected RUL in Fig. 7 following our previous reasoning in Sect. 73.3.

However, as we can see, the ground truth RUL, shown in Fig. 8 are different from the expected ground truth. Some testing data have long sequence length while at the same time having very short RUL. We thus anticipate that the model to perform more poorly on these data.
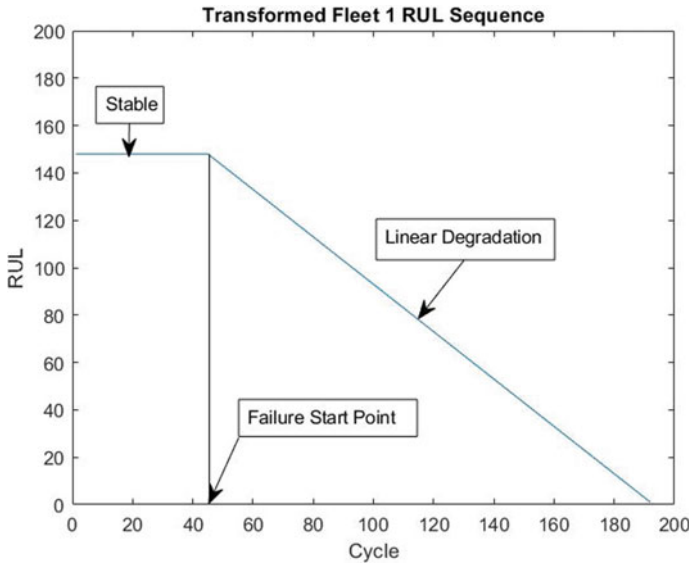
Fig. 3 Final HI assessment for fleet 1
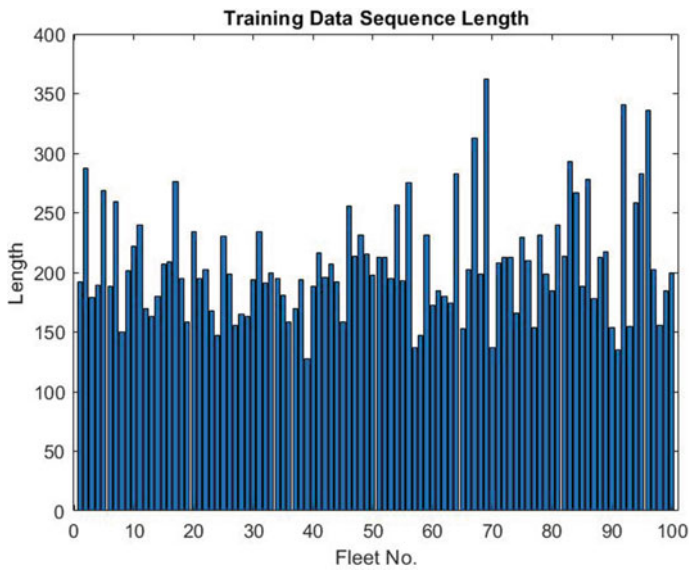
Fig. 4 Training data sequence length
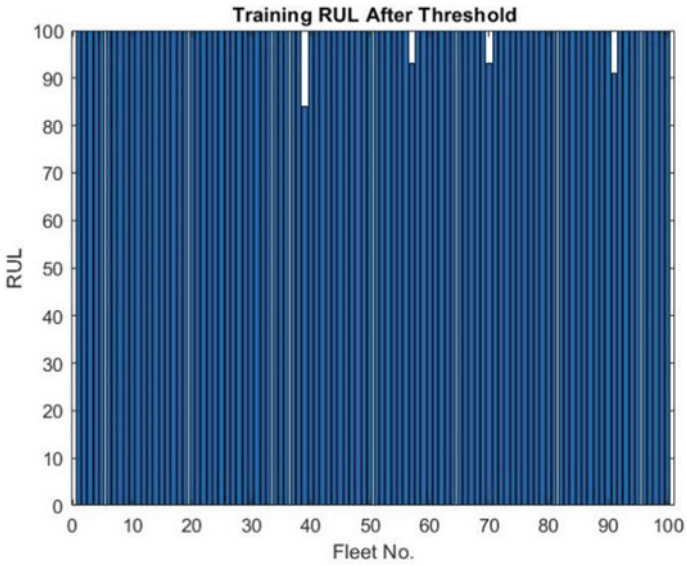
Fig. 5 Training RUL after threshold
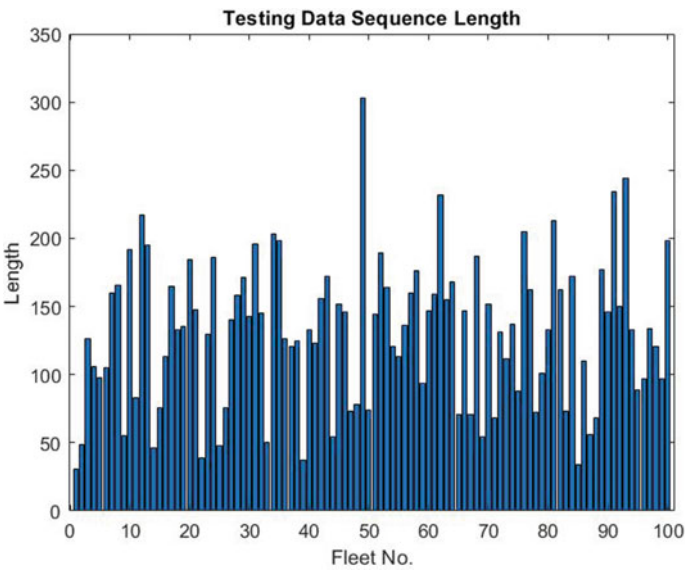
Fig. 6 Test data sequence length
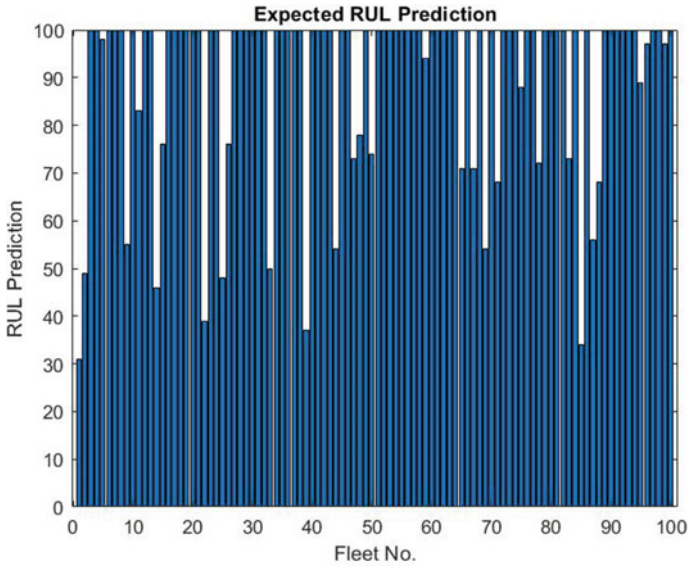
Fig. 7   Expected test data RUL prediction
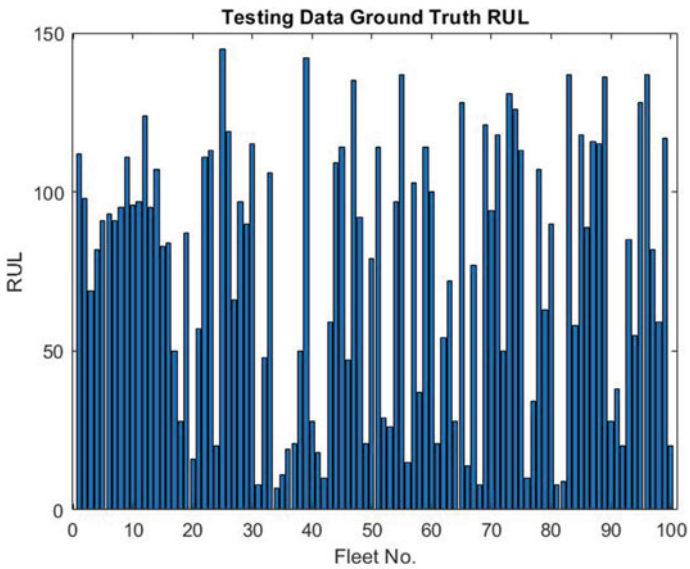
Fig. 8   Ground truth RUL
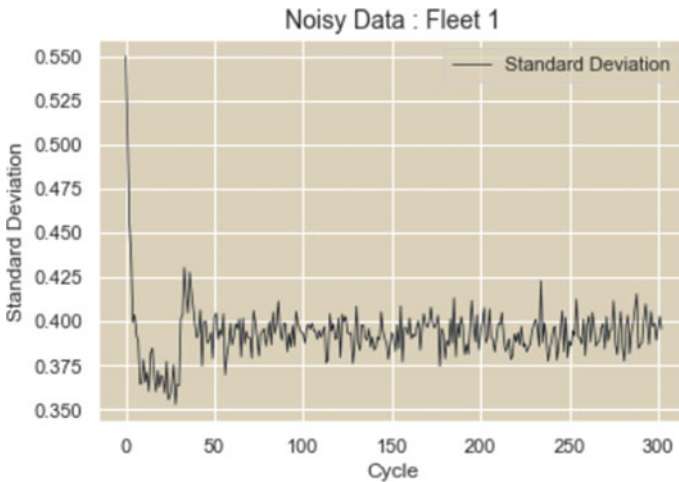
Fig. 9 Full noisy dataset

5 Result and Discussion

5.1 Aleatoric Uncertainty

Window size 2 was found appropriate for MOVAVG while 3 for EWMA. The uncertainty measurement according of the full dataset and respective methods are presented. For illustration purposes, we take the case of Fleet 1 as an example (Fig. 9).

5.2 RMSE Result

As we can see from the illustrations given, the lowest degree of uncertainty is accorded to SVD denoising in Fig. 10 while the highest uncertainty come from EWMA and EMD, shown respectively in Fig. 12 and 14. The curve of rolling standard deviation for SVD is also smoother than the rest of the methods. As for RMSE, again, SVD surpasses other denoising method in improving the probabilistic LSTM performance (Table 3).

MOVAVG and EWMA, as respectively shown in Fig. 11 and 12, cannot denoise the data properly as we are applying the same simplistic rule (window k and fixed weights) to denoise the whole data. As for Wavelet Transform and EMD, presented here in Fig. 13 and 14, the usage of default parameter setting, due to absence of domain expert, is obviously not sufficient to fulfil its intended function. To execute effective denoising in Wavelet Transform, manual manipulations need to be done first to determine the mother wavelet and decomposition level [9]. EMD denoising, on
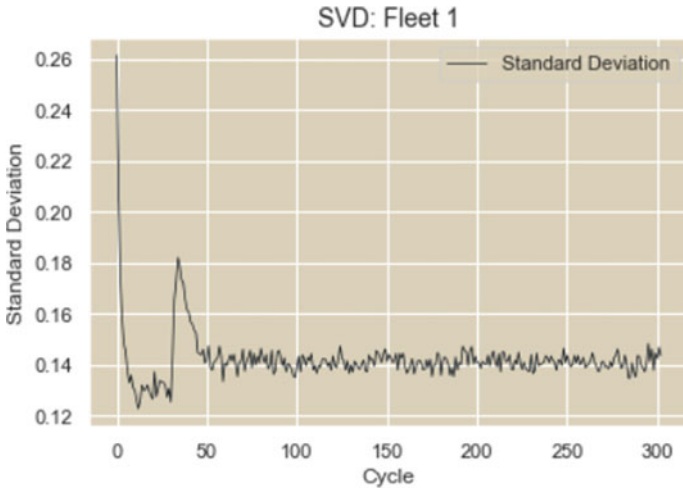
Fig. 10 SVD denoising

Table 3 RMSE results for each method

RMSE results

Method	Noisy data	SVD	MOVAVG	EWMA	WAVELET	EMD
RMSE	34.86	34.82	34.99	35.37	35.53	34.97
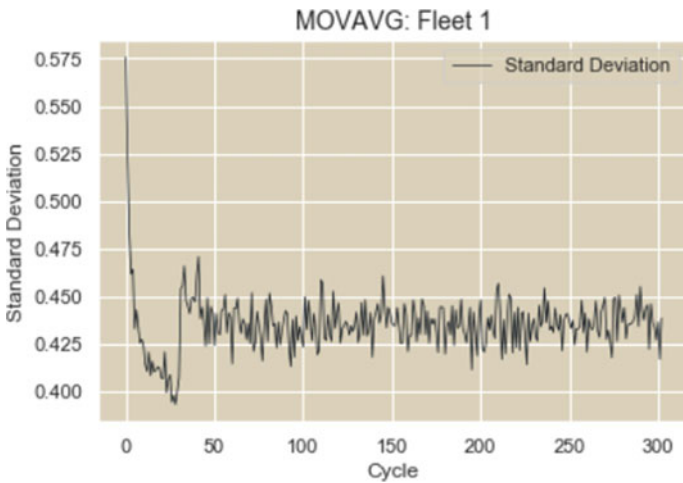
Fig. 11 Moving average
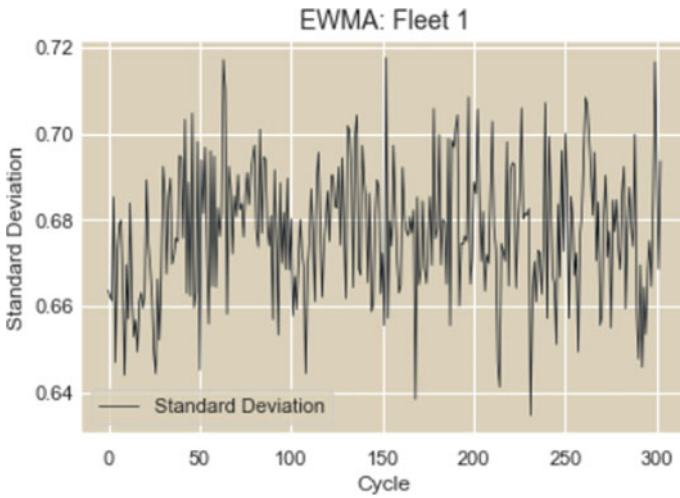
Fig. 12 Exponential weighted moving average
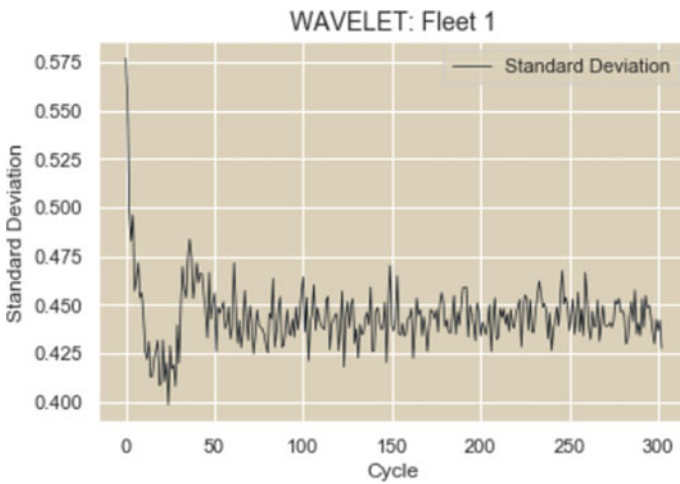
Fig. 13 Wavelet transform

the other hand, possess various decomposition modes that need deep understanding by the user to be successful [9]. SVD however treat each data sites individually by removing higher modes more prone to contamination, thus it denoise more effectively. We show the complete HI prediction for the original data and after SVD treatment in Fig. 15 and 16, respectively.
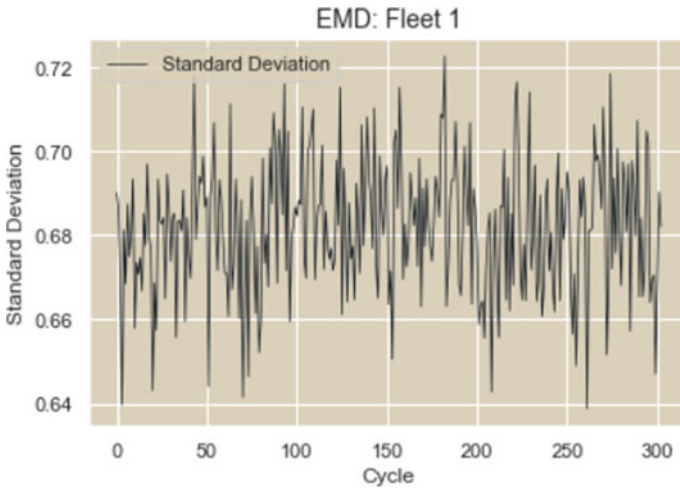
Fig. 14 Empirical mode decomposition denoising
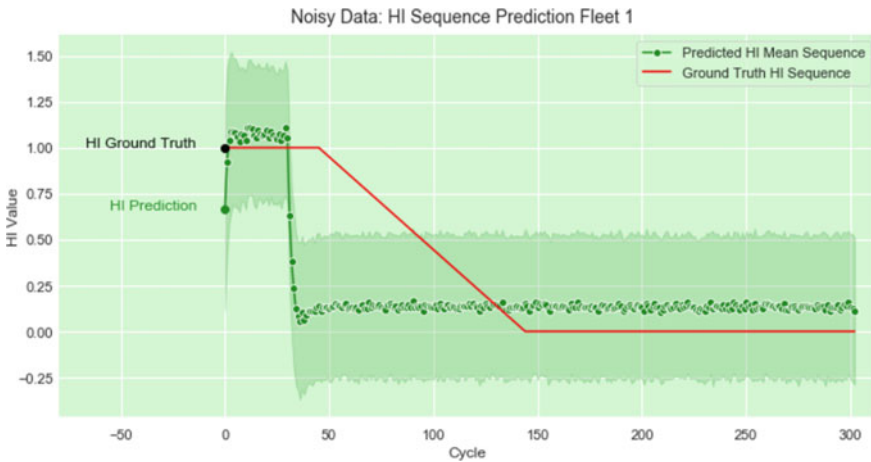
Fig. 15 Original data HI prediction with uncertainty

6 Conclusion

In this paper, SVD denoising treatment is applied to gas turbine failure dataset to reduce the Aleatoric uncertainty of failure prediction. A probabilistic LSTM is employed to predict the RUL distribution of the gas turbines based on sensor data. Our proposed method show superiority in reducing the prediction uncertainty compared to various other denoising techniques such as simple moving average, exponential weighted moving average, empirical mode decomposition and wavelet transform.
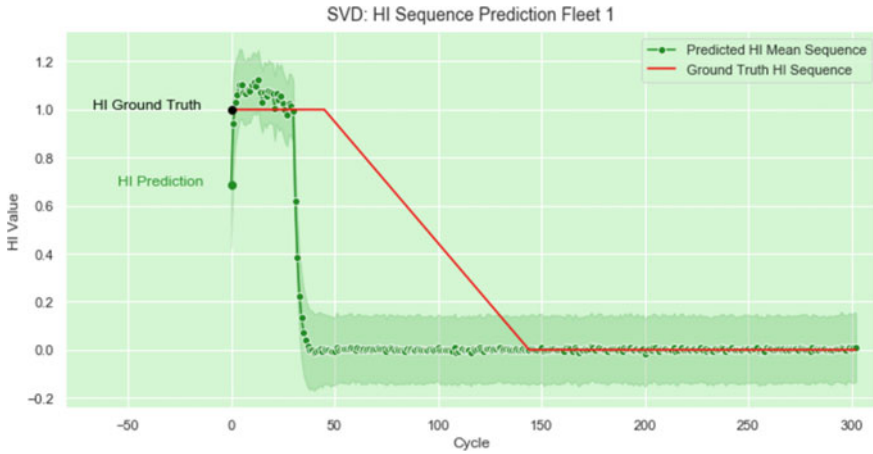
Fig. 16 HI prediction after SVD treatment

The results also indicate that SVD treatment improves the performance of prediction better than the other mentioned techniques.

References

1. Li R, Verhagen W, Curran R (2019) A systematic methodology for prognostic and health management system architecture definition. Reliab Eng Syst Saf 193:106598. https://doi.org/10.1016/j.ress.2019.106598
2. Vogl GW, Weiss BA, Helu M (2019) A review of diagnostic and prognostic capabilities and best practices for manufacturing. J Intell Manuf 30:79–95. https://doi.org/10.1007/s10845-016-1228-8
3. Qu Y, Ming X, Qiu S, Zheng M, Hou Z (2019) An integrative framework for online prognostic and health management using internet of things and convolutional neural network. Sensors 19:2338
4. Wang F, Pan S, Xiong Y, Fang H, Wang D (2017) Research on software architecture of prognostics and health management system for civil aircraft. In: 2017 international conference on sensing, diagnostics, prognostics, and control (SDPC). Shanghai, pp 510–513. https://doi.org/10.1109/SDPC.2017.102
5. Li X, Xie G, Liu H, Wang W (2018) Predicting remaining useful life of industrial equipment based on multivariable monitoring data analysis. In: 2018 Chinese automation congress (CAC). Xi'an, China, pp 1861–1866. https://doi.org/10.1109/CAC.2018.8623249
6. Zhang H, Hu C, Fan H, Zhang W, Gao Y (2015) A new remaining useful life prediction approach based on Wiener process with an adaptive drift. In: 2015 Chinese automation congress (CAC). Wuhan, pp 2052–2056. https://doi.org/10.1109/CAC.2015.7382842
7. Jun S (2016) Frequentist and Bayesian learning approaches to artificial intelligence. Int J Fuzzy Logic Intell Syst 16:111–118. https://doi.org/10.5391/IJFIS.2016.16.2.111

8. Alaa A, Schaar M (2020) Frequentist uncertainty in recurrent neural networks via blockwise influence functions
9. Kim M, Liu K (2020) A Bayesian deep learning framework for interval estimation of remaining useful life in complex systems by incorporating general degradation characteristics. IISE Trans. https://doi.org/10.1080/24725854.2020.1766729
10. Peng W, Ye Z, Chen N (2020) Bayesian deep-learning-based health prognostics toward prognostics uncertainty. IEEE Trans Industr Electron 67(3):2283–2293. https://doi.org/10.1109/TIE.2019.2907440
11. Shaker MH, Hüllermeier E (2020) Aleatoric and epistemic uncertainty with random forests. In: Berthold M, Feelders A, Krempl G (eds) Advances in intelligent data analysis XVIII. IDA 2020. Lecture notes in computer science, vol 12080. Springer, Cham. https://doi.org/10.1007/978-3-030-44584-3_35
12. Ståhl N, Falkman G, Karlsson A, Mathiason G (2020) Evaluation of uncertainty quantification in deep learning. In: Lesot MJ et al (eds) Information processing and management of uncertainty in knowledge-based systems. IPMU 2020. Communications in computer and information science, vol 1237. Springer, Cham. https://doi.org/10.1007/978-3-030-50146-4_41
13. Hüllermeier E, Waegeman W (2019) Aleatoric and epistemic uncertainty in machine learning: a tutorial introduction
14. Deng W, Ye B, Wu J, Wang X, Bao J, Chen C (2019) Sparse denoising of eddy current signals from conductive material defects based on K-SVD dictionary learning. In 2019 IEEE 8th data driven control and learning systems conference (DDCLS). Dali, China, pp 656–660. https://doi.org/10.1109/DDCLS.2019.8909047
15. Seyedebrahim M, Mansouri A (2017) Non-local means denoising based on SVD basis images. In: 2017 3rd international conference on pattern recognition and image analysis (IPRIA). Shahrekord, pp 206–210. https://doi.org/10.1109/PRIA.2017.7983047
16. Yang Y, Rao J (2019) Robust and efficient harmonics denoising in large dataset based on random SVD and soft thresholding. IEEE Access 7:77607–77617. https://doi.org/10.1109/ACCESS.2019.2921579
17. Zhao Z, Wang S, Wong D, Guo Y, Chen X (2020) The sparse and low-rank interpretation of SVD-based denoising for vibration signals. In: 2020 IEEE international instrumentation and measurement technology conference (I2MTC). Dubrovnik, Croatia, pp 1–6. https://doi.org/10.1109/I2MTC43012.2020.9129272
18. Wang M, Li Z, Duan X, Li W (2015) An image denoising method with enhancement of the directional features based on wavelet and SVD transforms. Math Probl Eng 2015:1–9. https://doi.org/10.1155/2015/469350
19. Zhang G, Xu B, Zhang K, Hou J, Xie T, Li X, Liu F (2020) Research on a noise reduction method based on multi-resolution singular value decomposition. Appl Sci 10:1409. https://doi.org/10.3390/app10041409
20. He Y, Gan T, Chen W, Wang H (2011) Adaptive denoising by singular value decomposition. IEEE Signal Process Lett 18:215–218. https://doi.org/10.1109/LSP.2011.2109039
21. Wang M, Yan W, Zhou S (2018) Image denoising using singular value difference in the wavelet domain. Math Probl Eng 2018:1–19. https://doi.org/10.1155/2018/1542509
22. Liu B, Liu Q (2020) Random noise reduction using SVD in the frequency domain. J Petrol Explor Prod Technol 10:3081–3089. https://doi.org/10.1007/s13202-020-00938-w
23. Zhao M, Jia X (2017) A novel strategy for signal denoising using reweighted SVD and its applications to weak fault feature enhancement of rotating machinery. Mech Syst Signal Process 94:129–147. https://doi.org/10.1016/j.ymssp.2017.02.036
24. Sherstinsky A (2018) Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network
25. Zheng S, Ristovski K, Farahat A, Gupta C (2017) Long short-term memory network for remaining useful life estimation pp. 88–95. https://doi.org/10.1109/ICPHM.2017.7998311
26. Epps B, Krivitzky E (2019) Singular value decomposition of noisy data: noise filtering. Exp Fluids 60:1–23. https://doi.org/10.1007/s00348-019-2768-4

27. Epps B, Krivitzky E (2019) Singular value decomposition of noisy data: mode corruption. Exp Fluids 60:1–23. https://doi.org/10.1007/s00348-019-2761-y

28. Wang Q, Zheng S, Farahat A, Serita S, Gupta C (2019) Remaining useful life estimation using functional data analysis. In: 2019 IEEE international conference on prognostics and health management (ICPHM). San Francisco, CA, USA, pp 1–8. https://doi.org/10.1109/ICPHM.2019.8819420

29. Ge Y, Sun L, Ma J (2019) An improved PF remaining useful life prediction method based on quantum genetics and LSTM. IEEE Access 7:160241–160247. https://doi.org/10.1109/ACCESS.2019.2951197

30. Ramasso E, Saxena A (2014) Performance benchmarking and analysis of prognostic methods for CMAPSS datasets. Int J Prognostics Health Manag 5:1–15

31. Saxena A, Goebel K, Simon D, Eklund N (2008) Damage propagation modelling for aircraft engine run-to-failure simulation. Int Conf Prognostics Health Manag. https://doi.org/10.1109/PHM.2008.4711414.M

32. Gao Y, Zhou J, Wu K, Zhao G, Hu C (2019) Construction method of turbine engine health indicator based on deep learning. In: 2019 prognostics and system health management conference (PHM-Qingdao). Qingdao, China, pp 1–6. https://doi.org/10.1109/PHM-Qingdao46334.2019.8943055; Wang Q, Zheng S, Farahat A, Serita S, Gupta C (2019) Remaining useful life estimation using functional data analysis. In: 2019 ieee international conference on prognostics and health management (ICPHM). San Francisco, CA, USA, pp 1–8. https://doi.org/10.1109/ICPHM.2019.8819420

33. Li J, Li X, He D (2019) Domain adaptation remaining useful life prediction method based on AdaBN-DCNN. In: 2019 prognostics and system health management conference (PHM-Qingdao). Qingdao, China, pp 1–6. https://doi.org/10.1109/PHM-Qingdao46334.2019.8942857

34. Detect small changes in mean using cumulative sum. MATLAB MathWorks® https://www.mathworks.com/help/signal/ref/cusum.html

35. Heimes F (2008) Recurrent neural networks for remaining useful life estimation. In: IEEE international conference on prognostics and health management

Reducing Uncertainty in Failure Prediction Using Singular Value Decomposition Feature Selection

Ahmad Kamal Mohd Nor, Srinivasa Rao Pedapati, and Masdi Muhammad

Abstract Uncertainty is one of the indicators to evaluate the prediction of deep learning models. In real-world prognostic domain, predicting failure with uncertainty provides users with powerful argument when important decisions related to safety, security and monetary must be made. While quantifying uncertainty has slowly becoming a norm in deep learning, no work has been dedicated in formulating the appropriate feature selection framework that lessens the prediction uncertainties. Much of the available techniques have only been proven to improve point estimate predictions, without considering its effect on uncertainty. In this paper, a feature selection method based on elimination of noisy data is proposed to reduce the Aleatoric uncertainty in a Remaining Useful Life (RUL) prediction problem. Singular Value Decomposition (SVD) technique is employed to denoise sensor data in SVD matrix by filtering higher SVD modes susceptible to contain noise. Then, the "cleaned" Signal to Noise Ratio (SNR) of each feature is calculated and ranked. Features with low SNR, thus with higher noise, are eliminated. We compare the uncertainty level and behavior between a full feature dataset and different percentage of features selected using SVD and SNR. The same comparison is done between our approach and other feature selection methods such as Pearson and Spearman correlations as well as F Regression. The results show that our approach achieved a lower uncertainty degree with generally better prediction performance than the other mentioned methods.

Keywords Uncertainty · Feature selection · PHM · RUL · CMAPSS

1 Introduction

Sudden failure is the worst imaginable situation involving complex engineering systems. It is a risk to be minimized to safeguard the interest of safety, security, and investment in the industrial environment. Throughout the decades, concerted efforts

A. K. M. Nor (✉) · S. R. Pedapati · M. Muhammad
Universiti Teknologi Petronas, 32610 Perak, Malaysia
e-mail: ahmad_18002773@utp.edu.my

have been put by researchers and reliability practitioners to improve operational longevity of industrial assets.

Recently, Prognostic and Health Management (PHM) has emerged as a strong force in providing frameworks for managing engineering system health [1–3]. PHM aims to predict impending failure, optimizing maintenance activities, reducing workload, improving safety, availability and reducing maintenance cost [4–7].

Remaining Useful Life (RUL) prognostic is one of the core activities in PHM [8, 9]. It consists of predicting the health state of engineering system often complex in nature. Currently, deep learning prognostic methods gain much attention due to the ease of deployment, the increasing data volume, and the sophistication of data transmission technology.

Deep learning methods can be divided into frequentist and probabilistic in nature. While the premier predicts or classifies point estimates, the latter produces distribution or interval predictions, thus predicting with uncertainty.

1.1   Aleatoric Uncertainty in Deep Learning

Uncertainty related to the quality of input data is called Aleatoric uncertainty. This uncertainty occurs when sensor data, used as input to deep learning model, is subjected to measurement error or noise contamination [10–13]. This uncertainty cannot be reduced by collecting more data [12, 13].

In the real-world applications, where ground truth RUL is not available, uncertainty is often the main indicator for users to trust the model's prediction. Security, safety, and investment strategies depend on prediction uncertainty.

1.2   Data Denoising Using SVD

Singular Value Decomposition (SVD) is a technique to represent signal in SVD matrix form. SVD matrix is composed of modes and singular values. When noisy data is represented in SVD matrix form, higher modes are found to be more prone to contamination while the lower modes remain relatively clean [14, 15]. By filtering the higher modes, the clean signal can be reconstructed.

1.3   Feature Selection's Impact on Uncertainty

Since Aleatoric uncertainty concerns the quality of input data, it is natural to assume that certain features contribute more to uncertainty than others. However, there have been no notable work dedicated to study this effect. Available feature selection techniques are suitable only to improve frequentist point estimates prediction without

considering the impact it can cause uncertainty. It is obvious that in the real-world deep learning applications, the lowest uncertainty in prediction is preferred.

In this paper, a feature selection technique to reduce Aleatoric uncertainty is proposed. Noisy data is denoised using SVD. Then, the Signal to Noise Ratio (SNR) for each clean feature is calculated and ranked. Signals with low SNR, i.e., high noise are eliminated according to certain percentage.

We demonstrate our approach in a gas turbine Remaining Useful Life (RUL) prediction problem. A probabilistic LSTM is employed to predict the gas turbine's RUL sequence distributions. We compare the impact of our approach on prediction uncertainty with full feature dataset. Comparison is also done with several others commonly used feature selection techniques such as Pearson and Spearman correlations and F Regression. We believe that this is the first such work concerning deep learning uncertainty reduction using feature selection. This paper is very important as uncertainty management is crucial for real-world deep learning application.

This paper is organized as follows: In Sect. 2, we mention the related literature to our work. In the next section, the methodology of SVD and SNR feature selection is explained. Then in Sect. 4, we describe the case study before we present and discuss our findings in Sect. 5. Finally, we present our concluding remarks in the last section.

2 Related Work

2.1 Uncertainty Reduction in Deep Learning

Most of the works concerning uncertainty in deep learning relate to its quantification. Very few work touches about uncertainty management. One such publication is [16] where Fuzzy Deep Neural Network (FDNN) is proposed to reduce data ambiguity and noise for classification task through the fusion of fuzzy rules and neural network logic representation. In Boskos et al. [17], Wasserstein ambiguity sets are employed in dynamic data collection process. Ambiguity sets enables the characterization of the true random variable's probability distribution's uncertainty, thus exploiting high confidence guarantees to minimize uncertainty. In Du et al. [18], to minimize the uncertainty in domestic solar energy harvesting, only predictions of photovoltaic (PV) output with large probability are chosen. These probabilities help to construct plausible scenarios enabling accurate home energy scheduling. In Rodemerk et al. [19] context-based indicators are used to reinforce the prediction confidence, i.e., minimize the uncertainty of car driver's behavior at intersections. This is done by combining context information, such as head pose and gaze direction with vehicle data to predict the intention of the driver.

2.2 SVD Denoising

SVD is a popular tool for denoising data without the intervention of domain expert. It is widely used in denoising image data. In Yang et al. [20] SVD, together with non-local self-similarity technique are used to denoise image data. Block matching techniques construct matrix based on similar patches that will be denoise by SVD. The denoised patches will be then be combined to reconstruct the clean image. In Leal et al. [21], sparse representation and SVD are employed to denoise MRI image for enhancing medical diagnosis. Sparse representations are obtained from the image subdivisions. Then, new subdivisions based on the representations are produced through SVD denoising process. These clean subdivisions will be used later to reconstruct the clean image. The same concept is applied in Yang et al. [22] where SVD is used to separate and denoise Computed tomography (CT) image components.

SVD is also used to denoise general noisy data. In Zhu et al. [23], an algorithm based on generalized S transform (GST) and SVD is employed to denoise echo signals in ultrasonic pulse-echo testing. The GST's function is to transform the echo signal to time–frequency matrix before going through SVD denoising treatment. In Zhang et al. [24], noisy signal is presented by SVD in Henkel matrix form where clean subspaces and noisy subspaces are separated. By only taking singular values residing in clean subspace, clean signal can be reconstructed. In Schanze [25] a signal compression and denoising method via objective transformation and singular values denoising is presented. The noisy signal is first mapped into matrix form where denoising and compression process are done. Then, inverse mapping is executed to reconstruct the clean signal. The application of this technique is demonstrated using biomedical signals.

3 Methodology

3.1 LSTM Architecture with Probabilistic Layer

LSTM architecture which is a powerful tool for sequential data modelling is employed. We execute Bayesian hyperparameters optimizations to find the best set of hyperparameter in the search space shown in Table 1 [26]. 500 Bayesian optimizations are done to minimize the validation error. For the sequential training targets, we use Health Index (HI) that corresponds to the normalized value of the RUL of each cycle. A Normal distribution layer is incorporated as the output layer of the LSTM network, outputting both mean and standard deviation of the HI prediction.

LSTM gating equation:

$$f_t = \sigma\left(W_f.X_t + R_f.h_{t-1} + b_f\right) \tag{1}$$

Fig. 1 Simplified turbofan diagram
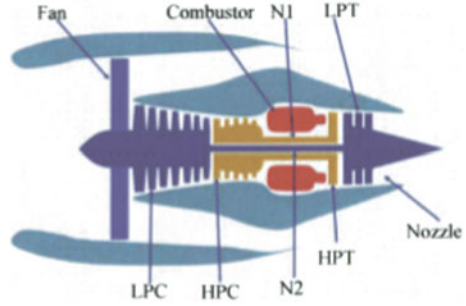
$$i_t = \sigma(W_i.X_t + R_i.h_{t-1} + b_i) \tag{2}$$

$$o_t = \sigma(W_o.X_t + R_o.h_{t-1} + b_o) \tag{3}$$

$$M_t = \varphi(W_c.X_t + R_c.h_{t-1} + b_c) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * M_t \tag{5}$$

$$h_t = o_t * \varnothing(C_t) \tag{6}$$

With W^* as the Input Weights, R^* as the Recurrent Weights, b^* as bias, and $\sigma, \varphi, \varnothing$ as activation functions.

In LSTM, gating mechanism enables the preserving and forgetting of data, resulting to a longer memory retention [27]. It consists of Input Layer, Hidden Layer and Output Layer with a recurrent mechanism enabling the output of the Hidden Layer to be added with new input. This permits the modelling of sequential data.

Equation (1)–(3) are Forget Gate, Input Gate and Output Gate equations. Equation (4) is used to calculate the candidate value that will be added to the new state "t" Eq. (5) together with the old state "t-1", regulated by the Forget Gate. Equation (6) represents the final output of LSTM. "." is matrix multiplication while "" is point to point multiplication [28].

3.2 Singular Value Decomposition (SVD)

The new SVD denoising method as presented in [14, 15] is used:

1. **A** is the matrix of clean data expressed as:

$$A \in \mathbb{R}^{T \times D}$$

with T, the time steps and D the number variables, i.e., column of features and with the condition that $T < D$.

2. A can be expressed in a reduced *singular value decomposition* form as:

$$A = U S V^{\mathsf{T}} = \sum_{k=1}^{T} u_k s_k v_k^{T} \qquad (7)$$

With $U \in \mathbb{R}^{T \times T}$ and $V \in \mathbb{R}^{D \times T}$ as orthogonal matrix and $S \in \mathbb{R}^{T \times T}$ as a diagonal matrix. The k_{th} SVD mode is composed of:

$$u_k \equiv U_{1:T,k}, \ (Left \ Singular \ Vector)$$

$$s_k \equiv S_{kk}, \ (Singular \ value)$$

$$v_k \equiv V_{1:D,k}, \ (Right \ Singular \ Vector)$$

3. In the same form, noisy data $\hat{A}$ can be express as:

$$\hat{A} \in \mathbb{R}^{T \times D}$$

and E is the matrix of noisy data containing random noise with standard deviation, $\in$ that can be expressed as:

$$E \in \mathbb{R}^{T \times D} E \in \mathbb{R}^{T \times D}$$

with their respective SVD representations and the same descriptions as mentioned before:

$$\hat{A} = \hat{U} \hat{S} \hat{V}^{\mathsf{T}} = \sum_{k=1}^{T} \hat{u}_k \hat{s}_k \hat{v}_k^{\mathsf{T}} \qquad (8)$$

$$E = \acute{U} \acute{S} \acute{V}^{\mathsf{T}} = \sum_{k=1}^{T} \acute{u}_k \acute{s}_k \acute{v}_k^{\mathsf{T}}$$

4. Writing $\hat{A}$ in function of A and E:

$$\hat{A} = A + E \qquad (9)$$

3.3 Noise Filtering with SVD

Lower SVD modes in noisy data stay quite clean, thus filtering the higher effected ones could give a cleaner signal. We assume the noise to be independent, identically distributed (i.i.d.) Gaussian noise.

1. We project noisy time series data $\tilde{A}$ as SVD matrix to obtain $\tilde{u}_k$, $\tilde{s}_k$ and $\tilde{v}_k$.
2. By fitting a *Marchenko-Pastur distribution* to the tail of noisy singular values $\tilde{s}$, we can calculate the measurement error $\in$. For full calculation, please refer to [14].
3. We can then estimate the root mean square error (RMSE) of the modes with:

$$
RMSE(\tilde{v}_k) \approx \min\left\{ \sqrt{\frac{2}{D}}, \frac{\in}{\tilde{s}_k} \left[\frac{D-w}{D} + \frac{w}{D} \sum_{\substack{m=1 \\ m \neq k}}^{T} \frac{\tilde{\lambda}_m\left(3\tilde{\lambda}_k - \tilde{\lambda}_m\right)}{\left(\tilde{\lambda}_m - \tilde{\lambda}_k\right)} \right]^{\frac{1}{2}} \right\}
$$

$$
\text{with} \tilde{\lambda}_k \equiv \tilde{s}_k \tag{10}
$$

4. For i.i.d Gaussian noise, the *spatial correlation parameter*, f and the *effective smoothing window width*, w, should be 1 [14]:

$$
f = w = 1
$$

5. Once RMSE of the modes obtained, we can calculate the parameter t_k, which represent the cleanness of a mode.

$$
t_k = \frac{log(RMSE(\tilde{v}_k)) - log\left(\sqrt{\frac{2}{D}}\right)}{log(RMSE(\tilde{v}_1)) - log\left(\sqrt{\frac{2}{D}}\right)} \tag{11}
$$

6. After that, we estimate the rank for minimum loss data reconstruction with the rank r_{min} as:

$$
r_{min} \equiv maximum\ of\ k\ such\ that\ t_k > 5\% \tag{12}
$$

7. We then reconstruct the clean singular values:

$$
\bar{s}_k = \begin{cases} \tilde{s}_k - \left(\in \hat{s}_k\right)^2 & (k < k_c) \\ 0 & (k_c \leq k) \end{cases} \tag{13}
$$

With $\in \prime \hat{s}_k$ as the *Marchenko-Pastur distribution* and k_c as the minimum index k that satisfies:

$$\tilde{s}_k < \in \prime \hat{s}_k \tag{14}$$

8. Finally, the clean signal can be reconstructed with r_{min} and $\bar{s}_k$ obtained previously:

$$\overline{A_r} = \sum_{k=1}^{r_{min}} \tilde{u}_k \bar{s}_k \tilde{v}_k \tag{15}$$

3.4 Signal to Noise (SNR) Calculation

For amplitude-based signal, the SNR can be calculated by Eq. (16) with S_{clean} as the clean signal and S_{noise} is the noise signal.

$$SNR_S = 20 log_{10}\left(\frac{S_{clean}}{S_{noise}}\right) \tag{16}$$

3.5 Performance Evaluation

The prediction uncertainty is presented by the RUL distribution's rolling standard deviation. Since the premier objective of the work is to evaluate the level and behaviour of uncertainty, comparisons are made between the full feature dataset with only 85 and 75% of the full features selected by SVD + SNR, Pearson, and Spearman correlations as well as F Regression feature selection techniques.

Secondly, we compare the root mean square error (RMSE) between the mean of RUL distribution and the ground truth [29, 30].

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^{M}\left(RUL_{truth}^{(i)} - Mean_{pred}^{(i)}\right)^2} \tag{17}$$

With $RUL_{truth}^{(i)}$ as the ground truth RUL for gas turbine i, $Mean_{pred}^{(i)}$ as the mean of RUL distribution for gas turbine i and M as the total number of gas turbine.

Table 1 Bayesian optimization search space

BayesOpt search space					
Hidden units	Dense layer size	Dropout rate	Mini batch size	Max epoch	Learning rate
10–1000	10–500	0.0–0.5	26–130	10–2000	5e-4 to 1e-3

4 Case Study

4.1 Gas Turbine Dataset

The data we use is from CMAPPS (Commercial Modular Aero Propulsion System Simulation) Turbofan run-to-failure dataset, published by Nasa Prognostic Centre (PCoE) of Ames Research Centre [31]. The diagram of the studied turbofan is shown in Fig. 1.

Run-to-failure time series data from 100 recorded turbofan degradations is used for training and testing. Each set of data corresponds to a turbofan whose health condition begin to deteriorate after certain cycle [31]. Each set comprises of Time (Cycle), 3 Operating Conditions (OC) and 21 sensors measurements as in presented in Table 2. The OC refers to different operating regimes combination of Altitude (O-42 K ft.), Throttle Resolver Angle (20–100), and Mach Number (0–0.84) [32]. High levels of noise are incorporated, and the faults encountered are hidden by the effect of various operational conditions [32]. Training data RUL targets are absent, only the ground truth RUL for each turbofan, are given. We thus need to do data preparation to select the appropriate sensors, visualize the sensor trends and extract each cycle RUL target.

4.2 Sensor Selection

From the 21 sensors mentioned above, only 14 strictly monotonic sensors, namely sensor 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 17, 20 and 21 are selected [33]. Monotonicity of these sensors represent single trending degradation contrary to irregular and unchanged sensor signals.

4.3 Piece-Wise Linear Degradation

To obtain RUL sequence labels for training, piece-wise linear degradation model is employed [34, 35]. Each fleet health is thus considered stable in the beginning until the start of deterioration which is the failure start point that initiate a linear degradation until failure.

Table 2 Sensors description

Sensor	Ref	Description	Unit
S1	T2	Total temperature fan inlet	^{0}R
S2	T24	Total temperature at LPC outlet	^{0}R
S3	T30	Total temperature at HPC outlet	^{0}R
S4	T50	Total temperature at LPT outlet	^{0}R
S5	P2	Pressure at fan inlet	psia
S6	P15	Total pressure in bypass-duct	psia
S7	P30	Total pressure at HPC outlet	psia
S8	Nf	Physical fan speed	rpm
S9	Nc	Physical core speed	rpm
S10	Epr	Engine pressure ratio (P50/P2)	N/A
S11	Ps30	Static pressure at HPC outlet	psia
S12	Phi	Ratio of fuel flow to Ps30	Pps/psi
S13	NRf	Corrected fan speed	rpm
S14	NRc	Corrected core speed	rpm
S15	BPR	Bypass ratio	N/A
S16	farB	Burner fuel–air ratio	N/A
S17	htBleed	Bleed enthalphy	N/A
S18	Nf_dmd	Demanded fan speed	rpm
S19	PCNfR_dmd	Demanded corrected fan speed	rpm
S20	W31	HPT coolant bleed	lbm/s
S21	W32	LPT coolant bleed	lbm/s

Table 3 SNR measurement

Sensor	S2	S3	S4	S7	S8	S9	S11
SNR	68.4	57.8	56.8	65.0	75.8	74.9	56.1
Sensor	**S12**	**S13**	**S14**	**S15**	**S17**	**S20**	**S21**
SNR	66.8	75.8	74.1	55.2	55.2	54.5	54.5

Table 4 SNR feature selection

SNR + SVD									
85%	S2	S3	S4	S7	S8	S9	S11	S12	S13	S14
75%	S2	S3	S4	S7	S8	S9	S11	S12		

Table 5 Pearson and spearman feature selection

Pearson and spearman correlation											
85%	S2	S3	S4	S7	S8	S9	S11	S13	S15	S17	S20	S21
75%	S2	S3	S4	S7	S8	S9	S15	S17	S20	S21		

Table 6 F regression feature selection

F regression											
85%	S2	S3	S4	S7	S8	S11	S12	S13	S15	S17	S20	S21
75%	S2	S3	S4	S7	S11	S12	S15	S17	S20	S21		

Each fleet is associated with time series corresponding to the total operational duration of the fleet and the last cycle indicate the final instant before failure. Thus initially, we model the RUL of a fleet to start with the value of the last cycle and degrade linearly until 0 as shown in Fig. 2.

The failure start point in each sensor is obtained by using Cumulative Sum (CUSUM) anomaly detection technique, which returns the first index of the upper and lower cumulative sums of each sensor signal that have drifted beyond 5 standard deviations above and below a target mean, indicating the initiating points of degradation [36]. Then, the mean of the failure start points for each sensor can then be calculated. From these means, we can calculate the mean of the means of the sensor's detection, which is our failure start point. Combining the linear degradation
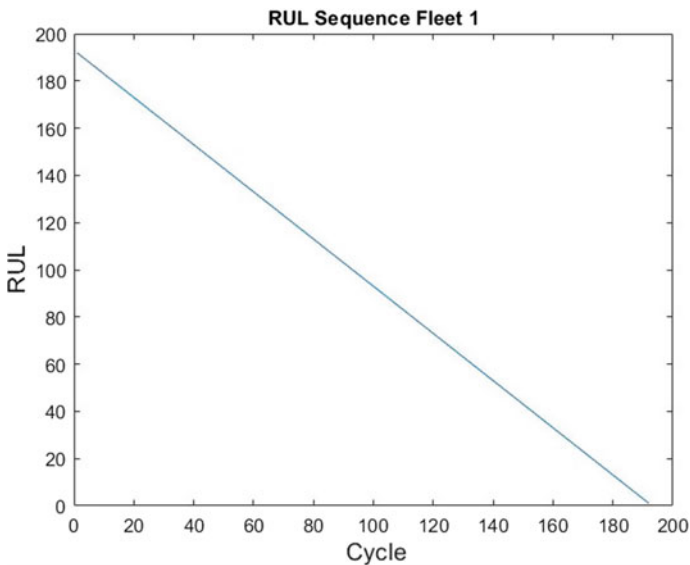
Fig. 2 Initial HI assessment for fleet 1
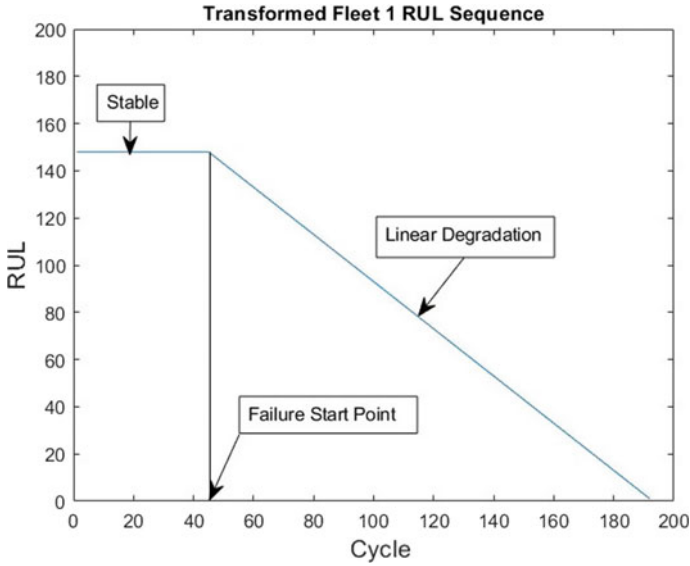
Fig. 3 Final HI assessment for fleet 1

obtained earlier and the failure start point, the transformed Fleet 1 RUL sequence is presented in Fig. 3.

4.4 Data Sequence Length

As mentioned before, we apply a piece wise linear degradation with maximum threshold base on the time series sequence length to estimate the RUL target for the training data. As such, the RUL value is proportionate to the sequence length like shown in Figs. 4 and 5. With the same reasoning, we can estimate the RUL value of the test data from its sequence length as shown in Figs 6 and 7.

However, looking at the ground truth RUL in Fig. 8, some of the RUL values are not proportionate to its data sequence length. We believe that the model will not perform well on these data sequences.

5   Results and Discussion

5.1   SNR Results and Feature Selection

See Tables 3, 4, 5, 6.
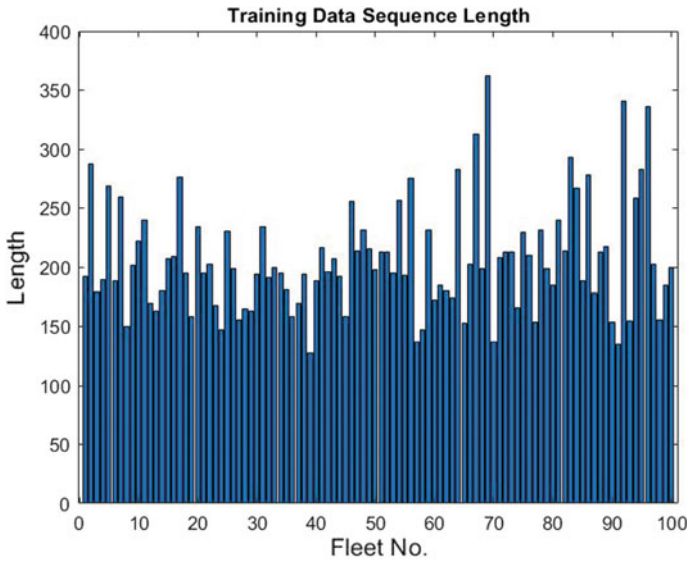
Fig. 4 Training data sequence length
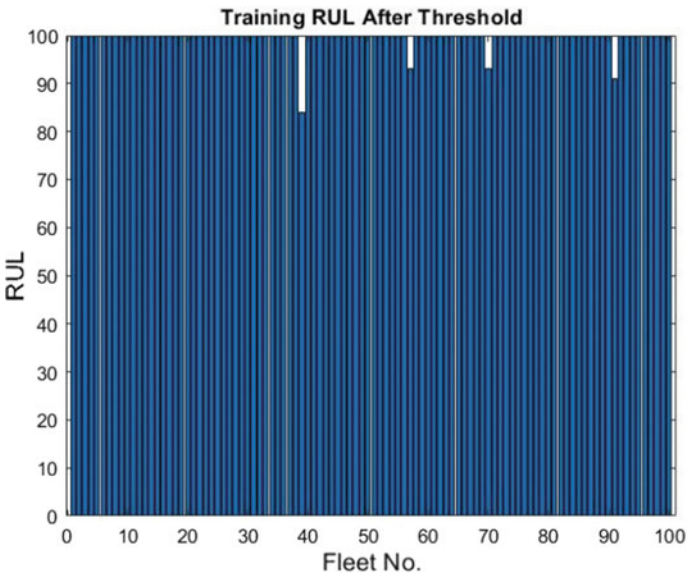
Fig. 5 Training data RUL
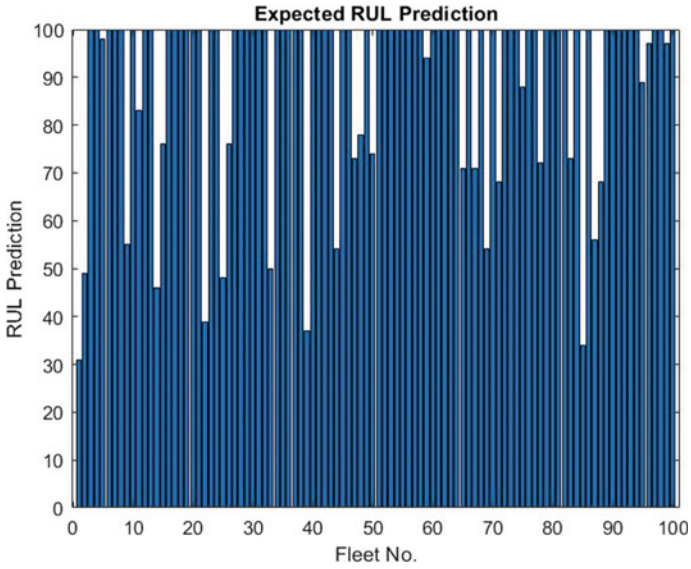
Fig. 6 Testing data sequence length
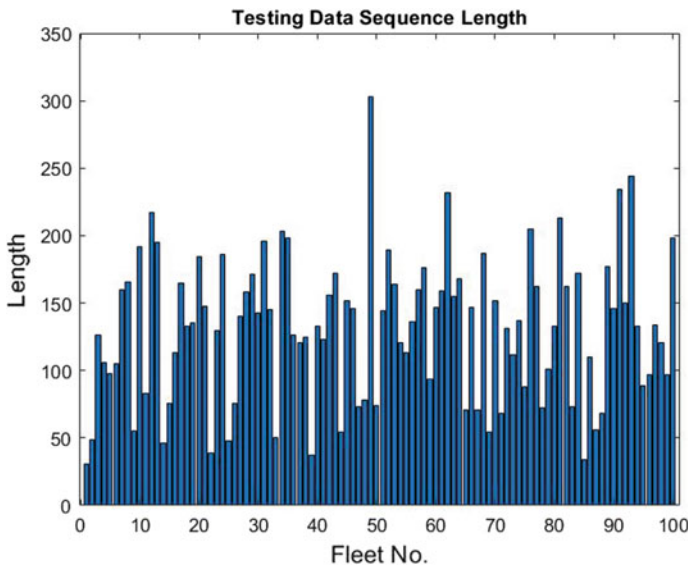
Fig. 7 Expected test data RUL prediction
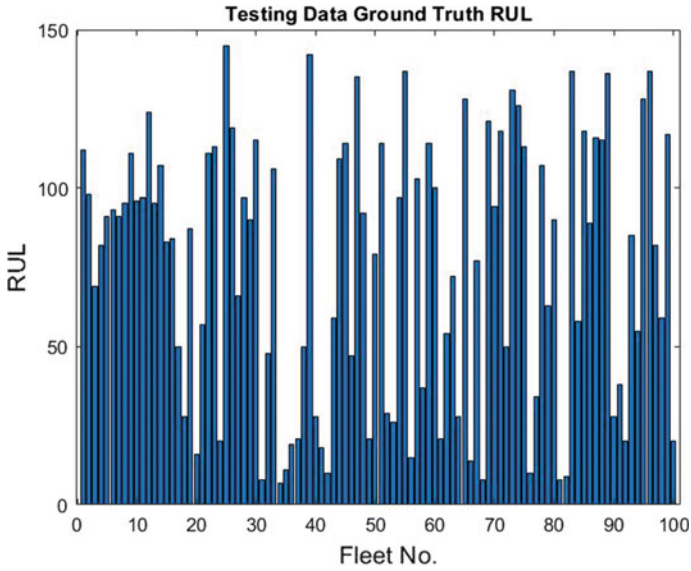
Fig. 8 Ground truth RUL
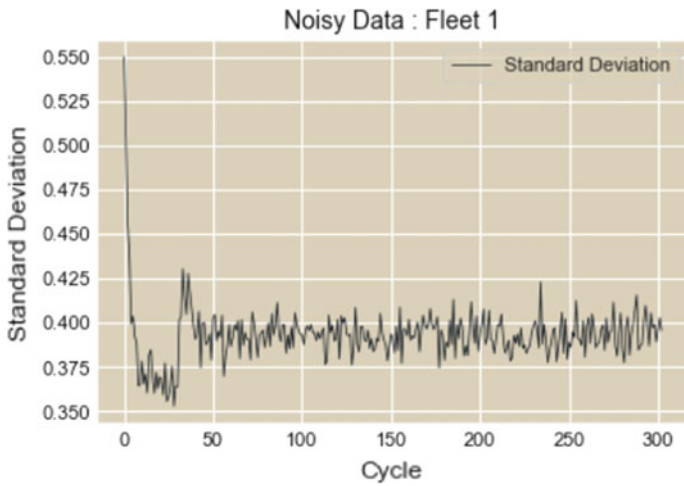
Fig. 9 Noisy data uncertainty

5.2 Aleatoric Uncertainty

The noisy data uncertainty is presented in Fig. 9. Here we present the result of our test. For illustration purposes, we only show the uncertainty result for Fleet 1.
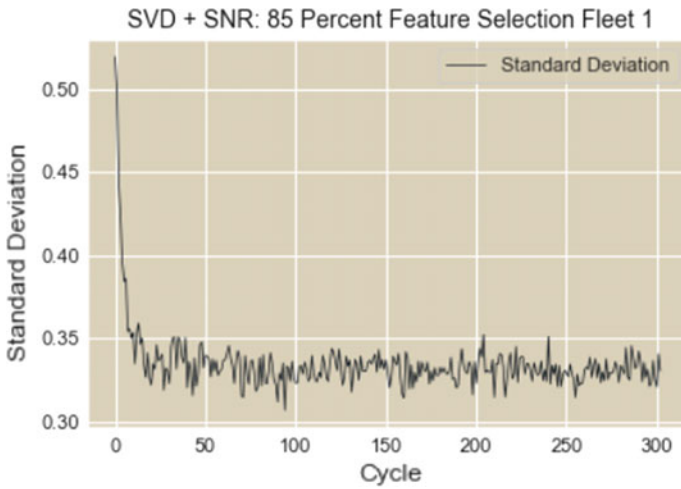
Fig. 10 SVD + SNR 85% features uncertainty

5.3 RMSE Results

The result of SVD + SNR with 85% features is shown in Fig. 10. As we can see in Fig. 11, SVD + SNR is the only technique that achieves the lowest uncertainty measurement in the order of 1e-3 with 75% features. Pearson and Spearman correlation feature selection with 85% features also produces lower uncertainty than the original data as shown in Fig. 12.
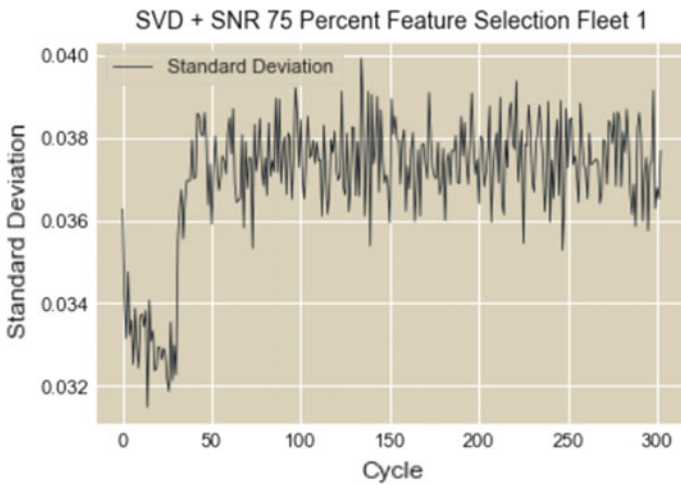
Fig. 11 SVD + SNR 75% features uncertainty
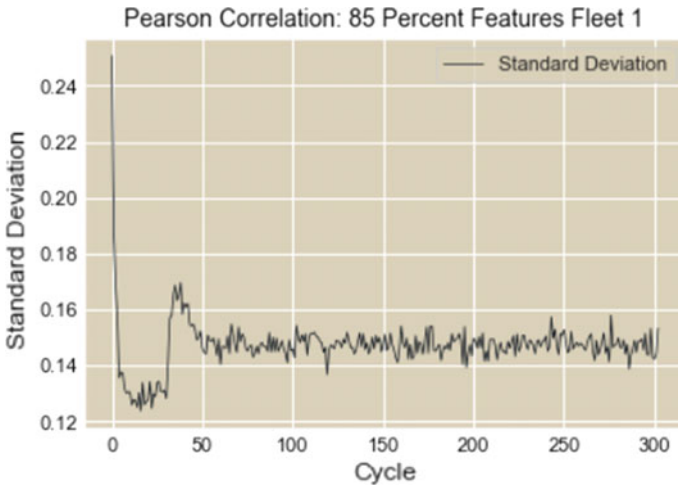
Fig. 12 Pearson and spearman correlation 85% features uncertainty

Table 7 RMSE results according to denoising treatment

Methods	Noisy Data	SVD 85%	SVD 75%	Pearson 85%	Pearson 75%
RMSE	34.86	34.26	34.48	34.37	34.89
Methods	**Spearman 85%**	**Spearman 75%**	**F reg 85%**	**F reg 75%**	
RMSE	34.37	34.89	35.01	35.19	

Looking at the RMSE result in Table 7, we can also see that the prediction performance is generally better with SNR + SVD than correlation-based feature selection methods. We can understand this phenomenon as only SVD + SNR that has the effect of denoising the data, thus improving the prediction uncertainty. In addition, SVD is known to be amongst the best denoising technique available. Thus, the effect of SVD denoising is really apparent with 25% less noisy features. Pearson and Spearman correlations only optimize the features in term of linear and nonlinear correlation, avoiding redundancy between correlated features. F Regression on the other hand, only evaluate the correlation between features and target. This also indicates that commonly used feature selection techniques such as Pearson and Spearman Correlation as well as F Regression could result to a higher prediction uncertainty like shown in Figs. 13 and 15. To have a clearer picture, we present in Figs. 16 and 17 the difference in uncertainty between prediction with noisy data and with SVD + SNR feature selection.
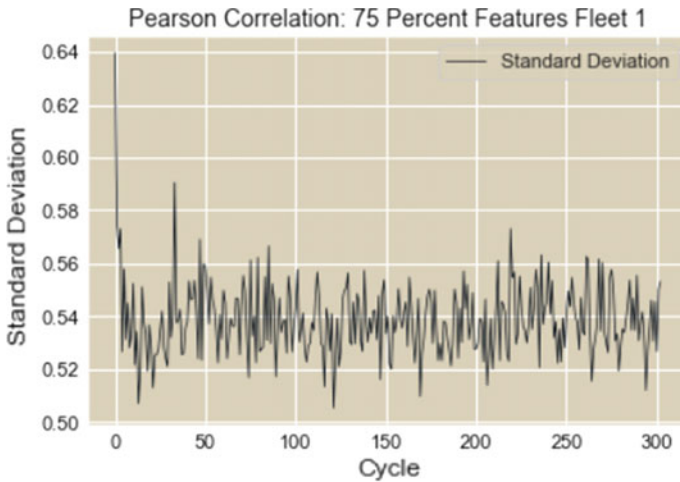
Fig. 13 Pearson and spearman correlation 75% features uncertainty
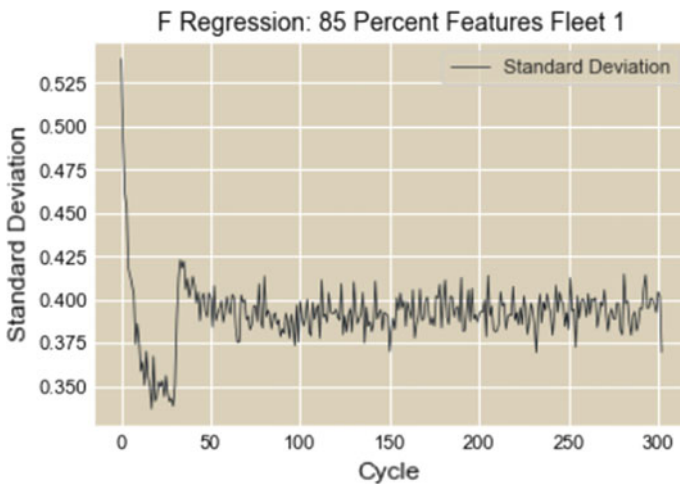
Fig. 14 F regression 85% features uncertainty

6 Conclusion

In this paper, we present a feature selection technique based on SVD + SNR where more noisy features are eliminated to reduce the Aleatoric uncertainty in an RUL prediction problem. A probabilistic LSTM that predicts RUL distributions of gas turbines is employed. From the evaluations that have been done, we demonstrate the superiority of this method compared to correlation-based feature selection techniques in reducing deep learning Aleatoric uncertainty. We also show that our feature
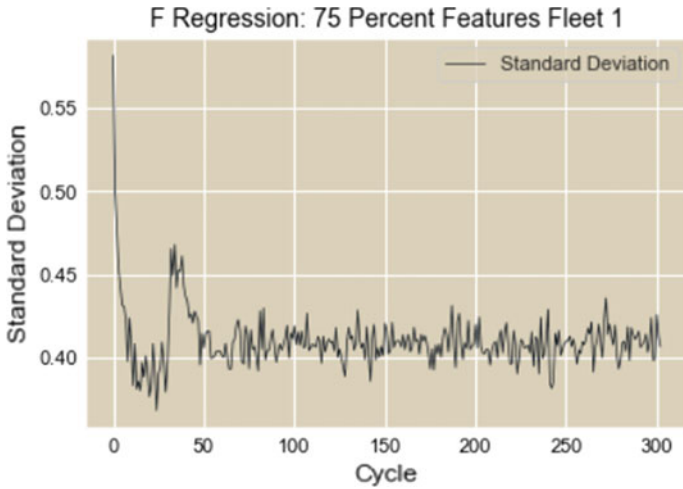
Fig. 15 F Regression 75% features uncertainty
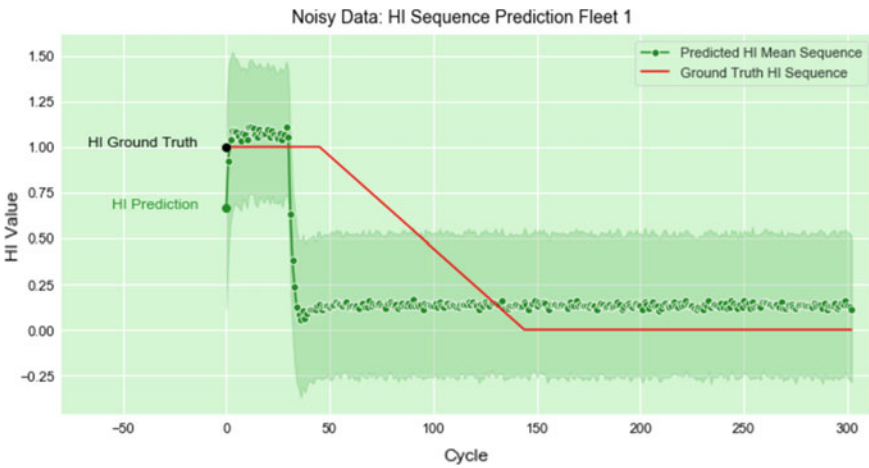
Fig. 16 Noisy data HI prediction with uncertainty

selection method yields a generally more accurate failure prognostic than the other tested techniques. Finally, this work illustrates that usual feature selection techniques popularly used such as Pearson and Spearman as well as F Regression could result in higher deep learning prediction uncertainty.
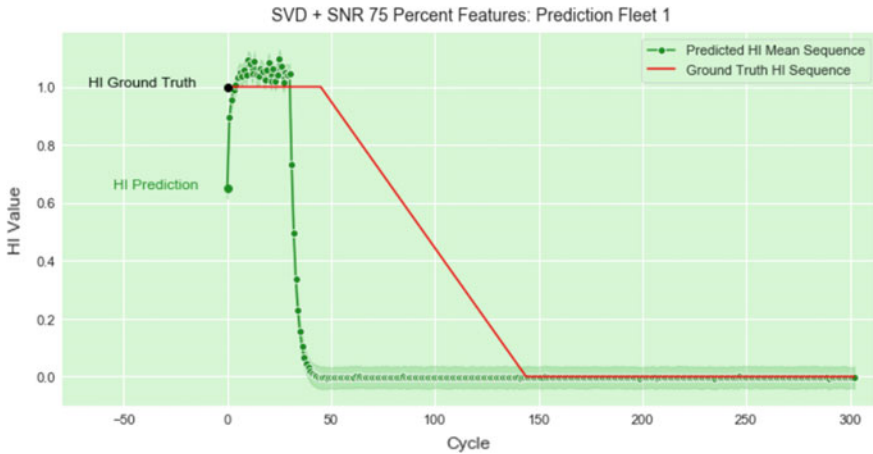
Fig. 17 HI prediction with uncertainty after SVD + SNR feature selection

References

1. Tsui KL, Chen N, Zhou Q, Hai Y, Wang W (2015) Prognostics and health management: a review on data driven approaches. Math Prob Eng 2015(793161):17. https://doi.org/10.1155/2015/793161

2. Xu M, Han M (2012) Research of prognostics and health management for Aircraft Propulsion System. In: Proceedings of the IEEE 2012 prognostics and system health management conference (PHM-2012 Beijing). Beijing, pp 1–4. https://doi.org/10.1109/PHM.2012.6228891

3. Zhang A, Cui L, Zhang P (2013) State prognostics based health management for complex systems. In: 2013 international conference on information technology and applications. Chengdu, pp 458–461. https://doi.org/10.1109/ITA.2013.111

4. Das S, Hall R, Herzog S, Harrison G, Bodkin M, Martin L (2011) Essential steps in prognostic health management. In: 2011 IEEE conference on prognostics and health management. Montreal, QC, pp 1–9. https://doi.org/10.1109/ICPHM.2011.6024332

5. Sutharssan T, Stoyanov S, Bailey C, Yin C (2015) Prognostic and health management for engineering systems: a review of the data-driven approach and algorithms. J Eng 7:215–222. https://doi.org/10.1049/joe.2014.0303

6. Wilkinson C, Humphrey D, Vermeire B, Houston J (2004) Prognostic and health management for avionics. In: 2004 IEEE aerospace conference proceedings (IEEE Cat. No.04TH8720), vol 5. Big Sky, MT, pp 3435–3447. https://doi.org/10.1109/AERO.2004.1368149

7. Shin I, Lee J, Lee JY et al (2018) A framework for prognostics and health management applications toward smart manufacturing systems. Int J Precis Eng Manuf-Green Tech 5:535–554. https://doi.org/10.1007/s40684-018-0055-0

8. Wang J, Wen G, Yang S, Liu Y (2018) Remaining useful life estimation in prognostics using deep bidirectional LSTM neural network. In: 2018 prognostics and system health management conference (PHM-Chongqing). Chongqing, pp 1037–1042. https://doi.org/10.1109/PHM-Chongqing.2018.00184

9. Wang H, Liu Y, Liu Z, Wang Z, Huang H (2013) Remaining useful life estimation for degradation and shock processes. In: 2013 international conference on quality, reliability, risk, maintenance, and safety engineering (QR2MSE). Chengdu, pp 1762–1764. https://doi.org/10.1109/QR2MSE.2013.6625917

10. Ghoshal B, Tucker A, Sanghera B, Wong WL (2019)Estimating uncertainty in deep learning for reporting confidence to clinicians when segmenting nuclei image data. In: 2019 IEEE 32nd international symposium on computer-based medical systems (CBMS). Cordoba, Spain, pp 318–324. https://doi.org/10.1109/CBMS.2019.00072

11. Shaker MH, Hüllermeier E (2020) Aleatoric and epistemic uncertainty with random forests. In: Berthold M, Feelders A, Krempl G (eds) Advances in intelligent data analysis XVIII. IDA 2020. Lecture notes in computer science, vol 12080. Springer, Cham. https://doi.org/10.1007/978-3-030-44584-3_35

12. Kendall A, Gal Y (2017) What uncertainties do we need in Bayesian deep learning for computer vision? In: Proceedings of the 31st international conference on neural information processing systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, pp 5580–5590

13. Depeweg S, Hernández-Lobato JM, Doshi-Velez F, Udluft S (2018) Decomposition of uncertainty in bayesian deep learning for efficient and risk-sensitive learning. ICML

14. Epps B, Krivitzky E (2019) Singular value decomposition of noisy data: noise filtering. Exp Fluids 60:1–23. https://doi.org/10.1007/s00348-019-2768-4

15. Epps B, Krivitzky E (2019) Singular value decomposition of noisy data: mode corruption. Exp Fluids 60:1–23. https://doi.org/10.1007/s00348-019-2761-y

16. Deng Y, Ren Z, Kong Y, Bao F, Dai Q (2017) A hierarchical fused fuzzy deep neural network for data classification. IEEE Trans Fuzzy Syst 25(4):1006–1012. https://doi.org/10.1109/TFUZZ.2016.2574915

17. Boskos D, Cortes J, Martinez S (2020) Data-driven ambiguity sets with probabilistic guarantees for dynamic processes. IEEE Trans Autom Control 1–1. https://doi.org/10.1109/TAC.2020.3014098

18. Du Z, Wang W, Zhang J, Zhang Y, Xu X, Liu J (2020) A data-driven home energy scheduling strategy under the uncertainty in photovoltaic generations. IEEE Access 8:54125–54134. https://doi.org/10.1109/ACCESS.2020.2980850

19. Rodemerk C, Winner H, Kastner R (2015) Predicting the driver's turn intentions at urban intersections using context-based indicators. In: 2015 IEEE intelligent vehicles symposium (IV). Seoul, pp 964–969. https://doi.org/10.1109/IVS.2015.7225809

20. Yang G, Wang Y, Xu B, Zhang X (2019) An image denoising algorithm based on singular value decomposition and non-local self-similarity. In: Vaidya J, Zhang X, Li J (eds) Cyberspace safety and security. CSS 2019. Lecture notes in computer science, vol 11983. Springer, Cham. https://doi.org/10.1007/978-3-030-37352-8_44

21. Leal N, Zurek E, Leal E (2020) Non-local SVD denoising of MRI based on sparse representations. Sensors 20:1536

22. Yang W, Hong JY, Kim JY, Paik SH, Lee SH, Park JS, Lee G, Kim BM, Jung YJ (2020) A novel singular value decomposition-based denoising method in 4-dimensional computed tomography of the brain in stroke patients with statistical evaluation. Sensors (Basel) 20(11):3063. https://doi.org/10.3390/s20113063 PMID:32481740;PMCID:PMC7309118

23. Zhu Y et al (2019) Denoising ultrasonic echo signals with generalized s transform and singular value decomposition. Traitement du Signal 36:139–145

24. Zhang X, Tang J, Zhang M, Ji Q (2016) Noise subspaces subtraction in SVD based on the difference of variance values. J Vibroeng 18:4852–4861. https://doi.org/10.21595/jve.2016.16745

25. Schanze T (2018) Compression and noise reduction of biomedical signals by singular value decomposition. IFAC-PapersOnLine 51:361–366. https://doi.org/10.1016/j.ifacol.2018.03.062

26. Ruben M (2014) BayesOpt: a Bayesian optimization library for nonlinear optimization, experimental design and bandits. J Mach Learn Res 15(1):3735–3739

27. Sherstinsky A (2018) Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network

28. Zheng S, Ristovski K, Farahat A, Gupta C (2017) long short-term memory network for remaining useful life estimation 88–95. https://doi.org/10.1109/ICPHM.2017.7998311
29. Zhang X et al (2019) Remaining useful life estimation using CNN-XGB with extended time window. IEEE Access 7:154386–154397. https://doi.org/10.1109/ACCESS.2019.2942991
30. Bruneo D, De Vita F (2019) On the use of LSTM networks for predictive maintenance in smart industries. In: 2019 IEEE international conference on smart computing (SMARTCOMP). Washington, DC, USA, pp 241–248. https://doi.org/10.1109/SMARTCOMP.2019.00059
31. Ramasso E, Saxena A (2014) Performance benchmarking and analysis of prognostic methods for CMAPSS datasets. Int J Prognostics Health Manag 5:1–15
32. Saxena A, Goebel K, Simon D, Eklund N (2008) Damage propagation modelling for aircraft engine run-to-failure simulation. Int Conf Prognostics Health Manag. https://doi.org/10.1109/PHM.2008.4711414.M
33. Gao Y, Zhou J, Wu K, Zhao G, Hu C (2019) Construction method of turbine engine health indicator based on deep learning. In: 2019 prognostics and system health management conference (PHM-Qingdao). Qingdao, China, pp 1–6. https://doi.org/10.1109/PHM-Qingdao46334.2019.8943055; Wang Q, Zheng S, Farahat A, Serita S, Gupta C (2019) Remaining useful life estimation using functional data analysis. In: 2019 IEEE international conference on prognostics and health management (ICPHM). San Francisco, CA, USA, pp 1–8. https://doi.org/10.1109/ICPHM.2019.8819420
34. Li J, Li X, He D (2019) Domain adaptation remaining useful life prediction method based on AdaBN-DCNN. In: 2019 prognostics and system health management conference (PHM-Qingdao). Qingdao, China, pp 1–6. https://doi.org/10.1109/PHM-Qingdao46334.2019.8942857
35. "Detect small changes in mean using cumulative sum" MATLAB MathWorks® https://www.mathworks.com/help/signal/ref/cusum.html
36. Heimes F (2008) Recurrent neural networks for remaining useful life estimation. In: IEEE international conference on prognostics and health management

Unsupervised Classification of Acoustic Emission Signal to Discriminate Composite Failure at Low Frequency

Noor A'in A. Rahman, Zazilah May, and Muhamad Shazwan Mahmud

Abstract The use of acoustic emission (AE) for damage assessment and detection technique in structural engineering is widely used and has earned a reputation as one of the reliable non-destructive techniques. AE source is produced based on the elastic wave propagation through the specimen which converted into the electrical AE signal by the AE sensors. Certain AE features belong to the signal allow their use to discriminate mode of damage in a composite material. However, the challenge encountered during analysis of AE signals attributed to the techniques like pattern recognition and classification method. In this paper, several orientation of laminated fiber specimens are undergoing tensile test. According to the information from tensile load test, significant features were monitored namely amplitude and energy in order to execute the classification method. The AE data are successfully cluster by k—means algorithm.

Keywords Acoustic emission · Composite fiber · Clustering · Low frequency monitoring

1 Introduction

Carbon fiber laminates composite are widely used in many structure and engineering application due to their excellent properties such as good corrosion resistance, high modulus elasticity, light weight accompanied by the high strength. Despite all these advances, the monitoring and prediction of the failure experience by a fiber composite remain major problems. Understanding damage on fiber composite is very challenging due to the complex process had occurred during the process.

N. A. A. Rahman (✉) · Z. May
Electrical and Electronics Department, Universiti Teknologi PETRONAS, 31260 Bandar Seri Iskandar, Perak, Malaysia

M. S. Mahmud
Mechanical Engineering Department, Universiti Teknologi PETRONAS, 31260 Bandar Seri Iskandar, Perak, Malaysia

In-service monitoring approach by Acoustic Emission (AE) is a reliable technique to monitor and distinguish these failures. AE events can be monitored with low frequency (audible sound range) up to high range frequency (ultrasonic range). When critical stress occurs on a structure, the strain energy stored is suddenly released thus create mechanical stress waves which spread concentrically around the source of occasion. The release energy can be detected with suitable sensors (mechanical information converted into the electrical signal).

AE from material damage in composite is relatively easy to monitor and record, but the efficient and effective analysis is still challenging [1]. One of the common analysis is clustering. Clustering can be divided into three major parts consists of choosing the significant features, identifying the right number of clusters and lastly the modifying technique of the clustering to achieve. Different AE features can be used for the first part namely amplitude, duration, energy, peak frequency and many more. Extra features can be calculated from combination of those features. For instance Bakhtiary and his co-workers [2] define their algorithm by ratio of strain energy and acoustic energy to identify the onset of delamination and its progression in composite fiber.

Whenever multiple phenomena are involved and thus different signal clusters, unsupervised pattern recognition is also sometimes used to address the problem of labeling the class/cluster. The term unsupervised pattern recognition is on the other hand used to describe the complete methodology consisting of procedures for cluster analysis, descriptor selection and cluster validity, whenever none of the information is available [3].

The k-means algorithm is one of the most widely used methods to cluster any signal including AE signal. Ech-Choudany et. al [4] applied the k-means combined with Incremental Clustering (IC) in order to construct a proper labelled learning database of AE signals. However, they conclude AE signals are more relevant to study in the frequency domain in order to extract more relevant descriptor.

A lot of works are available at finding the mode of failure on composite fiber with a wide range of frequency up to 600 kHz [1, 5]. Based on the dominant study on failure mechanism in the composite, matric cracking has occurred in almost low-frequency [6–8]. The resulting output from their research showed the AE event able to be clustered and based on well-defined frequencies. Rather than frequency, other AE waveform parameters should be therefore considered in clustering analysis. One of the main concerns during acquisition AE signal is capturing the signal by a proper frequency range of sensor. This work will be focused on a low frequency range of detection on composite failures. This can be done by using a proper AE sensor which having the working frequency range below than 100 kHz.

This paper presents the laboratory test results of AE measurement during tensile test on composite laminate fiber specimen. The AE features from laboratory test will be clustered by k-means algorithm. The unsupervised techniques are used for improving the clustering/classification of failures on composite fiber for automated clustering. Clustering generally requires user to identify the number of clusters in

Table 1 Specimen specification

Specimen type	Lay-ups	Length (mm)	Width (mm)	Thickness (mm)
T30	30°	320	250	2.5
T45	45°	320	250	2.5
T90	90°	320	250	2.5

second part of the process. There are many cluster validity techniques for determination the cluster. However, in this paper the mechanical failure of composite structure will be verified by the load diagram which obtain from the tensile testing.

2 Experimental Procedure

2.1 Materials and Specimen Preparation

Composite laminates were composed using 2 mm High Strength Carbon Fibre Sheet which a cured carbon fiber plate supplied by Easy Composites Ltd. Staffordshire, United Kingdom. High Strength Carbon Fibre Sheet is produced from in-house manufacturing proses with specially design platen press of 3 layers of XPEG XC130 from the same company. XPEG XC130 is epoxy based 2 × 2 twill weave prepreg carbon of Pyrofil TR30S high strength 3 k carbon. The specimens were prepared into three types of lay-up which were 30°, 45° and 90° lay-ups with final dimensions of (270 × 250 × 2.5) mm using a diamond coated abrasive cutter. Each lay-out was cut by referring to the initial 0° of the vertical direction of the first corner of 2 × 2 twill in weave. Prior to the tensile test, all specimens underwent pre-conditioning at 23 °C for 16 h in a desiccator to stabilize the humidity to satisfy the ASTM standard. Table 1 summarizes the lay-ups, dimension and abbreviated name used in the rest of paper.

2.2 Tensile Test

The mechanical testing was performed using a tensile test following the ASTM D3039/D, standard testing method to generate damage modes of composite material. The specimens were fixed between the jaws of the machine to undergo testing. These tests were conducted using a universal tensile machine (UTM) manufactured by Gotech Testing Machine Inc. (AI-7000L). The testing was used for inducing external force to the specimen at 15 kN capacity with a constant speed 2 mm/s, at room temperature. For each type of lay-up, at least five specimens were tested. The average value and the standard deviation were then recorded.

2.3 AE Equipment

Acoustic emission software AEWin and an acquisition system supplied by Physical Acoustic Corporation (PAC) with a sampling rate of 10 MHz were used to record AE event in this study. A piezoelectric sensor type R6I from PAC was used to sense event during tensile testing. The sensor was place in the middle of the specimen, coupled with grease to provide good acoustic coupling between them. The sensors had a resonance frequency of maximum 100 kHz. Pencil lead break (PLB) test was performed on the specimen in an arrangement that the source and the sensor is passes through in the middle of the specimen. PLB is a well-established procedure for generating simulated AE sources as recommended by ASTM standard (E976-99). Procedure were repeated for the rest of 15 specimens during tensile tests.

The typical AE waveform is shown in the following figure where the interested features are labelled (Fig. 1).

3 Methodology

A big data was collected during acquisition from the hardware. The hardware was set at 40 dB threshold value in order to cut off low amplitude signal related to environmental noise. The methodology used in the discrimination of failure modes in post composite laminates is by using unsupervised clustering techniques. The software used in implementation this methodology is KNIME Analytics Platform version 4.0.2. KNIME is increasingly used by researchers in various areas of data mining and machine learning research work. It is use graphical sequence based to add the node for each algorithm. The pattern recognition techniques selected for the present study is k-means. The data were normalized prior to the clustering procedure.

Fig. 1 Typical AE waveform generated on carbon fiber composite during acquisition
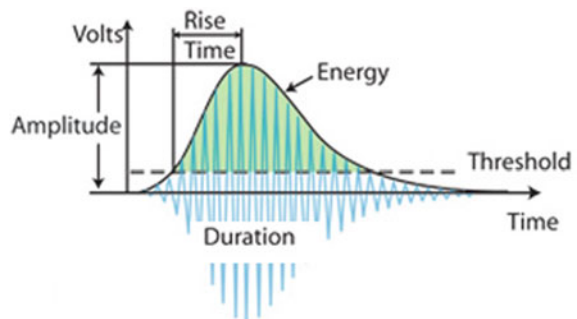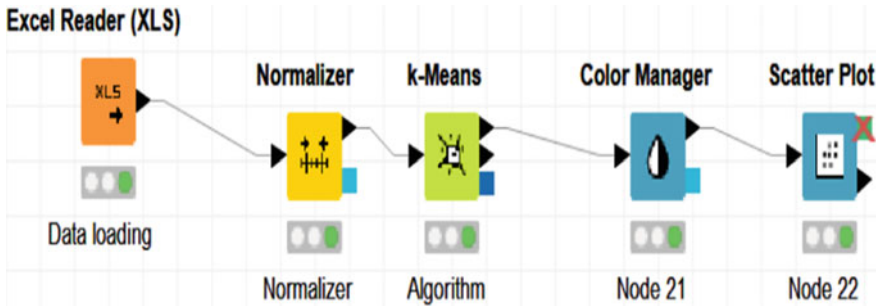
Fig. 2 *k*-means algorithm workflow in KNIME

3.1 *k-means*

The *k*-means algorithm aims to minimize the sum of the squared distances between all the vectors of a cluster and its center. The cluster number *k* is specified based on the mechanical test information. Process for executing the *k*-means algorithm in KNIME is shown in the following diagram (Fig. 2).

The *k*-means performs a crisp clustering that assigns a data vector to exactly one cluster. Thus, the algorithm terminates when the cluster assignments do not change anymore. Basically, the clustering algorithm uses the Euclidean distance on the selected attributes. The AE features information were normalized prior to the clustering process.

4 Result and Discussion

The investigation of the current effort has been carried out entirely experimentally by laboratory tests. The results are segregated to subsections to make it easier for understanding the insight of each effective variables.

4.1 *Correlations of AE Results with Tensile Test*

Figure 3 represents an evolution of stress–strain and AE features for each types of specimen. Practically no difference is observed between the stress–strain curve obtained in the different types of specimen. Each curve shows a linear behavior from initial up to the higher load applied, at the end of test. Then the stress drastically drops when the UTS is achieved (not shown in diagram). According to UTS values, state in Table 2, the average tensile strength (for all 5 specimen) increase by increasing the orientation degree of laminated fiber. T30 specimen reaches the break at the same of

Fig. 3 Load curve and
significant AE features
(energy and amplitude) for
each specimen
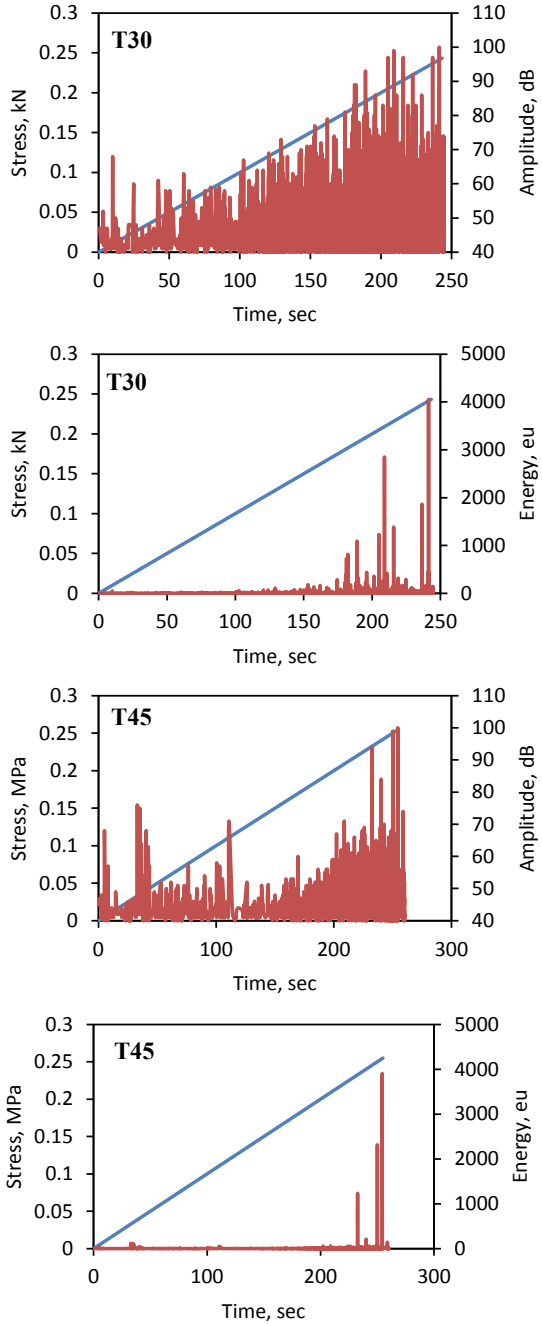
Fig. 3 (continued)
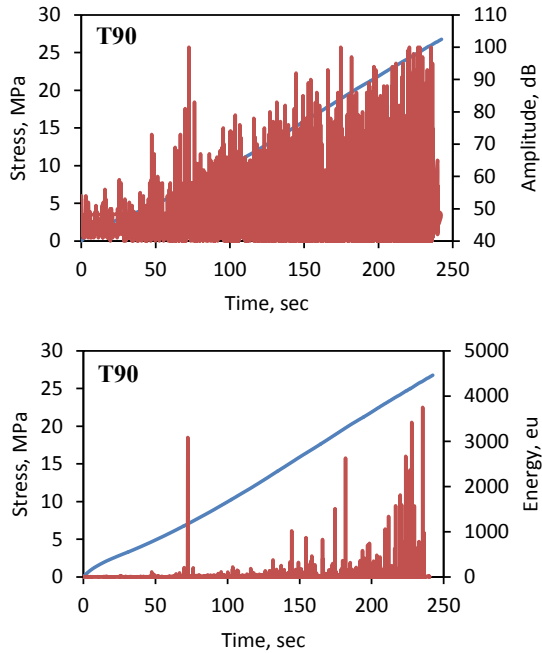
Table 2 Average ultimate tensile strength (UTS) and tensile strain for each test specimen

Specimen type	UTS (MPa)	Tensile strain (%)
T30	454.6	1.48
T45	484.4	1.62
T90	507.9	1.23

the T90 specimen followed by T45 later. Figure 3 also show the evolution of AE signal energy and amplitude during tensile load carried out on each specimen. Generally, the matrix cracking and some interfacial failures are recorded. Since the interested frequency (< 100 dB) are recorded. This is logical because, as the load increase, not only is there much more extensive matrix cracking but also many more delaminations occur at the previous loading steps. Moreover, as the loading continued, the crack propagation spread to the top, rupturing of the rest of specimen, while fiber pull-out gradually occurred.

According to Fig. 3, as expected the highest energy value recorded for all specimen are right before the specimen were break. Generally, the release elastic energy produced i.e. by matrix cracking and delamination is detected from the beginning of the tensile load test. Moreover, these failures producing a very 'loud' sound to be listened by the sensor. The highest recorded energy before fiber break for specimen T30, T45 and T90 are 4051 eu, T45 3094 eu and T90 4508 eu respectively. When the load increased, trends of the AE energy of the specimen were similar regardless the fiber orientation of the composite. Similar pattern is recorded by the amplitude

of AE signal. However, an almost linear behavior of AE peak amplitude is recorded in each of the specimen towards end of experiment. Unfortunately, the amplitude of AE signal is not much different for each laminated fiber. According to literature, the predominance of AE events occurring at an amplitude less than 60 dB possibly correlates with fiber breakage [9]. Thus, according to the test on this specimen, the fiber breakage occurs on the specimen mostly in the middle of the tensile load testing. In conclusion, it is evident that with continuous loading amplitude continued increasingly, up to the maximum value (approximately 100 dB) while approaching the UTS.

4.2 Clustering

AE cluster behavior were assessed and compared to each of the test specimen exposed to tensile loading. AE clusters were based on the normalized amplitude and energy feature of each AE event. Practically, the clustering is based on the damage evolution during tensile load testing. If one damage type is the exclusive source of an AE cluster, both damage and the accumulation of event cluster would start together [1]. In this study, the k value is based on the type of mechanical during tensile load. According to stress diagram, the mode of failures can be divided into two major cases namely delamination and matrix cracking (occurred at the beginning of test) and fiber breakage fiber pull-out towards end of experiment. As mention previously, only a low frequency of damage is counted, based on the type of AE sensor. Variations in the damage sources associated with the same mechanism most probably exist. The clustering results assigning the AE features (amplitude and energy) data as shown in Fig. 3 (Fig. 4).

5   Conclusion

Different laminated fiber orientation composite material tensile strength experiments were analyzed, and research on failures mode based on tensile load was conducted. Moreover, the clustering analyzes the potential AE features which can be indicates the mode of failure during monitoring. The following can be concluded:

(1)   The tensile strength of laminated fiber composite material is significantly increase in order T30 > T45 > T90.
(2)   The selected fiber orientation had almost similar behavior of amplitude and energy pattern in the laminated composite specimen.
(3) k-means algorithm able to cluster the amplitude and energy of AE signal by using KNIME data mining software.
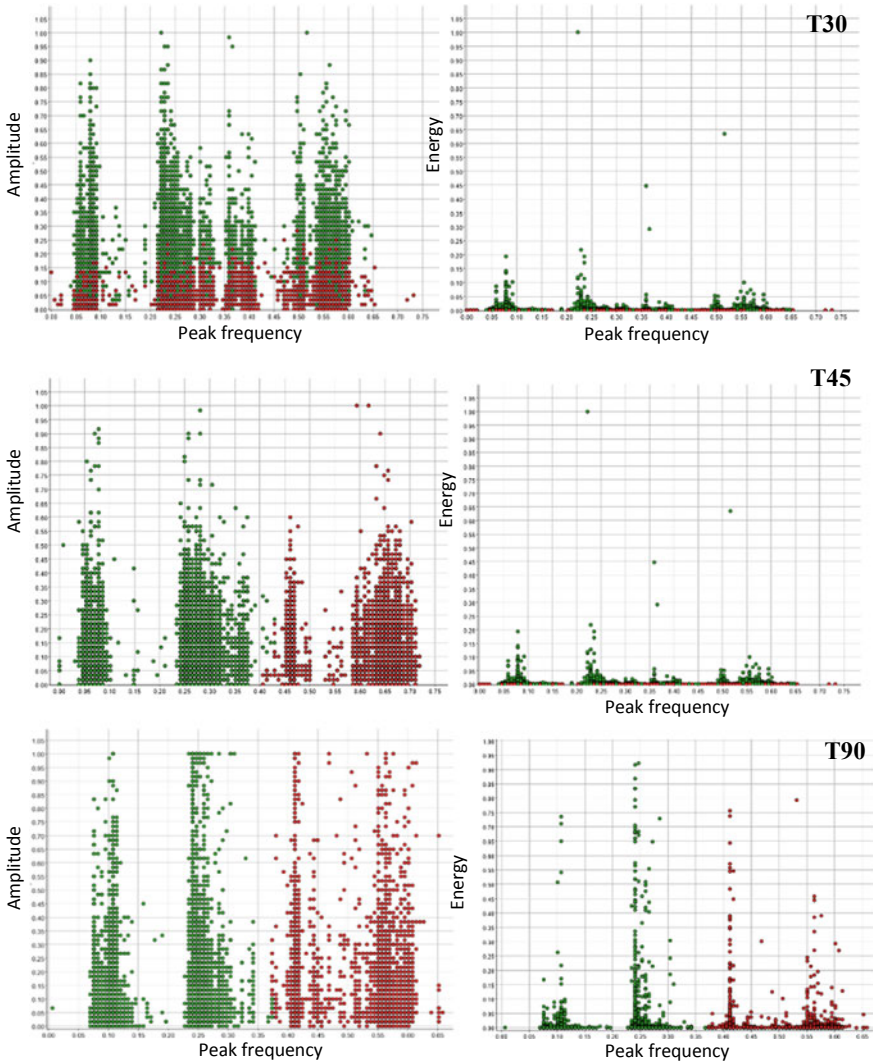
Fig. 4 k-means clustering output

References

1. Sawan HA, Walter ME, Marquette B (2015) Unsupervised learning for classification of acoustic emission events from tensile and bending experiments with open-hole carbon fiber composite samples. Compos Sci Technol 107:89–97
2. Bakhtiary Davijani AA, Hajikhani M, Ahmadi M (2011) Acoustic emission based on sentry function to monitor the initiation of delamination in composite materials. Mater Des 32(5):3059–3065
3. Calabrese L, Galeano M, Proverbio E, Di D, Cappuccini F, Donato A (2016) Monitoring of 13% Cr martensitic stainless steel corrosion in chloride solution in presence of thiosulphate by acoustic emission technique. Corrosion Science
4. Ech-Choudany Y, Assarar M, Scida D, Morain-Nicolier F, Bellach B (2017) Unsupervised clustering for building a learning database of acoustic emission signals to identify damage mechanisms in unidirectional laminates. Appl Acoust 123:123–132
5. Koruk H, Genc G (2015) Investigation of the acoustic properties of bio luffa fiber and composite materials. Materials Letter 157:166–168
6. Gutkin R, Green CJ, Vangrattanachai S, Pinho ST, Robinson P, Curtis PT (2011) On acoustic emission for failure investigation in CFRP: Pattern recognition and peak frequency analyses. Mech Syst Signal Process 25(4):1393–1407
7. Ramirez-Jimenez CR, Papadakis N, Reynolds N, Gan TH, Purnell P, Pharaoh M (2004) Identification of failure modes in glass/polypropylene composites by means of the primary frequency content of the acoustic emission event. Compos Sci Technol 64(12):1819–1827
8. De Groot PJ, Wijnen PA, Janssen RB (1995) Real time frequency determination of acoustic emission for different fracture mechanisms in carbon/epoxy composites. Compos Sci Technol 55:405–412
9. Gholizadeh S, Leman Z, Baharudin BTHT, Inayatullah O (2018) Acoustic emission analysis for characterisation of damage mechanisms in glass fiber reinforced polyester composite. Aust J Mech Eng 16(1):11–20

Finding the EEG Footprint of Stress Resilience

Rumaisa Abu Hasan, Syed Saad Azhar Ali, Tong Boon Tang, and Muhamad Saiful Bahri Yusoff

Abstract Work stress faced by adults can lead to decreased job performance, reduced mental and physical wellbeing, and other detrimental health problems. Researchers are reporting resilience as a key factor in determining a person's vulnerability towards mental stress disorders. Psychosocial measures of resilience conventionally use the self-assessment approach which is susceptible to potential biases caused by self-reporting and concerns of social stigma. With increasing emphasis of its role in mental health, researchers are using fMRI modality to identify the brain activity of stress resilience. But this approach is costly and lack practicality when evaluating stress resilience in daily tasks. The EEG modality provides a cost-efficient alternative with better practicality and high temporal resolution in studying the brain activity of stress resilience. However, EEG-based literatures on stress resilience are limited to brain activity during resting state. With reference to the cognitive affective conceptual stress model, we define stress resilience as an adaptation process, involving cognitive appraisal, physiological arousal and coping behaviour, that utilizes individual resources to cope with stress. This paper proposes an approach to identify the features of EEG-neural correlates of stress resilience through brain rhythms, hemispheric asymmetry and brain network.

Keywords Resilience · Stress · EEG feature

R. A. Hasan (✉) · S. S. A. Ali · T. B. Tang
Centre for Intelligent Signal and Imaging Research (CISIR), Electrical and Electronic Engineering Department, Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Perak, Malaysia
e-mail: rumaisa_19000937@utp.edu.my

M. S. B. Yusoff
Department of Medical Education, School of Medical Sciences, Universiti Sains Malaysia, Kelantan, Malaysia

1   Introduction

In year 2002, the World Health Organization published a monograph concerning the alarming mental health problem as one of the leading causes of disability and disease world-wide [1]. The increase in work-related stress among adults is a raising concern throughout the world [2–4]. This problem has been recognized as detrimental in both poor and rich countries, and persist in population regardless of age, gender and social strata. In Malaysia, the percentage of clinically-diagnosed adults with stress-disorders have increased [5], with a 29.9% overall prevalence of work-related stress [2]. This indicates that one of three adults are suffering from work-related stress. Despite being an "unseen" condition, work-related stress has its pathogenicity [2] and leads to mental health deterioration [6].

1.1   *Role of Resilience in Mental Health*

Stress resilience is a measure of an individual's vulnerability towards stress in perceiving adverse events as minimally threatening and developing adaptive physiological and psychological responses [7, 8]. Recent findings from researches on brain responses towards stress have reported resilience as a key factor for perceiving and reacting positively, both emotionally and behaviorally, towards stressful conditions. Davydov et al. identified resilience as a defense mechanism of the body that protects it from the negative effects of stress response [9]. It is a measure of how the body adapts or maladapts to internal and external stressors [10]. Adapting towards stress build resilience and has been associated with positive emotions and good emotion-regulating abilities [11]. Connor and Davidson emphasized the need to quantify and validate the measure of resilience as it can be used in gauging efficiency of psychotherapy and to investigate which stress coping strategies are adaptive and maladaptive to the individual [10].

1.2   *Limitation of Psychological Assessment for Stress Resilience*

Researchers in the psychosocial fields have proposed the multidimensional characteristic of resilience [7, 9–11]. The resilience is viewed as an integration of multiple constructs as it combines the biology, psychology and social factors of an individual [9]. Although these assessments have been reported to have valid psychometric properties in measuring resilience and has been used in literatures on different cohorts, there is a limitation of non-physiological approach which only considers characteristics or level of resilience and not the resiliency adaptation process [10]. In addition, self-assessment approach in psychotherapy has the disadvantage of being vulnerable

to potential biases caused by self-reporting and concerns of social stigma. In studies which use associated factors to measure resilience, researchers are required to use a number of measuring instruments to ensure that the multidimensional characteristic of resilience is addressed [12, 13]. Hence, there is a need for a physiological-based assessment of resilience that allow objective monitoring of the body responses towards stressful life events.

1.3 Brain Activity of Stress Resilience

As resilience is the adaptation of mental and physical responses to stress, the brain also exhibits neural activities that reflect this mechanism. When continuously exposed to stress, a region within the brain limbic system known as amygdala undergoes structural changes through dendritic growth and spine synapses formation, whereas medial prefrontal cortex has been observed to undergo dendritic shortening [14]. Neural activity within the prefrontal cortex has also been reported to influence the reactivity of amygdala in emotional arousal [15]. The hypothalamic–pituitary–adrenal (HPA) axis is another neural circuitry which plays a key role in determining the degree of individual stress resilience from the gestational stage [8]. Stress stimuli alter the neuroendocrine activity of this axis, triggering the cascading effects of neuronal modulation and hormonal secretion that affects the brain. Moreover, the reciprocal communication of both the HPA axis and medial prefrontal cortex with the dorsal raphe nucleus forebrain region highlights the functional implication of prefrontal cortex on the HPA axis [8]. Franklin et al. (2012) highlighted an interesting hypothesis on the role of ventral medial prefrontal cortex in acquiring resilience through the cognition of progressive learning. These findings on the functional connectivity of frontal activity and both amygdala and the HPA axis emphasize the relevance for a brain-based measure for stress resilience as this characteristic has been strongly associated to vulnerability to mental illness [8, 9].

1.4 Limitation of FMRI-Based Application on Stress Resilience

Findings from functional magnetic resonance imaging (fMRI) –based literatures have identified the association of stress resilience with the anterior cingulate cortex [16] and prefrontal cortex [13, 17, 18]. However, the use of fMRI modality limits its applicability in clinical and research purposes due to its immobility and high cost required for the machine, procedures and setup [19]. The procedure of collection fMRI data itself involves noisy scanner and are highly susceptible to physiological noise due to motion [20]. The fMRI scanner limits the movement of a person as they are required to be in supine position throughout the procedure and therefore cannot

be used in collecting data involving any task performance. In a review by Thibault et al., the minimum initial setup cost for fMRI modality is USD 500,000 whereas the maximum initial setup cost for electroencephalography (EEG) modality is only USD 50,000 [21]. In addition, the EEG modality has the advantage of being user-friendly especially for claustrophobic patients that cannot use the fMRI, portable and has high temporal resolution [21]. Findings from fMRI studies highlight the significance of the anterior cingulate and prefrontal cortices for resilience in both resting [16] and building [13, 17, 18] states. As brain activity in these locations can be monitored using EEG-modality, this suggests the relevance of EEG-modality in identifying the neural correlates for stress resilience. By identifying the EEG-neural correlates for stress resilience, it can be used as a risk marker for mental illnesses at a much higher applicability rate in both clinical and non-clinical populations.

1.5   EEG-Based Approach in Identifying Stress Resilience

The role of resilience in stress coping is evident [8, 16–18]. However, the EEG-based literature investigating the stress resilience is still at its infancy despite the increasing findings from fMRI studies highlighting the role of prefrontal cortex in resilience. Findings from studies that focus on cohorts with specific life experience may not be generalized to normal healthy populations. The association of EEG-brain waves with resilience at resting state of the brain may also differ when the brain is responding to a stressful event. These limitations highlight the need to identify EEG-neural correlates for resilience in a stress-inducing state. These findings can further be used in resilience training program for healthy population.

1.6   Rationale of Study

The identification of EEG-neural correlates of stress resilience can be used as brain markers to assess risk of individual in developing mental illness. By establishing the correlation between these EEG-neural correlates for stress resilience with established psychological assessments, the brain markers can be used in cognitive-behavioural therapy in both clinical and non-clinical populations. With the use of EEG-modality, these findings can be further explored in research focusing on resilience as it is a more cost-efficient alternative than the fMRI modality.

2   Proposed Methodology

In order to identify EEG-neural correlates for resilience building, there is a need to present the brain with a stress stimulus. This can be observed in the stress stimuli

presented to subjects in the fMRI-based literatures [13, 17, 18]. The stimuli trigger the stress resilience mechanism which attempts to cope with the brain activity changes induced by stress. Although many studies on stress resilience using the EEG modality evaluated the brain activity at resting state [11, 22, 23], these findings provide directions for further analysis in identifying the features for stress resilience.

2.1 Current Findings of EEG for Stress Resilience

Compared to the literature on stress responses, the EEG brain markers of stress resilience are less studied [23]. Paban et al. (2019) has observed significant correlation between resilience level and brain network flexibility in the delta, alpha and beta waves. Flexibility index is a measure of how often one node (i.e. a brain region) within a network changes its affiliation with a module (i.e. a set of strongly connected brain regions performing specific function [24]) over time [23, 25, 26]. Flexibility of delta, alpha and beta brain networks during 10 min of resting with eyes closed was negatively correlated with the psychological resilience score [23]. Network flexibility of specific brain regions significantly decreased with higher resilience score, suggesting that the brain core network is less flexible in resilient subjects.

Zhang et al. (2018) also investigated the association of resilience of network at resting-state of EEG signal between patients with major depressive disorders and healthy participants [22]. In the study, network resilience reflected the degree of tolerance when the functional brain network is subjected to unstable events. Patients with MDD had lower inter-nodal mean and positive correlation coefficients in the alpha wave of the functional brain network, which suggests weaker density of node-pairing connections compared to healthy participants. This finding strengthens the postulation by Paban et al. (2019) that people with higher psychological resilience level have a more stable rich-club network.

In a study investigating biomarker for attentional control using theta/beta ratio, Putman et al. observed that higher power density of the ratio is significantly associated with lower self-reported trait attentional control [27]. Findings from this study was concluded to contribute in predicting an individual's resilience to effects of stress-inducing cognitive performance anxiety on the self-reported state attentional control. Although this conclusion is unclear as to its association with stress resilience, the beta band may be linked to the brain activity reflecting stress resilience as observed by [23].

Curtis and Cicchetti studied the resilience and its associated hemispheric EEG asymmetry during resting eyes-open and closed in maltreated and non-maltreated children [11]. Findings suggested that resilient children had greater left hemispheric activity in the central cortical area, C3 and C4. Asymmetrical activity in the parietal cortices, favoring P3 location, was also observed in non-maltreated children.
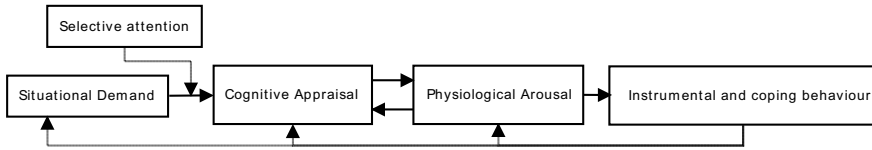
Fig. 1 Cognitive affective conceptual stress model adapted from [7]

2.2 Stress Resilience as a Flexible Adaptation Process

The cognitive affective stress management training program is a brief six-session stress program designed to promote resilience using the "induced affect" procedure by imagining stressful situations [7]. The procedure allows patients to experience and learn to control the negative effects of stress, and thus enhance stress resilience. The theoretical framework of this training program can be explained using the cognitive affective conceptual stress model [28], shown in Fig. 1, that relates the cognition, physiological responses and behaviour towards stress [7]. When a person faces a stressful situation, the brain responds by paying selective attention to important stimuli of the situation and appraising how demanding and relevant the situation is to their well-being, and what resources do they have to cope with the situation. Physiological arousal is the response to what has been perceived earlier and involves the mobilization of resources to deal with the stressful situation. During this process, the level of emotional intensity is also fed back to cognitive appraisal and therefore influence the appraisal component. In response to the cognitive appraisal and physiological arousal, a person takes behavioural actions as a coping mechanism towards the stressful situation. The actions can orient towards palliative emotion-regulating or direct problem-solving techniques [28–30].

Throughout this process, stress resilience can be defined as the flexible adaptation process where an individual mobilizes their resources to healthily cope with current and future adversities. We hypothesize that the EEG-features extracted from these three components correlate with the psychological measure of stress resilience: cognitive appraisal, physiological arousal, and behavioural tendencies.

2.3 Proposed EEG-Signal Analysis

With reference to the cognitive affective conceptual stress model, we suggest inducing stress to observe the elicitation of stress resilience using EEG brain activity. This paper proposes the pretest–posttest quasi experimental study design where the brain activity will be recorded from subjects at three time-points; (1) at relax-state as-signed as control condition, (2) during stress-induced training as intervention condition, and (3) at relax-state post-intervention. This study design is similar to the within-subjects experiment where the same subjects are assigned to both control and

intervention conditions [31]. The sample size estimated for this experiment is N = 29, calculated as follows [32]:

$$N = \frac{2\left(z_\alpha + z_\beta\right)^2 (1 + (n-1)\rho)}{n\delta^2}$$

$$N = 29$$

where
N = sample size.
Z_α = 1.96 at level confidence of 95%
Z_β = 0.842 at 80% power.
n = 3, number of time points.
ρ = 0.5, assumed correlation of repeated measure.
δ = 0.6, effect size.

To address our hypothesis, the flow of EEG-signal analysis shown in Fig. 2 is proposed for our study. We aim to identify the prominent EEG bands and hemispheric asymmetries that may reflect the adaptation process of stress resilience. We also aim to identify the brain networks between stress-inducing event and stress-induced training using EEG source-space connectivity method.

Fig. 2 Proposed approach for EEG features of stress resilience
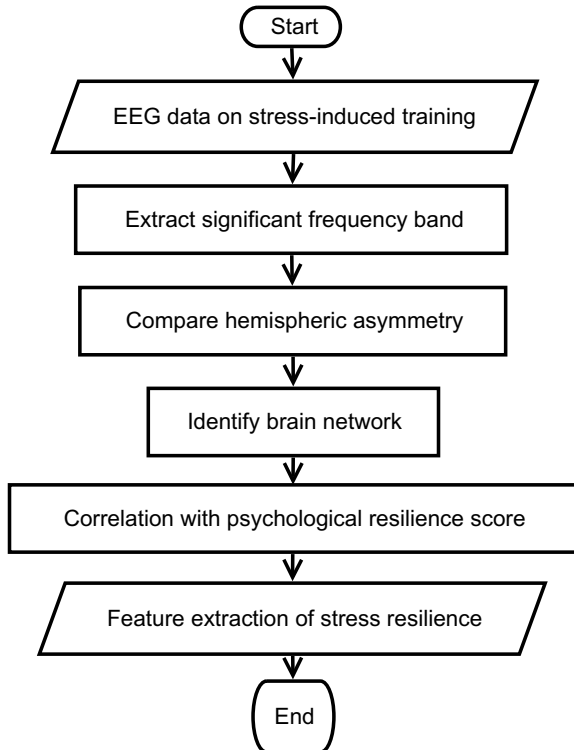
3   Expected Findings of EEG Footprint for Stress Resilience

In understanding the neural activity of resilience, our approach of identifying the EEG features from brain rhythm, hemispheric asymmetry and brain network can be used to provide a physiological measure for resilience. The brain rhythms and network extracted from EEG signal have been strongly documented in literatures to reflect different brain activities such as aversive response towards stress [33–35] and coping behaviour [34, 35]. Hence, we hypothesize that resilience at resting state and during training will elicit different EEG bands, and that these significant bands are functionally link and form the brain networks of stress resilience.

4   Summary

Prolonged stress in working environment can lead to reduced quality of job performance and quality of life. Literatures have found that mental health is associated with lower stress level and higher resilience. As resilience is defined as the ability to cope and adapt to stressful situation, there is a need to identify the EEG-neural correlates of this psychological parameter when stress in being induced. In addition to its advantages of being non-invasive, having high temporal resolution, low cost and ease-of-portability, the EEG modality has been extensively used in monitoring stress responses. Through this study, we would be able to identify the features of stress resilience and correlate them to the conventional psychological measures. Similar to the application of EEG-feature of stress response, these potential features may be used in clinical practices for risk assessment, diagnosis and interventions.

References

1. Harnois G, Gabriel P (2002) Mental health and work: impact, issues and good practices. World Health Organization
2. Kassim MSA, Ismail A, Ismail R (2018) A review of occupational stress prevalence and its predictors among selected working populations in Malaysia. Malays J Publ Health Med 18(2):1–6
3. Fink G (2016) Stress: Concepts, Cognition, Emotion, and Behavior, ser. In: Handbook of Stress Series. Academic Press, vol. 1
4. Feltoe G, Beamish W, Davies M (2016) Secondary school teacher stress and coping: insights from Queensland, Australia. Int J Arts Sci 9(2):597–608
5. Menon J (2016) Depression on the rise as Malaysians burn out from stress, expert warns. MalayMail. [Online]. Available: https://www.malaymail.com/news/malaysia/2016/03/24/depression-on-the-rise-as-malaysians-burn-out-from-stress-expert-warns/1086087

6. Thye TSLL (2016) Work stress and mental health. New Straits Times. [Online]. Available: https://www.nst.com.my/news/2016/04/142074/work-stress-and-mental-health

7. Smith RE, Ascough JC (2016) Promoting emotional resilience: cognitive-affective stress management training. The Guilford Press, New York

8. Franklin T, Saab B, Mansuy I (2012) Neural mechanisms of stress resilience and vulnerability. Neuron 75(5):747–761. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0896627312007532

9. Davydov DM, Stewart R, Ritchie K, Chaudieu I (2010) Resilience and mental health. Clin Psychol Rev 30(5):479–495. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0272735810000437

10. Connor KM, Davidson JR (2003) Development of a new resilience scale: the connor-davidson resilience scale (CD-RISC). Depression Anxiety 18(2):76–82. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/https://doi.org/10.1002/da.10113

11. Curtis WJ, Cicchetti D (2007) Emotion and resilience: a multilevel investigation of hemispheric electroencephalogram asymmetry and emotion regulation in maltreated and nonmaltreated children. Dev Psychopathol 19(3):811–840

12. Burton NW, Pakenham KI, Brown WJ (2010) Feasibility and effectiveness of psychosocial resilience training: a pilot study of the ready program. Psychol Health Med 15(3):266–277, pMID: 20480432. [Online]. Available: https://doi.org/10.1080/13548501003758710

13. Keynan JN, Cohen A, Jackont G, Green N, Goldway N, Davidov A, Meir-Hasson Y, Raz G, Intrator N, Fruchter E, Ginat K, Laska E, Cavazza M, Hendler T (Jan 2019) Electrical fingerprint of the amygdala guides neurofeedback training for stress resilience. Nat Hum Behav 3(1):63–73. [Online]. Available: https://doi.org/10.1038/s41562-018-0484-3

14. McEwen BS (2006) Protective and damaging effects of stress mediators: central role of the brain. Dialogues Clin Neurosci 8(17290796):367–381. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3181832/

15. Goodman RN, Rietschel JC, Lo L-C, Costanzo ME, Hatfield BD (2013) Stress, emotion regulation and cognitive performance: the predictive contributions of trait and state relative frontal EEG alpha asymmetry. Int J Psychophysiology 87(2):115–123. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167876012005934

16. Kong F, Wang X, Hu S, Liu J (2015) Neural correlates of psychological resilience and their relation to life satisfaction in a sample of healthy young adults. NeuroImage 123:165–172. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811915007296

17. Peres JF, Foerster B, Santana LG, Fereira MD, Nasello AG, Savoia M, Moreira-Almeida A, Lederman H (2011) Police officers under attack: resilience implications of an fMRI study. J Psychiatr Res 45(6):727–734. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022395610003225

18. Reynaud E, Guedj E, Souville M, Trousselard M, Zendjidjian X, Khoury-Malhame ME, Fakra E, Nazarian B, Blin O, Canini F, Khalfa S (2013) Relationship between emotional experience and resilience: an fmri study in fire-fighters. Neuropsychologia 51(5):845–849. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0028393213000134

19. Meir-Hasson Y, Keynan JN, Kinreich S, Jackont G, Cohen A, Podlipsky-Klovatch I, Hendler T, Intrator N (May 2016) One-class fmri-inspired EEG model for self-regulation training. PLOS ONE 11(5):e0154968. [Online]. Available: https://doi.org/10.1371/journal.pone.0154968

20. Scarapicchia V, Brown C, Mayo C, Gawryluk JR (2017) Functional magnetic resonance imaging and functional near-infrared spectroscopy: insights from combined recording studies. Frontiers Human Neurosci 11:419. [Online]. Available: https://www.frontiersin.org/article/https://doi.org/10.3389/fnhum.2017.00419

21. Thibault RT, Lifshitz M, Birbaumer N, Raz A (2015) Neurofeedback, self-regulation, and brain imaging: clinical science and fad in the service of mental disorders. Psychother Psychosom 84(4):193–207. [Online]. Available: https://www.karger.com/DOI/https://doi.org/10.1159/000371714

22. Zhang M, Zhou H, Liu L, Feng L, Yang J, Wang G, Zhong N (2018) Randomized EEG functional brain networks in major depressive disorders with greater resilience and lower rich-club coefficient. Clin Neurophysiol 129(4):743–758. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1388245718300324

23. Paban V, Modolo J, Mheich A, Hassan M (2019) Psychological resilience correlates with EEG source-space brain network flexibility. Netw Neurosci 3(2):539–550. [Online]. Available: https://doi.org/10.1162/netn_a_00079

24. Bassett DS, Wymbs NF, Porter MA, Mucha PJ, Carlson JM, Grafton ST (2011) Dynamic reconfiguration of human brain networks during learning. Proc National Acad Sci 108(18):7641–7646. [Online]. Available: https://www.pnas.org/content/108/18/7641

25. Mattar MG, Betzel RF, Bassett DS (July 2016) The flexible brain. Brain 139(8):2110–2112. [Online]. Available: https://doi.org/10.1093/brain/aww151

26. Braun U, Schäfer A, Walter H, Erk S, Romanczuk-Seiferth N, Haddad L, Schweiger JI, Grimm O, Heinz A, Tost H, Meyer-Lindenberg A, Bassett DS (Sept 2015) Dynamic reconfiguration of frontal brain networks during executive cognition in humans. Proc National Acad Sci US Am 112(26324898):11678–11683. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4577153/

27. Putman P, Verkuil B, Arias-Garcia E, Pantazi I, van Schie C (Jun 2014) EEG theta/beta ratio as a potential biomarker for attentional control and resilience against deleterious effects of stress on attention. Cogn Affect Behav Neurosci 14(2):782–791. [Online]. Available: https://doi.org/10.3758/s13415-013-0238-7

28. Lazarus RS, Folkman S (1984) Stress. Springer Publishing Company, Appraisal and Coping

29. Kyriacou C (2001) Teacher stress: directions for future research. Educ Rev 53(1):27–35

30. Research H (2015) Supporting teachers during times of change and stress. Hanover Research

31. Jhangiani RS, Chiang I-CA, Price PC (2015) Research Methods in Psychology. BCcampus Open Education

32. Hedeker D, Gibbons RD, Waternaux C (1999) Sample size estimation for longitudinal designs with attrition. J Educ Behav Stat 24:70–93

33. Seo S-H, Lee J-T (2010) Stress and EEG. In: Crisan M (ed) Convergence and hybrid information technologies. Intech, ch 27, pp 413–426

34. Brouwer A-M, Neerincx MA, Kallen V, van der Leer L, ten Brinke M (2011) EEG alpha asymmetry, heart rate variability and cortisol in response to virtual reality induced stress. J Cyberther Rehabil 4(1):27–40

35. Flo E, Steine I, Blågstad T, Grønli J, Pallesen S, Portas CM (2011) Transient changes in frontal alpha asymmetry as a measure of emotional and physical distress during sleep. Brain Res 1367:234–249. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0006899310021566

Modeling and Tuning of PID Controller for Continuous Stirred Tank Reactor

A. Suguna, S. N. Deepa, and N. Rajasingam

Abstract Continuous Stirred tank reactor is a chemical reactor system which exhibits complex non-linear dynamic characteristics. The quality of final product is based on the design of the controller. The mathematical modeling of CSTR is designed based on first principle method. Conventional PID controllers Ziegler-Nichols, Tyreus-Luyben, Cohen-Coon and IMC based PID have been implemented and the performance analysis of different PID tuning methods is done. The performance of the PID controller is analyzed in MATLAB simulation.

Keywords CSTR · PID controller · Tuning

1 Introduction

Continuous stirred tank reactor which is widely used equipment mainly in the chemical industry, exhibits highly nonlinear behavior. A CSTR is a exothermic irreversible reaction, a fluid stream is continuously fed to the reactor. Since the reactor is perfectly mixed the exit stream has the same concentration and temperature as the reactor fluid. The jacket surrounding the reactor has feed and exit streams. The jacket is assumed perfectly mixed, energy passes through the reactor walls into the jacket, removes the heat generated by reaction. The mathematical model is developed from material balance equation [1]. The two state variables are the reactor concentration and reactor temperature. PID control has three modes i.e., proportional, integral and derivative. The proportional term is proportional changes for error (the difference between the set point and process variable) to the control output. The integral term examines the process variable over time and it corrects the output by reducing the offset from

A. Suguna
Government College of Technology, Coimbatore, India

S. N. Deepa
Anna University Regional Campus, Coimbatore, India

N. Rajasingam (✉)
Dr. N. G. P. Institute of Technology, Coimbatore, India

process variable. Derivative control mode monitors the rate of change of the process variable and therefore changes the output when there are unusual variations. Each parameter of the three control functions adjust to obtain the desired performance from the process [2]. Easy implementation and maintenance due to their simple structure, PID controllers are the most extensively used controllers in the process control industry. Different types of controllers are available like sliding mode control, robust control. But the PID controllers still used in industries, Conventional tuning methods of PID controller used for this CSTR process.

2 Mathematical Model of CSTR

The mathematical model for the Continuous Stirred Tank Reactor is derived from mass balance and energy balance equation [3]. Mass balance equation is given by the rate of accumulation for material is equal to difference between rate of material inflow and outflow.

$$\frac{dVC_A}{dt} = FC_{A0} - FC_A - rV \tag{1}$$

Energy balance equation is given by

$$d\left(v_P C_p (T - Tref)\right)/dt = F_\rho C_p \left(T_f - Tref\right) - F_\rho C_p (T - Tref) \\ + (-\Delta H)V_r - UA(T - T_i) \tag{2}$$

State variable form of equations can be written as

$$f_1(C_A, T) = \frac{dC_A}{dt} = \frac{F}{V}\left(C_{Af} - C_A\right) - r \tag{3}$$

$$f_2(C_A, T) = \frac{dT}{dt} = \frac{F}{A}\left(T_f - T\right) + \left(\frac{-\Delta H}{p\mathbb{C}_p}\right)r - \frac{UA}{v_p C_p}(T - T_i) \tag{4}$$

The reaction rate per unit volume r (Arrhenius expression) is given by

$$r = K_0 exp\left(\frac{-\Delta E}{RT}\right)C_A \tag{5}$$

State variables are concentration and temperature. At steady state $\frac{dC_A}{dt} = 0$ and $\frac{dT}{dt} = 0$

$$f_1(C_A, T) = 0 = \frac{F}{V}\left(C_{Af} - C_A\right) - K_0 exp\left(\frac{-\Delta E}{RT}\right)C_A \tag{6}$$

$$f_2(C_A, T) = 0 = \frac{F}{V}(T_f - T) + \left(\frac{-\Delta H}{p\mathbb{C}_p}\right)K_0 exp\left(\frac{-\Delta E}{RT}\right)C_A - \frac{UA}{v_p C_p}(T - T_i)$$

(7)

The controlled variable is temperature which is changed by regulating the coolant flow rate. CSTR is a non-linear system, the linearization has be done using Jacobian matrix form

$$A = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} -\frac{F}{V} - K_S & -K_S' C_{AS} \\ \frac{\Delta H}{p\mathbb{C}_p}K_S & -\frac{F}{V} - \frac{UA}{V_p C_p} + \frac{(-\Delta H)}{p\mathbb{C}_p}K_S' C_A \end{pmatrix}$$

$$K_S = K_0 exp\left(\frac{-E}{RT}\right)$$

(8)

$$k_S' = \frac{\partial K_S}{\partial T}$$

$$B = \begin{pmatrix} \frac{\partial f_1}{\partial u_1} \\ \frac{\partial f_2}{\partial u_2} \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{UA}{v_p C_p} \end{pmatrix}$$

(9)

where, x_1- Reactor concentration, x_2- reactor temperature and u_2- coolant flow rate.

Substituting the steady state operating data the state space matrices can be obtained as follows

$$A = \begin{bmatrix} 0.0121 & -0.0000 \\ 2.2291 & -0.0080 \end{bmatrix} B = \begin{bmatrix} 0 \\ 7 \end{bmatrix} C = \begin{bmatrix} 0 & 1 \end{bmatrix} D = [0]$$

The Steady state operating data are available in [4].

3 PID Controller

3.1 Ziegler-Nichols Tuning (ZN)

This paper presents ZN closed loop tuning method. It requires the determination of the ultimate gain and ultimate period [5] By adjusting the controller gain (Ku) till the system goes sustained oscillations (ultimate gain) this can be achieved, while maintaining the integral time Constant (Ti) at infinity and the derivative time constant (Td) at zero. Note down ultimate gain-Ku (gain at oscillation occurs) and ultimate period—Pu. The tuning parameters Proportional gain, Integral Time and Derivative Time are calculated from prescribed ZN tuning rules as mentioned in the Table.1 [6].

Table 1 Tuning rules of PID controller

Tuning methods	K_P	T_I	T_d
Ziegler and Nichols	$0.6K_u$	$0.5P_u$	$0.125P_u$
Tyreus- Luyben	$\frac{K_u}{2.2}$	$2.2P_u$	$\frac{P_u}{6.3}$
Cohen and coon	$\frac{1}{K}\frac{T}{\tau}\frac{4}{3}\frac{\tau}{4T}$	$\tau\frac{\left(32+\frac{6\tau}{T}\right)}{\left(13+\frac{8\tau}{T}\right)}$	$\tau\frac{4}{\left(11+\frac{8\tau}{T}\right)}$
IMC based PID	$\frac{(\tau+0.5T)}{K_p(\lambda+0.5T)}$	$\tau+0.5T$	$\frac{\tau T}{2\tau+T}$

3.2 Tyreus-Luyben Method (TL)

B.D Tyreus and W.I Luyben in 1997, proposed this method. To calculate the ultimate gain Ku and ultimate period Pu the same Ziegler Nichols tuning procedure have to be followed. Controller parameters calculated from the rules given below.

3.3 Cohen and Coon Method (CC)

The other name of Cohen and Coon method is Process reaction curve method, the step input is given to the open loop response of the system, the S shaped curve obtained. From this curve gain, dead time and time constant calculated, Substituting this parameters into the tuning rules controller tuning parameters computed.

3.4 IMC

To model CSTR process by fitting the open-loop step test data as a first order function with time delay, the internal model control-PID (IMC-PID) tuning rules have the advantage of using a single tuning parameter to achieve a clear tradeoff between closed-loop performance and robustness to model inaccuracies.

The IMC-PID controller provides good set-point tracking but has a sluggish disturbance response [7]. The different tuning rules for PID controller as shown in Table 1.

4 Simulation and Results

Ziegler-Nichols, Tyreus-Luyben, Cohen-Coon and IMC based PID methods conventional PID parameters are determined using the tuning rules given in Table 1. The values of PID controller parameters for different tuning rules are identified in Table

Table 2 Tuning parameters of PID controller

Tuning methods	K_P	T_I	T_d
ZN	10.8	1.5	0.375
TL	12.1	1.9	0.25
CC	3.7	1.5	0.0388
IMC based PID	2.109	12.607	0.0157

Table 3 Time domain specifications of PID controller

Tuning methods	Rise time	Peak overshoot	Settling time
ZN	0.23	25.8	2
TL	0.0852	30.8	1.5
CC	0.0359	33.6	0.488
IMC based PID	0.121	5.61	0.41

2. The unit step response for the close loop control of CSTR is given to get the output response.

Closed loop response parameters are analyzed using the time domain specifications of all PID tuning methods and shown in Table 3. To use these empirical tuning rules to CSTR process, we need to fit the data to a first order with dead time. First order with dead time system obtained the approximation of the CSTR by giving step signal input.

Figures 1, 2, 3 and 4 represents the closed loop response of the CSTR process based on the PID tuning parameters of Ziegler-Nichols, Tyreus—Luyben, Cohen-Coon and IMC based PID methods.

Fig. 1 Response of CSTR-Ziegler Nicholas
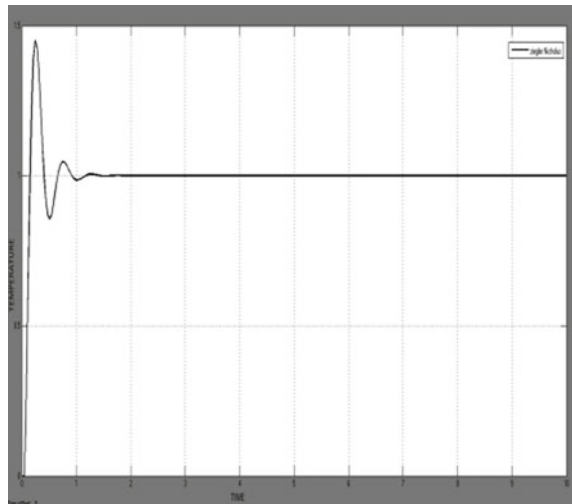
Fig. 2 Response of
CSTR-Tyrus Luyben
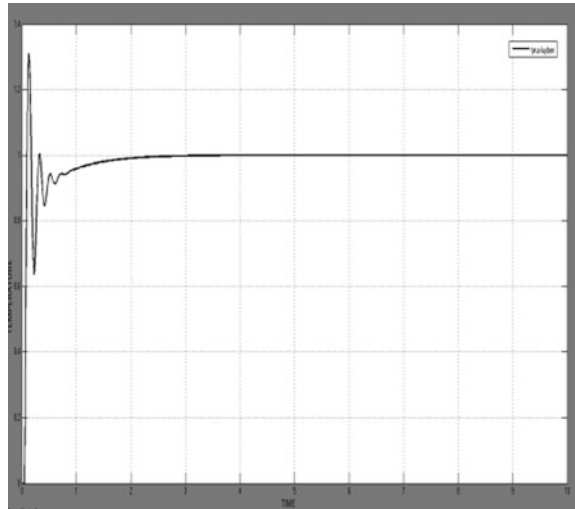
Fig. 3 Response of
CSTR-Cohen and Coon
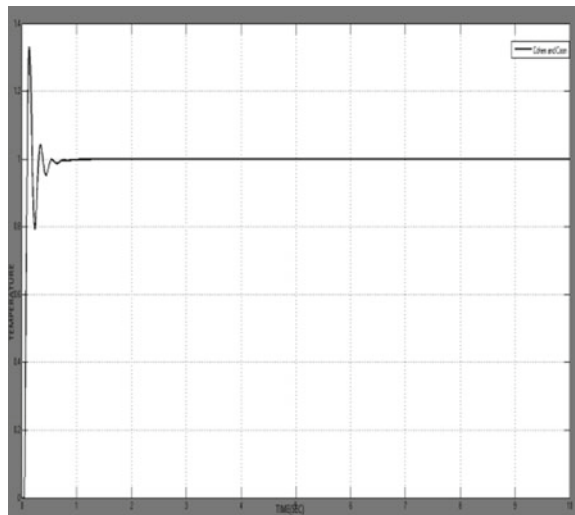
5 Conclusion

Comparison of different tuning methods of PID controller for temperature control
of Continuous Stirred Tank Reactor has been implemented. The tuning methods
are implemented in the CSTR process transfer function obtained from first order
principle method. Simulation results are analyzed by giving the step input to get the
output response using MATLAB. These tuning methods are analytical; Based on the

Fig. 4 Response of CSTR-IMC based PID
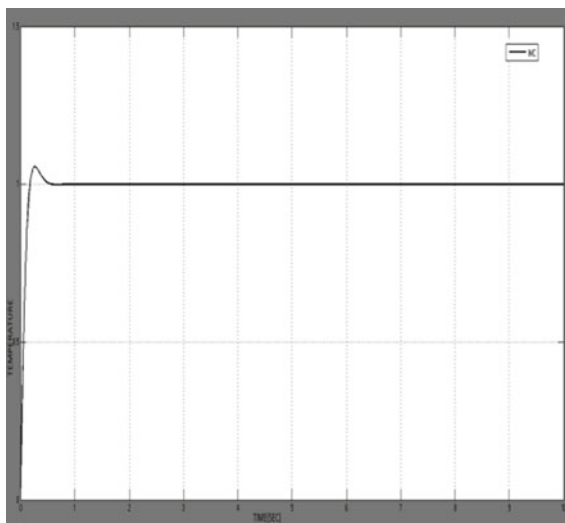
analyze IMC based PID controller provide good transient response for the CSTR process.

References

1. Ogata K, Yang Y (2002) Modern control engineering. Prentice hall, India
2. Tyreus BD, Luyben WL (1992) Tuning PI controllers for integrator/dead time processes. Ind Eng Chem Res 31(11):2625–2628
3. Bequette BW (2003) Process control: modeling, design, and simulation. Prentice Hall Professional
4. Kamala N (2011) Studies in modeling and design of controllers for a nonideal continuous stirred tank reactor
5. Cominos P, Munro N (2002) PID controllers: recent tuning methods and design to specification. IEE Proc-Control Theory Appl 149(1):46–53
6. George S (1984) Chemical process control: An introduction to theory and practice. PTR Prentice Hall, Inc. 3(9015):03983
7. Shamsuzzoha M, Lee M, Lee J (2005) IMC-PID controller tuning for improved disturbance rejection of unstable time delay processes. Theories and application of Chemical Engineering, vol 11

Adaptive Security for Cognitive Robotic Process Automation in Enterprise Computing Using AI-Powered Analytical Engine

Mohamed Ibrahim Beer Mohamed and **Mohd Fadzil Hassan**

Abstract The robotic process automation (RPA) in enterprise computing refers to software bots that are capable of mimicking most of the human–computer interactions to carry out day-to-day business operations. This RPA is targeted for automating rule-based repetitive and high-volume tasks with higher accuracy which eventually reduces the operation cost and processing time. The cognitive automation which is being developed as part of enterprise computing automation best utilizes Artificial Intelligence (AI) and Machine Learning (ML) techniques to enhance the process model of RPA in terms of improved accuracy, reliability, and consistency in taking intelligent business decisions with no or minimal human interventions. The majority of the communications between computing nodes and bots are being provisioned through service-oriented enterprise computing interfaces, which are built based on Service-Oriented Architecture (SOA). The SOA by itself does not possess any security layer and it defaults to the Open Systems Interconnection (OSI) model for security that is inadequate in this modern era of process automation and interfacing. In this paper, the security concerns of Cognitive RPA for enterprise computing are analyzed, and a novel approach is presented for adding-up the security layer for Cognitive RPA. This is an adaptive approach which works on the predict-prevent-learn pattern for effective proactive security as differed from traditional reactive security, where the Artificial Intelligence (AI) techniques are used for 'predict' part for predicting potential security threats, the Artificial Neural Networks (ANN) techniques are applied for 'learn' part on unsupervised learning of anticipated security vulnerabilities, and security prevention algorithms are equipped for 'prevent' part to defend against the security threats on Cognitive RPA systems.

Keywords Artificial Intelligence · RPA · Security · Cognitive automation · ANN

M. I. Beer Mohamed (✉) · M. F. Hassan
Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, 32610
Seri Iskandar, Malaysia

1 Introduction

The robotic process automation (RPA) is being focused on enterprise computing to minimize the operational cost and processing time. The cognitive RPA allows enterprises to automate tasks that usually need human intervention for interpretation and decision making. As a result of Cognitive RPA, the processes which have been performed by humans are assigned to software modules, named "bots". These "bots" are equipped with Artificial Intelligence (AI) techniques for real-time analysis and thus to take cognitive decisions resulting in minimal or no human interactions. While RPA is more focused on data-intensive processes, the Cognitive RPA concerns with logic-based outcomes [1].

RPA is effective for the automating processes which are (i) rule-based, (ii) repeated/pre-defined trigger, (iii) definitive inputs/outputs, and (iv) large-scale in volume. With the Cognitive RPA, an 'intelligent' part is added to these processes which result in the following significant benefits [2, 3]. Moreover, software bots are easy to deploy and integrate with existing business application modules seamlessly.

- *Minimal human interventio*n: The AI-powered software bots replaced human interactions for performing intellectual tasks.
- *Better accuracy*: The results of software bots are more accurate in analytics and decision makings on big data consistently.
- *Improved compliance*: The software bots are more controlled and operated in compliance with organizational regulations and policies.
- *Quick wins on cost savings*: As a quick win, most of the organizations experience more than 60% cost saving in less than 12 months of time after implementation of Cognitive RPA [4]. This cost savings rate increased over the years as only it requires minimum operating costs.
- *Scalability*: Comparing to the human-based operating model, bots are scalable on an as-needed basis where they can be added or reduced based on workload.

A comprehensive cognitive RPA bot should be equipped with an array of functional and non-functional entities such as reliability, scalability, speed, light-weighted, manageability, intelligence, and more importantly security [3, 4].

Despite the extended benefits, cognitive ROA introduces a new attack surface to enterprise computing in addition to the existing security risks of integration on process automation which results in the paramount importance of having RPA security. Applying security for cognitive RPA is not as simple as implementing a reactive way of the security establishment. It requires that the security should be applied across the entire product life cycle from design, architecture, interfacing, and implementation to underlying business operations satisfying essential security goals, including integrity, traceability, confidentiality, and control.

In this paper, the security concerns of the cognitive RPA are studied from an enterprise computing perspective. As a novelty, a security component named "Intelligent Security Engine (ISE)" is constructed based on Artificial Intelligence (AI) techniques for risk prediction and Artificial Neural Networks (ANN) for unsupervised learning

of security vulnerabilities for Cognitive RPA development and implementation, and the same is presented as the adaptive way forward security layer for Cognitive RPA.

The remaining sections of this paper are organized as the rationale and motivation of the study is briefed in Sect. 2, the proposed AI-based security approach is presented in Sect. 3, a brief discussion on the Proof-of-Concept is given in Sect. 4, and Sect. 5 concludes the paper.

2 Rationale for the Study

This section briefs on the rationale and motivation behind the conducted research study. The security concerns and the potential vulnerabilities of the cognitive robotic process automation are reviewed in terms of available literature, and the mandatory need for a proactive security model is highlighted as the traditional reactive security model is inadequate in this modern era of cognitive automation with the advent of related software and hardware technologies.

2.1 Security Risks in Cognitive RPA

In addition to the common enterprise application integration risks, the potential risks for cognitive RPA are listed below [5, 6].

- *Privilege escalation*: It is the risk of abusing the given access privilege given to a bot.
- *Data tampering*: It is the risk of disclosing sensitive data that are tampered with during inter-process communication.
- *Denial of Service*: It is the risk of stressing with the bot with a greater number of fake requests so that it will deny or delay in servicing to even legitimate requests.
- *Application vulnerabilities*: It is the risk of enacting the security threat using the existing vulnerabilities in the application components and/or supported software modules that are deployed on bots.
- *Abusing the digital identity and access control*: It is the risk of illegitimately applying digital federated identity and access privileges that are given to specific bots.
- *Data identification and protection*: It is both business and operations risk of sensitive data handling by robotics and automation tools adhering to compliance with the governance body.

2.2 Security Vulnerabilities on Cognitive RPA

In addition to the technology and implementation constraints, multiple business and managerial aspects could lead to security vulnerabilities, including the following [7, 8].

- Inadequate security controls on governing bots
- Lack of business continuity on workflow management
- Single point of failure, service unavailability, and machine errors on processing
- Unfortunate vulnerabilities in data processing and insufficient data protection
- Inconsistency on results among processes
- Inability of handing exceptions during the automation process
- Insufficient authentication/authorization controls on inter-process communications
- Unpredictable patterns on compliance and regulatory implications

2.3 Adaptive Security Approach for Cognitive RPA

The potential security threats for RPA are increasing as the attack surface is widening due to the advent of the latest automation and supported technologies. Hence, the security for RPA became a mandatory element in the business operating model. The traditional reactive way of security threat identification and perimeter defense strategies are obsolete right now [9–11]. It should be the paradigm shift with continuous monitoring and response to threats, shifting from reactive to a proactive approach to handling and preventing security attacks with the support of Artificial Intelligence (AI) and related technologies.

An adaptive security approach applies a layered way of detecting and preventing security threats. Hence it is focused to sense, detect, and respond to the prevention of security threats before the potential vulnerabilities are being used by the attackers on performing attacks for the exploitation of enterprise assets.

The adaptive security approach focuses on security aspects before, during, and post-execution of automated processes, targeted for (i) reducing the attack surface and attack velocity, (ii) real-time event-based monitoring on identifying security breaches, (iii) assurance of reliability and availability of automation resources, (iv) ensuring reliability in automation processes, (v) prevention of data and identity loss, and (vi) anticipated security thereat for earlier prevention.

The artificial neural networks (ANN) approach differs from conventional computation techniques in problem-solving as the conventional techniques apply the standard procedural and algorithmic approach with a pre-defined set of instructions where the ANN is being used to solve the problems which cannot be expressed exactly in terms of algorithmic steps. Security for cognitive-based solutions is one kind of such problem that can be effectively solved only through AI-based analytics and implementation [12].

3 Proposed Security Approach

The proposed adaptive security approach is built on the Predict-Prevent-Learn pattern, as depicted in Fig. 1. In the 'predict' stage, the security risks are identified and anticipated in the underlying cognitive RPA implementation, the 'prevent' stage hardens the system defending against the identified and anticipated security attacks without jeopardizing the industrial security standards. The 'learn' stage built the knowledge base on the security risks and vulnerabilities which is sourced by unsupervised learning processes.

The adaptive approach is applied and built as a software module, named "Intelligent Security Engine (ISE)," which can be attached to any computing node with minimal configuration. Basically, this ISE is an AI-powered analytical engine equipped with threat modeling and prevention. The cross-cutting view of ISE is given in Fig. 2 that applies the existing security standards for authentication/authorization, analytical rule-based decision engine, organization-specific security policies, and token management which are powered by AI and ANN implementations.

The data for analytical decision making on the severity of security threats and protection are provisioned by knowledge base which is sourced by Artificial Neural Networks (ANN) learning. The computational model of an artificial neuron is inspired by the biological neuron. A comparison of a biological neuron and an artificial neuron is given in Fig. 3, where the biological neuron works on electrochemical signals, and the artificial neuron works based on the mathematical model. In a biological neuron, the cell body is called "Nucleus", branching input as "Dendrite", and branching output as "Axon". The synapses connect dendrites to axons for propagating electrochemical signals through the nucleus. The neuron fires only its input

Fig. 1 Adaptive security approach for cognitive RPA
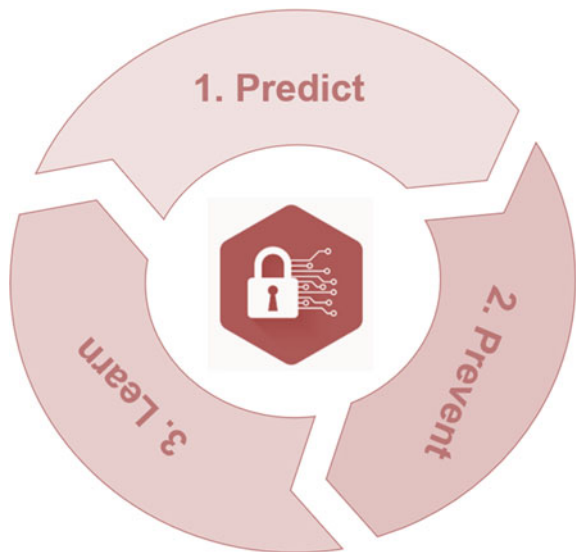
Fig. 2 Cross-cutting view of Intelligent Security Engine (ISE)
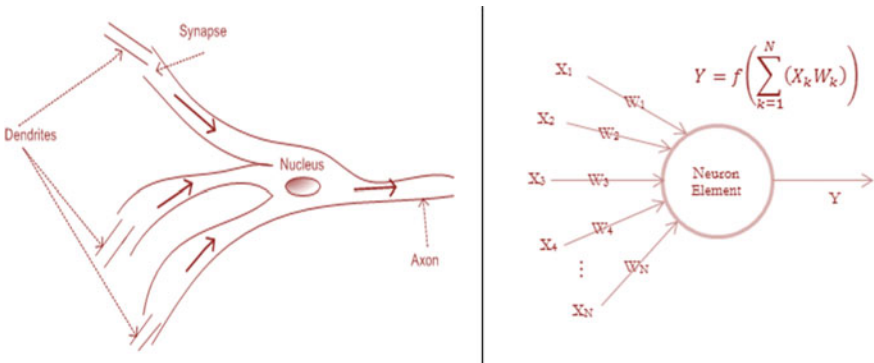
Fig. 3 Biological neuron versus artificial neuron

signals meet the threshold limit in a short period. The transmitted electrochemical signal from dendrites to an axon via the nucleus will act as an input dendrite signal for another neuron.

The artificial neuron collects input signals as values that are being multiplied by the factor, named weights which represent the strength of the respective input signals. Then, these values are fed to the mathematical function (aka transfer function) which determines whether to activate the neuron or not. If the weights to the input signals are high, then the corresponding neuron is strongly connected. The mathematical function is equivalent to the nucleus in a biological neuron, and the weights are similar to synapses which can hold positive, negative, or neutral (zero) values. The learning in ANN is the process of adjusting these weights for desired functionality such as threat identification and pattern matching. For this research study, the hyperbolic tangent sigmoid mathematical model is taken as a transfer function which has the representation as below.

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

The gradient of each weight concerning the desired output is measured as, $\frac{\partial E}{\partial w_{ij}}$ where E represents the sum of input values which are multiplied by their corresponding assigned weights, and w_{ij} represents the actual assigned weight on an edge which is connecting node i to node j. The backpropagation performs this gradient measurement on each edge and adjusts the weights incrementally using stochastic gradient descent as $w_{ij} = w_{ij} - \eta \frac{\partial E}{\partial w_{ij}}$ for all the nodes in the defined neural network.

The process flow of the proposed AI-powered threat prevention solution involves a systematic approach as outlined in Fig. 4.

The proposed security component "ISE" as depicted in Fig. 2 embeds this process model as given in Fig. 4. This software component can be easily added with any RPA processing nodes through an interceptor design pattern with a simple configuration where the input/output messages will be intercepted for threat monitoring and prevention. Though this model is good enough on learning the existing security vulnerabilities and threats, the training part will require a significant amount of time to build the knowledge base. Instead, the data mining techniques can be used to construct the initial knowledge base where it will be further refined by neural nets.

4   Result and Discussion

For proof-of-concept (PoC) purposes, the proposed security approach is implemented in a large-scale banking environment with real-time data on cognitive robotic process automation of customer onboarding and cheque clearance modules. The new to bank (NTB) customers are onboarded to the bank by the automated bots without any human intervention, and bots are automated to perform the entire customer
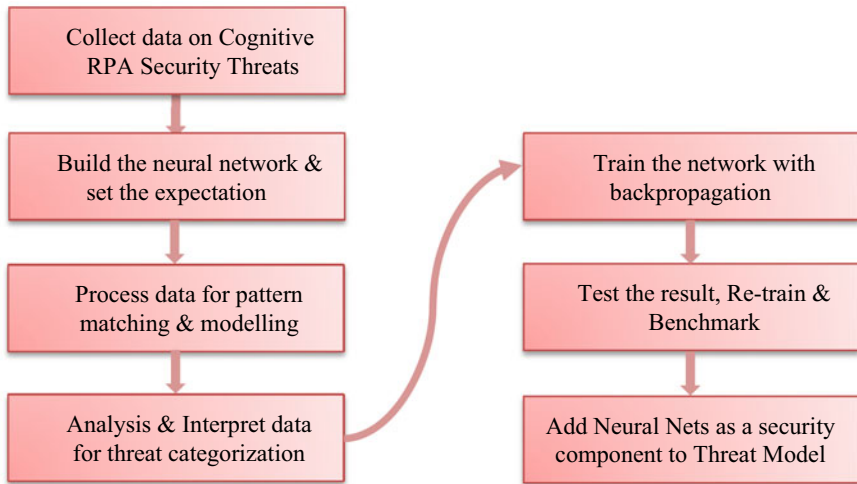
Fig. 4 Process flow of the proposed AI-powered security approach

on-boarding journey from due diligence checking, fraud monitoring, KYC updates, opening up the banking accounts, and registration for debit/credit cards, up to instruct the customer upon the account is ready for operation. The cheque clearance bot performs an intelligent check on signature matching, verification of the given details in the cheque, validation of transactions, and perform the fund transfer.

The artificial neural network model with back-propagation is chosen for the learning part and implemented for fraud and security risk assessment purposes, as part of PoC implementation. This ANN-based model produces reliable results when compared to other forecasting mechanisms.

As obtained from the conducted PoC, the statistical data on security attack and prevention ratio with existing (without AI) and proposed (AI-powered) security solutions on the chosen attack vector on cognitive RPA is given in Table 1. The ratio is represented as a failure of protection to protect from security attacks, for example, the ratio of 0.6000: 0.4000 represents 60% of failure on protection and 40% of protection from the underlying security attacks.

Table 1 Analysis of proof-of-concept (PoC) result

S.No	Attack vector	Existing security	Proposed security
1	Privilege escalation attack	0.5771: 0.4229	0.1083: *0.8917*
2	Data tampering attack	0.5469: 0.4531	0.1854: *0.8146*
3	Denial of Service attack	0.3442: 0.6558	0.0964: *0.9036*
4	Man-in-the-middle attack	0.5922: 0.4078	0.2357: *0.7643*
5	Abusing digital identity	0.6230: 0.3770	0.1323: *0.8677*
6	Broken access control	0.2123: 0.7877	0.2066: *0.7934*

By analyzing the PoC results, the observed sensitivity rate falls above 75% for the proposed AI-powered security solution. Except for the security protection for broken access control, the protection rate of the proposed security solution is at least 35% higher than the existing security solutions. Moreover, a higher protection rate for broken access control can be achieved through customized machine learning processes. The graphical representation of the obtained result as given in Table 1 is portrayed in Figs. 5, 6 and 7. With this analysis, it is clear that the proposed AI-powered security solution for cognitive RPA supersedes the existing non-AI based security solutions with a higher protection rate.
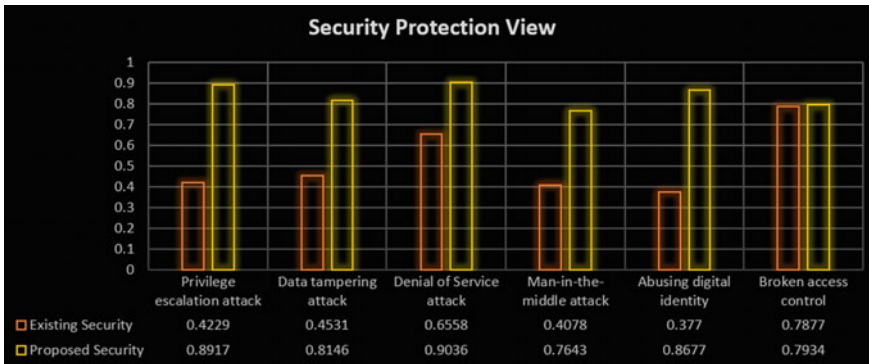
Fig. 5 Security protection view of PoC result
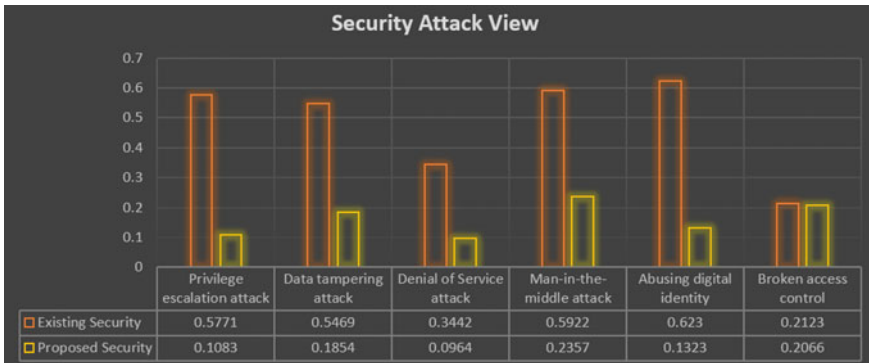
Fig. 6 Security attack view of PoC result
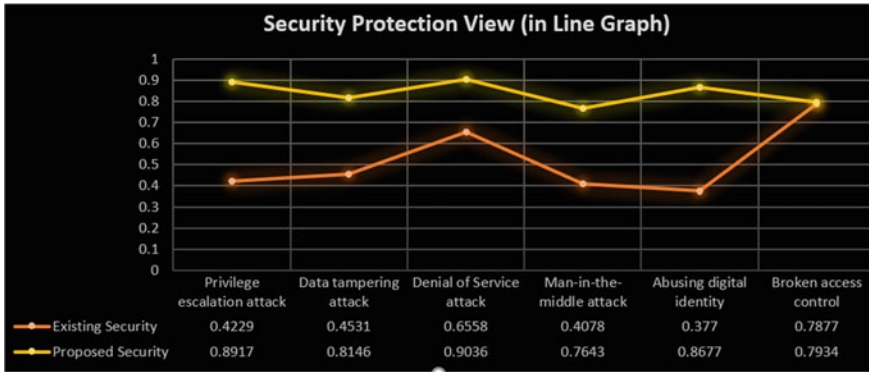
Fig. 7 Security protection view of PoC results in a line graph

5 Conclusion

The cognitive robotic process automation is the modern trend in enterprise computing and automation targeted for cost reduction and higher throughput in a minimum amount of time. With cognitive RPA, the automated tasks are getting completed with more accuracy and expected scalability. However, the security boundaries are getting expanded due machine to machine interfacing and end-to-end automation of business processes without any human interventions. In this paper, the major security risks on cognitive RPA are studied in the view of available literature, a navel security software component is constructed and presented in an adaptive way forward approach. This security approach is equipped with artificial intelligence and related technologies and best suited for this current era of business process automation. The proposed approach is practically implemented and tested as part of proof-of-concept, the obtained results are analyzed, and proved that the proposed AI-powered security solution provides at least 35% higher security than the existing non-AI based security solutions.

References

1. Martínez-Rojas A, Barba I, Enríquez JG (2020) Towards a Taxonomy of Cognitive RPA Components. In: The proceedings of Springer international conference on business process management, pp 161–175
2. Lamberton C, Brigo D, Hoy (2017) Impact of robotics, RPA, and AI on the insurance industry: challenges and opportunities. J Financ Perspect 4(1)
3. Teli R, Prasad SK (2020) Delivering value in procurement with robotic cognitive automation (RCA) Services. In: The IGI global proceedings of robotic systems: concepts, methodologies, tools, and applications, pp 1773–1785
4. Danger JL, Guilley S, Hoogvorst P, Murdica C, Naccache D (2012) Low-cost countermeasure against RPA. In: The proceedings of Springer international conference on smart card research

and advanced applications, pp 106–122

5. Syed R, Suriadi S, Adams M, Bandara W, Leemans S (2020) Robotic process automation: contemporary themes and challenges. Computers in Industry, pp 1–55

6. Zaharia-Radulescu AM, Pricop CL, Shuleski D, Ioan AC (2017) RPA and the future of workforce. In: Proceedings of the international management conference, vol 11(1), pp 384–392

7. Santos F, Pereira R, Vasconcelos JB (2019) Toward robotic process automation implementation: an end-to-end perspective. Bus Process Manage J

8. Leshob A, Bourgouin A, Renard L (2018) Towards a process analysis approach to adopt robotic process automation. In: IEEE 15th international conference on e-business engineering (ICEBE), pp 46–53

9. Hassan MF (2016) Construction of customizable SOA security framework using artificial neural networks. Jurnal Teknologi 78:69–75

10. Rudman R, Bruwer R (2016) Defining Web 3.0—opportunities and challenges. The Electronic Library

11. Willcocks L, Hindle J, Lacity M (2019) Keys to RPA success. Executive Research Report, Knowledge Capital Partners

12. Park S, Park H ANN based intrusion detection model. In: Workshops of the Springer international conference on advanced information networking and applications, pp 433–437

Multidirection Features Based Image Inpainting with Color-Direction Patch-Sparsity

B. Vidhya and M. Nikhil Madhav

Abstract To preserve the texture clarity, coherence of structure and surrounding consistence of the inpainted areas of an image in a better way, patch sparsity based image inpainting approach through color direction is proposed. To approximate the multi direction characteristics of a degraded image, super- wavelet transform is utilized in this method. It combines with color details to build the weighted color-direction distance (WCDD) for the determining the difference among two patches. WCDD is used to interpret the color direction structure sparsity which gives a more satisfying order of filling and more matching candidate patches for selection. Then, the target regions are sparsely characterized by the numerous candidate patches under surrounding consistency by considering the multi direction and color spaces constrictions together. Experimental results prove that the removal of scratch, text and objects are performed effectively through the proposed inpainting method.

Keywords Image inpainting · Multi direction feature · Color direction structure sparsity · Wavelet transform

1 Introduction

Image inpainting is the method of restoring the target regions that are scratched, missed or damaged. This process was first introduced to reconstruct the deteriorated old pictures or artworks. It is also called as image disocclusion or completion. Image inpainting has become a dynamic research subject in the area of computer vision and image processing since the recent advancement of digital image usage. The ultimate goal of image inpainting is to restore the lost region of a damaged picture in an unnoticeable manner. Nowadays image inpainting are widely used for reconstructing old pictures and restoring digital images.

B. Vidhya (✉)
Dr. N. G. P Institute of Technology, Coimbatore, India

M. N. Madhav
Sri Shakthi Institute of Engineering and Technology, Coimbatore, India

The main classification of image inpainting methods are: the exemplar based, the sparse based and the diffusion-based techniques. Partial Differential Equations (PDE) are used in the diffusion based approach where the missing regions are filled by the diffusion of image details from the source regions. The filling of missing regions was achieved by the diffusion of the image details along the isophote direction by Bertalmio et al. [1]. Later Bertalmio et al. [2] presented the Navier–Stokes equivalence in fluid dynamics into the process of image inpainting. T. F. Chan et al [3] utilized the Total Variation (TV) model for reconstructing the damaged regions. Then, T. F. Chan et al [4] anticipated the curvature-driven diffusion (CDD) equivalence to improve the TV model that underperformed in the connectivity principle.

P-Harmonic energy minimization was introduced in the TV model for filling the target regions by Zhang et al. [5], while other methods normally use surrounding pixels to fill the target pixels. Telea used first marching method to complete the target region by filling the known pixels of the neighbour from exterior to interior region [6]. Weickert et.al [7] proposed a unified framework for interpolation built on elliptic partial differential equations. Image was restored by Takeda et al. [8] through the association of the nonparametric statistics field and expanded kernel regression concepts. Many examinations were conducted by the researchers in the process of image inpainting to attain an improved restoration result.

Reconstruction of non textured missing regions that are smaller in size achieved by the diffusion-based algorithms have proven results. But its performance is poor for smaller textured or structured portions because of inadequacy of semantic texture/structure combination, and are unsatisfactory for larger missing regions.

The issues in image inpainting methods were addressed by the image sparse representation. Here the illustration of the image is performed by the sparse union of an over complete dictionary using DCT, wavelet, Contourlet, wavelet and so on [9–13] and the approximation of the missing pixels are corrected adaptively by the sparse representation. As an illustration, the damaged picture is separated into the texture layers and cartoon layers, that are correspondingly sparsely represented by Curvelet transforms and the Discrete Cosine Transform [9].

The problems in image inpainting can be addressed by the sparse representation of the image. Guleryuz [14] and Fadili et al. [15] characterized an image by its sparse combination using DCT, wavelets, Contourlet, Curvelet in the overcomplete dictionary approach. Pixels that are missing are calculated dynamically by restoring these sparse representation periodically. The target region or a damaged region in an image is segregated into textural regions and structural regions that are characterized sparsely into Curvelet Transform and Discrete Cosine Transform respectively [13]. Fadili et al. [16] presented an approach of expectation maximization (EM) for image inpainting in Bayesian framework for restoring the coefficients by the indorsing of prior penalty sparsity. All the above methods have limited generalization of algorithms, that are improved by the adaptive dictionary learning approach.

The proposed approach concentrates on the image inpainting technique based on matching performed by the patch sparsity. Here, multi-direction feature (MDF) using patch sparsity for inpainting the image is proposed for preserving the clarity of the textures and coherence of the structures. Patch sparse representation and structure

sparsity are used to represent the patch sparsity in the images. Based on the concept of algorithms used in the image inpainting algorithms, super-wavelet transform is implemented for extracting the multi-direction feature of an image.

It will be blended with the color details to form a weighted color-direction distance (WCDD) to estimate the homogeneity of patches. The value of WCDD calculated between the known and target patches also focuses on preserving the structure coherence through the color-direction structure sparsity (CDSS) function. For yielding subspace for patch sparse representation, numerous reference patches are checked using WCDD.

Additionally, an enhancement calculation with local patch uniformity limitations in both the color and the multi direction spaces is created to attain sparse linear combination factors of reference patches. The ultimate aim of this technique is to preserve the consistency of the inpainted regions with respect to the neighboring patches and to enhance the clarity of the structures and the textures of the images.

The key contributions of this article are: (i) to present the multi-direction feature to the image inpainting algorithm; (ii) to implement weighted color-direction interval to explore appropriate reference patches; (iii) to claim CDSS to get a strong order of filling; and (iv) to integrate the multi-direction and color limitations into the optimization measure for getting superior inpainting outcomes. This paper is ordered in the following order. Section 2 explains the method of extracting multi-direction characteristics of an image and the suggested exemplar based inpainting technique. Experimental results are shown in Sect. 3 and possible future work is proposed in the conclusions.

2 Proposed Algorithm

2.1 Extraction of Multi-direction Features in Image Inpainting

Image inpainting is the process of filling the target regions based on the details of neighboring regions. Image inpainting is mainly performed to preserve the structural consistency and surrounding consistency. If the details of the source regions known is higher, then the outcomes of the inpainting process will be better. For corrupted images, gradient features or divergence characteristics were used by the existing methods of image inpainting. Image consists of varied direction characteristics, so the gradient function that defines only 2 directions will not be able to completely define the details of an image.

The multidirectional and multiscale characteristics can be estimated using super-wavelet transform. Hence in the proposed algorithm, for obtaining the multidirectional attributes of an image, the super-wavelet transform is used. Among the different types of super-wavelet transforms, Curvelet transform [17], is selected

in the proposed system as it has proven results of extracting valid details accurately than other transforms like Contourlet transform [18] (CT), Non-subsampled Contourlet transform [19] (NSCT). The method of extracting eight direction matrices is explained with an illustration as given below.

Images are considered in YUV color space compared with RGB color space for coding efficiency. Curvelet forward transform is taken for the Y component of the image (IY) for obtaining the multi direction and multi scale decomposition. The coefficient matrix set of an image IY, is given by the Curvelet forward transform (T^+) of the Y component of the image.

$$Q = T^+(IY) \tag{1}$$

where $Q = \{Q_{s,d}\}$ gives the coefficient matrix set of image IY, and s and d in the matrix indicates the scale and the direction of the image respectively. If the multi scale curvelet coefficient is illustrated, it will resemble the structure as shown in the Fig. 1, where the first scale square matrix is present at the centre of the figure.

The subsequent squares observed from inside to outside of the structure represents the second to fifth scale coefficient matrices in the increasing order. Varied number of Curvelet coefficient matrix can be observed for all scales of coefficient matrix except for first scale. This is because of the reason that first scale coefficient matrix represents the low frequency characteristics whilst other scale coefficients are high frequency components. The second to fifth scale coefficient matrices comprises of the direction matrices of the size sixteen, thirty two, thirty two and sixty four respectively. As a starting point to figure out the coefficient matrix of every scale layer, 45° is considered as a reference point and is measured in anti-clockwise direction. Figure 2 shows the representations of fourth scale layer of the image consisting of 32 coefficients matrices.

Fig. 1 Distribution and direction of Curvelet coefficients
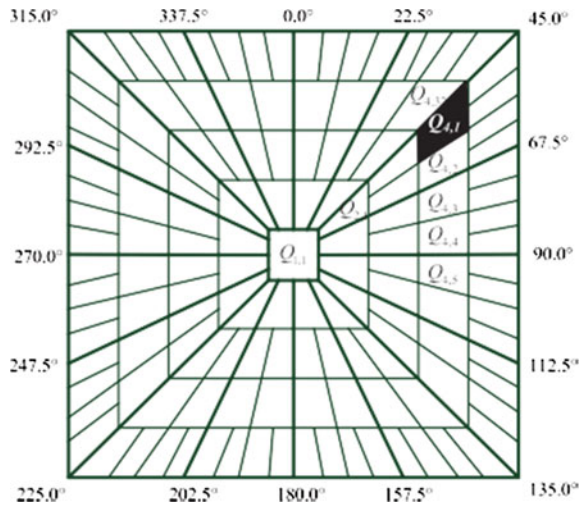
Fig. 2 Partition sets of directional Curvelet coefficients
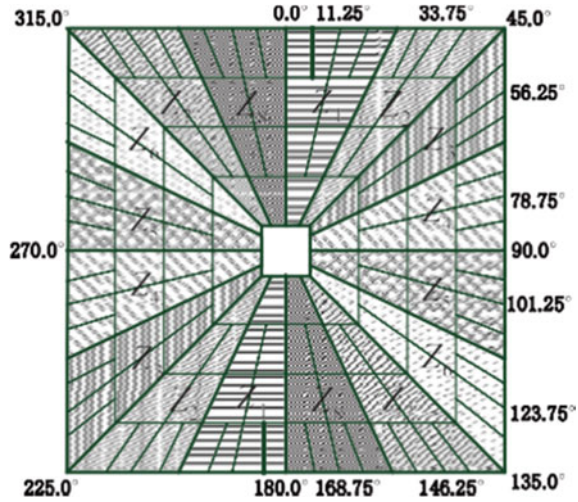
{Qs, d} values as *{Q4, 1}, {Q4, 2}, {Q4, 3},…* denotes the distinct coefficient matrices for varied directional characteristics of an image. These coefficient matrices are used to calculate the Multi Direction Features of an image. The second to fifth scale layers of the Curvelet coefficient matrices are divided into eight sets based on the direction of angles given by 157.5°, 135°, 112.5°, 90°, 67.5°, 45°, 22.5° and 0° in anti- clockwise direction that are represented by Z_1, Z_2,..., and Z_8 respectively.

By using the following rule, nth directional multi-direction matrix A_n is given by

$$A_n = T^- (Cn(Q)) \qquad (2)$$

where $T -$ represents Curvelet inverse transform, n = 1, 2, … 8 and *Hn (Q)* represents the coefficient matrix set.

$$Cn(Q) = \begin{matrix} \{Qsd\}, Qs, d \in Zn \\ 0, \text{otherwise} \end{matrix} \qquad (3)$$

This equation provides eight directional matrices A_n for eight directions. Image size IY.

and nth directional multi-direction matrix A_n are of same size and every point A_n (i, j) represents the direction intensity corresponding to the image pixel IY (i, j) that is observed in the nth direction. If the value of A_n (i, j) is, the higher, then the value of the pixel IY (i, j) will have richer feature details in the nth direction.

According to the direction partition rule based on the Fig. 1, the direction for every coefficients set is assigned with the higher limit of the angle interval. Example, if the coefficients set is given by (0°–22.5°), then the direction is assigned to be 22.5°. Similarly for (22.5°–45°) and (157.5°–180°) coefficient set direction is 45° and 0° respectively. Eight directional matrices are obtained by applying the above rule as
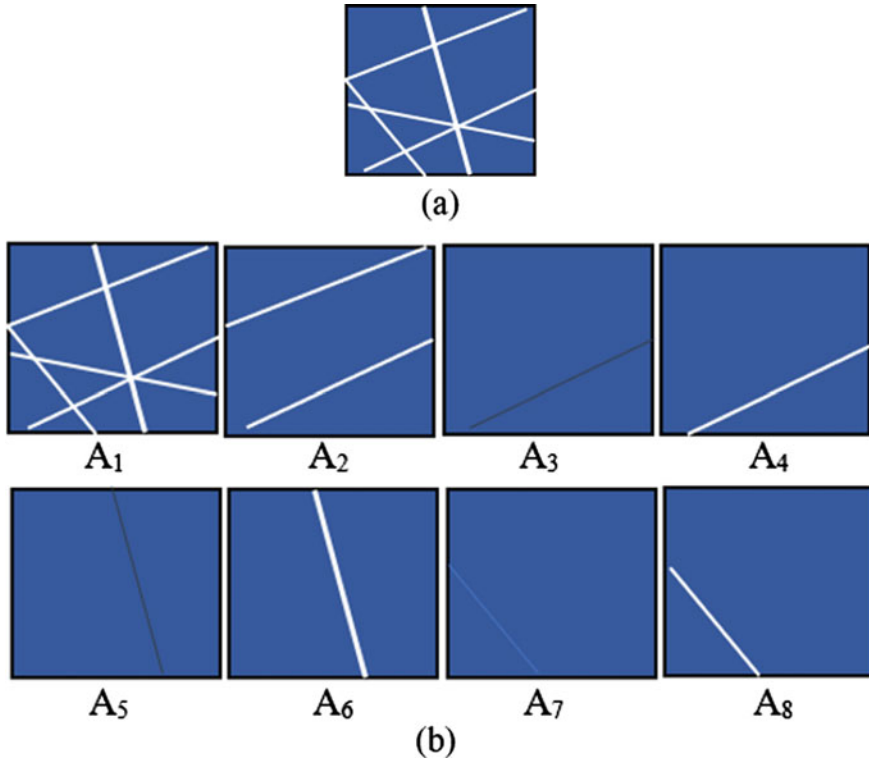
Fig. 3 **a** Original image. **b** The multi-direction matrices A_n of eight directions

illustrated in the Fig. 3b. The multi-direction matrices An of the first row of Fig. 2b illustrates the images obtained through the direction features at 0°, 22.5°, 45° and 67.5° from left to right and second row of illustrates the images obtained through the direction features at, the image at 90°, 112.5°, 135° and 157.5° from left to right respectively. For understanding the direction partition rule, consider the first image A_1, where the angle of the upper white stripe is around 175° but the angle of the lower white stripe is nearly 160°. This image is assigned with direction 0° according to the partition rule and only distinguished direction features are obtained.

The direction features that are extracted from the previous step are applied to the image inpainting technique. Initially the multi directional features are integrated with the color details to obtain the WCDD for estimating the difference between neighboring patches.

Then the color direction structure sparsity is obtained through the patch selection procedure for calculating the filling order. Using WCDD procedure, the best candidate patches are selected and multi directional features are combined with color consistency constraints for the representation of patch sparsity in the optimization equation.

2.2 Distance Measurement

Accurate process of image inpainting can be obtained only if the distance between the two patches is captured correctly. Distance measurement is a very important step because it is used to select the best matching patch as well as to find the filling order through the estimation of CDSS value. These two steps are necessary to perform image inpainting through patch matching process.

To estimate the difference of two patches, the multi color features are combined with the color details. Consider a corrupted image I that has a missing region Ω. If Ψp and Ψq are the two patches that are centred at pixels p and q respectively. Using super wavelet transform, A_n can be estimated for sparse representation coefficients and the preliminary image I_p that is repaired can be obtained by the considering the size of the neighborhood N(p) as 25×25. A local search criterion is used to find the multiple candidate patches. The size of the patch Ψp that is centred at pixels p and the size of the multi directional patch Γp that is centred at pixels p in A_n are same. Similarly the size of the multi directional patch Γq and Ψq are same.

To estimate the difference between the patches Ψp and Ψq, a weight color direction distance (WCDD) is performed which is denoted by dw that is given by the following expression.

$$dw = \eta \,.\, d\,(\Psi p,\ \Psi q,\,) + (1 - \eta)\,.\, d(\Gamma p, \Gamma q,\,) \qquad (4)$$

where η represents the weight coefficient that balances the details of the color and direction features. The function d (Ψp, Ψq) represents the mean square distance between Ψp and Ψq. To estimate WCDD, only the known values of the patches Ψp, Ψq and Γp, Γq are considered in the Eq. (4) which makes the estimation to be precise for measuring the difference of color and direction features and WCDD better.

To illustrate the performance of the proposed WCDD algorithm in image inpainting, the test results were compared with SSD technique as shown in Fig. 4c, d. In this demonstration, all the same procedures were followed except the application of SSD and proposed WCDD algorithm for image inpainting to fill the missing region by the best matching patches. Figure 4c shows some structures in the inpainted image and WCDD method give better results of image inpainting. Similarly, these tests were conducted over wide number of images for analysing the performance of the proposed system with SSD method. Table 1 gives the PSNR value of the five different test images for the SSD and WCDD method. From the results, it is proven that the proposed WCDD algorithm has higher PSNR value and reproduces the image that is more similar to the original image than SSD method. From the results of Table 1, it is inferred that the proposed WCDD method provides better PSNR value and used for filling the filling order through CDSS function.
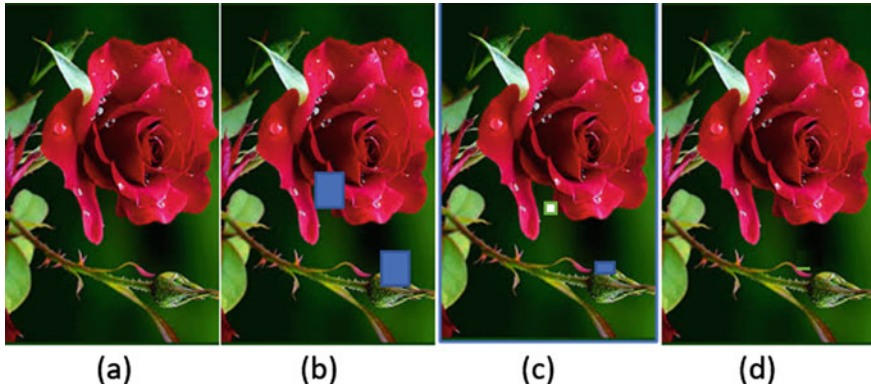
Fig. 4 Analysis of Image inpainting techniques, **a** Original image, **b** corrupted image, **c** SSD, **d** WCDD

Table 1 Analysis of image inpainting method using PSNR

Method	SSD	WCDD
Rose image (Fig. 4)	30.67 dB	38.15 dB
Plane image	33.51 dB	38.97 dB
Duck image	32.45 dB	37.98 dB
Baby image	36.02 dB	38.91 dB
Water falls	34.71 dB	39.12 dB

2.3 Estimation of Filling Order

The structure patch must be filled first and texture of the patches must be preserved whist filling the missing patches from the neighboring pixels. To accomplish this, a sturdy filling order must be calculated which is usually defined by the product of color direction structure sparsity S(p) and the confidence term C(p). This product is termed as the priority function V(p).

$$V(p) = S(p) \cdot C(p) \tag{5}$$

This procedure is performed to make sure that priority of filling the patch is given to the structure patch compared to other patches by S(p) parameter. S(p) parameter is used to preserve the texture and smooth patches in the patch filling.

A. Criminisi et.al [16, 17] defined the confidence term C(p) as a degree of reliable details neighboring the pixel in an image. Z. Xu and J. Sun [20] measured the confidence of a patch present at the structure by introducing the structure sparsity. This produced a better filling order as that of the method explained in [16, 17, 21, 22]. But few constraints were identified like patches in complex structural and

textural regions or the missing regions with less or absence of nearby patch details or consistent confidence value is not provided by the structure sparsity.

To overcome the above limitations, a color direction structure sparsity (CDSS) based on weighted color-direction distance (WCDD) is used to isolate the structures and the textures of an image. A weighted color-direction similarity function built on WCDD is given by the following expression

$$X_{p,k} = \frac{1}{Z(p)} \exp\left(-\frac{d_w}{\sigma^2}\right) \tag{6}$$

where Z(p) is a normalization factor and summation of $X_{p,k}$ will be equal to 1, σ = 5 and d_w value is obtained from Eq. 4. The similarity function gives the similarity measures between two patches that are present in the known and missing regions. The color-direction structure sparsity (CDSS) S(p) is expressed by the following way.

$$S(p) = (1 - \theta)\sqrt{\sum_k X_{p,k}^2} + \theta \tag{7}$$

where θ is a correcting factor that can take up the value between 0 and 1. In the proposed system, value of θ is assigned as 0.21 after evaluating using different values. θ is mainly used to rise the minimum value of S(p) to fill the target patches exactly as texture patch propagates poorly. CDSS uses color and MDS details to estimate the similarity of the patches like the structure sparsity techniques and also more accurate results are obtained.

3 Experimental Results

Various experiments were conducted to analyse the performance of the proposed method. Initially the performance of image inpainting is analysed for different super wavelet transforms like Curvelet and CT for four and eight direction features as shown in the Fig. 5. From the results, it has been observed that the eight direction curvelet transform provides higher PSNR value compared to other methods.

The proposed method is also applied for the application of block removal. Figure 6 is used to illustrate the results of image inpainting technique for the purpose of object removal. The image is artificially corrupted by introducing objects or blocks in the original image.

In this study, about 7 methods are used to remove the objects. But in the Fig. 6c, d the reconstructed image from Exempler based image inpainting and the proposed method is alone showcased. IN the analysis of output produced, it was found that the TV-P method, sparsity method, Xu method, Wangs method produced the images
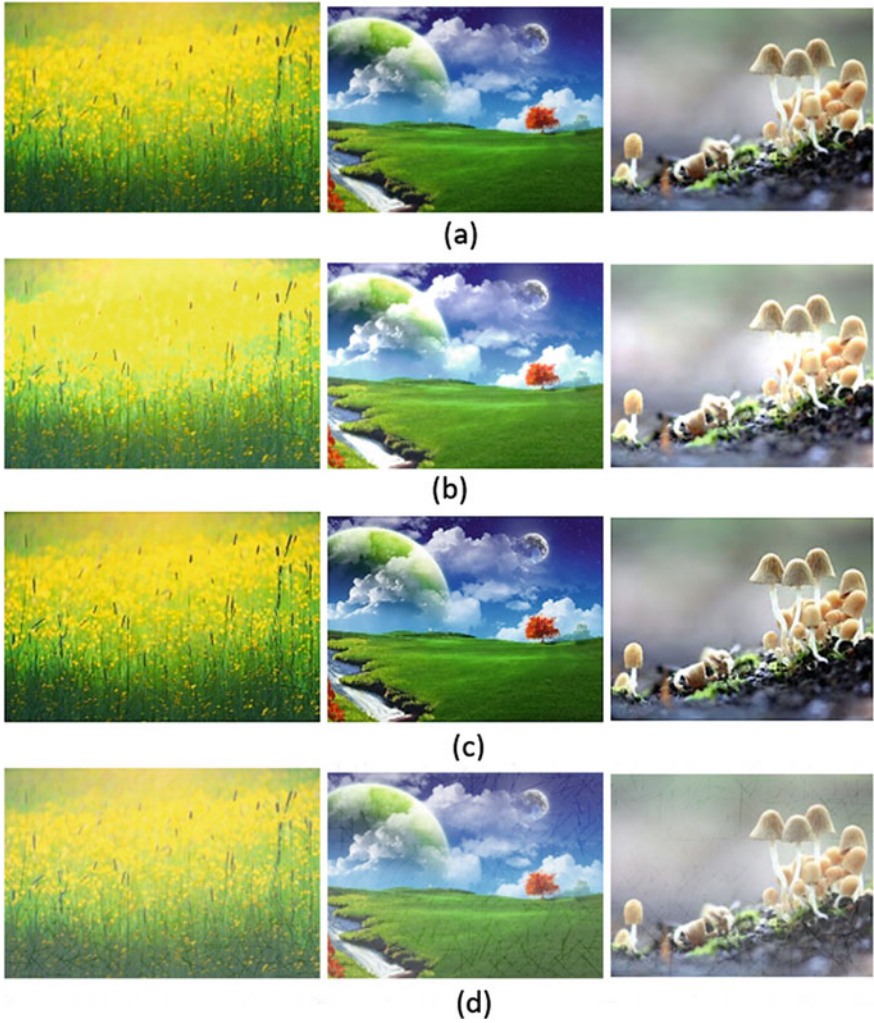
Fig. 5 Comparative results of image inpainting by wavelet transforms with directions. **a** Four direction feature extracted by Curvelet Transform and. **b** Contourlet transform –CT. **c** Eight direction feature extracted by Curvelet Transform and. **d** CT

which has the traces of objects in the restored images. But the proposed method, produced an image similar to the original image with higher PSNR values.
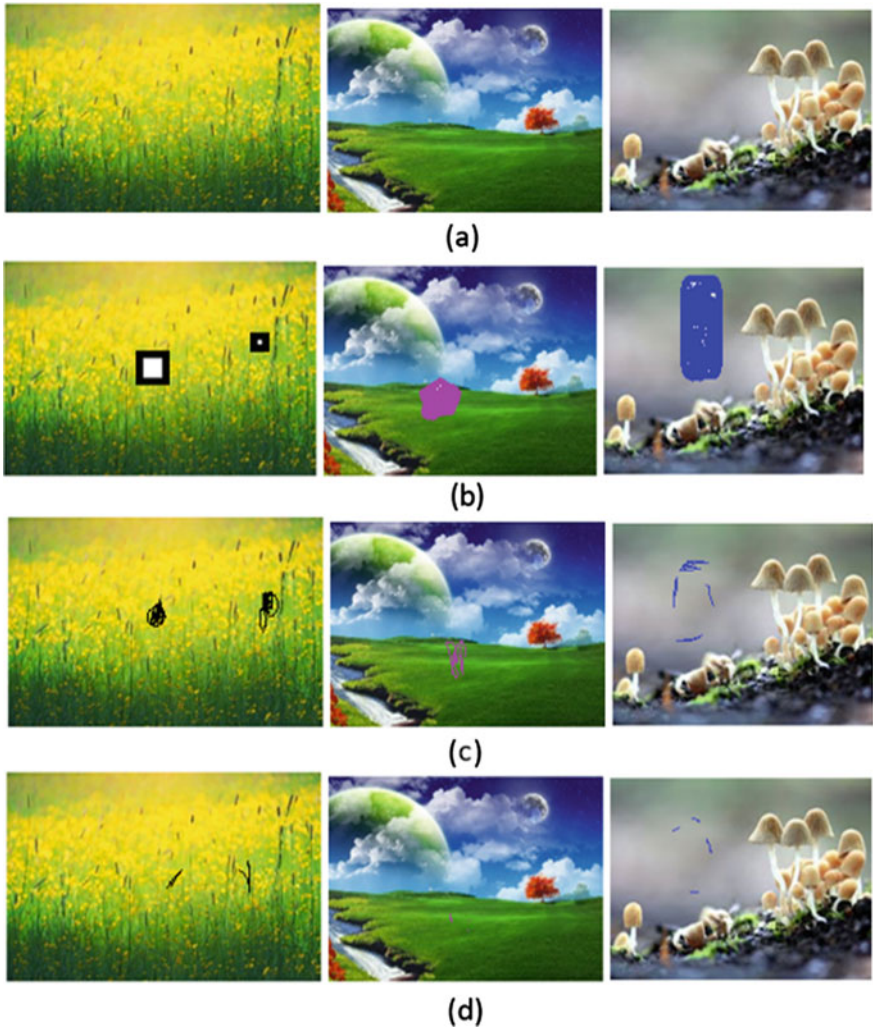
Fig. 6 Analysis of results of image inpainting techniques for object removal. **a** Original image. **b** corrupted image. **c** Exempler method [23]. **d** Proposed method

4 Conclusion

An image inpainting algorithm is proposed in this article to preserve the structural and textural coherence characteristics along with the neighboring consistency through process of adopting color and multi directional features in the process of image inpainting. The proposed method can be used for various applications like object removal, text and scratch removal. First, the multidirectional features are introduced in the process of image inpainting and WCDD is estimated to measure the distance

and similarity of the patches. Then to find the filling order of the patches, CDSS which is a function of structure sparsity is used. To achieve sharp inpainting results, limitations of the color and multi-direction features are combined in the optimization measure. It has been proved that the proposed method provides better structural coherence over the other existing inpainting techniques.

References

1. Bertalmio M, Sapiro G, Caselles V, Ballester C (2000) Image inpainting. In: Proceeding 27th Annual Conference ACM SIGGRAPH, pp 417–424
2. Bertalmio M, Bertozzi AL, Sapiro G (Dec 2001) Navier–Stokes, fluid dynamics, and image and video inpainting. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp I-355–I-362
3. Chan TF, Shen J (2001) Mathematical models for local nontexture inpaintings. SIAM J Appl Math 62(3):1019–1043
4. Chan TF, Shen J (2001) Nontexture inpainting by curvature-driven diffusions. J Vis Commun Image Represent 12(4):436–449
5. Zhang HY, Wu B, Peng QC, Wu YD (2007) Digital image inpainting based on p-harmonic energy minimization. Chin J Electron 3(3):525–530
6. Telea A (2004) An image inpainting technique based on the fast marching method. J. Graph. Tools 9(1):23–34
7. Weickert J, Welk M (2006) Tensor field interpolation with PDEs. Visualization and processing of tensor fields. Springer, Berlin, Germany, pp 315–325
8. Takeda H, Farsiu S, Milanfar P (2007) Kernel regression for image processing and reconstruction. IEEE Trans Image Process 16(2):349–366
9. Miyoun J, Bresson X, Chan TF, Vese LA (2011) Nonlocal Mumford-Shah regularizers for color image restoration. IEEE Trans Image Process 20(6):1583–1598
10. Wen Y-W, Chan RH, Yip AM (2012) A primal–dual method for total-variation-based wavelet domain inpainting. IEEE Trans Image Process 21(1):106–114
11. Mainberger M et al (2012) Optimising spatial and tonal data for homogeneous diffusion inpainting. Scale space and variational methods in computer vision. Springer, Berlin, Germany, pp 26–37
12. Yu G, Sapiro G, Mallat S (2012) Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity. IEEE Trans Image Process 21(5):2481–2499
13. Elad M, Starck J-L, Querre P, Donoho DL (2005) Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). Appl Comput Harmonic Anal 19(3):340–358. 9
14. Guleryuz OG (2006) Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising—Part I: theory. IEEE Trans Image Process 15(3):539–554, 11 Mar 2006
15. Fadili MJ, Starck JL (Sep 2005) EM algorithm for sparse representation- based image inpainting. In: Proceedings IEEE International Conference Image Processing, pp II-61–II-64. 10
16. Criminisi A, Perez P, Toyama K (Jun 2003) "Object removal by exemplar- based inpainting. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp II-721–II-728
17. Criminisi A, Perez P, Toyama K (2004) Region filling and object removal by exemplar-based image inpainting. IEEE Trans Image Process 13(9):1200–1212

18. Wu J-Y, Ruan Q-Q (2009) A novel exemplar-based image completion model. J Inf Sci Eng 25(2):481–497
19. Do MN, Vetterli M (2005) The contourlet transform: An efficient directional multiresolution image representation. IEEE Trans Image Process 14(12):2091–2106
20. Xu Z, Sun J (May 2010) Image inpainting by patch propagation using patch sparsity. IEEE Trans Image Process 19(5):1153–1165, CT
21. Guleryuz OG (2006) "Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising—Part II: adaptive algorithms. IEEE Trans Image Process 15(3):555–571, 12 Mar 2006
22. Wong A, Orchard J (Oct 2008) A nonlocal-means approach to exemplar-based inpainting. In Proceeding 15th IEEE International Conference Image Processing, pp 2600–2603
23. Zhang Q, Lin J (2012) Exemplar-based image inpainting using color distribution analysis. J Inf Sci Eng 28(4):641–654

Classification of Sub-frequency Bands Based Two-Class Motor Imagery Using CNN

Muhammad Ahsan Awais , **Mohd Zuki Yusoff** , **and Norashikin Yahya**

Abstract EEG has been primarily used in both clinical and research applications. Brain-computer system (BCI) is one of the leading EEG research applications that offer special users a new means of communication. Previous studies have reported the occurrence of MI patterns in mu and beta rhythms, but that does not provide in-depth knowledge of the frequency range. This paper focuses on the classification of 2-class Motor Imagery using several frequency sub-bands in the mu and beta range. "EEG motor imagery dataset from the Physionet database," has been used for validation purposes. Although this data includes both imagery and real movements, we have just used the imagination data. Data is collected from 109 healthy subjects, but we have only used the first 15 subjects in the study. The study aims to divide the data into multiple frequency bands to study the motor imagery classification behaviour over different frequencies. Afterward, a CNN-based deep learning model with two convolutional layers has been used to classify the left and right classes for different types of same data. The study seeks to compare the results from various sub-frequency bands.

Keywords EEG · BCI · CNN · Motor imagery

1 Introduction

The number of people with disabilities who have lost their mobility is substantial. In addition to their damaged body, they have a working sound brain in their bodies. The brain should be used to allow them the freedom to move easily by using their minds to power the wheelchair. The brain-computer interface offers a forum for people to create a communication relationship between users and computers [1]. It helps us to interact with the outside world using our brains independently. The BCI framework does not require any actual muscle movement commonly involved in the

M. A. Awais (✉) · M. Z. Yusoff · N. Yahya
Centre for Intelligent Signal and Imaging Research (CISIR), Electrical and Electronic Engineering Department, Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia
e-mail: muhammad_18001588@utp.edu.my

communication process. The worldwide scientific community is designing numerous BCI-based technologies, supporting patients who are unable to speak and provide mobility. Several tools are available for capturing electrical signals that match every person's mental or physical reaction. These techniques include electrocorticography (ECoG), functional near-infrared spectroscopy (fNIRS), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) [2, 3].

Electroencephalography is non-invasive, electrophysiological monitoring that records electrical neural activities of the brain using small metal discs (i.e., electrodes). Most of the brain-computer interface systems incorporate the EEG technology because of its portable, cost-effective and non-invasive nature. Each action or thought emits a special brain signal that can vary between individuals.

Throughout the work, the brain signal undergoes signal treatment (pre-processing) and signal processing (feature extraction and classification) to provide the state machine's feature for several applications such as a wheelchair. The integration of robotic wheelchair and EEG technologies will enhance mobility, boost the quality of lifestyle and reduce caregiver expenses for differently-abled people. To provide basic mobility for the disabled people to move around by themselves, brainwave controlled electrical wheelchair is the best solution. There are many factors like BCI strategy (Motor Imagery, SSVEPs, etc.), EEG acquisition system, feature extraction and classification schemes on which the quality of any BCI wheelchair depends [4].

Significant effort has been made by researchers in the field of EEG based BCI wheelchair on which BCIs are categorized into different types based on the EEG brain activity patterns. These types are P300 component of event-related potentials (ERPs), steady-state visual evoke potentials (SSVEP), Motor Imagery (Event-related desynchronization/synchronization) [5]. For each category, a different type of experimental paradigms is developed and utilized by researchers to collect the EEG data.

The most common and feasible type of EEG datasets for BCI wheelchair lies in the category of motor imagery. Motor imagery can be interpreted as a dynamic state in which a person mentally simulates an action without actually performing it physically [6]. Researchers have collected MI-based EEG data under multiple classes. The data is divided into different classes based on actions (left-hand movement, feet movement, etc.) against which imagery data has been collected.

Studies [7, 8] revealed the existence of MI patterns in the mu (8–13 Hz) and beta band (13–32 Hz). But this does not provide in-depth knowledge about the MI patterns. For this purpose, we have proposed a system in which the mu and beta frequency range (i.e., 8–32 Hz) has been divided into several sub-frequency bands to study the effect of frequency bands on classification accuracy.

2 Methodology

The research will develop a brain-computer interface system which aims to introduce the use of several sub-bands of motor imagery-based EEG data. CNN with frequency sub-bands seeks to achieve better performance, reliability and flexibility.

2.1 Dataset

EEG MI dataset from the Physionet database has been used in the proposed system [9]. In producing these records, the BCI2000 instrumentation system has been used. The data collection comprises above 1500 recordings of one and two minutes from 109 subjects. The dataset incorporates both actual and motor imagination tasks from the subjects. The experimental paradigm comprised 14 runs, where subjects performed different tasks including rest state, the right fist/left fist movement, the imagination of right fist/left fist movement, both fists/ both feet movement, and the imagination of both fists/ both feet movement. The EEG signals were recorded from 64 electrodes as per the international 10–10 system. The data was sampled at 160 Hz for all subjects. In this study, only the MI trials (imagination of left and right fist movement) of the first 15 subjects were used. For motor imagery, subjects were instructed to perform the imagination task as the target stimulus appears on the left side of the screen or the right side of the screen. The subjects imagine the fist to be opened and closed before the target stimulus disappears. Then the subject goes to the rest state. Every subject recorded three sessions for each type of task. The single-session comprises 7–8 random trials of each class, i.e., left or right movement imagery. Each trial is carried out for 4 secs, followed by the rest period of 4 s $\pm$ 5%.

2.2 Pre-processing

It is essential to remove the noise or unwanted components from the signal. The undesired signals in our case can be EMG signals (real hand movement) and the frequency components of the EEG signal other than the motor-imagery task.

The 15 subject's data has been filtered using BrainStorm in MATLAB. Initially, the data has been passed through 3 steps including, DC offset correction, notch filter at 60 Hz in order to eliminate the electrical interference, and the bandpass filter from 8 to 32 Hz, to eliminate other frequencies like EMG (50–250 Hz) and other unwanted signals.

2.3 Sub-band Frequency Division

The pre-processed data from the previous stage has been divided into nine sub-frequency bands using the same platform, i.e., Brainstorm, in order to study the data in detail. The sub frequency bands include (8–12 Hz), (12–16 Hz), (16–20 Hz), (20–24 Hz), (24–28 Hz) and (28–32 Hz), (8–16 Hz), (16–32 Hz) and (8–32 Hz).

2.4 Convolutional Neural Networks (CNN)

In this study, CNN based deep learning approach has been adopted to identify the effect of MI pattern on different frequency ranges.

CNN is a Deep Learning algorithm that can take in any input and process it efficiently to classify the different classes. CNN uses a system that has been designed for reduced processing requirements. The layers of a CNN consist of an input layer, an output layer, and a hidden layer containing multiple convolutional layers, pooling layers, fully connected layers and normalization layers.

The overall 15 subject's data is divided into training (10 subjects) and validation (5 subjects).

The CNN model (Fig. 1) proposed in this research comprises 2 convolutional layers with 5 kernels per layer. 1st layer performs convolution and the output data is passed through batch normalization. ReLU activation function is used in the model and the output is reduced in the max-pooling layer. The same process is applied to the 2nd layer as well. The output of the 2nd layer is passed through the dropout layer in which we have used the default value, i.e., 50%. The data is fed to the fully connected layer after flattening it into 1-D. The output is generated by a softmax layer with many neurons that fit the number of categories in the data to be classified. In our case, the output is classified as the left and right class.

The CNN model is composed of 2 convolutional layers and the model is trained for 10 epochs. The initial learning rate is set to 0.001, while the validation frequency is 100.

The stated CNN model has been used for two different cases. In case 1, the model has been implemented on all the 64 channels using 9 sub-frequency bands stated in Sect. 80.3. Whereas in Case 2, all 9 sub frequency bands have been implemented using only 3 channels (i.e., C3, Cz and C4) from the central region of the brain.

3 Result and Discussion

The CNN model has been implemented on the basis of two cases. Table 1 illustrates the results derived from both cases (i.e., use of 64 channels data and the use of only 3 EEG channels). The accuracy of the model is calculated over 10 epochs.

Fig. 1 Block diagram of the proposed CNN model
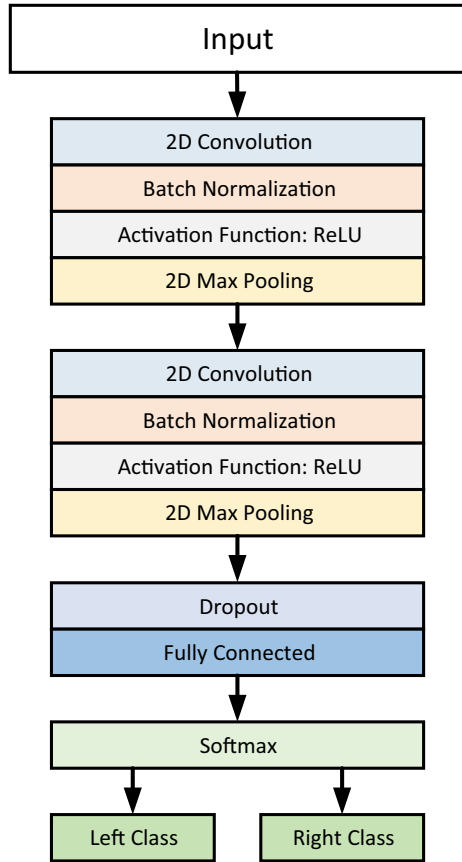
Table 1 Average validation accuracy against sub-frequency bands

Frequency Band (Hz)	Accuracy (64 channels) (%)	Accuracy (3 channels) (%)
08–16	73.29	73.03
08–32	65.97	68.46
16–32	61.11	59.91
08–12	77.93	77.82
12–16	76.88	75.39
16–20	61.92	61.01
20–24	63.21	61.24
24–28	63.33	61.95
28–32	62.08	62.27
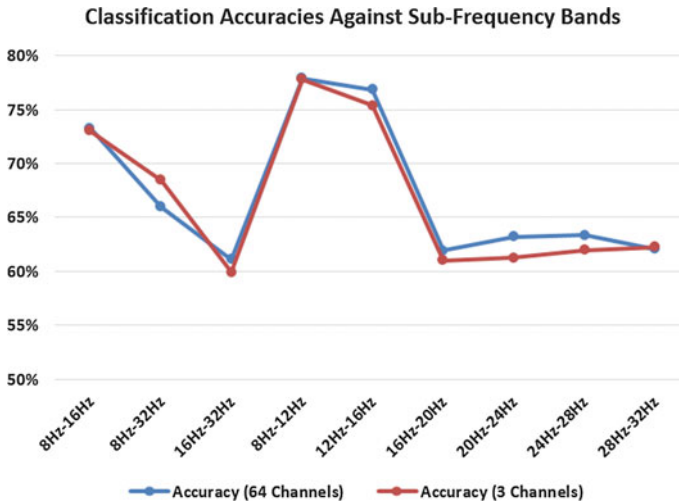
Fig. 2 Classification comparison of different sub-frequency bands for both cases

In case 1, the maximum accuracy, i.e., 77.93%, was reported by the 8–12 Hz frequency band. In contrast, the frequency band 16–32 Hz resulted in the lowest validation accuracy, i.e., 61.11%. Whereas in case 2, the maximum accuracy, i.e., 77.82% has been reported by 8–12 Hz. On the other hand, the minimum validation accuracy for case 2 has been obtained by the sub-band 16–20 Hz. Figure 2 provides a graphical representation of the classification comparison of cases 1 and 2.

Among 9 different frequency bands, the 8–16 Hz main band and its two sub-bands, including 8–12 Hz and 12–16 Hz, has shown better results for 2-class MI prediction. Figure 2 also demonstrates that frequency bands ranging from 16–32 Hz, including sub-bands and the main band, has given the lowest classification accuracies. Thus, it shows that there is no effect of dividing the frequency if the main band is not giving a better performance. One of the major limitations of getting low classification accuracies might be the use of a small dataset. Deep learning algorithms are enhancing efficiency by reducing complexity. But still, it lacks the ability to handle the small datasets with the same level of efficiency being addressed in the case of large EEG datasets.

4 Conclusion

The objective of this research is the analysis of sub-frequency bands for the development of the BCI system using 2-class motor imagery. CNN has been used for the classification of two classes using several sub-frequency bands for 15 subjects' data. The maximum classification accuracy does not exceed 77.93% as the data is considered to be a small dataset in order to be processed in a deep learning algorithm. In

the future, we will be using a sufficient dataset for deep learning algorithms in order to achieve the maximum classification results effectively.

References

1. Abiri R, Borhani S, Sellers EW, Jiang Y, Zhao X (2019) A comprehensive review of EEG-based brain-computer interface paradigms. J Neural Eng 16(1):011001
2. Biasiucci A, Franceschiello B, Murray MM (2019) Electroencephalography. Curr Biol 29(3):R80–R85
3. Awais MA, Yusoff MZ, Yahya N, Qamar MU(2020) Brain controlled wheelchair: a smart prototype. In: J Phys Conf Series 1529:042075
4. Andrade MK et al.,(2020) An EEG brain-computer interface to classify motor imagery signals. In: Biomedical signal processing. Springer, pp 83–98
5. Aoh Y et al (2019) Event-related desynchronization/synchronization in Spinocerebellar Ataxia Type 3. Front Neurol 10:822
6. Padfield N, Zabalza J, Zhao H, Masero V, Ren J (2019) EEG-based brain-computer interfaces using motor-imagery: techniques and challenges. Sensors 19(6)
7. Tacchino G, Coelli S, Reali P, Galli M, Bianchi AM (2020) Bicoherence interpretation, in EEG, requires signal to noise ratio quantification: an application to sensorimotor rhythms. IEEE Transactions on Biomedical Engineering
8. Tariq M, Trivailo PM, Simic M (2020) Mu-Beta event-related (de) synchronization and EEG classification of left-right foot dorsiflexion kinaesthetic motor imagery for BCI. Plos one 15(3):e0230184
9. EEG Motor Movement/Imagery Dataset https:/www.physionet.org/physiobank/database/eeg mmidb/. Last Accessed 25 Oct 2020

Prediction of Methane Hydrate Formation Rate in Multiphase System using Artificial Neural Network

Wan Adli Nuaim Bin Wan Ahmad Nadzri, Omar Nashed, Bhajan Lal, Khor Siak Foo, and Khalik Mohd Sabil

Abstract The research of the hydrate formation has advanced over the last few decades. Several models have been developed to understand the hydrates formation kinetics and conditions. In this study, computer based model is used to predict the gas hydrate formation rate. Since the hydrae formation is stochastic phenomenon and it is common to get inconsistent data, Artificial Neural Network (ANN) model has potential to outstand other conventional kinetic models. ANN used to predict the methane hydrate formation rate in multiphase system. The liquid phase composed of water + drilling oil + nonionic surfactants was used to form methane hydrates at pressure 8.80 MPa and temperature of 274.15–277.15 K. This research would essentially assess the effectiveness of ANN model for the kinetic modeling of the formation of gas hydrate from the acquired regression analysis. The result of this research revealed that ANN model with 16 number of hidden neurons had a better prediction as the highest regression value R was found to be 0.9956.

Keywords Gas hydrate · Artificial neural network · Formation rate

W. A. N. B. W. A. Nadzri · B. Lal (✉) · K. S. Foo
Chemical Engineering Department, Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Perak, Malaysia
e-mail: bhajan.lal@utp.edu.my

O. Nashed
Petroleum Engineering Department, College of Energy Technologies, Jikharrah, Libya

K. S. Foo
PTTEP, Level 26–30, Tower 2, Petronas Twin Towers, Kuala Lumpur City Centre, 50088 Kuala Lumpur, Malaysia

K. M. Sabil
PETRONAS Research Sdn Bhd, Kawasan Institusi Bangi, Malaysia, Jalan Ayer Itam, 43000 Bandar Baru Bangi, Selangor, Malaysia

1   Introduction

Gas hydrates are an ice-like inclusion compound, which mainly consist of low molecular weight gas molecules entrapped into hydrogen bonded water cages [1, 2]. Gas hydrate formation in gas production may hinder the flow and eventually lead to serious operational problem and huge financial loss. Despite imposing negative impacts in the oil and gas industries, gas hydrate has some potential application in some industries [3, 4]. Gas hydrate-based technology could be applied for water desalination, gas separation, transportation, and storage [5]. Not to forget that, gas hydrates are available naturally at deep marine and permafrost area [6, 7].

Several modeling studies available are focusing mainly on the thermodynamics properties of gas hydrate, e.g. phase equilibria and molecular structures [8]. However, less papers are available in the literature discussing and predicting the time-dependent behaviour of gas hydrate formation and dissociation [9]. This is due to the naturally occurring intrinsic stochastic phenomena of gas hydrate which lead to difficulty in developing a well-established model [2, 10, 11]. Some of the hydrate growth kinetic models available in the literature include reaction kinetics based model [12], mass transfer based model [13, 14], combination of reaction kinetics and mass transfer [15, 16], heat transfer based model [17] as well as model based on combination of heat transfer and reaction kinetic [18]. Gas hydrate formation kinetics are very complex process, and it is almost difficult to display a definite outcome from the kinetic model as hydrates form are very uncertain and formation is based on various driving forces [19]. In addition, the gas hydrate growth rate depends on the phases composition as well as testing conditions. Recently, artificial neural network (ANN) has attracted a massive attention due to its ability to deal with scattered and big data to solve complex problems. ANN is a technique used nowadays for prediction of data set which use the structure and function of neuron [9, 20–22]. In gas hydrate studies, most of the papers utilizing ANN focusing on the prediction of the phase equilibrium. However, less research conducted on the kinetic of the hydrate formation.

Zahedi et al. used two different methods with 203 experimental data points from literature to predict the hydrate formation temperature. It was found that the best predictor of hydrate formation temperature was MLP with seven neurons [21]. Moreover, Chapoy et al. estimated hydrate dissociation pressure in the presence of inhibitor. In their studies, 19 input variables with 35 neurons in one hidden layer were trained, but various numbers of hidden layers and neurons have not been examined to find the best structure in the neural network [23].

In this work, ANN was employed to predict methane hydrates formation rates in multiphase system in the presence of nonionic surfactants. The HLB values were used to identify the surfactant and used as input along with temperature and surfactants' concentration.

2 Methodology

The data used in this work were collected from experimental study of the gas hydrate formation kinetics in multiphase system (water + oil + gas) by using 6 different chemicals: Span 20, Span 40, Span 80, Tween 20, Tween 40, and Tween 80. The experimental data is presented in Table 1. The network was trained to minimize the errors between the desired target values and the values computed from the model based on iterative gradient using MATLAB. The inputs are the types of non-ionic surfactant represented by Hydrophilic-Lipophilic Balance (HLB), temperature, and concentration. Meanwhile, the output is the rate of CH_4 hydrate formation. The data set were split into two subset which were for training as well as for validation/testing. By default, the percentage of each set were 70% training, 15% testing and 15% validation. For the training algorithm, it was selected according to the quantity of the data set so that the network could simulate a task with minor error. Usually, the Levenberg–Marquardt algorithm is the most used algorithm to automatically adjusted the biases and weights of the neural network [9].

However, for this model, Bayesian Regularization algorithm was selected because this algorithm could result in good generalization for difficult, small, and noisy

Table 1 Raw experimental data from laboratory

Sample	HLB	Concentration (wt%)	Temperature (°C)	Rate of gas hydrate formation (mol/min)
Sp20	8.6	1	4	0.021247
Sp20	8.6	2	4	0.016504
Sp40	6.7	1	4	0.030413
Sp40	6.7	2	4	0.030926
Sp80	4.3	1	4	0.006789
Sp80	4.3	2	4	0.016452
Tw20	16.7	1	4	0.004801
Tw20	16.7	2	4	0.004642
Tw40	15.6	1	4	0.004651
Tw40	15.6	2	4	0.004477
Tw80	15	1	4	0.005277
Tw80	15	1.5	4	0.004744
Tw80	15	2	4	0.004837
Tw80	15	2.5	4	0.004719
Tw80	15	3	4	0.004433
Tw 80	15	2.5	1	0.001705
Tw80	15	2.5	2	0.002763
Tw80	15	2.5	3	0.003638
Tw80	15	2.5	4	0.004719

datasets which suits our input data sets. Next, the number of hidden layers were manipulated to get the least error. The number of hidden layers were determined through a process of trial an error [22]. In this paper, the number of hidden layers were decided through the optimization of regression. A correlation coefficient (R) was chosen to evaluate the prediction as 1 means a perfect correlation while 0 indicates for random correlation.

3 Result and Discussion

Dividing the data to training, validation, and testing data set was the first step for applying ANN prediction. This ANN model was built from 19 number of data of six chemicals, each taken from different temperature, HLB value and concentration. In ANN model, the number of hidden neurons were manipulated starting from 1 to 20 hidden neurons. Trial-and-error results showed that optimal hidden neurons 16. Figure 1 illustrates the ANN architecture with respect to performance function.

Regression analysis was carried out to determine the correlation between the experimental and predicted data by the ANN models. The highest R is found to be 0.9956 for the ANN model with 16 number of neurons. It is evident there is excellent agreement between the predicted and experimental data for the same training conditions as shown in Fig. 2. However, ANN model used in this work could have insufficient number of data with only 3 inputs. Additional data might be required to get more practical ANN model to predict the gas hydrate formation rate. Additionally, another limitation of ANN model prediction is that the model valid only within the studied range of HLB, concentration and temperature for the prediction of gas hydrate formation rate. Consequently, ANN model is not applicable for extrapolation.

Lastly, it is worth mentioning that using too many hidden neurons increases the complexity of the model which lead to over-fitting. In other hands, different problem occurs when the network is assumed to has very few hidden neurons causing the network unable to fit or in other words, under-fitting. The number of hidden neurons in the hidden layers determines to a network's capability to compute the output results.
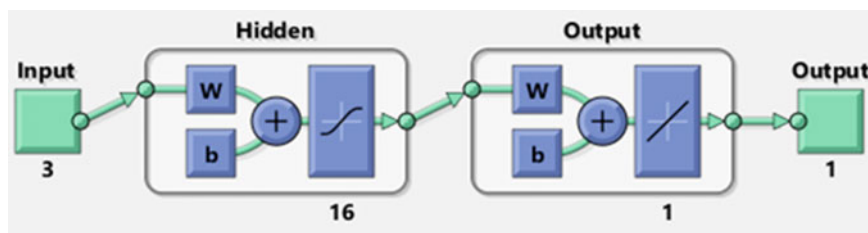
Fig. 1 Multilayer Perceptron with 16 hidden neurons
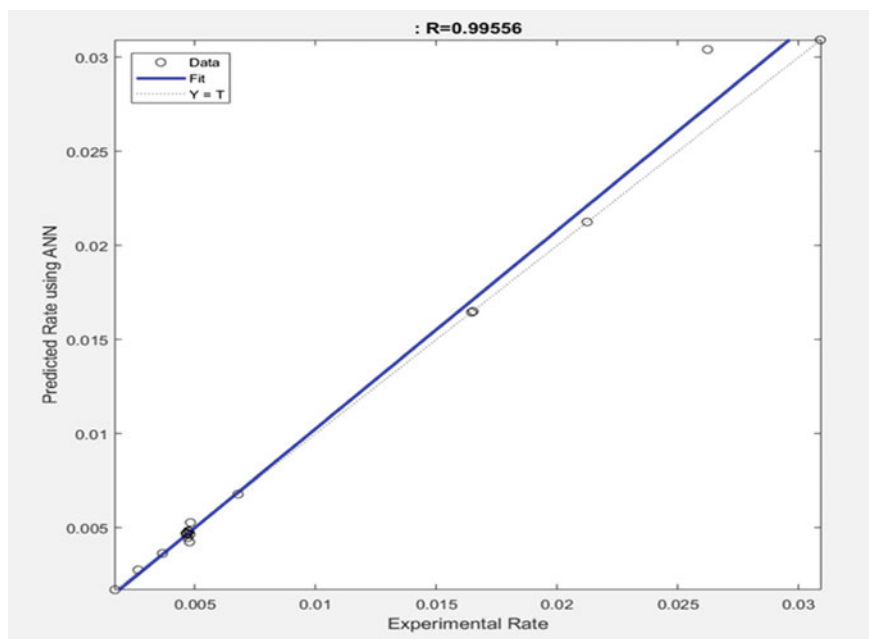
Fig. 2 Correlation coefficient between experimental and predicted data by ANN with 16 number of neurons

4 Conclusion

ANN was used to perform the numerical study of kinetic behavior of methane gas hydrate formation rate in the presence of nonionic surfactants. It was found that the predicted results by ANN with 16 number hidden neurons is the closest to the experimental data. The correlation coefficient value for 16 hidden neurons is 0.99556. Therefore, ANN could be an efficient method for predicting the hydrate formation rate.

References

1. Bavoh CB, Lal B, Keong LK (2020) Introduction to gas hydrates, in chemical additives for gas hydrates. In: Lal B, Nashed O (eds). Cham, Springer International Publishing, pp 1–25. https://doi.org/10.1007/978-3-030-30750-9_1
2. Yin Z, Khurana M, Tan HK, Linga P (2018) A review of gas hydrate growth kinetic models. Chem Eng J 342:9–29. https://doi.org/10.1016/j.cej.2018.01.120
3. Yin Z, Chong ZR, Tan HK, Linga P (2016) Review of gas hydrate dissociation kinetic models for energy recovery. J Nat Gas Sci Eng 35:1362–1387. https://doi.org/10.1016/j.jngse.2016.04.050

4. Yaqub S, Lal B, Mellon NB, Sufian BS (2018) Effect of the natural green materials on methane hydrate formation kinetics. IOP Conf Ser: Mater Sci Eng, 458:012074. https://doi.org/10.1088/1757-899x/458/1/012074

5. Nashed O, Lal B, Shariff AM, Sabil KM (2020) Gas hydrate promoters. In: Lal B, Nashed O (eds), Chemical additives for gas hydrates. Cham, Springer International Publishing, pp 47–65. https://doi.org/10.1007/978-3-030-30750-9_3

6. Khan MS, Lal B, Bustam MA (2020) Gas hydrate inhibitors. In: Lal B, Nashed O (eds), Chemical additives for gas hydrates. Cham, Springer International Publishing, pp 27–46. https://doi.org/10.1007/978-3-030-30750-9_2

7. Yaqub S, Lal B, Shariff AM, Mellon NB (2019) Unraveling the effect of sub-cooling temperatures on the kinetic performance of biopolymers for methane hydrate. J Nat Gas Sci Eng 65:68–81. https://doi.org/10.1016/j.jngse.2019.03.002

8. Manakov AY, Penkov NV, Rodionova TV, Nesterov AN, Fesenko EE Jr (2017) Kinetics of formation and dissociation of gas hydrates. Russ Chem Rev 86:845–869. https://doi.org/10.1070/rcr4720

9. Foroozesh J, Khosravani A, Mohsenzadeh A, Mesbahi AH (2014) Application of artificial intelligence (AI) in kinetic modeling of methane gas hydrate formation. J Taiwan Inst Chem Eng 45:2258–2264. https://doi.org/10.1016/j.jtice.2014.08.001

10. Naeiji P, Varaminian F, Rahmati M (2019) The kinetic modeling of methane hydrate growth by using molecular dynamic simulations. Int J Heat Mass Transf 142:118356. https://doi.org/10.1016/j.ijheatmasstransfer.2019.07.006

11. Sayani JK, Pedapati SR, Lal B (2020) Phase behavior study on gas hydrates formation in gas dominant multiphase pipelines with crude oil and high CO_2 mixed gas. Sci Rep 10:1–2. https://doi.org/10.1038/s41598-020-71509-6

12. Vysniauskas A, Bishnoi PR (1985) Kinetics of ethane hydrate formation. Chem Eng Sci 40:299–303. https://doi.org/10.1016/0009-2509(85)80070-1

13. Englezos P, Kalogerakis N, Dholabhai PD, Bishnoi PR (1987) Kinetics of formation of methane and ethane gas hydrates, vol 42, pp 2647–2658. https://doi.org/10.1016/0009-2509(87)87015-X

14. Skovborg P, Rasmussen P (1994) A mass transport limited model for the growth of methane and ethane gas hydrates. Chem Eng Sci 49:1131–1143. https://doi.org/10.1016/0009-2509(94)85085-2

15. Hashemi S, Macchi A, Servio P (2007) Gas hydrate growth model in a semibatch stirred tank reactor. Ind Eng Chem Res 46:5907–5912. https://doi.org/10.1021/ie061048+

16. Henning RW, Schultz AJ, Thieu V, Halpern Y (2000) Neutron diffraction studies of CO_2 clathrate hydrate: formation from deuterated ice. J Phys Chem A, Article 104:5066–5071. https://doi.org/10.1021/jp0001642

17. Freer EM, Selim MS, Sloan ED Jr (2001) Methane hydrate film growth kinetics. Fluid Phase Equilib 185:65–75. https://doi.org/10.1016/S0378-3812(01)00457-5

18. Bollavaram P, Devarakonda S, Selim MS, Sloan ED Jr (2000) Growth kinetics of single crystal sII hydrates. Elimination of mass and heat transfer effects, Annals of the New York Academy of Sciences 912:533–543. https://doi.org/10.1111/j.1749-6632.2000.tb06808.x

19. Partoon B, Sahith J, Lal B, and Maulud ASB (2020) Gas hydrate models. In: Lal B, Nashed O (eds), Chemical additives for gas hydrates. Cham: Springer International Publishing, pp 67–85. https://doi.org/10.1007/978-3-030-30750-9_4

20. Sahith J, Pedapati SR, Lal B (2019) Application of artificial neural networks on measurement of gas hydrates in pipelines. Test Eng Manag 81:5769–5774

21. Zahedi G, Karami Z, Yaghoobi H (2009) Prediction of hydrate formation temperature by both statistical models and artificial neural network approaches. Energy Convers Manage 50:2052–2059. https://doi.org/10.1016/j.enconman.2009.04.005

22. Ghavipour M, Ghavipour M, Chitsazan M, Najibi SH, Ghidary SS (2013) Experimental study of natural gas hydrates and a novel use of neural network to predict hydrate formation conditions. Chem Eng Res Des 91:264–273. https://doi.org/10.1016/j.cherd.2012.08.010
23. Chapoy A, Mohammadi AH, Richon D (2007) Predicting the hydrate stability zones of natural gases using artificial neural networks. Oil Gas Sci Technol 62:701–706. https://doi.org/10.2516/ogst:2007048

Chemometrics Analysis and Wavelength Biomarker Identification Using Fourier Transform Infrared Spectroscopy for Lard Adulteration

Muhammad Aadil Siddiqui, Mohd Haris Md Khir, Gunawan Witjaksono, Muhammad Junaid, Saeed Ahmed Magsi, and Abdul Saboor

Abstract Food authenticity is a major concern that is closely linked to public health and safety in the period in which the long food supply chain is the norm. One of the key problems facing customers in meat markets is the detection of non-halal meat. While several studies have been carried out to identify biomarkers for adulterated meats such as lard, still these studies are at an early stage, and there is no authenticated biomarker available which can produce specific identification as a result. The purpose of this research is to utilize the technique of Fourier Transform Infrared Spectroscopy (FTIR) to analyze pig, cow, lamb, and chicken fat to identify potential biomarkers for the identification of lard. FTIR studies have shown that chicken and lard fat have particular peaks relative to lamb and beef fats at wavenumbers 1600, 1750, 2750 and 2920 cm^{-1}. Furthermore, PCA shows promising clustering and correlation between the species which can lead to development of a model for successful adulteration detection.

Keywords Near infrared spectroscopy · FTIR · Lard · Halal · Chemometric

1 Introduction

In the food industry around the world, food authenticity is a significant concern. Food fraud is still a problem because of the glut of packaged food with a long supply chain on the market. First coined by Spink and Moyer which stated that food fraud is defined as alteration of the true food labeling ingredients in which costly, fewer accessible raw materials are substituted by cheaper alternatives [1]. Some of these adulterations may impact the public mass financially only, but others may be more seriously affected by some adulteration, i.e. food allergy [2], food poisoning, religious views [3], etc. Although food tampering is not a recent issue, some of these tampering was quite poisonous. For instance, incidences such as addition of sawdust to make white bread

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022
R. Ibrahim et al. (eds.), *International Conference on Artificial Intelligence for Smart Community*, Lecture Notes in Electrical Engineering 758,
https://doi.org/10.1007/978-981-16-2183-3_82

[4], formula milk adulteration with melamine [5], adulteration of consumable oil with non-consumable [6]. Not only dangerous to health, but some incidences also involved misrepresentation of food ingredients, such as adulteration of meat with horse meat in 2013 in the UK [7]. Because the price difference between pork and other red meats, halal and kosher consumers are always concerned with the authentication of halal and kosher foods [8]. Global halal market is worth billions, and it is projected to increase to Trillions by 2021 [9]. Therefore, the industry is very interested to strengthen consumer's trust in the halal brand. Market confidence is necessary to protect this [10]. Several authentication techniques can be employed to ensure halal and kosher brand food products adhere to the standards, such as DNA based methods [11]. Majority of approaches still require considerable time for preparing sample or getting vague results due to extreme sensitivity to adulterants which results in overall poor accuracy. Nowadays, spectroscopy based on vibrations is popularized due to its effectiveness for adulteration detection in various food [12]. This is because the simplicity of this nondestructive method reduces time and cost for overall experiment. FTIR spectroscopy can separate, easily and with high accuracy. For instance, pork meat in meatball broth [13], chocolate [14], vegetable oils [15], etc. Some research starts to veer to Near Infrared (NIR) spectroscopy mainly because the feasibility of this would open possibilities to make food authentication instrumentation setup to be portable [16]. However, the research done on this reported lower accuracy in discrimination of adulterants. Although FTIR is quick and relatively inexpensive with easier sample preparation and non-destructive process [17], it is still limited in portability due to its bulky equipment. The aim of this research paper is to improve the accuracy of NIR spectrometer by means of optimizing post-processing analysis, and to quantify the level of lard adulteration in mixtures.

1.1 Food Authentication Based on Vibrational Spectroscopy

In food safety, the production of food authentication is important since this is the mechanism that will verify whether a certain food follows its label definition. [18]. This can include the origin of the food (species, gene, geographical), production method (conventional, organic, free range, etc.), and processing method (frozen, etc.). Vibrational spectroscopy is based on the idea that inside molecules, atom-to-atom bonds vibrate at frequencies that can be represented and are therefore subject to measurement by the laws of physics. In molecules, during the compression step, the respective electron clouds of the two bound atoms, as well as the charges on the nuclei, limit the approach of the nuclei, forming an energy barrier. The bond will inevitably split at the extension of the stretch when the vibrational energy exceeds the energy of dissociation. Table 1 shows the different stretching frequencies.

Table 1 Important IR stretching corresponding to their respective frequency and bond [19]

Chemical bond	Spectrum location (Cm^{-1})	Absorbance strength
$C \equiv N$	2260–2220	Medium
$C \equiv C$	2260–2100	Medium to weak
$C = C$	1680–1600	Medium
$C = N$	1650–1550	Medium
⬡	1500–1430	Strong to weak
$C = O$	1780–1650	Strong
$C - O$	1250–1050	Strong
$C - N$	1230–1020	Medium
$O - H$ (alcohol)	3650–3200	Strong
$O - H$ (carboxylic acid)	3300–2500	Strong
$N - H$	3500–3300	Medium, broad
$N - H$	3300–2700	Medium

1.2 Chemometrics

The use of statistical transformation techniques such as Principal Component Analysis (PCA), a classical statistical approach for translating a dataset's attributes into a new collection of uncorrelated attributes called main components (PCs), is the basis of chemometrics techniques. The principle is that this allows the data to be reduced to a smaller number of dimensions, with low loss of information, simply by discarding some PCs. Each PC is a linear combination of the original inputs and each PC is orthogonal, reducing the collinearity problem. Chemometrics is based on Beer's law that directly relates the concentration of a sample with the intensity in absorbance. Beer's law states that the absorbance of a key band of an analyte is proportional to its concentration. The law can be expressed as.

$$A = \varepsilon \, cd \tag{1}$$

where A is the spectra observances, ε is the molar absorption coefficient in lt/mol cm, c is the concentration in dm^2/mol, and d is the cell thickness in cm.

2   Materials and Methods

2.1   Meat Sample

Samples were purchased from a local market at Teronoh, Malaysia. After that, the meat was cut into small element (1 cm × 1 cm) and keep it at –2C°.

2.2   Preparing Mixture Samples

Lard was adulterated with fats of lamb, Beef and Chicken to obtain a series of standard sets of ten pure and thirty samples containing 10–50% v/v of lard in Lamb, Beef, Chicken samples. This method follows according to Rohman et al. [18]. The mixture samples are mixed according to Table 2.

Table 2   Meat Samples with composition percentage for each species

Mixture samples	Pork%	Lamb%	Beef%	Chicken%	Number of samples
1	10	90			2
2	20	80			2
3	30	70			2
4	40	60			2
5	50	50			2
6	10		90		2
7	20		80		2
8	30		70		2
9	40		60		2
10	50		50		2
11	10			90	2
12	20			80	2
13	30			70	2
14	40			60	2
15	50			50	2
Total mixture samples					30

2.3 Data Analysis

The data analysis was performed using MATLAB and Spectrograph. Spectrograph 1.1 has been used to extract information from spectrum results, where the data is pre-processed as needed. To further examine the findings from pre-processing, MATLAB R2017b was used. To analyze the consistency of lard adulteration, the Principle Components Analysis (PCA) technique was used.

3 Results and Discussion

3.1 FTIR Spectra of Pure Fats

The FTIR Spectra of pure fats is shown in Fig. 1. This spectrum consists of four regions: the first region from 4000 to 2500 cm^{-1}, the second region from 2500 to 2000 cm^{-1}, the third region from 2000 to 1500 cm^{-1}, and finally the fingerprint region from 1500 to 800 cm^{-1}. For second region, there are no peaks showing no functional group present in this range. Moreover, 1st and 3rd region have the potential to differentiate between the sample species.

Figure 2 represents PCA results for all the samples, at 1-D PCA projection, some of the samples such as Chicken and Pig are already starting to cluster together, even though there is still overlap between samples of different species. PCA results at 2-D projection shows that Chicken and Pig samples clustered together without overlap, Lamb samples have one outlier, but the beef samples proved challenging since results are still scattered. The 3-D PCA shows better results though with all pig samples lined around third PCA Component.
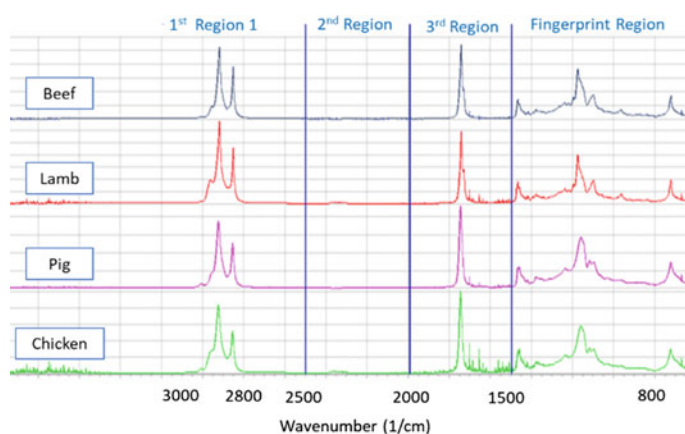
Fig. 1 FTIR Spectrum of all the meat species with region identification
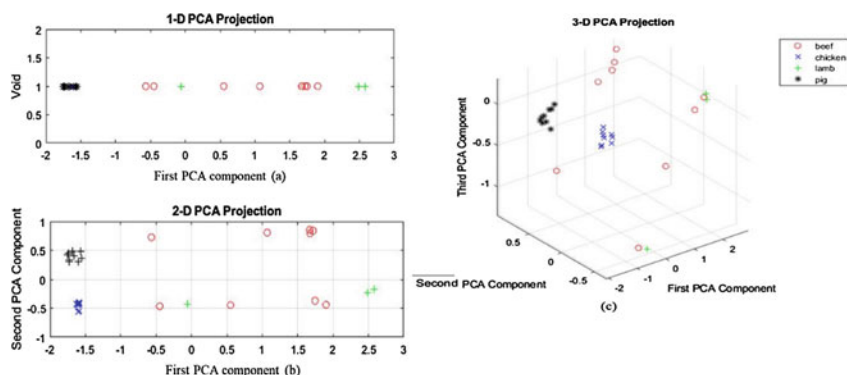
Fig. 2 1D, 2D and three-dimensional representation of chicken, beef, lamb, and lard samples using principal components

4 Conclusion

Identification of lard discrimination from chicken, lamb, and cow fats was demonstrated in this paper. FTIR studies have shown that by having a peak at a wavenumber 1600 cm^{-1}, while other species of fats do not have peak at this stage, lard differentiates itself from other species of fats. In addition, shoulder peak was seen for lard in the range of wavenumber 2950–2990 cm^{-1}, while cow and lamb fats showed sharp peaks, and double peaks showed chicken. In comparison to the usual peaks seen for the other fats, at wavenumber 2750 cm^{-1} a reversed peak was detected for lard. For future work we can further develop this method by analyzing the adulterated samples as well.

References

1. Spink J, Moyer DC (2011) Defining the public health threat of food fraud. J Food Sci 76:R157–R163
2. Guardian T (2017) Allergic teenager's death after eating kebab was accidental, rules coroner. The Guardian.
3. Li DK (2017) Toddler allergic to dairy dies after pre-school serves him grilled cheese. In: New York Post, ed
4. Barlass T (2015) Child aged 10 dies after drinking coconut drink as importer admits label charges. In: The Sydney Morning Herald ed
5. FSA (2013, 7/2/2018) Timeline on horse meat issue. Available: http://webarchive.nationalarchives.gov.uk/20150403184406/, http://www.food.gov.uk/enforcement/monitoring/horse-meat/timelinehorsemeat
6. Rohman A, Che Man YB (2012) Analysis of pig derivatives for Halal Authentication Studies. Food Rev Int 28:97–112
7. Tähkäpää S, Maijala R, Korkeala H, Nevas M (2015) Patterns of food frauds and adulterations reported in the EU rapid alert system for food and feed and in Finland. Food Control 47:175–184

8. Abaitua Borda I, Philen RM, Posada de la Paz M, Gomez de la Camara A, Diez Ruiz-Navarro M, Gimenez Ribota O et al (1998) Toxic oil syndrome mortality: the first 13 years. Int J Epidemiol 27:1057–1063

9. Regenstein JM, Chaudry MM, Regenstein CE (2003) The kosher and halal food laws. Compr Rev Food Sci Food Saf 2:111–127

10. Chuah L-O, He XB, Effarizah ME, Syahariza ZA, Shamila-Syuhada AK, Rusul G (2016) Mislabelling of beef and poultry products sold in Malaysia. Food Control 62:57–164

11. Reuters T (2016) State of The Global Islamic Economy Report 2016/17

12. Barnett J, Begen F, Howes S, Regan A, McConnon A, Marcu A et al (2016) Consumers' confidence, reflections and response strategies following the horsemeat incident. Food Control 59:721730

13. Schmutzler M, Beganovic A, Böhler G, Huck CW (2015) Methods for detection of pork adulteration in veal product based on FT-NIR spectroscopy for laboratory, industrial and on-site analysis. Food Control 57:58–267

14. Vlachos A, Arvanitoyannis IS, Tserkezou P (2016) An updated review of meat authenticity methods and applications. Crit Rev Food Sci Nutr 56:1061–1096

15. Danezis GP, Tsagkaris AS, Camin F, Brusic V, Georgiou CA (2016) Food authentication: techniques, trends & emerging approaches. TrAC Trends Analyt Chem 85:123132

16. Kurniawati E, Rohman A, Triyana K (2014) Analysis of lard in meatball broth using Fourier transform infrared spectroscopy and chemometrics. Meat Sci 96:94–98

17. Meza-Márquez OG, Gallardo-Velázquez T, Osorio-Revilla G (2010) Application of mid-infrared spectroscopy with multivariate analysis and soft independent modeling of class analogies (SIMCA) for the detection of adulterants in minced beef. Meat Science 86:511–519

18. Rahmania H, Sudjadi, Rohman A (2015) The employment of FTIR spectroscopy in combination with chemometrics for analysis of rat meat in meatball formulation. Meat Sci 100:301–305

19. Bruice PY (2016) Organic Chemistry, 8th edn. PEARSON

Deep Learning Approach for Divergence Behavior Detection at High Density Crowd

M. U. Farooq, M. N. Mohamad Saad, Y. Saleh, and S. Daud Khan

Abstract At high-density crowd gatherings, people naturally escape from the region where any unexpected event happens. Escape in high-density crowds appears as a divergence pattern in the scene and timely detecting divergence patterns can save many human lives. In this paper, we propose to physically capture crowd normal and divergence motion patterns (or motion shapes) in form of images and train a shallow convolution neural network (CNN) on motion shape images for divergence behavior detection. Crowd motion pattern shape is obtained by extracting ridges of Lagrangian Coherent Structure (LCS) from the Finite-Time Lyapunov Exponent (FTLE) field and convert ridges into the grey-scale image. We also propose a divergence localization algorithm to pinpoint anomaly location(s). Experimentation is carried out on synthetic crowd datasets simulating normal and divergence behaviors at the high-density crowd. Comparison with state-of-the-art methods shows our method can obtain better accuracy for both divergence behavior detection and localization problems.

Keywords Divergence · FTLE · LCS · Motion estimation · Image shape

1 Introduction

Divergence detection at the high-density crowd is a tough task due to several challenges involved in high-density crowd videos e.g., few pixels available per head, extreme occlusion, cluttering and noise, and perspective problems, etc. If crowd

M. U. Farooq (✉) · M. N. Mohamad Saad
Center for Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi PETRONAS, Tronoh, Malaysia

Y. Saleh
Electrical Engineering Department, Faculty of Engineering and Islamic Architecture, Umm Al-Qura University, Makkah, Saudi Arabia

S. Daud Khan
Department of Computer Science, National University of Technology, Islamabad, Pakistan
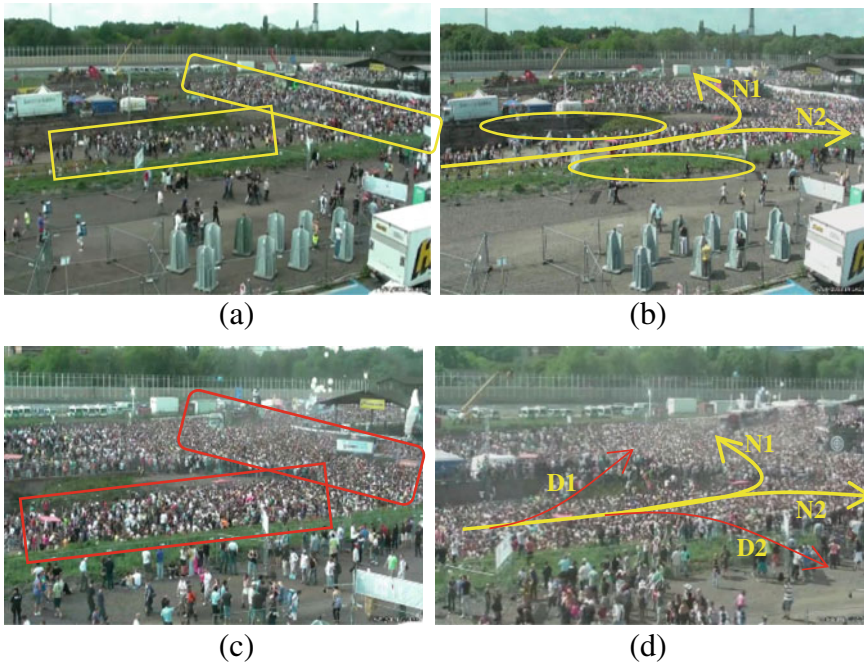
Fig. 1 Demonstration of crowd divergence at Love parade 2010: **a** crowd walking under normal conditions with low density **b** crowd walking paths N1, N2 under normal conditions **c** high-density crowd within the same region **d** crowd diverging through paths D1 and D2

divergence is not detected earlier at its development stage, it may lead to larger disasters like a stampede. Figure 1 shows an example of high-density crowd divergence behavior (Love parade 2010 musical festival [1]) where divergence eventually leads to disastrous stampede. Figure 1a, b demonstrate high-density crowd normal behavior following paths N1 and N2, whereas in Fig. 1c, d, a critical situation is shown where the incoming crowd is blocked by a stationary crowd and diverge through D1, D2 paths. Such divergence situations are common in mass gatherings where the whole crowd is marching towards a common destination and with an increase of density, ends up with a half-stationary half-moving crowd segments that result in divergence behavior.

Previous divergence detection methods [2, 3] learn manual motion features for every individual in the crowd from optical flow (OPF) including location, direction, magnitude, etc. An inherent problem with such methods is with an increase of crowd density, it is almost impossible to capture individual-level motion information and one must learn global crowd features. Later several methods have been developed to capture global crowd motion information e.g., optical flow with pathline trajectories [4–7], pathlines with Lagrangian particle analysis [8], streakflow [9–13], etc. These methods performed well in capturing crowd global motion information under normal

behavior scenes only. Unfortunately, no results are reported in the literature for abnormal behavior detection at very high-density crowd levels.

In this work, we solve divergence detection in the high-density crowd by directly capturing crowd global motion in form of images and learn crowd normal and divergent motion shapes through a neural network that predicts crowd behavior for the unknown scene. We also propose a novel divergence localization algorithm to pinpoint divergence location with the help of a bounding box. Finding a source of divergence can help to efficiently deploy crowd management staff right at the critical locations.

2 Related Work

Motion is one of the key ingredients in the crowd scene analysis and the success of the behavior prediction scheme greatly relies on the efficiency of the motion estimation (ME) method. Therefore, we provide a comprehensive review of ME techniques and the corresponding abnormal behavior detection methods with emphasis on their capabilities for 'high' density crowded scenarios. OPF is considered to be one of the most fundamental motion flow model [14–17] that has been widely employed for motion estimation [18, 19], crowd flow segmentation [20], behavior understanding [21–23] and tracking in the crowd [24]. However, OPF methods suffer from various problems like motion discontinuities, lack of spatial and temporal motion representation, variations in illumination conditions, severe clutter and occlusion, etc.

To overcome problems of OPF ME, researchers employ particle advection concepts from fluid dynamics into the computer vision domain [8] and obtain long-term "motion trajectories" under the influence of the OPF field. We et al. [7] employ chaotic invariants on Lagrangian trajectories to determined either the behavior of the crowd is normal or not. They also perform localization of anomaly by determining the source and size of an anomaly. Unfortunately, no results were reported for the high-density crowd. Similarly, Ali et al. [8] obtain Lagrangian Coherent Structures (LCS) from particle trajectories by integrating trajectories over a finite interval of time termed as Finite-Time Lyapunov exponent (FTLE). LCS appears as ridges and valleys in the FTLE field at the locations where different segments of the crowd behave differently. Authors perform crowd segmentation and instability detection in the high-density crowd using LCS in FTLE, however actual anomalies of the high-density crowd like crowd divergence, escape behavior detection, etc. are not performed. Similarly, authors in [10, 11] obtain particle trajectories using high accuracy variational model for crowd flow and perform crowd segmentation tasks only. Mehran et al. [9] obtain streakflow by spatial integration of streaklines that are extracted from particle trajectories. For anomaly detection, they decompose streakflow field into curl-free and divergence-free components using the Helmholtz decomposition theorem and observe variations in potential and streak functions used with SVM to detect anomalies like crowd divergence/convergence, escape behavior,
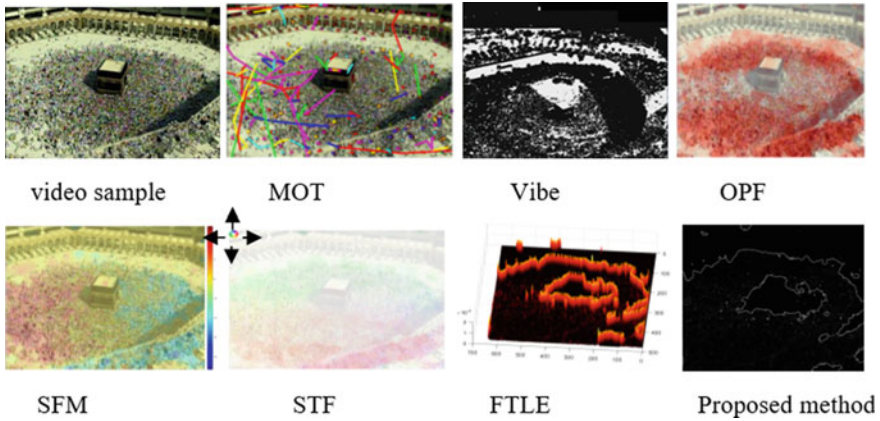
Fig. 2 High-density crowd motion estimation by state-of-the-art methods

etc. However, results are reported for anomaly detection and segmentation at low-density crowd and efficacy is still questionable for anomalies at the high-density crowd. Eduardo et al. [25] obtain long-range motion trajectories by using the farthest point seeding method called streamline diffusion on streamlines instead of spatial integration.

Behavior analysis is performed by linking short streamlines using Markov Random Field (MRF). However, only normal behavior detection and crowd segmentation results are reported. Although particle flow methods discussed above are better candidates for ME of the high-density crowd, but they are rarely employed for abnormal behavior detection at high density crowded scenes. Figure 2 provides a comparison of ME methods for high-density crowd performing Tawaf around Kabbah. Conventional object tracking based ME methods [26, 27] (Fig. 2b, c) works best at low crowd density but completely fails at high crowd density. The OPF method from Brox et al. [15] can estimate motion at high density but motion information is short-term. SFM [28] method can provide better motion estimation in low-density crowd areas but at high density, the performance of SFM also degrades. Streakflow [9] method also performs similarly to the SFM method at a high-density crowd. Unfortunately, all these methods are unable to provide a clean motion-shape for the crowd. FTLE method [9] (Fig. 2g) produce clear ridges at cowd boundaries and can be best to describe high-density crowd motion. Therefore, in this work, we utilize the FTLE method to obtain crowd motion-shape and translate it to a single channel greyscale image (Fig. 2h) for both normal and abnormal behavior analysis. Our framework for divergence detection is shown in Fig. 3 (top portion). It consists of two main phases: Phase 1: low-level FTLE feature extraction and conversion into a grey-scale motion shape image; Phase 2: behavior classification using a CNN. Motion shape images are also used for divergence localization process.
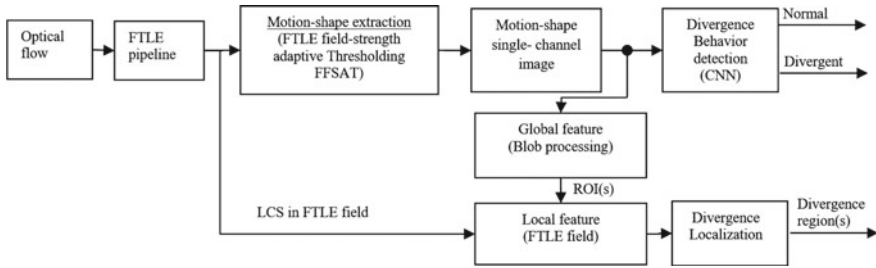
Fig. 3 Framework for divergence behavior detection in high-density crowd

3 Divergence Detection with Motion Shape and Deep Convolution Neural Network

3.1 Data Preparation

Due to the unavailability of a very high-density crowd dataset with divergence behavior, we generate synthetic data by simulating crowd in Massmotion software [29]. We model two crowd scenarios: Stampede at Loveparade 2010 and Tawaf around Kabbah. Example snapshots of normal and divergence crowd behaviors are shown in Figs. 4 and 5.

3.2 Global Motion Estimation and Shape Extraction

In this work, high-density crowd motion is computed by the Finite-Time Lyapunov Estimation (FTLE) method [8, 30]. Lagrangian Coherent Structure (LCS) appears as ridges in the FTLE field where two crowd segments behave differently. We extract LCS from FTLE field FTLE using the field-strength adaptive thresholding (FFSAT) scheme and convert it into a grey-scale image. At every integration step in the FTLE pipeline, maximum Eulerian distance (dmax) is calculated between LCS absolute peak value and average FTLE field strength, and a threshold (ffsat_thr) is set for dmax (65% in our work). LCS values crossing ffsat_thr are extracted and converted into a single-channel grey-scale image. FFSAT algorithm ensures only strong magnitude LCS values from the FTLE field are extracted and noise is filtered out.

3.3 Deep Network for Divergence Detection

A deep CNN network developed for normal and divergence classes in the high-density crowd is shown in Fig. 6.

a. Normal behavior

b. Diverging crowd

c. K13 view

d. K5 view

e. K12 view
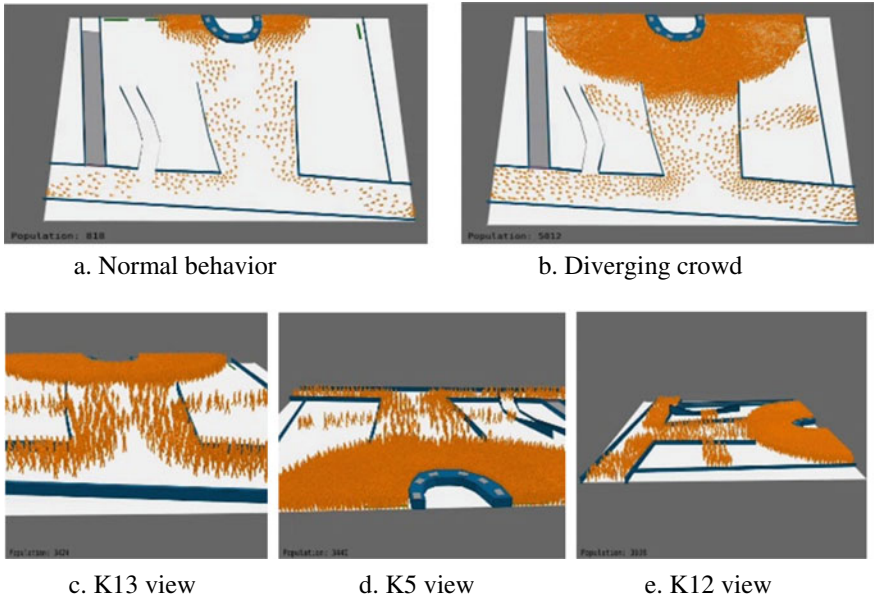
Fig. 4 Synthetic crowd data for Love parade 2010 disaster: **a**, **b** are camera-top views; **c**, **d**, **e**- same crowd with perspective views

a. Normal behavior

b. Diverging behavior
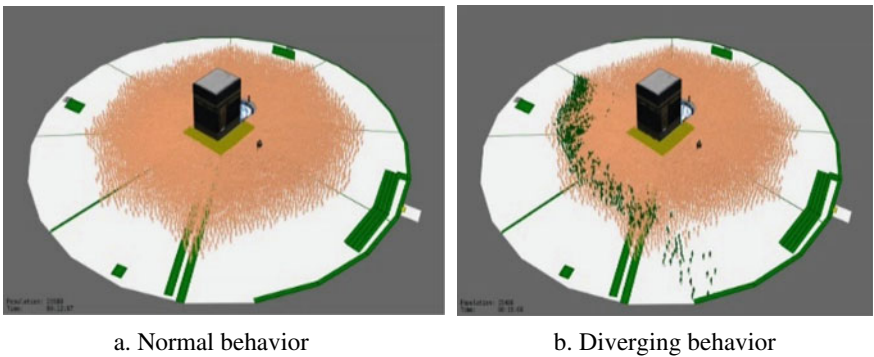
Fig. 5 Synthetic crowd data for Kabbah Tawaf—Normal behavior and divergent crowd

The greyscale image is first rescaled to 50×50 pixels at the input layer. A convolution layer is used (24 filters) with ReLU activation. The purpose of using a large number of convolution filters is to ensure all important receptive fields of CNN are excited about a given motion-shape. ReLU is adopted as the activation function because of its good performance for CNNs [31] and Max pooling is used for each 2×2 region. Finally, two fully connected layers are used and the softmax layer is used for the classification of normal or divergent behavior.
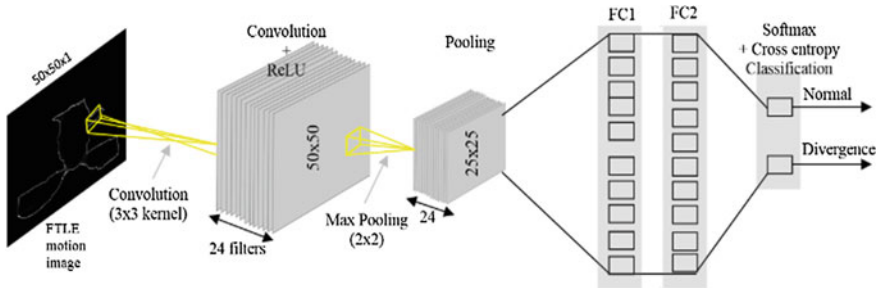
Fig. 6 Deep CNN for divergence behavior detection

3.4 Divergence Localization Algorithm

We propose a novel divergence localization algorithm that analyzes changes in motion shape blob to search for the region of divergence. It was noticed that motion-shape also exhibit undesired local variations (Fig. 7 top row) that could lead to false divergence region detection. These changes are occurred due to the to-and-fro motion experienced by the crowd at high densities crowd [32]. As these oscillatory motions propagate and reach the crowd boundary, the shape does not remain consistent in every frame. Whereas the initial occurrence of divergence also appears as a small shape change and progressively increases in size (as shown in Fig. 7 bottom row). To cater to undesired local shape changes, a blob processing pipeline is implemented shown in Fig. 8.
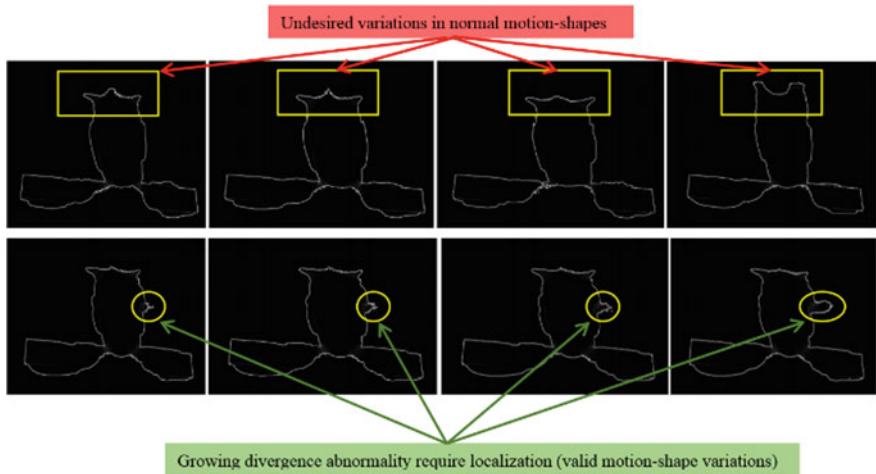
Fig. 7 Top row: Undesired motion-shape variations due to crowd oscillatory motion; Bottom row: Real shape change due to divergence

Fig. 8 Baseline blob extraction pipeline for normal and divergence behaviors

Baseline blob extraction pipeline extracts a baseline blob from the normal and divergence motion shapes and input to divergence localization algorithm is shown in Table 1. The divergence localization algorithm indicates divergence location(s) with the bounding box.

4 Experimentation Results

We evaluate proposed methods of crowd divergence behavior detection and divergence localization using crowd datasets of Love parade and Kabbah (data preparation details in Sect. 83.1). A detailed qualitative and quantitative analysis is provided for both methods on two selected scenarios. We also compare our methods with OPF from Brox et al. [15] by converting the OPF field in binary images.

4.1 Divergence Behavior Detection

For divergence behavior detection at two scenarios, the crowd is simulated to diverge from 25 different locations in each scenario and 1000 motion-shape images are captured (total images for 25 divergence locations $= 25 \times 1000 = 25,000$ divergence images for each scenario). One thousand images for each divergence location are generated to train CNN with minor local motion changes contributed by crowd oscillatory motion. Similarly, 2500 images are generated for normal crowd behavior. The dataset for each scenario is split into two parts: randomly 20 divergence locations data $(20 \times 1000 = 20,000$ images) are used for training/validation purposes, whereas the remaining random 5 divergence locations data (completely unseen to CNN) is used for prediction. Figure 9 provides a confusion matrix of divergence behavior detection for both scenarios and performance is compared with the OPF method.

For both the Love parade and Kabbah scenario, our method can achieve 100% accuracy. However, in both scenarios, OPF was able to detect approx. 50% of divergence behaviors only. Motion-shapes obtained through the OPF method are not as smooth and consistent as produced by our method; hence OPF performance degradation is evident.

Table 1 Algorithm for divergence localization with bounding box

Algorithm: Divergence localization using the bounding box	
Init	**N**: Total video frames, **M**: Total divergence frames to analyze=50, **s_n**:No. of frames for spatial averaging =5, **t_n**: No. of frames for temporal averaging = 5, **st_thr**: FTLE field spatio-temporal threshold = 0.5, **blb_cnt**: number of small blobs to process = 5, **i_t**: transition frame index
Step 1	**Search normal to divergence-starting Transition frame (global features analysis)**

	FOR frames i=1:*N*
	   -Obtain FTLE motion-shape images for two frames i(previous) and i+1(current),
	     -Run blob processing pipeline (fig. 8) on previous and current frames FTLE images and obtain large single blobs for both frames
	     -Subtract the current large blob from the previous large blob to obtain several smaller blobs
	     -Sort blb_cnt small blobs according to the area in descending order
	     -Blob shape analysis for ROI
	      FOR blob_num=1: blb_cnt
	         Compare blob shape with shape-list [square, rectangle, triangle, pentagon, hexagon, semi-circle or nearly circular]
	            IF blob shape matches with any of shape-list
	              Transition frame 'i_t' with the first ROI, Exit this loop and go to step 2
	                 ELSE
	                   Select next blob in area sorted list
	          END
	         -Increment frame number i
	    END

Step 2	**Bounding box search for divergence location (global features analysis)**

	FOR frames j=i_t:i_t+M
	   - Obtain FTLE motion-shape images for two frames j(previous) and j+1(current)
	   - Run blob processing pipeline (fig. 8) on previous and current frames FTLE images and obtain large single blobs for both frames
	   - Subtract current large blob from previous large blob to obtain several smaller blobs
	   - Sort blb_cnt small blobs according to the area in descending order
	   - **Blob shape analysis for ROI**
	     FOR blob_num=1: blb_cnt
	        IF blob shape matches with any of shape-list
	             Mark this blob as possible ROI and update ROI_list
	          ELSE
	            Select next blob in area sorted list
	     END
	   - Jump to step 3 to obtain divergence location bounding box coordinates for all ROI of the current frame
	   - Increment j
	    END

(continued)

Table 1 (continued)

Step 3	FTLE field search for refined divergence location bounding box (local features analysis)
	FOR ROI=1:length(ROI_list)

FOR ROI=1:length(ROI_list)
- **Spatial Averaging**: For each ROI boundary pixel, perform 8-connected spatial averaging of FTLE field (for all frames from j-s_n to j+s_n frames)
- **Temporal averaging:** For each ROI boundary pixel in j-t_n to j+t_n frames, perform temporal averaging.
- FOR all ROI boundary spatio-temporal averaged values

      IF spatio-temporal averaging value >= st_thr,

        keep pixel as valid

      ELSE

        discard pixel

   END
- Obtain bounding box coordinates of current ROI by calculating maximum and minimum (x,y) coordinates from selected valid pixels list in above step

END

Return bounding box coordinates for all ROIs in ROI_list

Fig. 9 Confusion matrices for divergence behavior detection: Love parade scenario **a** Proposed method **b** OPF method; Kabbah scenario **c** Proposed method **d** OPF method

4.2 Divergence Localization

The performance of the divergence localization algorithm is evaluated by calculating the Intersection over Union (IoU) area of the predicted bounding box and ground truth bounding box for each divergence region. Ground truth bounding boxes are obtained by hand labeling divergence regions of each abnormal frame. IoU is calculated using Eq. (1).

$$IoU = \frac{Area\ of\ overlap}{Area\ of\ union} \qquad (1)$$

Generally, an IoU score greater than 0.5 (50% overlap) is considered a good prediction by any bounding box (b. box) detection algorithm [33]. In this work, the IoU score is calculated for N post i_t frames. IoU score of six selective frames (out of N = 50 post i_t frames) for the Love parade scenario is shown in Fig. 10. The green color b.box represents ground truth and the red color b.box represents prediction by our algorithm. The Final IoU score is obtained by averaging N frames IoU scores. The average IoU score of our algorithm for the Love parade scenario is 0.501 (50% overlap). We also perform divergence region detection using OPF motion images. The average IoU score with the OPF method is found to be 0.15 (15% overlap) which proves our method performs well than OPF for divergence
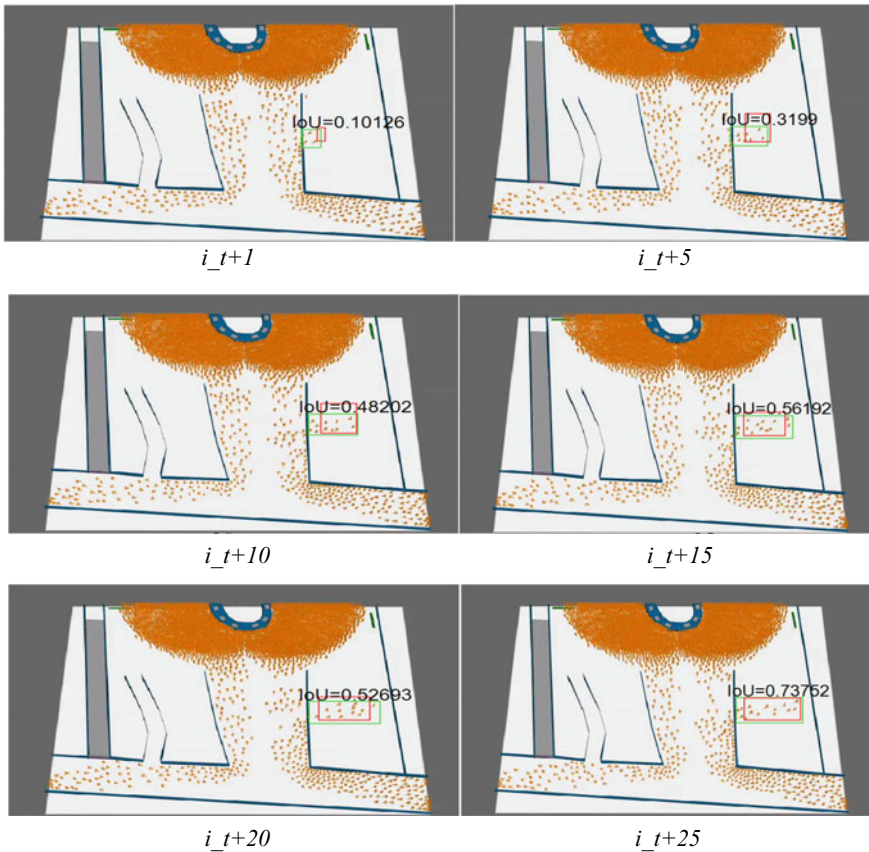
Fig. 10 Divergence region localization with our proposed method. IoU scores shown for six post *i_t* frames

region localization. Similarly, the average IoU score for the Kabbah scene with our algorithm is 0.63 (63% overlap) and 0.18 (18% overlap) for the OPF method.

5 Conclusion

In this work, we propose a deep CNN-based divergence behavior detection framework that extracts high-density crowd motion shapes in form of images to train deep CNN. Experimentation results show that the proposed method can achieve close to 100% accuracy for divergence detection in challenging Loveparade and Kabbah crowding scenarios. Similarly, a novel divergence region detection algorithm efficiently detects divergence regions with IoU of more than 50%. However, we notice there are few limitations of our proposed methodology of converting crowd motion into images using the FTLE method. Motion shape analysis is inefficient in the situations when a crowded segment in high density gets stationary due to any reason. Since there is no more movement at the stationary crowd segment, FTLE is unable to predict crowd motion shape at static crowd portions and results in incomplete or broken motion-shapes. Therefore, for our framework to work efficiently, the crowd needs to keep moving (for consistent motion-shape) that is always not true. Secondly, in the FTLE method, LCS ridges appear only at crowd boundaries, if any anomaly takes place at interior portions of the crowd (far from crowd boundaries towards the center), FTLE is unable to provide any information there. Therefore, in future work, we shall improve our method by incorporating spatial and temporal crowd density variations to capture static crowd behavior. And predict crowd behavior in all segments of the crowd, either crowd is stationary or in motion.

References

1. Helbing D, Mukerji P (2012) Crowd disasters as systemic failures: analysis of the Love Parade disaster. EPJ Data Sci 1(1):1–40. https://doi.org/10.1140/epjds7
2. Chen CY, Shao Y (2015) Crowd escape behavior detection and localization based on divergent centers. IEEE Sens J 15(4):2431–2439. https://doi.org/10.1109/JSEN.2014.2381260
3. Wu S, Wong HS, Yu Z (2014) A bayesian model for crowd escape behavior detection. IEEE Trans Circuits Syst Video Technol 24(1):85–98. https://doi.org/10.1109/TCSVT.2013.2276151
4. Andrade EL, Blunsden S, Fisher RB (2006) Modelling crowd scenes for event detection. Proc Int Conf Pattern Recognit 1:175–178. https://doi.org/10.1109/ICPR.2006.806
5. Courty T, Corpetti N (2007) Crowd Motion Capture. Comput Animat Virtual Worlds 18(September 2007):361–370
6. Nam Y, Hong S (2014) Real-time abnormal situation detection based on particle advection in crowded scenes. J Real-Time Image Process 10(4):771–784. https://doi.org/10.1007/s11554-014-0424-z
7. Wu S, Moore BE, Shah M (2010) Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, pp 2054–2060

8. Ali S, Shah M (2007) A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In: IEEE conference on computer vision and pattern recognition, pp 1–6

9. Mehran R, Moore BE, Shah M (2010) A streakline representation of flow in crowded scenes. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) LNCS 6313(PART 3):439–452. https://doi.org/10.1007/978-3-642-15558-1_32

10. Wang X, Gao M, He X, Wu X, Li Y (2014) An abnormal crowd behavior detection algorithm based on fluid mechanics. J Comput 9(5):1144–1149. https://doi.org/10.4304/jcp.9.5.1144-1149

11. Wang X, Yang X, He X, Teng Q, Gao M (2014) A high accuracy flow segmentation method in crowded scenes based on streakline. Opt Int J Light Electron Opt 125(3):924–929. https://doi.org/10.1016/j.ijleo.2013.07.166

12. Wang X, He X, Wu X, Xie C, Li Y (2016) A classification method based on streak flow for abnormal crowd behaviors. Opt Int J Light Electron Opt 127(4):2386–2392. https://doi.org/10.1016/j.ijleo.2015.08.081

13. Huang S, Huang D, Khuhro MA (2015) Crowd motion analysis based on social force graph with streak flow attribute. J Electr Comput Eng 2015. https://doi.org/10.1155/2015/492051

14. Horn BK, Schunck BG (1981) Determining optical flow. Artif Intell 17(1981):185–203

15. Brox T, Papenberg N, Weickert J (2004) High accuracy optical flow estimation based on a theory for warping. In: Computer Vision - ECCV 2004, vol 4, no May, pp 25–36. https://doi.org/10.1007/978-3-540-24673-2_3

16. Lucas BD, Kanade T (1981) An iterative image registration technique with an application to stereo vision. Proc Imaging Underst Work 130:121–130

17. Fortun D, Bouthemy P, Kervrann C, Fortun D, Bouthemy P, Kervrann C (2015) Optical flow modeling and computation: a survey. Comput Vis Image Underst 134:1–21

18. Lawal IA, Poiesi F, Anguita D, Cavallaro A (2016) Support vector motion clustering. IEEE Trans Circuits Syst Video Technol X(X):1–1. https://doi.org/10.1109/TCSVT.2016.2580401

19. Cheriyadat AM, Radke RJ (2008) Detecting dominant motions in dense crowds. IEEE J Sel Top Signal Process 2(4):568–581. https://doi.org/10.1109/JSTSP.2008.2001306

20. Ali S, Shah M (2007) A lagrangian particle dynamics approach for crowd flow simulation and stability analysis

21. Hu MHM, Ali S, Shah M (2008) Learning motion patterns in crowded scenes using motion flow field. In: 2008 19th international conference pattern recognit, pp 2–6. https://doi.org/10.1109/ICPR.2008.4761183

22. Solmaz B, Moore BE, Shah M (2012) Identifying behaviors in crowd scenes using stability analysis for dynamical systems. IEEE Trans Pattern Anal Mach Intell 34:2064–2070. https://doi.org/10.1109/TPAMI.2012.123

23. Chen DY, Huang PC (2011) Motion-based unusual event detection in human crowds. J Vis Commun Image Represent 22(2):178–186. https://doi.org/10.1016/j.jvcir.2010.12.004

24. Hu W, Xiao X, Fu Z, Xie D, Tan T, Maybank S (2006) A system for learning statistical motion patterns. IEEE Trans Pattern Anal Mach Intell 28(9):1450–1464. https://doi.org/10.1109/TPAMI.2006.176

25. Pereira EM, Cardoso JS, Morla R (2016) Long-range trajectories from global and local motion representations. J Vis Commun Image Represent 40:265–287. https://doi.org/10.1016/j.jvcir.2016.06.020

26. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: EEE computer society conference on computer vision and pattern recognition (CVPR'05), pp 886–893

27. Barnich O, Van Droogenbroeck M (2011) ViBe: a universal background subtraction algorithm for video sequences. IEEE Trans Image Process 20(6):1709–1724. https://doi.org/10.1109/TIP.2010.2101613

28. Mehran R, Oyama A, Shah M (2009) Abnormal crowd behavior detection using social force model. In: IEEE computer society conference on computer vision and pattern recognition work. CVPR Work 2:935–942. https://doi.org/10.1109/CVPRW.2009.5206641.

29. https://www.oasys-software.com

30. Shadden SC, Lekien F, Marsden JE (2005) Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows. Phys D Nonlinear Phenom 212(3–4):271–304. https://doi.org/10.1016/j.physd.2005.10.007
31. Zeiler MD et al (2013) On rectified linear units for speech processing New York University, USA Google Inc ., USA University of Toronto , Canada. In: IEEE international conference on acoustic speech and signal processing (ICASSP 2013), pp 3–7
32. Krausz B, Bauckhage C (2012) Loveparade 2010: automatic video analysis of a crowd disaster. Comput Vis Image Underst 116(3):307–319. https://doi.org/10.1016/j.cviu.2011.08.006
33. Ahmed F, Tarlow D, Batra D (2015) Optimizing expected intersection-over-union with candidate-constrained CRFs. In: Proceedings of the IEEE international conference on computer vision, vol 2015, pp 1850–1858. https://doi.org/10.1109/ICCV.2015.215

Anomaly Localization at High-Density Crowd Using Motion Shape Image (MSI)

M. U. Farooq, M. N. Mohamad Saad, S. Daud Khan, and Y. Saleh

Abstract Anomaly localization plays a critical role if a disaster occurs in a high-density crowd to efficiently rescue the crowd from the right location. This paper enriches anomaly localization by introducing new localization features, in addition to the 'only' localization feature of source/start point detection existing literature offers. New features introduced include crowd density estimation in localized regions and direction/angle of a localized region. A motion shape image (MSI) based approach is introduced for localization and features detection and experimentation is performed on benchmark and our proposed high-density crowd datasets.

1 Introduction

It is well known that automated anomaly detection in surveillance videos has attracted a lot of attention due to efficient methods developed in the past. However, anomaly detection alone is less effective without localizing it. Anomaly localization is critical information that can help to efficiently deploy rescue-resources at the right place to save many precious lives.

There have been several methods developed in the past focusing anomaly localization at low to medium density crowd [1–3], however, localizing anomaly at high-density is challenging and there are not many methods available in the literature [4–7].

M. U. Farooq (✉) · M. N. Mohamad Saad
Center for Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi PETRONAS, Tronoh, Malaysia

S. Daud Khan
Department of Computer Science, National University of Technology, Islamabad, Pakistan

Y. Saleh
Electrical Engineering Department, Faculty of Engineering and Islamic Architecture, Umm Al-Qura University, Makkah, Saudi Arabia

889

Therefore, this work aims to propose an anomaly localization method at high-density crowd gatherings. Representative anomaly localization methods are discussed next.

Wu et al. [4] localize anomaly in the crowd by finding the position of abnormal representative trajectories. Wu et al. [5] localize crowd escape anomaly by determining the divergence center locations using a reversible jump Mteropolis-Hasting algorithm. Similarly, Chen et al. [1] localize escape behavior anomaly by applying K-nearest search and distance segmentation methods on optical flow features of magnitude, speed, and direction. Chaker el at. [6] localize anomaly using a social network model that separates rare actions from dominant actions, however, their moving object in a crowd, not localizing crowd anomaly. Mehran et al. [7] localize crowd divergent or convergent behavior by computing potential functions from streakflow. Ongun et al. [2] localize crowd anomaly by clustering finite-time Lyapunov exponent heat maps and applying thresholding on clusters to detect an anomaly, however, anomaly localization details are not clearly stated in this work. Ren et al. [3] perform anomaly detection using an entropy model and anomaly is localized by detecting high behavior entropy (BE) value pixels in an image.

The methods discussed above suffer from two types of important limitations: (1) Majority of existing art localize anomalies at low or medium-density crowds, to the best of our knowledge, no method is developed to localize anomalies at high/very high-density crowds. (2) the localization information is limited i.e., all methods only show the source or start of the anomaly. There is no updated information on an anomaly at the localized region later in the video. Hence, limits the application of existing methods in real-life anomaly localization tracking/updates.

In this paper, we propose a motion shape image (MSI) based approach for the anomaly localization method. The proposed method not only locates start/source points of an anomaly but also provides more useful pieces of information about anomaly localized region throughout the video and hence is useful for practical anomaly localization at high-density crowds. Specifically, the main contributions of this paper are as follows:

1. An anomaly localization method, that generates localization masks for normal as well as abnormal crowd segments,
2. Provides anomaly localization updates each frame i.e., if anomaly location is moved in the video from initial detection, the method will provide an updated location of the anomaly at later times,
3. Crowd density estimate is provided inside anomaly localized regions,
4. The angle and direction of the localized region are estimated with the help of the angle of the localized anomaly region measured about the normal crowd segments.

In this work, we focus on escape or divergence anomaly at a high-density crowd. However, the definition of divergence in the existing literature is not clear and we define the divergence of the crowd as the 'significant change in crowd motion pattern from its normal walking motion pattern'. Usually a high-density crowd exhibit two types of normal motion patterns i.e., straight and circular. And the significant

change in motion pattern from the normal straight or circular patterns in any direction is considered as divergence. In the next subsection, the localization method for divergence anomaly at the high-density crowd is discussed in detail.

2 Proposed Method

A detailed block diagram of our anomaly detection and localization framework for divergence behavior at a high-density crowd is shown in Fig. 1. The top portion of the block diagram shows an anomaly detection network using an MSI. The performance evaluation of the proposed anomaly detection network is discussed in detail in [8]. The proposed anomaly detection network outputs scores for two classes i.e., normal and anomalous (divergent). A divergent class score greater than a pre-defined threshold triggers the divergence localization process, shown at the bottom of Fig. 1.
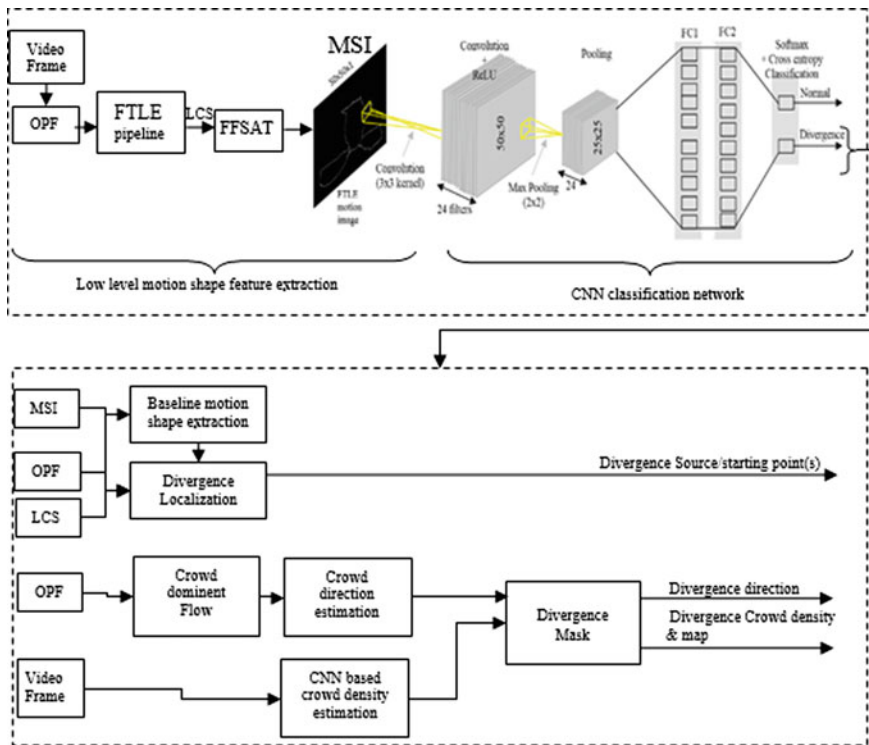
Fig. 1 Divergence detection and localization framework

2.1 Divergence Localization and Source/start Point Detection

Our divergence localization scheme is simple, it takes the difference of incoming MSI from a reference motion shape as shown in Fig. 2. The differencing process is started upon the indication of divergence detected by CNN and the motion shape at that time instance is saved as a reference shape for the subsequent subtraction process. However, motion shapes obtained at every integration step are not similar and exhibit both local and global shape variations.

Figure 3 shows samples of shape variations for normal (top row) and divergence (bottom row) behaviors for the crowd at the Loveparade scenario (crowd on the ramp entering from left and right tunnels and heading towards ramp top for exiting). Regions marked with squares in Fig. 3 (upper row) are motion shape variations experienced during normal behavior frames. Simply differencing normal and divergence motion shapes (in the presence of shape variations) produce many undesired blobs that are not actual divergence sources and can lead to false divergence-source detections.

As mentioned above, shape variations occur locally and globally at the high-density crowd. Local motion shape variations are due to to-and-fro motion [9] and global crowd motion shape changes are due to the segment(s) of the crowd become 'stationary' and results in no FTLE field. For the former problem, the crowd naturally starts oscillating left and right at their central axis, generating crowd waves. As the waves reach boundaries, the crowd naturally expands and shrink. As a result, the FTLE field at the crowd boundary also expands/shrinks causing local (minor) motion shape variations.
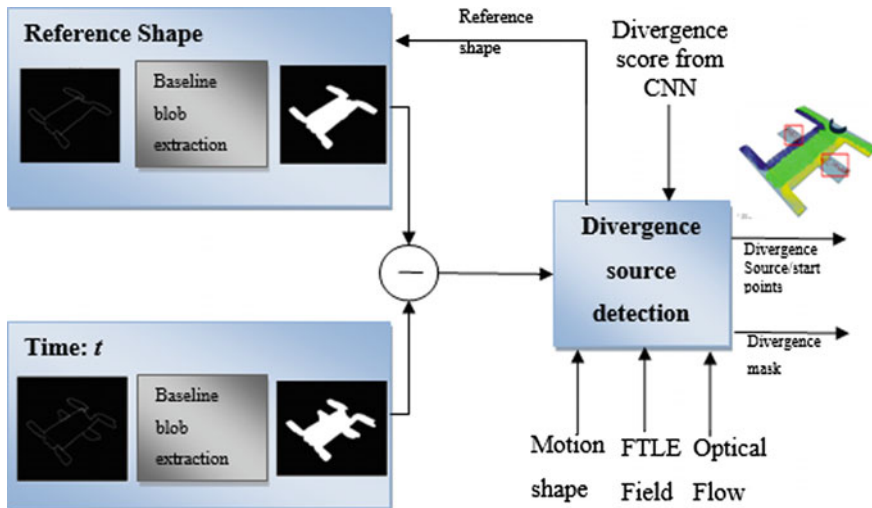
Fig. 2 Divergence localization method for source/starting point detection and divergence mask
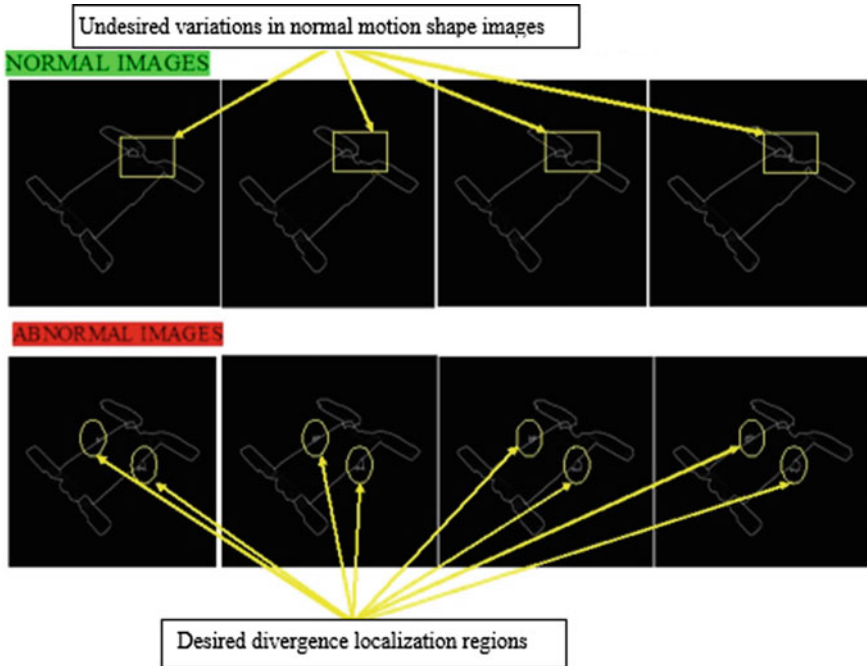
Fig. 3 Top row: undesired motion-shape variations due to crowd oscillatory motion; Bottom row: real shape change due to divergence

Global motion shape variation occurs if any crowd segment stops that reduce crowd velocities to zero, results in no OPF/ FTLE field in that crowd region. Thus, LCS ridges in the FTLE field vanishes from stationary crowd segments and only appear at the moving crowd segments, causing an undesired global motion shape variation. Due to the undesired changes in motion shape, raw MSI's cannot be utilized for further analysis and we need to extract baseline motion shape for both normal and divergence behaviors. A baseline motion shape represents a shape that is most common throughout all the specific behavior frames. Figure 4 shows the pipeline for baseline motion shape extraction that extracts baseline motion shape for both normal and abnormal video frames.

Baseline motion shapes obtained for normal and divergent motion patterns differ in the shape in the way that divergent motion shape will contain extra blobs representing diverging crowd. The reference normal baseline motion shape and baseline divergent motion shape are then subtracted that generates smaller blobs. The resulting small
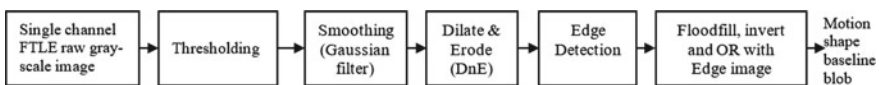
Fig. 4 Pipeline for baseline motion shape extraction for normal and divergence behaviors
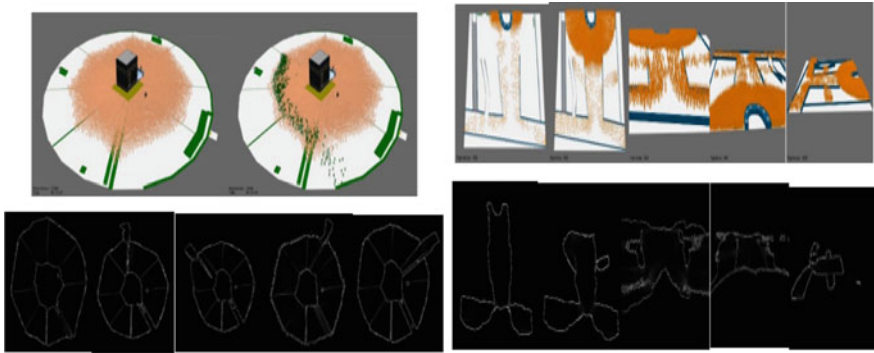
Fig. 5 Left side: Crowd circulatory normal motion with crowd diverging in the top second image (lower row shows corresponding MSI's); right side images for synthetic love parade dataset crowd diverging from different locations (lower row are corresponding MSI's)
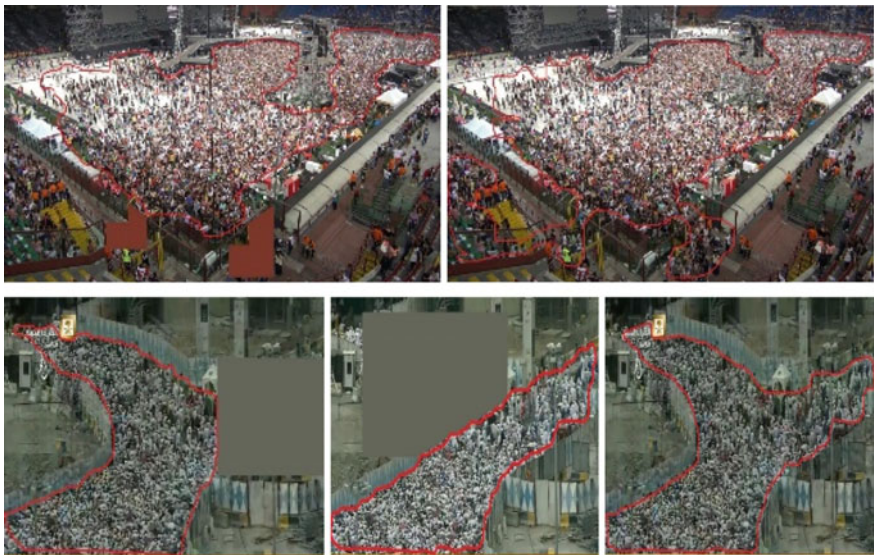
Fig. 6 Top row: MELAN dataset with crowd diverging from left and right gates; bottom row: PILGRIM dataset with two normal scenarios and last image showing divergent crowds for two normal crowd scenarios

blobs contain the actual divergent region(s) and many noisy blobs. The noisy blobs are filtered out by applying temporal and spatial filtering on OPF and FTLE fields on current and past N_{filter} frames.

The blobs obtained from the above process represent divergent regions or divergence masks around the diverging crowd. However, we need to find an exact starting point or source of divergence from the divergent mask detected above. Initially, the

contour of the divergence mask is obtained and the Euclidean distance of contour pixels from normal baseline motion shape pixels is computed. The contour pixels having a distance less than a threshold is/are marked as divergence source or starting point(s). The experimentations have shown that a threshold of 10 pixels is most optimal for the majority of scenes and can represent correct source points. Euclidean distances more than 10 pixels are generally far from the actual source point and can represent false source point detections.

2.2 Crowd Density Estimation (CDE) in the Localized Mask

Our anomaly localization process discussed above primarily generate a localization mask around the diverging crowd. Crowd density inside the divergence mask is an important indicator of the severity of divergence i.e., a high crowd density crowd inside the divergence mask indicates a life-threatening situation that forces the crowd to take a divergence path, etc. In this work, crowd density is estimated using the CSRNet method from [10] (Keras implementation is available at [11]). The model is pre-trained on the ShanghaiTech dataset and further fine-tuned on our PILGRIM dataset and on high-density crowd images from Hajj (a detailed discussion on crowd datasets is provided in section III). To estimate the crowd density inside the divergence mask, initially, a density map and crowd density count is generated by CSRNet for the whole image. The density map and crowd count are then cropped according to the divergence mask produced by the divergence localization algorithm.

2.3 Direction and Angle of the Localized Mask

The diverging crowd direction and the angle of divergence are critical information for first responders to effectively manage the crowd. To quantify the crowd divergence information, crowd "dominant motion" needs to be computed. Dominant motion represents the motion of the major segments of the crowd in the video. In the case of divergence, we have to compute dominant flows for both normal and divergent crowd segments to show how far the divergent crowd moves away from the normal walking crowd. In this work, dominant motion flow is computed using the method from [12]. Summarizing the method of [12] for dominant motion flow, initially, video is divided into several segments and particle advection is performed to obtain small size raw tracklets in each segment. Raw tracklets provide short duration motion information and crowd starting (source) and ending (sink) points cannot be determined. Local long tracks are identified in each video segment by using an unsupervised hierarchical clustering algorithm that utilizes the Longest Common Sub-sequence (LCS) similarity measure to combine smaller tracklets into local large trajectories. To find dominant flow throughout the video, local trajectories from different video segments are also combined using the same clustering and similarity method to produce global

dominant flows. This method is also able to identify the source and sink point of a dominant motion.

The above algorithm is run to obtain dominant motion flows in both normal and divergence masks. To quantify the divergent dominant motion from normal dominant motion, we need to compute the mean angle for both dominant flows and the angle difference between the two flows would show how far the divergent crowd moves away from the normal crowd. A dominant motion trajectory represents the positions of a particle P_i it takes from source to sink in video and there is associated angle information ($\emptyset_i$) at every position of a particle. The particle angle ($\emptyset_i$) of each particle P_i in the dominant flow trajectory is computed using OPF velocities given in Eq. 1.

$$\emptyset_i = \tan^{-1}\left(\frac{v}{u}\right) \tag{1}$$

where u and v are optical flow horizontal and vertical velocities respectively.

It is important to notice here that the OPF used in the particle advection process is computed at an evenly spaced grid whereas particle positions of dominant flow trajectory are at off-the-grid indexes. And we do not have OPF at off-grid particle positions. Therefore, linear interpolation is implemented to obtain OPF at off-grid particle positions and $\emptyset_i$ is computed from interpolated OPF u and v velocities. The angle $\emptyset_i$ computed above lies in the range of $[0, \pi]$. To know the exact quadrant in which a particle lies, $\emptyset_i$ is mapped to a unit circle with the angle from $[0, 2\pi]$. We use the method in [13] to map the linear angle into a circular angle on the unit circle. The circular angle shows the actual quadrant of a particle it lies in. A similar process is repeated for all particles by mapping all on the unit circle. Then a mean circular angle is computed that represents the mean direction of a dominant flow. The mean angle for both normal and divergent dominant flows is computed and the divergence angle is calculated by taking the difference of two mean angles.

3   Proposed Crowd Dataset

The majority of existing datasets containing divergent scenes consist of either low-density crowd or contain normal behavior only A new large-scale dataset is developed to evaluate our method under realistic high-density crowd divergence scenarios. Mainly we construct three datasets: (1) Synthetic (SYN) dataset (2) MELAN CONCERT dataset and (3) PILGRIMS dataset. A summary of the datasets is provided in Table 1 (more details for the proposed dataset can be seen at [8]).

Table 1 Video datasets comparison (existing benchmark and ours)

Dataset	No, of Frames		No.of scenes	Resolution	Frame rate	Density level	Type (Real/synthetic)	Normal motion type
	Normal	Abnormal						
UCF	110	140	1	480×360	25	Low	Real-Ped	Straight
UMN	500*	100*	11	320×240	30	Low	Real-Ped	Circular
PETS2009	52*	52*	8	768×576	7	Low-medium	Real-Ped	Straight
NGSIM	250*	241*	8	640×480	10	Low	Real-Traffic	Straight
OUR-SYN	10000*	10000*	100	1280×720	24	High	Syn-Ped	Straight/circular
OUR-MELAN CONCERT	553*	553*	4	1920×1080	50	High	Real-Ped	Straight
OUR-PELGRM	1500*	1500*	3	1280×1080	50	Hight	Real-Ped	Straight

*Number of frames averaged for the redpective number of scenes (normal abnormal)

4 Experimentation and Discussion

4.1 Parameters Settings

Divergence class score value from CNN lies between 0 and 1 and the threshold at the class score is set to 0.65 (65%) to trigger the localization detection process. Residual blobs are filtered out through temporal averaging filtering and spatially through connected neighbors filtering. The number of images for temporal averaging N_{filter} is set to 10 i.e., averaging previous 10 frames OPF and FTLE fields. Spatial filtering is applied by performing an 8-connected neighbor on residual blob(s).

4.2 Localization–Qualitative Evaluation

In qualitative evaluation, we compare our proposed divergence localization scheme with two state-of-the-art divergence localization methods [5] and [1]. The method in [5] estimates divergent centers by placing potential destination points over normal image sequences. For the escape scene, a set of divergent points is initialized and the foreground velocity patches start from the divergent points reach potential destinations are marked as divergent centers. The Method in [1] analyzes intersections of foreground velocity vectors and uses the distance segmentation method and *knn* search to locate the divergent center. Intersections are obtained with the assumption that there are three moving objects in the neighbor of the desired divergent center and straight lines of moving objects should intersect to declare the point as the divergent center.

A qualitative comparison of three methods is shown in Fig. 7. Divergent centers are marked with triangles for three methods and ground truth. Scenarios in the first two rows of Fig. 7 depict divergent center detections at the low-density crowd. Existing state-of-the-art methods accurately detect divergent center's at low crowd density whereas our method shows few false detections at the low crowd density divergence. At low crowd density, the crowd motion shape is broken and produces many residuals after the difference from the reference motion shape. The residual blobs cause marking of false divergent centers by our method. Also, since our method analyzes variations at crowd boundaries, the divergence points detected by our method at low crowd density are slightly offset from the ground truth. The last two images in the third row of Fig. 7 show divergence center detections at the high-density crowd. It can be seen that existing art failed to detect divergence sources at the high-density crowd.

The method at [5] analyzes variations in foreground velocity patches from high (in non-escape) to low (in escape) case. However, at high density, foreground velocity concentration is the same in both behaviors, no divergence center points are reaching potential destinations. However, the method is still able to detect divergence center points at the crowd boundary as patches at boundaries experience variations in

Fig. 7 Divergence localization comparison with state-of-the-art methods. Legends: Red triangle [5], Green triangle [1], Orange triangle our method, Blue triangle Ground truth. The top two rows show divergence localization at low-density scenes. False divergent center detections by our method at 2nd, 3rd image at the first row and 1st image, second row. Existing art failed to detect divergent centers at high density (last two images in the third row) while our method can accurately detect divergent centers

velocity. Similarly, the method in [1] shows poor performance at high-density divergence. Assumption of three moving objects and solving three straight-line equations is possible at low density but, at the high-density crowd. As there can be hundreds of people in the neighbor of a divergent center and solving straight line equations for hundreds of moving objects is computationally very expensive. The method with existing model settings is unable to detect actual divergent center locations at high density. However, the model can be improved by obtaining intersections of a large number of neighboring moving objects with reasonable compute. False detections by existing art on the high-density crowd can be seen in the last two images of the third row in Fig. 7. Our method perform well at high density scenario and can idenfity divergence points at correct locations.

As mentioned earlier our localization algorithm also identifies divergence regions (or divergence mask) evolves temporally. Figure 8 shows three examples of divergence mask detected by our algorithm. The first row is a crowd divergence sequence taken from the PETS2009 dataset, the second row shows people diverging from the normal path of walking over the zebra crossing, and the last row shows people
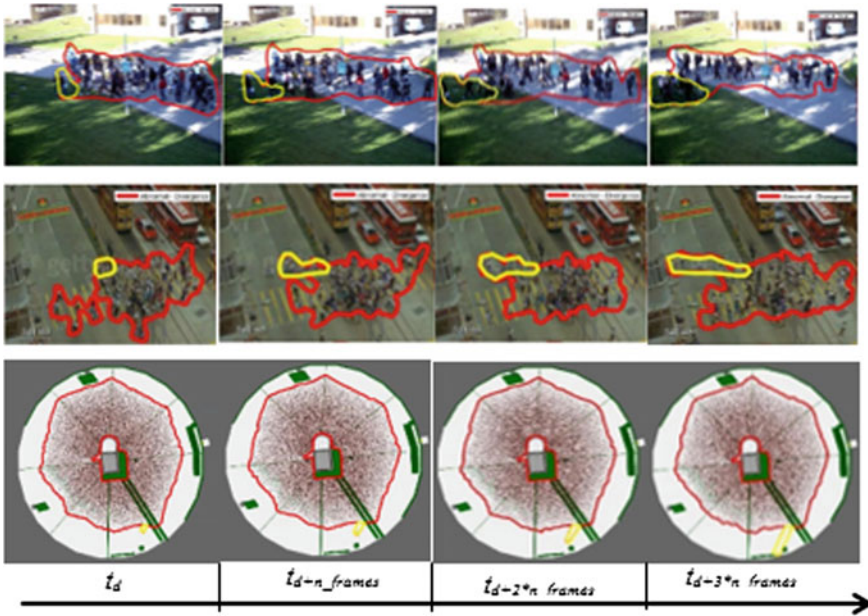
Fig. 8 Divergence region mask changes over time. Divergence regions: first row PETS2009; second-row UCF; and third-row Synthetic Kabbah datasets. The first column is the region at the time of divergence detection (t_d) by CNN. The second column represents divergence region detection n_frames after t_d, the third column represents divergence region detection 2*n_frames after t_d, and so on. The value of n_frames is different for each sequence

diverging from circular Tawaf after completing seven laps of Tawaf. Images in columns show the temporal progression of divergence where images are taken every n_frames after the time instant (t_d) divergence is detected by CNN. Divergence mask size can be seen increasing over time indicating variations in the size of divergence and also shows the direction in which divergence is leading.

4.3 Localization—Quantitative Evaluation

As mentioned earlier our divergence localization algorithm not only identify divergence source points but also detect region(s) of divergence. To qualitatively evaluate the performance of our algorithm, we compute Intersection over Union (IoU) between the predicted divergence region and ground truth divergence region. Ground truth regions are obtained by hand labeling divergence regions at each abnormal frame. IoU score is calculated using Eq. (2),

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{2}$$

where,

Area of overlap is the overlap area between the predicted region and ground truth region.

Area of union is the area encompassed by both the predicted bounding box and ground truth bounding box.

IoU score greater than 0.5 (50% overlap) is generally considered a good prediction by algorithm [14]. Two samples of divergent regions detected overlaid with ground truth divergence region are shown in Fig. 9. Again it is clear from Table 2 that at low-density datasets (UMN, PETS2009, NGSIM), the divergence shape is not smooth and is broken, resulting in a low IoU score, whereas the IoU score at high densities is better meaning our algorithm is better able to detect divergent regions at high crowd densities. IoU scores for divergent datasets used in this work are provided in Table 2.
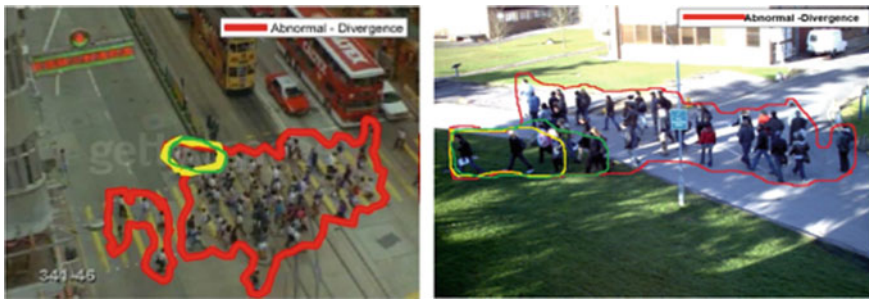
Fig. 9 Examples of divergence region by our method (yellow) compared to ground truth (green)

Table 2 IoU score obtained by our method for divergence regions

Dataset	IoU score
UCF	0.70
UMN	0.55
PETS2009	0.63
NGSIM	0.68
SYN-KABBAH	0.82
SYN-LOVEPARADE	0.89
CONCERT	0.76
PILGRIM	0.80

* IoU score is averaged for dataset contain multiple sequences

4.4   Crowd Density Estimation Inside Localization Mask

We use pre-trained CSRNet from [10] on high-density crowd images from Kabbah and Hajj and perform experimentation on benchmark datasets. Figure 10 shows a crowd density map and crowd count inside divergence masks for various datasets. The crowd count inside the divergence mask is compared with the ground truth count.

4.5   Angle of Localization

In this section, experimentation is performed on divergence at low-density datasets (UCF and PETS2009) and divergence at high-density crowd datasets (PILGRIM and CONCERT).

Figure 11 shows the normal crowd motion mask (green color) and divergent crowd mask (red color) for UCF. Tracklets obtained after the particle advection process are also plotted on the same image (different segments tracklets superimposed on a single image). Noisy tracklets whose length is shorter than a given threshold δ are removed by computing the Euclidean distance between start and end points i.e., $\left(x_i^1, y_i^1 \right) - \left(x_i^T, y_i^T \right)_2 < \delta$. Dominant flows inside each mask are also shown with start and endpoint market with circle and arrow respectively.

One dominant flow from the normal mask and a single dominant flow of the divergent mask is selected for subsequent computation of divergence angle. Initially, the particle angle $\emptyset_i$ for all particles in a dominant flow are computed from interpolated u and v velocities of OPF. $\emptyset_i$ are then mapped on the unit circle to find the exact quadrant in which the dominant flow lies. Figure 11 shows the mapping of each particle circular angle $\emptyset_i$ of a dominant flow onto the unit circle (first column) with a mean circular angle (angles in radian). It also shows the quadrant in which particular dominant flow lies and hence shows the direction of the normal and diverging crowd. Similarly, the second column represents a circular angle histogram (in degrees) with a mean circular angle plotted as well. The angle of divergence (Δ) is computed by taking the difference of mean circular angle of normal dominant flow ($\overline{\emptyset}_n$) and mean circular angle of divergent dominant flow ($\overline{\emptyset}_d$) i.e., $\Delta = \left| \overline{\emptyset}_n - \overline{\emptyset}_d \right|$. Figure 11 also shows Δ computed for UCF divergent scenario found to be diverging crowd drifts aways 24.82° from the actual path of the normal walking crowd. Results for PETS, PILGRIM, and CONCERT datasets are shown in Fig. 12.

5   Conclusion

In this paper, we introduce a more extended view of looking into anomaly localization to make this information more useful for real-life rescue resource deployments. a localization mask is initially generated and features of source/starting point, crowd
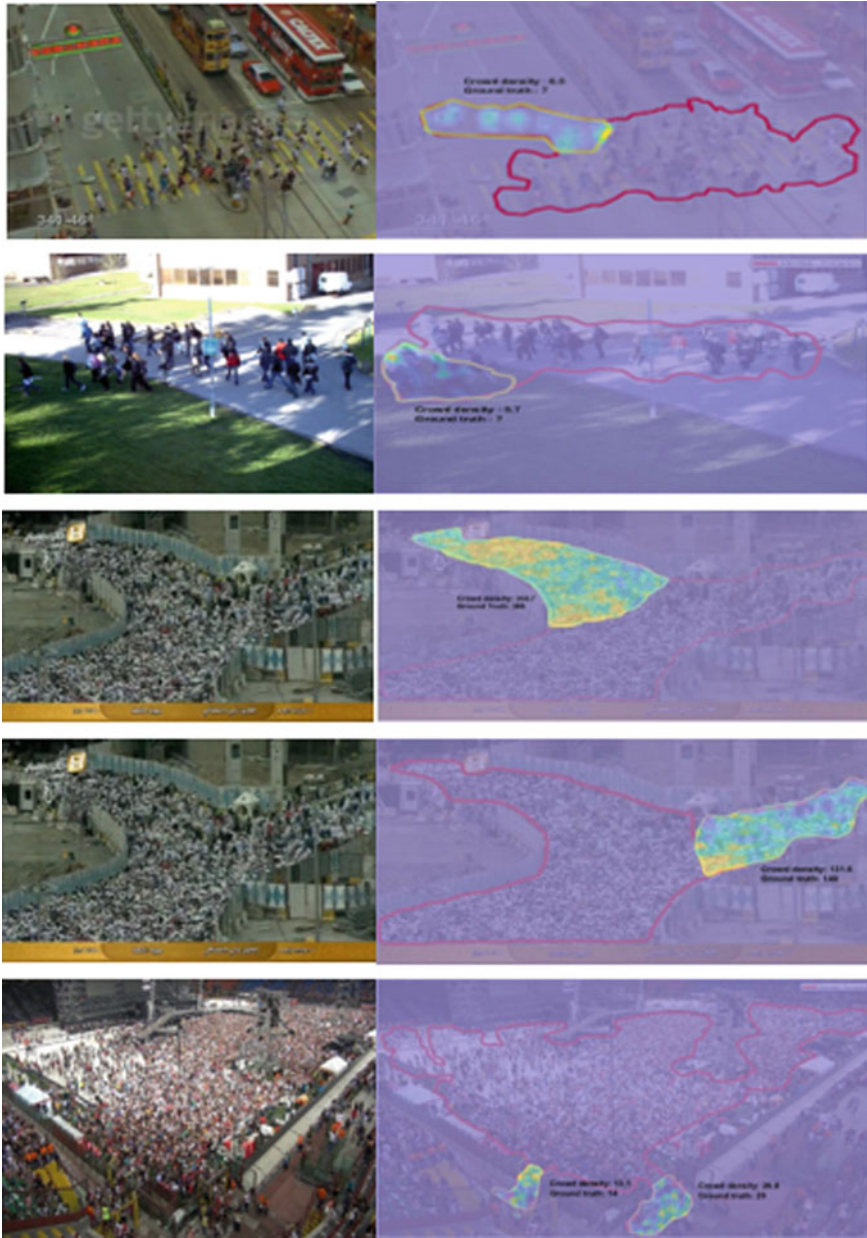
Fig. 10 Crowd density map and density count inside divergence mask. 1st row: UCF; 2nd row: PETS2009; 3rd row: PILGRIM left-divergence; 4th row: PILGRIM right divergence; 5th row: two divergences at CONCERT
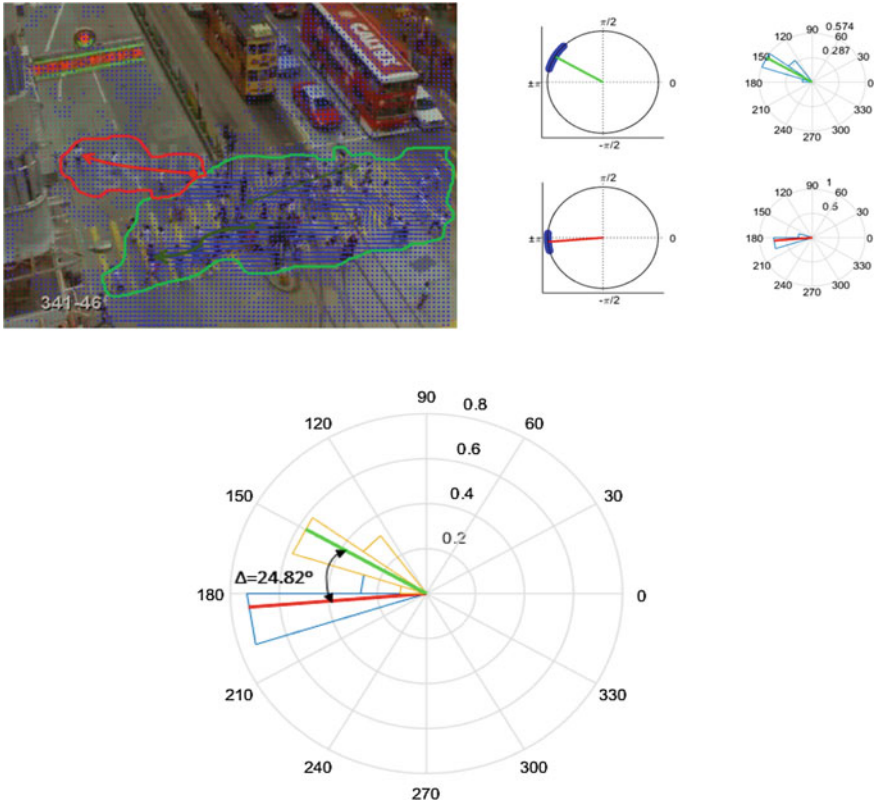
Fig. 11 Divergence (localized) direction and angle estimation for UCF dataset

density estimation, and direction/angle of the localized region are computed. Experimentation results show that the proposed method outperforms the existing state-of-the-art method for source/starting point detection at the high-density crowd and CDE and direction/angle estimations are close to the ground truth.
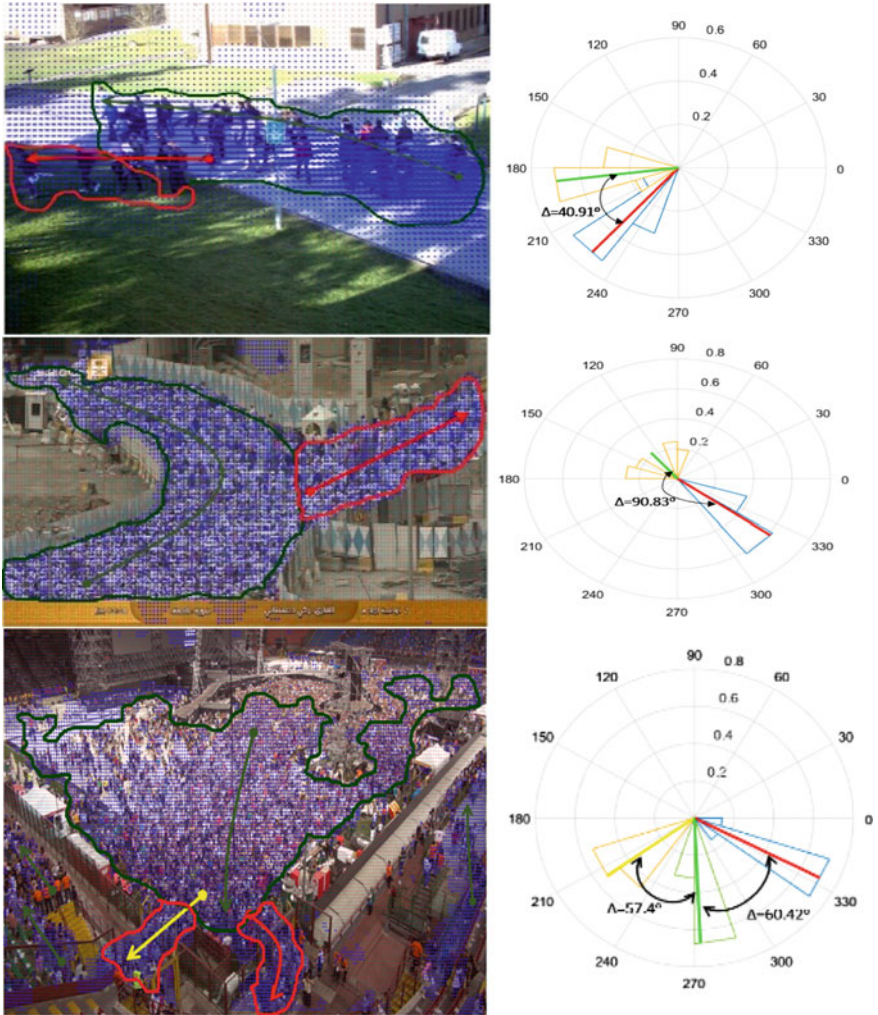
Fig. 12 Direction and angle identification of localized region for PETS (1st row), PILGRIM (2nd row), and MELAN (3rd row)

References

1. Chen CY, Shao Y (2015) Crowd escape behavior detection and localization based on divergent centers. IEEE Sens J 15(4):2431–2439. https://doi.org/10.1109/JSEN.2014.2381260

2. Ongun C, Temizel A, Temizel TT (2014) Local anomaly detection in crowded scenes using finite-time lyapunov exponent based clustering. 11th IEEE international conference advance video signal-based surveillance, AVSS, pp 331–336. https://doi.org/10.1109/AVSS.2014.691 8690.

3. Ren WY, Li GH, Chen J, Liang HZ (2012) Abnormal crowd behavior detection using behavior entropy model. In International conference on wavelet analysis and pattern recognition, pp 212–221. https://doi.org/10.1109/ICWAPR.2012.6294781

4. Wu S, Moore BE, Shah M (2010) Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, pp 2054–2060

5. Wu S, Wong HS, Yu Z (2014) A bayesian model for crowd escape behavior detection. IEEE Trans Circ Syst Video Technol 24(1):85–98. https://doi.org/10.1109/TCSVT.2013.2276151

6. Chaker R, Al Aghbari Z, Junejo IN (2017) Social network model for crowd anomaly detection and localization. Pattern Recognit 61:266–281. https://doi.org/10.1016/j.patcog.2016.06.016

7. Mehran R, Moore BE, Shah M (2010) A streakline representation of flow in crowded scenes. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics), vol 6313 LNCS, no PART 3, pp 439–452. https://doi.org/10.1007/978-3-642-15558-1_32

8. Muhammad Farooq U (2020) Divergence behavior detection and localization using Motion Shape Image (MSI). Universiti Teknologi PETRONAS

9. Krausz B, Bauckhage C (2012) Loveparade 2010: automatic video analysis of a crowd disaster. Comput Vis Image Underst 116(3):307–319. https://doi.org/10.1016/j.cviu.2011.08.006

10. Li Y, Zhang X, Chen D (2018) CSRNet: dilated convolutional neural networks for understanding the highly congested scenes. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit February: 1091–1100. https://doi.org/10.1109/CVPR.2018.00120

11. https://github.com/ZhengPeng7/CSRNet-Keras

12. Khan SD, Bandini S, Basalamah S, Vizzari G (2016) Analyzing crowd behavior in naturalistic conditions: Identifying sources and sinks and characterizing main flows. Neurocomputing 177:543–563. https://doi.org/10.1016/j.neucom.2015.11.049

13. Berens P, Velasco MJ (2009) Journal of statistical software. Wiley Interdiscip. Rev. Comput. Stat. 1(1):128–129. https://doi.org/10.1002/wics.10

14. Ahmed F, Tarlow D, Batra D (2015) Optimizing expected intersection-over-union with candidate-constrained CRFs. Proc IEEE Int Conf Comput Vis Inter 1850–1858. https://doi.org/10.1109/ICCV.2015.215

EEG Based Brain Controlled RC Car with Attention Level

Muhammad Talha Ejaz (ORCID), **Ammara Zahid** (ORCID), **and Muhammad Mudassir Ejaz** (ORCID)

Abstract The brain-controlled interface is gaining popularity in the academic and research industry due to its promising performance in many fields. In this research work, a prototype is mentioned to help the Quadriplegic patients move the wheelchair from their thinking decision. The proposed method used the BCI technology that takes the input data in an EEG signal using a Neurosky headset. The signal is then processed and classified into five significant actions. The movement of the remote-controlled (RC) car is controlled through attention levels. We incorporate the Internet of Things (IoT) for wireless communication that transmits the data using an android application. For safety purposes, we have designed a mobile application that allows the user to control the RC car manually. The proposed method was implemented on the RC car, and the attention level of 5 different subjects was recorded for 3 min. A video of our experiments can be found at https://www.youtube.com/watch?v=UzPGdy54AZw.

Keywords EEG · BCI · Nuerosky · Quadriplegia

1 Introduction

Spinal Cord Injuries (SCI) are increasing rapidly. According to the study of the National Spinal Cord Injury Statistical Center (NSCISC) at UAB. The most prominent cause of SCI is vehicular accidents, which are approx. 39.3%, then by falling 31.8% and so on. One of the US studies shows that around 1.25 million died in road

M. Talha Ejaz (✉)
Department of Mechatronics Engieering, PAF—Karachi Institute of Economics and Technology, Karachi, Pakistan

A. Zahid
Department of Biomedical Engineering, NED University of Engineering and Technology, Karachi, Pakistan

M. Mudassir Ejaz
Department of Electrical and Electrincs Engineering, Universiti Teknologi PETRONAS (UTP), Seri Iskandar, Malaysia

accidents in the world [1]. Quadriplegia is generally caused by road accidents, which becomes a global catastrophe and leading cause of death by injury. Quadriplegia is one of the spinal cord injuries in which the lower body becomes paralysis. It is caused by damage to the spinal cord at a level of C1–C7 or sometimes brain injury.

Quadriplegia caused several complications, including pain, blood clots, pressure sores, related injuries, spastic muscles, respiratory problems, automatic dysreflexia, and loss of bladder and bowel control. The most important complication is the loss of the motor sensing control of limbs and sometimes hands also. These complications make a person handicapped. Different types of wheelchairs, including manual hand-controlled, hand gesture-controlled, voice recognition control, eye motion, finger movement control, and remote control, are available in the market. Still, some of them are not feasible for Quadriplegia patients, and some are expensive.

Brain-controlled wheelchairs are designed for those patients who are unable to use the powered wheelchairs for mobility. After the successful advancements in robotics, artificial intelligence, and sensor technology, more advanced methods come into existence for designing a smart wheelchair. EEG based controlled wheelchairs are very popular nowadays as they are cost-effective and easy to operate. EEG or electroencephalography is a process to record and detect brain activity in terms of signals in real-time. The human brain comprises 100 billion neurons, and signals are transmitted from one neuron to another through electric impulses. Alpha, Beta, Theta, Delta, and Gamma are five different waveforms found in the human brain, and each one is different from another in terms of frequency and amplitude. Electrodes are attached on a scalp, and conductive gel is used to get better results to record or monitor an EEG signal. It takes around 30 min for the setup, which is not suitable to drive a wheelchair. To consider this issue, Neurosky is being used to detect the brain signals, which does not need any gel before using it, and this method is fast and cost-effective. However, this method is not accurate as detecting the signals using electrodes with gel.

Brain-Computer Interface (BCI) is an approach where a physical device is operates using brain activity. Attention level and eye blink are the main methods used in the BCIs application. In [2–4], attention level was used to analyze human activities such as sleep patterns and recognized the degree of human attention. When a person blinks an eye, a deflection in a waveform is detected. By capturing that deflection, several applications were proposed in [5–8]. In [9, 10], methods were discussed for the wheelchair for quadriplegia patients using the head movement and retina movement utilizing the image processing and accelerometer technique. These methods were not suitable for a person, as it causes fatigue while moving the head. A neuro handset is used to record the attention level for driving a wheelchair that mitigates the issues. It makes a system faster and allows the patient to control the wheelchair by focusing the directions.

In this paper, we proposed a conceptional idea where RC car is drive using the attention level. This method helps the handicapped especially quadriplegia patients to move from one place to another on their own. We also incorporate IOT in the proposed method, where signals are transmitted to an android application.

The rest of the paper is organized as follows. Section 2, covers the background knowledge of different types of EEG signals and recording methods. The proposed methodology and results are discussed in Sects. 3, 4 respectively. Section 5 covered the conclusion.

2 Background

Electroencephalography, abbreviated as EEG, is a process to monitor the brain activity of a brain. Signals are taken using the electrodes attached to a scalp. Different EEG headsets are used to monitor the different number of channels. It is currently widely used for Brain Controlled Interface (BCI) applications as it is a non-invasive method with high time resolution and a low cost. Countless research [11, 12] has been conducted for the Motory Imagery BCIs, where users imagine the movements in a brain, generates a signal that capture by an EEG headset. More the channels in the headset mean more accurate results will be. The cells in the brain called neurons produced an action potential by communicating with each other. The opening and closing of the ion channels (Na^+ and K^+) in the neuron's membrane cause this discharge known as the action potential.

One main advantage of using an EEG headset is that it is less weight and a portable device, allowing users to wear it and control the wheelchair. However, two main challenges are faced by EEG signals, such as artifacts caused by the noise and pose determination. EEG signals are mostly disturbed when the noise is present in the surroundings, and it also affects the position of a body.

Alpha wave (α) is generated when a person is relaxed and profoundly found in the frontal lobe or at the temporal lobe. It is more activated when a person thinks while eyes are open, and the range of the frequency starts from 7.5 Hz to 12 Hz. Beta wave (β) has the second-highest frequency, ranging from 12 to 30 Hz, and activated during the stress and working a hazard scenario. It is further classified into two bands, such as $\beta1$ and $\beta2$. It is also visible during sleep and rapid eye movement. Theta wave (θ) is mainly shown when a person is emotionally disturbed and the time when a person awake from a deep sleep. It lies between 3.7 and 7.5 Hz. Delta wave (δ) has the lowest frequency bands, and it is activated when a person becomes unconscious or thinks very deeply during sleeping. The frequency range of this waveform iss 0.5–3.5 Hz [13]. A comparison is mentioned in Table 1.

Signal acquisition of MI BCIs is not expensive, but it's a challenging task. EEG signals are not stable, which, due to which sometimes it fails to classify the wave, and many classifiers fail to consider time-series information [14]. The general structure of MI BCIs application is depicted in Fig. 1.

Table 1 Frequency description of EEG bands

Type	Frequency (Hz)	Location	Normally
Alpha (α)	7.5–12	Posterior regions of head, both sides, higher in amplitude on dominant side. Central sites (c3–c4) at rest	Relaxed/reflecting closing the eyes
Beta (β)	12–30	Both sides, symmetrical distribution, most evident frontally. Low amplitude waves	Alert/working active, busy or anxious thinking, active concentration
Theta (θ)	3.7–6.7	–	Young children drowsiness or arousal in older
Delta (δ)	0.5–3.5	Frontally in adults, posteriorly in children; high amplitude waves	Adults slow wave sleep in babies
Gamma (γ)	30–100	–	Certain cognitive or motor functions
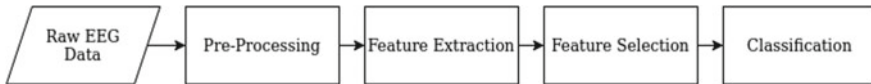
Fig. 1 A signal processing flowchart of MI BCIs applications

3 Methodology

The novelty of the proposed method is that Neurosky was used as a sensor. It is an EEG headset for acquiring the EEG signal. This setup not reduced the cost of the overall system but increase efficiency. Neurosky is a one channel band with a sampling rate of 512 Hz. It is the only sensor used to drive an RC car in four directions: right, left, forward, and backward. The specification of Neurosky (Mind wave) is mentioned in Table 2.

Neurosky is placed on a frontal side of a head precisely at the FP1 location. It has a dry electrode, which senses the EEG and gives the value during attention and eye blink. The Neurosky and the site where it is placed is shown in Fig. 2.

We dealt with the attention level of a person. Attention was calculated by acquiring a signal from a device and passed to the signal processing unit. EEG is recorded as a sampling frequency of 512 Hz with a 16-bit quantized level. Fast Fourier Transform (FFT) is being used to convert time signal into a frequency domain with a low pass filter with a cutoff frequency of 50 Hz.

Table 2 Frequency description of EEG bands

Channels	Electrode type	Bandwidth	Sampling rate	Communication	Weigth
1	Dry	3–100 Hz	512 Hz	Bluetooth	90 g
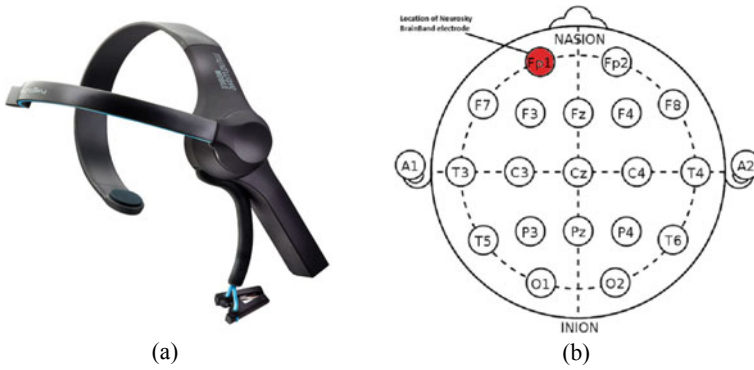
Fig. 2 **a** shows the Neurosky headset and **b** depicts the position where Neurosky attached to the scalp

3.1 Design Flow

In this section, we are discussing the hardware and software flow of the proposed method. (Appendix shows the hardware and software flow of a system). The flow starts by calibrating the headset, as all the movement depends upon the data acquired from the headset. After obtaining the data, it was propagated to the micro-controller to perform the actions. L298N motor driver was also attached to the microcontroller that received the command and then powered the motors for the actions. The actions were also monitored in the mobile application, which was designed in an android studio for this particular application. For communication between the RC car and application was done by the Wi-Fi module (ESP8266). In the proposed method, four basic movements were considered: forward, backward, left, and right. One more additional command, stop, was considered for the safety purpose. The RC car moved when the attention level was greater than 10. The actions were classified as mentioned in Table 3.

Table 3 Action classification according to the attention level	Attention level	Action
	$10 \leq$ Attention level	Stop
	$10 \leq$ Attention level ≤ 40	Forward
	$40 \leq$ Attention level ≤ 50	Reverse
	$50 \leq$ Attention level ≤ 80	Right
	$80 \leq$ Attention level ≤ 100	Left

4 Experimental Results

We had conducted the experiments on the RC car, we named it as SpyShip. Figure 3 shows the model of SpyShip. The communication between the headset and actuators of the car was done wirelessly using a Bluetooth module. The attention level was measured for three minutes of 5 subjects, and they were asked to focus on the directions to move the car. Figure 4 depicts each subject's attention level curves at each time step. The result shows that the proposed method can capture the attention level without any disturbance and then derived the car in real-time.

Figure 5 shows that the right direction was chosen mostly by all the subjects and then the forward direction. The maximum value of the right direction taken by a subject is 108. Whereas for forwarding is 75, backward is 52, and for the left is 62. The mean value of each direction of five different subjects is calculated and depicted in Table 2. The average mean value of the right direction is 83.4, and the p-value of the data set is 0.0000367. Figure 6 shows the Android application interface and the action selected by the subject.

The subjects mostly have their attention level between 50 and 80. It can also be seen that the proposed method can capture the data in real-time and then classify

Fig. 3 SpyShip, side view of the RC car
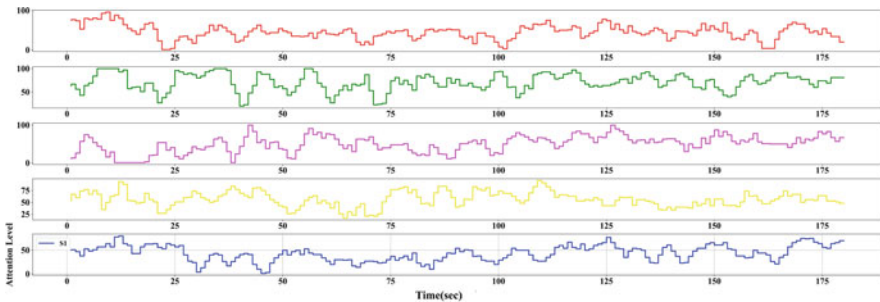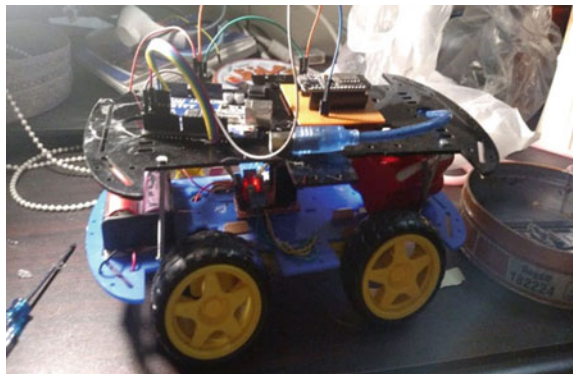
Fig. 4 Attention level of each subject at each time step
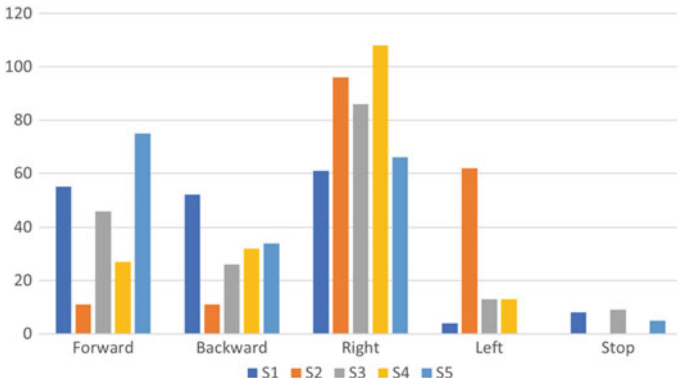
Fig. 5 Illustration of actions selections by each subject

(a)   (b)
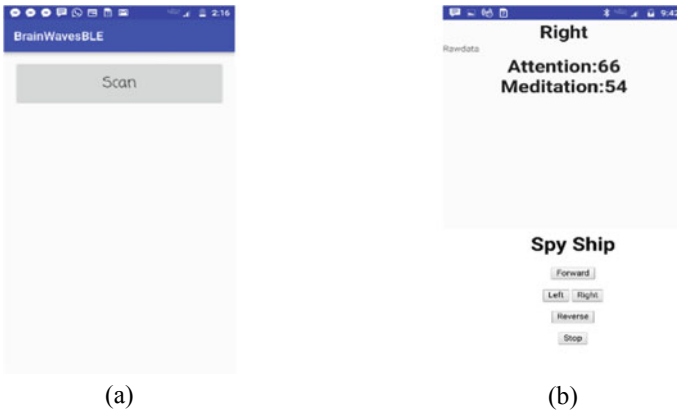
Fig. 6 a shows the Andriod application interface and **b** depicts the action selected by the subject

it accordingly. For wireless communication, we have designed an android app that transmits and receives the data in real-time. It was possible by using the NodeMCU that supports ESP8266 SOC. It acts as a LAN HTTP server or LAN web server. This research work used to convert IP base links into serial communication to process in both station and hotspot mode. (see Appendix).

5 Conclusion

In this article, a conceptual idea is presented that can be used for Quadriplegic patients to control the wheelchair using the EEG signals. The signals have extracted using the Neurosky headset placed at the FP1 position of the user's head. Actions are classified based on attention levels such as forward, backward, left, right, and stop. The proposed method was tested on five different subjects and results shown that all the subjects were able to classify the actions according to the attention level. An IoT-based approach was also used in the proposed method that transmits and receives the data wirelessly with the help of NodeMCU. An android app was also developed that shows the actions, and again, a user can control the car manually if malfunctions happen. In the future, the idea will be implemented in the wheelchair and also incorporate the GPS for tracking.

Appendix

See Figs. 7 and 8; Tables 4 and 5.
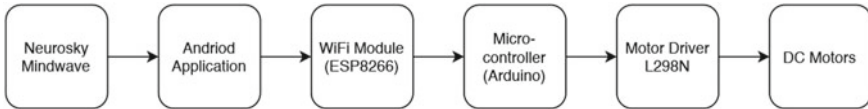
Fig. 7 Hardware flow

Fig. 8 Software flow
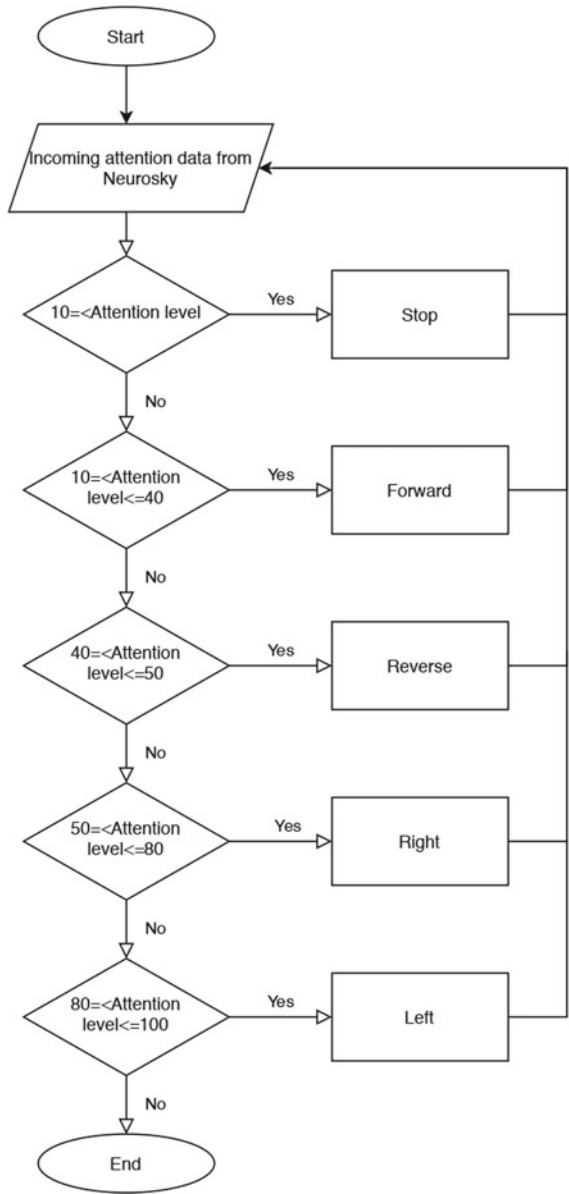
Table 4 Station mode IP base link to serial characters

No	Links	Serial output
1	http://192.168.1.12/?pin=Forward	F
2	http://192.168.1.12/?pin=Backward	B
3	http://192.168.1.12/?pin=Right	R
4	http://192.168.1.12/?pin=Left	L
5	http://192.168.1.12/?pin=Stop	S

Table 5 Hotspot mode IP base link to serial characters

No	Links	Serial output
1	http://192.168.43.12/?pin=Forward	F
2	http://192.168.43.12/?pin=Backward	B
3	http://192.168.43.12/?pin=Right	R
4	http://192.168.43.12/?pin=Left	L
5	http://192.168.43.12/?pin=Stop	S

References

1. World Health Organization Website (2017) [Online]. Available: http://www.who.int/mediacentre/factsheets/fs358/en/
2. Liu NH, Chiang CY, Chu HC (2013) Recognizing the degree of human attention using EEG signals from mobile sensors. Sensors 13(8):10273–10285
3. Van Hal B, Rhodes S, Dunne B, Bossemeyer R (2014) Low-cost EEG-based sleep detection. In: Proceeding of 36th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 4571–4574
4. Poltavski DV (2015) The use of single-electrode wireless EEG in biobehavioral investigations. Methods Mol Biol 1256:375–390
5. Jiang Y; Lee H, Li G, Chung W-Y (2016) High performance wearable two-channel hybrid BCI system with eye closure assist. In: Proceeding. of 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 5869–5872
6. Ang MS, Zhang ZG, Hung YS, Mak JNF (2015) A user-friendly wearable single-channel EOG-based human-computer interface for cursor control. In: Proceeding of IEEE 7th international IEEE/EMBS conference on neural engineering (NER), pp 565–568
7. Varela M (2015) Raw EEG signal processing for BCI control based on voluntary eye blinks. In: Proceeding. of IEEE thirty fifth central american and panama convention (CONCAPAN XXXV), pp 1–6
8. Szibbo D, Luo A, Sullivan TJ (2012) Removal of blink artifacts in single channel EEG. In: Proceeding of annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 3511–3514
9. Baig MZ, Aslam N, Shum HPH, Zhang L (2017) Differential evolution algorithm as a tool for optimal feature subset selection in motor imagery EEG. Expert Syst Appl 90:184–195
10. Oikonomou VP, Georgiadis K, Liaros G, Nikolopoulos S, Kompatsiaris I (2017) A comparison study on EEG signal processing techniques using motor imagery EEG Data. In: Proceeding of the IEEE 30th international symposium on computer-based medical systems (CBMS), Thessaloniki, Greece, 22–24 June 2017
11. Cheng D, Liu Y, Zhang L (2018) Exploring motor imagery EEG patterns for stroke patients with deep neural networks. In Proceeding of the IEEE international conference on acoustics, speech and signal processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018

12. Awais MA, Yusoff MZ, Yahya N, Ahmed SZ, Qamar MU (2020) Brain controlled wheelchair: a smart prototype. J Phys: Conf Seri 1529(4):042075
13. Khan D, Kamel N, Muzaimi M, Hill T (2020) Effective connectivity for default mode network analysis of alcoholism. Brain Connectivity
14. Guo X, Wu X, Gong X, Zhang L (2013) Envelope detection based on online ICA algorithm and its application to motor imagery classification. In Proceeding of the 6th international IEEE/EMBS conference on neural engineering (NER), San Diego, CA, USA, 6–8 November 2013

Automated Pill and Syringe Dispenser

R. Paviya, S. Prabakar, K. Porkumaran, and Abu Bakar Sayuti Saman

Abstract This paper presents the Automated Fabrication of Pill and Syringe Dispenser. As it is necessary to provide medication for the elderly people on time a medication remainder is already developed. So, in addition to help the diabetic patient, automatic pill and syringe dispenser is recommended. It is designed in such a way that, based on the real time clock it dispenses the pill and syringe for the patients on required time and it acts as a reminder. In addition to this, automatic syringe dispenser is developed, which loads the insugen automatically. The major components of the combination of pill and syringe dispenser are DS3231 RTC, Arduino UNO, Servo motor, LCD, LED, Buzzer, and in addition to this, the software called BLYNK, is used to pass the information through Bluetooth and the medication is dispensed on the required time for the elderly ones. And the components used for automatic syringe dispenser are Arduino UNO, lead screws, L293d, DVD stepper motor board, servo motor, piezo buzzer.

Keywords Medication dispenser · Automatic dispenser · Pill and Syringe · Micro controller · Servo motor · Alzheimer disease · Android application · Syringe dispenser

R. Paviya (✉)
N.G.P Institute of Technology, Kalapatti road, Coimbatore, India

S. Prabakar
Sona College of Technology, Salem, India
e-mail: srisornaprabu@gmail.com

K. Porkumaran
Sri Sairam Engineering College, Chennai, India

A. B. S. Saman
Universiti Teknologi PETRONAS, Seri Iskandar, Perak, Malaysia
e-mail: sayuti@utp.edu.my

1   Introduction

As there are many elderly individuals who need constant help, to take their medications on time and to prevent any illness timing is must. So Pill Dispenser has been developed. Also, many elders face problems like Alzheimer's, poor vision and Diabetics etc.., and most of the death occur due to the intake of wrong medications.

So, in addition to help the diabetic patient, Syringe dispenser is combined with the Pill dispenser. In this combination of pill and syringe dispenser the patient themselves has to load the insugen. So, the extension of Syringe dispenser is made. Where the syringe is automatically loaded to the insugen and reminds the people to take their drug on time.

1.1   Pill Dispenser

These are items which release pills at specified times.

- These devices are proposed in a motive to help the people with Alzheimer disease. The main function of pill dispenser is to remind the patients to take their medications on time.
- When the predetermined time comes to take the medication, the device automatically releases the pill and it sounds a loud warning signal that it is time to take the medication.

1.2   Syringe Dispenser

These are items which release syringe at specified times. And automatic Syringe dispenser is developed. This mainly helps in loading the insugen for the required amount and it dispenses the syringe at the predefined time. This is done with the help of stepper motor, servo motor, lead screws, L293d IC, 3D print software for holding the syringe.

- Their purpose is to help the Diabetic patients those who are insulin dependent, to take their medications on time.
- It dispenses Syringe at a set interval of time, especially for the Diabetic patient.

2   Literature Survey

(i) "Smart drugs: Improving Healthcare using Smart pill box for Medicine remainder and monitoring system" Diaa Salama Abdul Minaam, Mohamed Abd-elfattah [1].

The paper "Smart drugs: Improving Healthcare using Smart pill box for Medicine remainder and monitoring system" published in the journal Science Direct in 2018. This has the concept, design and creation of pill box prototype which has nine separate sub boxes to solve the deficiency in medical area. It enables the care takers to take the determined amount of pill and timing to take pill on every day. It requires the client to stack the box consistently in order to refill it with medicines.

(ii) "Smart Medication Dispenser" Suraj shinde, Nitin bange, Monika kumbhar, Snehal patil [2].

The paper "Smart Medication Dispenser" published in the International Journal of advanced research in Electronics and Communication Engineering (IJARECE) in 2017. This uses 4×4 matrix keypad, GSM module, LCD display and real time clock (RTC) and this helps to provide information automatically to patients to take their right dosages according to the prescription at appropriate time.

(iii) "Construction of a Smart Medication Dispenser with High Degree of Scalability and Remote Manageability" Jugeon Pak, Keehyun Park [3].

The paper "Construction of a Smart Medication Dispenser with High Degree of Scalability and Remote Manageability" published in the Journal of Biomedicine and Biotechnology in 2012. In this each medication have a medication tray dispenser. The time is set in RTC and when the pre-determined time reaches the time, it gives some kind of notification and then the user has to push a button and the medication is dispensed from the medication tray (MDT).

3 Methodology for Both Pill and Syringe Dispenser

3.1 Introduction

For the combination of pill and syringe dispenser same methodology is used and this dispenses the syringe and pill and it reminds the patient to take their drug on time and in this, the patient has to load the insulin into the needle.

3.2 Block Diagram

Construction: For both the pill and syringe dispenser, same methodology is used. The Arduino UNO has two-way input, one from the power supply and another from the 'BLYNK' app via Bluetooth. The time is set in BLYNK as well as in RTC and it is compared. When the RTC (DS3231) reaches the BLYNK time, it starts dispensing the medications on the predefined time. Thus both the pill and syringe is released for the patients use, for the required time (Fig. 1).

Working: Once the power supply is fed into Arduino UNO and the information from BLYNK is received, and it sends the signal to DS3231 and it again receives as
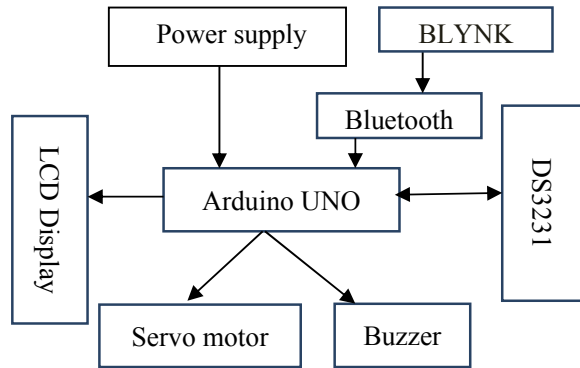
an input to the Arduino UNO and then the Arduino processes the program to run a servo in order to dispense the pill and syringe on the required time. An Alarm gives off with the buzzing action which is also controlled by the Arduino and then the LED glows.

4 Methodology for Automatic Syringe Dispenser

4.1 Introduction

In this, the extension of syringe dispenser is made as, the Syringe holder is made by using the 3D printer and the syringe moves forward and backward as the syringe holder is placed on the CD driver setup. This setup consists of the lead screws with stepper motor which helps in the linear motion and IC L293d is used, it mainly helps to run 2 DC motors with same IC. It can control both the speed and direction.

4.2 Block Diagram

Construction: Arduino UNO is connected to the IC L293d and also connected to the stepper motors of CD drive. The syringe holder is made by using the 3D printer and the servo setup moves front and back with the help of lead screws in CD driver. And the syringe gets dispensed and it starts to buzzer. This process continues (Fig. 2).

Working: When the power supply is fed into Arduino UNO and then it initiates the IC L293d and the IC starts moving the CD driver motor forward and backward with the help of lead screws and the syringe moves forward into the insugen bottle and then the back holder of syringe moves backward, so that the insugen is sucked inside the syringe and then both the setup moves backward and then the micro servo

Fig. 2 Block diagram of automatic syringe dispenser

```
                              ┌─────────┐
                              │  Start  │
                              └────┬────┘
                                   │
              ┌────────────────────┴────────────────────┐
              │      Initialize Arduino UNO and L293d    │
              └────────────────────┬────────────────────┘
                                   │
                              ◇ Timer              NO
                              ◇ If t=y? ─────────────┐
                                   │ YES
        ┌──────────────┐    ┌──────────────┐    ┌──────────────┐
        │ Servosetup1, │───▶│ Servo setup  │───▶│  Servo run   │
        │  2 forward   │    │  2 reverse   │    │              │
        └──────────────┘    └──────────────┘    └──────────────┘
        ┌──────────────┐    ┌──────────┐    ┌──────────────────┐
        │   Buzzer     │───▶│   Stop   │◀───│ Dispense syringe │
        └──────────────┘    └──────────┘    └──────────────────┘
```
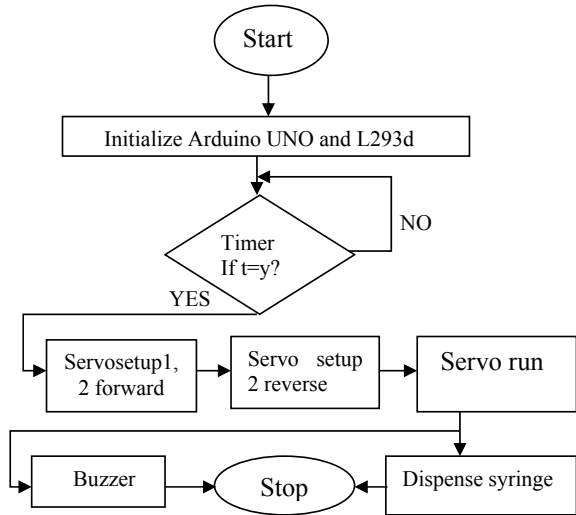
pushes the syringe out and it starts buzzing off and notifies the patient to take the syringe.

5 Results and Discussion

5.1 Overall Performance

The overall performance of this device is good. First the combination of both pill and syringe dispenser (Fig. 3) has been developed. This helps to remind the patients or the care taker to take the drug on time. It dispenses the Pill followed by the Syringe. In this the syringe is dispensed without suction of insugen. So an extension of syringe dispenser is done (Fig. 4). In this extension of syringe dispenser, it automatically sucks the insugen with the help of the CD driver setup and IC. And the syringe gets dispensed and alarm goes on and notifies the patient. This mainly helps the patient in not missing their medications and it will not lead to any medication error.

6 Future Improvements and Scope

There are some improvements in this device to be made in future. It includes automatic refilling of syringe after usage, sending message to the caretaker if the medication is not taken this can be done using the GPS, fixing the insugen bottle with the needle, and also an input panel will be set so that the amount of insugen to be sucked
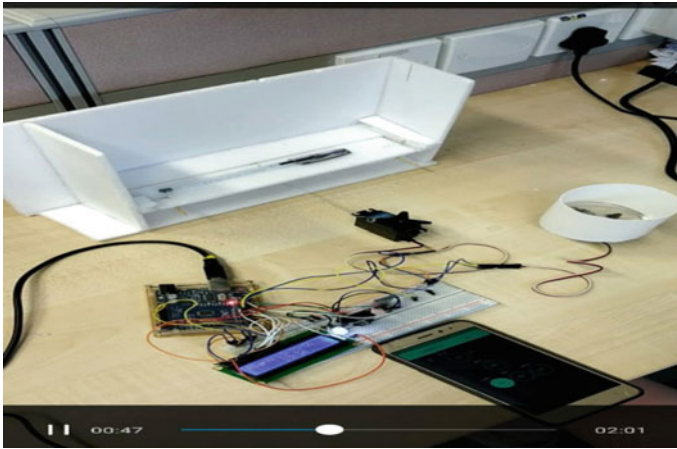
Fig. 3 Combination of pill and syringe dispenser
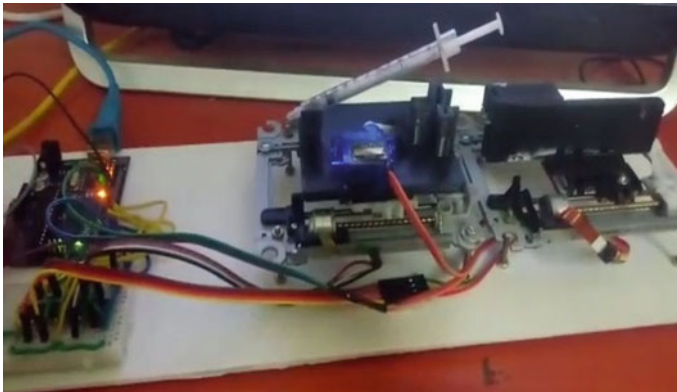
Fig. 4 Automatic syringe dispenser

by the needle can be entered easily and time can be set easily by the care taker. This would be very useful for the patients and also it reduces stress for the care takers.

7 Conclusion

There are many pill boxes available which are bit complicated and most of them are costlier. But the combination of pill and syringe dispensers is not available, so it was developed. In this fabrication of pill and syringe dispenser, with the help of BLYNK app the time is set and when the time reaches the pre-determined time the pill gets dispensed followed by the syringe. In this the syringe is not loaded with insugen.

So, the extension of this process has been developed with the help of CD drivers and IC L293d the syringe automatically loads the insugen and it is released with the help of micro servo. This is user friendly and it helps the patients in not missing the medications. It is less weight and portable, it reduces the medication errors. This can be used in hospitals, homecare, and other old age homes. The main goal of the system is to provide healthy, tension free life to those who are taking pills regularly and to provide it at an affordable cost. This product is easy to design. So, thus it requires less maintenance.

References

1. Abdul Minaam DS, Abd-Elfattah M (2018) Smartdrugs: improving healthcare using smart pill box for medicine reminder and monitoring system. Sci Dir 29 November (2018)
2. Shinde S, Bange N, Kumbhar M, Patil S (2017) Smart medication dispenser of a proceedings paper: IJARECE
3. Pak J, Park K (2012) Construction of a smart medication dispenser with high degree of scalability and remote manageability. J Biomed Biotech

Multi-classification of Brain Tumor Images Using Transfer Learning Based Deep Neural Network

Pramit Dutta, Khaleda Akhter Sathi, and Md. Saiful Islam

Abstract In recent advancement towards computer-based diagnostics system, the classification of brain tumor images is a challenging task. This paper mainly focuses on elevating the classification accuracy of brain tumor images with transfer learning based deep neural network. The classification approach is started with the image augmentation operation including rotation, zoom, horizontal flip, width shift, height shift, and shear to increase the diversity in image datasets. Then the general features of the input brain tumor images are extracted based on a pre-trained transfer learning method comprised of Inception-v3. Finally, the deep neural network with 4 customized layers is employed for classifying the brain tumors in most frequent brain tumor types as meningioma, glioma, and pituitary. The proposed model acquires an effective performance with an overall accuracy of 96.25% which is much improved than some existing multi-classification methods. Whereas, the fine-tuning of hyper-parameters and inclusion of customized DNN with the Inception-v3 model results in an improvement of the classification accuracy.

Keywords Image augmentation · Transfer learning · Inception-v3 · Deep neural network · Brain tumor classification

1 Introduction

A brain tumor is defined as an uncontrolled and unnatural growth of neural cells. According to the world cancer report, in this year around 18,020 adults may die from primary brain cancerous disease [1]. Therefore, the early classification of brain tumors into their particular types plays an imperative role to treat the tumor efficiently for reducing the human death rate. In this case, the implication of deep learning methods to classify the tumor images can accelerate the treatment process more effectively. Recently numerous researches have been conducted on deep learning-based classification method to increase the classification accuracy of the brain tumor

P. Dutta · K. Akhter Sathi · Md. Saiful Islam (✉)
Department of Electronics and Telecommunication Engineering, Chittagong
University of Engineering and Technology, Chittagong 4349, Bangladesh

images. For instance, Sajjad et al. [2] employed an unsupervised learning method called the convolutional neural network (CNN) algorithm for the classification of the brain tumor images in different classes. The accuracy was found to be almost 94.58% for the classification of the multiple categories of tumors. Moreover, Amin et al. [3] employed a fusion process using the discrete wavelet transform (DWT) method to extract a more informative tumor region. Then the noise removal process was applied based on a partial differential diffusion filter (PDDF) before segmentation. After that, the CNN model was utilized for classifying the tumors as cancerous and non-cancerous. In another study, Sultan et al. [4] developed an approach for the multi-classification of brain tumors. The classification of the brain tumors images in multiple classes was based on the CNN algorithm. The accuracy was found to be almost 96.13% for the classification of the multiple categories of tumors. Abiwinanda et al. [5] proposed a classification model based on CNN for the multi-classification of brain MRI images. The architecture of the CNN model consisted of different layers i.e., convolution, max-pooling, flattening, and fully connected one hidden layer. The classification method based on CNN provided a classification accuracy of approximately 94.68%.

The major contribution of this paper is to develop an efficient classification method using CNN with the aid of the inception-v3 transfer learning method. Moreover, a suitable learning rate and batch size are employed to make the designed model robust and also faster and smooth the training process. In addition, the effectiveness of the designed method is analyzed by comparing it with some existing classification methods in terms of classification accuracy.

The rest of the paper is organized as follows: Sect. 2 provides the methodology of the proposed system with a detailed explanation of each of the steps. Section 3 represents the results with analysis and the comparative study is also conducted in this Section. Finally, Sect. 4 shows concluding remarks.

2 Proposed Multi-classification Approach

The process of the proposed multi-classification method is divided into three steps as illustrated in Fig. 1. In the first step, the image preprocessing is performed using the image augmentation technique to increase the amount of total image dataset. After the preprocessing step, the image features are automatically generated by employing
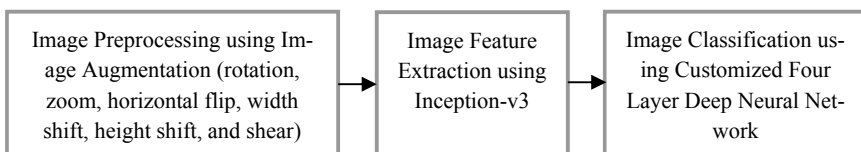
Fig. 1 The process of proposed classification method

the Inception-v3 transfer learning method. Finally, the extracted features are feed into the modified DNN to classify the three brain tumor types.

2.1 Tumor Dataset

The dataset used for this model is obtained from 233 patients with three categories of brain tumor images at different slices (a) 994 axial images, (b) 1045 coronal images, and (c) 1025 sagittal images. This T1-weighted contrast-enhanced image dataset is provided by Cheng [6]. It comprised of 3064 brain tumor MRI images. The datasets are formulated with 1426 meningioma images, 708 glioma images, and 930 pituitary images. Each of the images has a size of 512×512 in pixels. For decreasing the computational time and dimensionality the images are resized to 150×150 pixels. Then the total image datasets are splitting into training (80% of the total dataset) and validation (20% of the total dataset) dataset. The validation dataset is employed to estimate the proposed classifier model. Moreover, the details explanation of each step of the classifier model are described as following.

2.2 Image Augmentation

Before feeding the classifier model, the image augmentation process is performed because of the lower number of the image dataset. For this case, various operation including rotation, zoom, horizontal flip, width shift, height shift, and shear is performed to enhance the diversity of the brain tumor dataset.

2.3 Classifier Model Using Inception-v3 and DNN

The architecture of the proposed classifier model is comprised of Inception-v3and customized deep neural network as shown in Fig. 2a to classify the brain tumor classes effectively.

2.3.1 Feature Extraction Using Inception-v3

After image augmentation, the features are automatically extracted by using the Inception-v3 based transfer learning model. The model utilizes two or three layers of a small convolutional layer based on factorized convolution operation instead of a large convolution layer that reduces the parameter without reducing the efficiency of the model. The factorization process is represented by the block as shown in Fig. 2a. This model also employed a grid size reduction technique for mapping the features

Fig. 2 **a** The architecture of proposed classification model, **b** The detail of block, and **c** The detail of block_1

as shown in Fig. 2b. For this factorization process, the 42 layers Inception-v3 model with fewer parameters is much more efficient than the VGGNet [7]. Moreover, the default weights are used for this particular model. This model contains many filters to detect simple feature that is very effective for classification problem. For this reason, the images are convoluted to extract the desired features.

2.3.2 Classification Using DNN

In order to make the classifier more effective, the output of the Inception-v3 model is coordinated with the customized DNN that results in fine tuning of the transfer learning model. Figure 3 shows the network architecture of DNN with four layers where the first layer is used to attend the output of the inception model and the second layer consisting of 1024 neurons and the next layer is utilized to dropout. Finally, the output layer is designed to generate the output with 3 neurons that represents the brain tumor types of glioma, meningioma, and pituitary respectively.

For training the network, the Adam is employed as an optimizer and categorical cross-entropy is utilized as a loss function with the learning rate of 0.00003 and batch size of 32. The training process of the designed model is performed in the

```
flatten (Flatten)              (None, 11520)        0           mixed8[0][0]

dense (Dense)                  (None, 1024)         11797504    flatten[0][0]

dropout (Dropout)              (None, 1024)         0           dense[0][0]

dense_1 (Dense)                (None, 3)            3075        dropout[0][0]
===========================================================================
Total params: 22,475,427
Trainable params: 22,454,051
Non-trainable params: 21,376
```

Fig. 3 The architecture of deep neural network

Google Colab environment using the Graphics Processing Unit (GPU). The training process takes about 6 min for 19 epochs (callback stop) with 20 s per epoch. During the training, the model occupied 3.06 GB RAM and 2.53 GB GPU in the colab environment.

3   Results Analysis

After completing the training process, the loss and accuracy of the designed classifier model are evaluated at different epochs as shown in Fig. 4. The training and validation loss are found 0.0614 and 0.1468 that ensure optimum performance of the classifier model since the validation loss shows a decreasing outline with the increasing of the number of epochs. Moreover, the accuracy is obtained 97.80% and 96.25% for training and validation sets that ensure a good fit of the classifier model because both sets show an increase of accuracy value for every epoch.
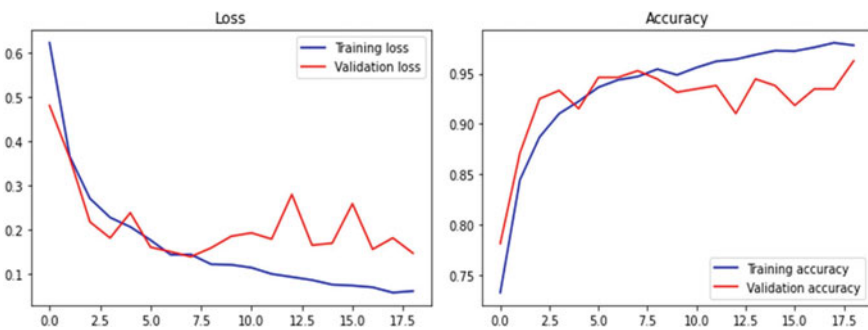
Fig. 4 The loss and accuracy of the proposed classification model

Table 1 Precision and recall of the proposed classifier set at different epochs

Epoch	At epoch 1	At epoch 4	At epoch 7	At epoch 10	At epoch 13	At epoch 16	At epoch 19
Training precision	0.7859	0.9166	0.9490	0.9524	0.9644	0.9726	0.9816
Validation precision	0.8094	0.9373	0.9522	0.9343	0.9161	0.9195	0.9639
Training recall	0.6663	0.9013	0.9412	0.9466	0.9621	0.9718	0.9755
Validation recall	0.7276	0.9266	0.9429	0.9282	0.9086	0.9135	0.9592

For evaluating the performance of the proposed model, various metrics including precision, recall are obtained at different epochs shown in table 1. The precision and recall of the training set are found 98.16% and 96.39% and for validation set are found 97.55% and 95.92% respectively. Moreover, the F1 Score for training and validation set is calculated 97.85% and 96.15% respectively depending on the following equation.

$$\text{F1 Score } = 2 \times \frac{\text{precision * Recall}}{\text{Pricision + Recall}} \times 100\% \tag{1}$$

Table 2 shows the comparison of the proposed work with some existing research works based on the classification method and performance parameter. From this table, it is shown that the accuracy is quite improved by using the transfer learning-based CNN classifier proposed in this work. Therefore, the proposed method is proficient to significantly improve the multi-classification accuracy comparing with other models presented in the table.

Table 2 Comparison of the proposed model with the existing model

References	Classification method	Accuracy (%)
[1]	Modified VGG-19 CNN	94.58
[4]	CNN	96.13
[5]	CNN	94.68
[8]	Deep transfer learning with ResNet-50	95.23 ± 0.6
[9]	VGG-16 with LSTM	84.00
This work	Inception-v3 with customized DNN	**96.25**

4 Conclusion

This paper proposed a method of multi-classification of brain tumor images based on inception-v3 transfer learning with customized DNN in MRI images. The fine-tuning transfer learning model is employed after the image augmentation can improve the accuracy of the classifier for precisely classifying the tumor images as glioma, meningioma, and pituitary respectively. The proposed deep neural network-based classifier provides satisfactory performance in terms of accuracy of 96.25%. Therefore, the improved classification accuracy has a great impact on the computer aided diagnosis of brain cancer that assists the physicians to make an exact decision for the treatment of the neurological patient.

References

1. Siegel RL, Miller KD, Jemal A (2020) Cancer statistics. CA A Cancer J Clin 70:7–30 https://doi.org/10.3322/caac.21590
2. Sajjad M, Khan S, Muhammad K, Wu W, Ullah A, Baik SW (2019) Multi-grade brain tumor classification using deep CNN with extensive data augmentation. Journal of computational science 1(30):174–182
3. Amin J, Sharif M, Gul N, Yasmin M, Shad SA (2020) Brain tumor classification based on DWT fusion of MRI sequences using convolutional neural network. Pattern Recogn Lett 1(129):115–122
4. Sultan HH, Salem NM, Al-Atabany W (2019) Multi-classification of brain tumor images using deep neural network. IEEE Access. 27(7):69215–69225
5. Abiwinanda N, Hanif M, Tafwida Hesaputra S, Handayani A, Mengko TR (2018) Brain tumor classification using convolutional neural network. In: World congress on medical physics and biomedical engineering 2018. Springer, Singapore, pp 183–189
6. Cheng J (2017) Brain tumor dataset (version 5) 2017. https://doi.org/10.6084/m9.figshare.1512427.v5.
7. Tsang S-H (2018) Review: inception-v3 — 1st Runner Up (Image Classification) in ILSVRC 2015. https://sh-tsang.medium.com/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c. Accessed 10 September 2018
8. Talo M, Yildirim O, Baloglu UB, Aydin G, Acharya UR (2019) Convolutional neural networks for multi-class brain disease detection using MRI images. Comput Med Imaging Graph. https://doi.org/10.1016/j.compmedimag.2019.101673
9. Shahzadi I, Tang TB, Meriadeau F, Quyyum A (2018) CNN-LSTM: cascaded framework for brain Tumour classification. In: 2018 IEEE-EMBS conference on biomedical engineering and sciences (IECBES) 2018 Dec 3. IEEE, pp 633–637

A Meta Model Based Particle Swarm Optimization for Enhanced Global Search

Rasel Ahmed [ID]**, Shuhaimi Mahadzir** [ID]**,
and Nor Erniza Mohammad Rozali** [ID]

Abstract The performance of a metaheuristic algorithm depends on the appropriate selection of its behavioral parameters. A good selection of parameters increases the search ability of an algorithm and avoids premature convergence. Particle swarm optimization (PSO) is swarm-based metaheuristic algorithm, which needs few parameter adjustments and less computational time. Meta-optimization has been used to tune the parameters and to get better results. Previously, authors applied meta optimization techniques to specific problems to tune the parameters and to get better results for specific case studies in different fields, but the application of meta optimization in benchmark functions are limited. The present study proposes meta optimization-based PSO to minimize the computational effort required for manual trial and error-based parameter selection. The proposed algorithm is tested for 14 benchmark functions (8 unimodal and 6 multimodal), and its efficiency and robustness are assessed via statistical analysis. The algorithm outperforms other renowned established algorithms (GA, PSO), and its performance remains consistent with increasing modality and dimensionality.

Keywords Particle swarm optimization · Genetic algorithm · Evolutionary algorithms · Swarm intelligence · Computational intelligence · Metaheuristic algorithms · Meta optimization · Stochastic optimization

1 Introduction

Population-based evolutionary algorithms belong to a higher level of stochastic techniques, which do not need detailed information, high computational resources and expert knowledge of a system [1, 2]. They improve the quality of a solution step by step by utilizing the basic features such as selection, variation, updating velocity and position [3, 4]. These all involve parameters; whose values need to be chosen properly to achieve the best results for a specific problem. Parameters could be set in two

R. Ahmed (✉) · S. Mahadzir · N. E. Mohammad Rozali
Department of Chemical Engineering, Universiti Teknologi Petronas, 32610 Perak, Malaysia

ways: tuning of parameters and parameter control [5]. Traditionally, researchers used to do fine tuning of algorithm parameters manually to get better performance of the algorithm. Shi and Eberhart conducted studies to manually improve the performance of algorithms [7, 8]. Some researchers analyzed particle swarm optimization (PSO) and showed how the parameters affect the convergence of particles to the global optimum [9, 10]. However, these approaches are sometimes not systematic and have limited scope. Another widely used path followed by researchers is to develop new variants of algorithms [11, 15]. Two common ways to this are: hybridize one algorithm with another for better performance and controlling the particles contraction in the swarm [11–13]. This article presents a meta-optimization of PSO and its evaluation on 8 unimodal and 6 multi-modal benchmark functions (Table 1) to test both of its exploration (global search) and exploitation (local search) ability. The proposed

Table 1 Details of 8 unimodal and 6 multi-modal benchmark functions [15]

Sl. No	Function name	Function	Dimension	Global optima	Search range	Initialization range
	Uni-modals					
1	Sphere	F1	30	0	[−100, 100]	[−100, 100]
2	Rosenbrock	F2	30	0	[−5, 10]	[−5, 5]
3	Dixon and price	F3	30	0	[−10, 10]	[−10, 10]
4	Zakharov	F4	30	0	[−5, 10]	[−5, 10]
5	Beale	F5	2	0	[−4.5 ,4.5]	[−4.5, 4.5]
6	Rotated_Hyper_Ellipsoid	F6	30	0	[−100, 100]	[−100, 100]
7	Branin	F7	2	5/4Pi	[−5, 15]	[−5, 15]
8	Modified schaffer3	F8	30	0.001566	[−100, 100]	[−100, 100]
	Multi−modals					
9	Levy	F9	30	0	[−10, 10]	[−10, 10]
10	Schwefel	F10	30	0	[−500, 500]	[−500, 500]
11	Ackley	F11	30	0	[−32, 32]	[−32, 16]
12	Griewank	F12	30	0	[−600, 600]	[−600, 200]
13	Rastrigin	F13	30	0	[−5.12, 5.12]	[−5.12, 2]
14	Shubert	F14	2	−186.7	[−10, 10]	[−10, 10]

algorithm showed significant improvement in terms of quality of solution and convergence and outperforms other two well-known robust optimization algorithms such as the genetic algorithm (GA) and PSO.

2   Meta Model Based Particle Swarm Optimization

Meta optimization is a systematic way of finding optimal parameters of an algorithm for a specific optimization problem [6]. Meta optimization is a two-step algorithm, where the super swarm (i.e., outer algorithm) tries to find the optimum values of algorithm parameters for use by the sub swarm (i.e., inner second algorithm) to optimize the objective function under investigation. Getting the optimum values of algorithm parameters by this way is an established concept in the literature to significantly improve the performance of an algorithm. Meta optimization-based GA, DE and PSO have already been applied to different fields of engineering. The main elements of a meta optimization for PSO are a) super swarm, b) sub swarm and c) objective function [14]. Super swarm particles are related to the sets of sub swarm, in order to determine the quality of the former; the sub swarm is run up to the number of pre-defined iterations, then the optimum final values are transferred to the super swarm. The super swarm parameter values are selected based on the analysis from the ranges in the literature, namely, $S = 20$ to 70, $w = -0.5$ to 0, $c1 = -0.5$ to 0 and $c2 = 0$ to 4 [6]. The details procedure of the meta model based PSO algorithm has been shown in the following flow chart (Fig. 1).

3   Experimental Result and Discussion

To demonstrate the robustness of our meta optimized particle swarm optimization algorithm, we tested it against fourteen renowned benchmark functions as minimization problem. Among them the first eight functions are unimodal (F1–F8) and the last six functions (F9-F14) are multi-modal. Among these functions F11 has a narrow valley, F12 also has a narrow global optimum and many trivial local optima, F10 is a complex multi-modal function with many local optima. The parameters values assigned to the three algorithms are similar such as the iteration number and particle numbers are considered as 100, number of dimensions is 30 for all problems where F5, F7, F14 are considered as 2-dimensional problems (except few inherent properties of GA). The search area (lower and upper bounds) are kept similar for all three algorithms. Each algorithm has been repeatedly run for 10 times and the mean, standard deviation and the best (minimum) values are calculated from the MATLAB simulation environment. The details summary of the benchmark functions and their comparative performances are summarized in Tables 1 and 2.

The unimodal benchmark functions (F1–F8) are well known for testing exploitation ability of an algorithm, whereas the exploration ability of an algorithm is judged
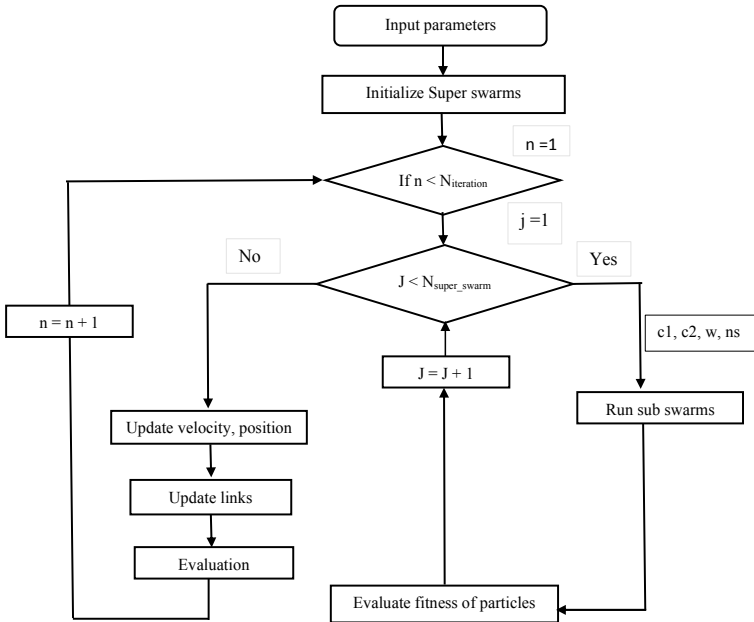
Fig. 1 Working principles of meta model base PSO [14]

by multi-modal (F9–F14) functions. In addition to this the local optima avoidance of an algorithm can also be validated by multi-modal (F9–F14) functions. The proposed algorithm outperforms other two algorithms in terms of exploitation, exploration and local minima avoidance. In case of the convergence of algorithms, meta PSO shows better convergence ability than PSO and GA for 13 of the 14 benchmark functions, it also outstripped other two robust algorithms for both 30 dimensional and 2 dimensional problems simultaneously.

Among the 14 benchmark functions, for 8 benchmark functions the best (minimum) value achieved by the meta PSO is better than other algorithms best values. The meta PSO reached the exact global optimal solution for 6 (F5, F7, F8, F12, F13, F14) functions where GA, PSO reached the exact global optima for 4 and 5 of these function but the GA, PSO couldn't reach the global solution for the function F12. For few functions F4, F6, F11 meta PSO performs significantly better than the GA, PSO. It has been noticed from the figures (Figs. 2, 3, 4, 5, 6, 7 and 8) that for all the considered cases meta PSO converge faster than GA, PSO except for F10 where GA performs better than both PSO meta PSO. In terms of the mean, best and standard deviation value (Table 2), the values that we get from the meta PSO is better than the other two algorithms, which means that the results generated by meta PSO is more accurate and stable. In terms of total computational time the meta PSO requires more time to complete 10 runs (with 100 iteration each run) than GA, PSO. It happened because of the complex nature of meta PSO algorithm and it's wide spread search ability where information is exchanged between the subswarms and super swarms.

Table 2 Statistical analysis and comparison of result for Algorithms

Func	Name of Par	GA	PSO	Meta-PSO	Func	Name of par	GA	PSO	Meta-PSO
F1	Mean	1.3889	5.56E−07	0.0012462	F8	Mean	0.0022568	0.0015669	0.0015669
	Std	0.99309	2.53 E−08	0.0021579		Std	0.0010892	3.21 E−17	0
	Best	0.49275	5.28 E−07	3.11 E−07		Best	0.0015688	0.0015669	0.0015669
F2	Mean	89.0693	25.7472	18.3969	F9	Mean	1.50 E−32	1.4998 E−32	1.4998 E−32
	Std	22.8912	0.60685	1.3046		Std	0	0	0
	Best	66.3609	25.0824	17.1719		Best	1.50 E−32	1.50 E−32	1.4998 E−32
F3	Mean	8.0144	0.67802	0.54096	F10	Mean	3820.7129	5382.7844	3913.8624
	Std	4.3274	0.019657	0.22578		Std	458.5762	883.9675	370.0767
	Best	3.2016	0.66667	0.2803		Best	3326.9552	4402.6275	3604.3912
F4	Mean	1.42 E−32	2.63 E−58	1.67 E−88	F11	Mean	0.29216	1.3116	0.068682
	Std	2.30 E−32	3.68 E−58	2.90 E−88		Std	0.06441	0.32935	0.029537
	Best	2.04 E−34	4.55 E−59	2.41 E−97		Best	0.2231	0.9313	0.034592
F5	Mean	5.00 E−13	0	0	F12	Mean	0.031249	0.0049307	0
	Std	8.66 E−13	0	0		Std	0.039196	0.0042701	0
	Best	7.33 E−21	0	0		Best	0.007396	0	0
F6	Mean	80.3431	1.36 E−05	3.15 E−11	F13	Mean	0	0	0
	Std	126.9302	1.76 E−05	5.18 E−11		Std	0	0	0
	Best	5.8177	1.42 E−07	3.83 E−13		Best	0	0	0
F7	Mean	0.39789	0.39789	0.39789	F14	Mean	−186.7309	−186.7309	−186.7309
	Std	0	0	0		Std	2.01 E−14	2.01 E−14	0
	Best	0.39789	0.39789	0.39789		Best	−186.7309	−186.7309	−186.7309
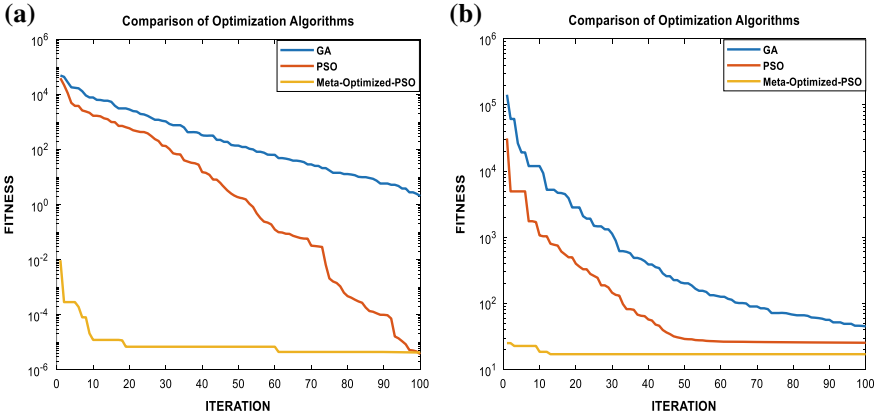
Fig. 2 Convergence curve for Sphere and Rosenbrock benchmark functions
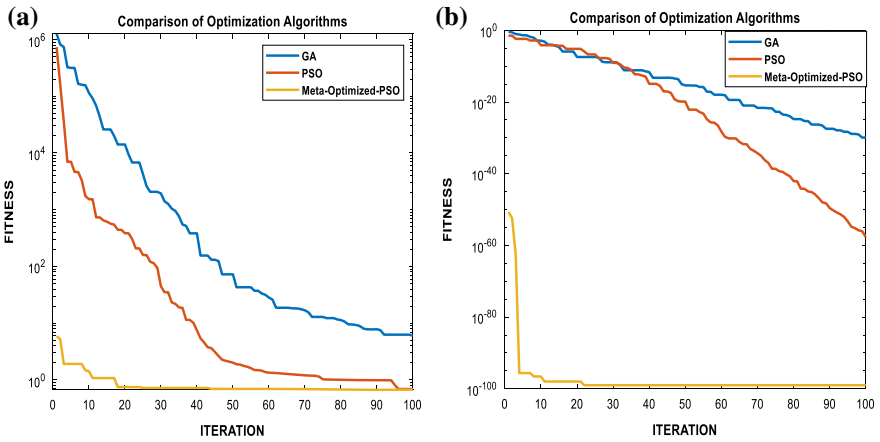
Fig. 3 Convergence curve for Dixon Price and Zakharov benchmark functions

It agreed with the no free launch theorem, that if we want to emphasize any specific property of any algorithm, we need to sacrifice some other properties as a tradeoff.

4 Conclusion

The main objective of this article is to analyze the performance of meta model based PSO algorithm in case of optimization benchmark functions. It outperforms the other algorithms in terms of efficiency, convergence and robustness, it also overcomes the limitations of manually selecting optimal behavioral parameters. The proposed
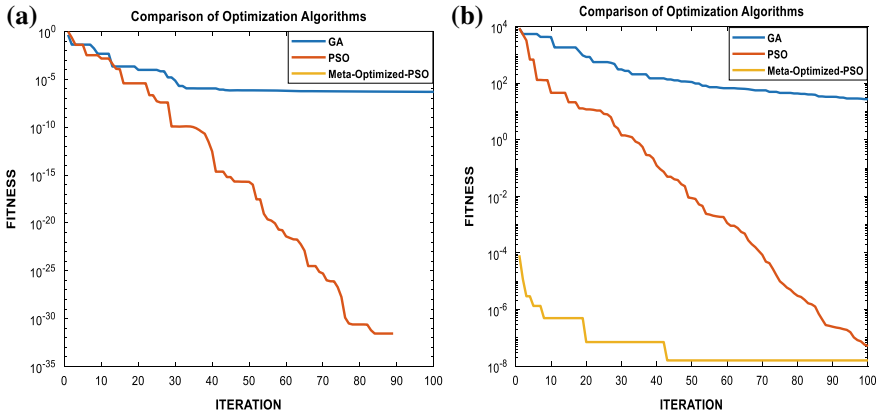
Fig. 4 Convergence curve for Beale and Rotated_hyper_ellipsoid benchmark functions
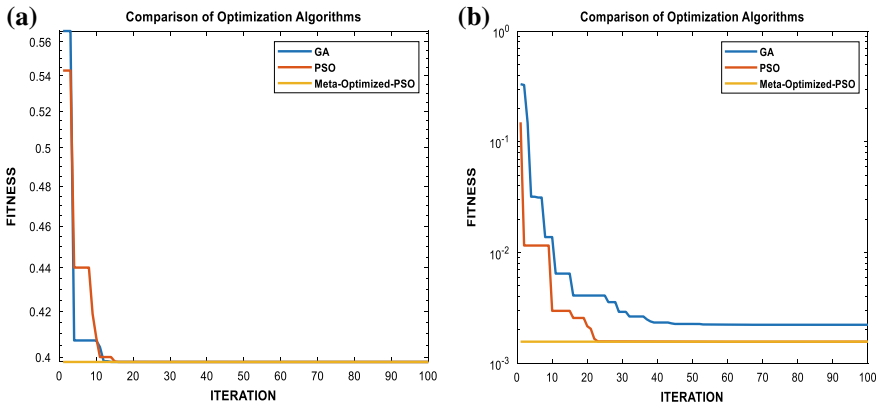
Fig. 5 Convergence curve for Branin and Modified Schaffer3 benchmark functions

meta optimization increases the exploration and exploitation ability of the algorithms by continuously updating the parameter values and by maintaining a good balance between exploration and exploitation, that makes the meta model based PSO algorithm to outperform other renowned algorithms and keep its performance consistent with increasing modality and dimensionality.
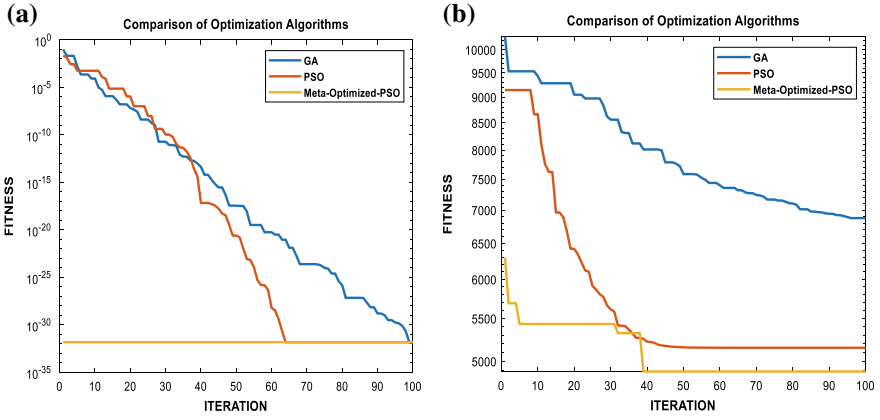
Fig. 6 Convergence curve for Levy and Schwefel benchmark functions
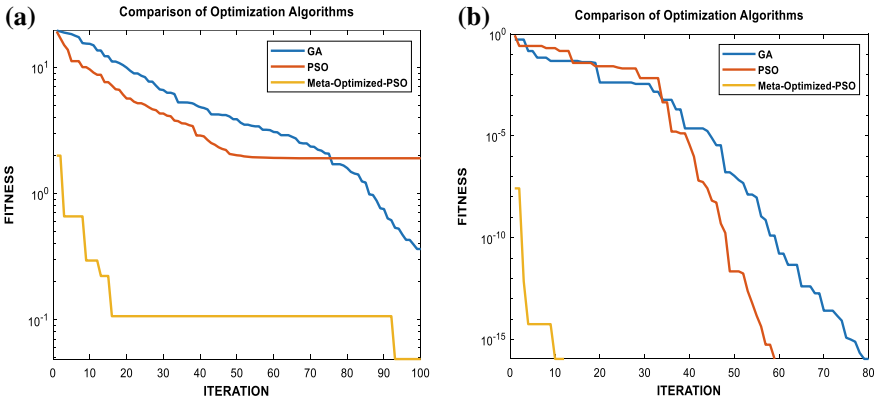
Fig. 7 Convergence curve for Ackley and Griewank benchmark functions

(a)

(b)
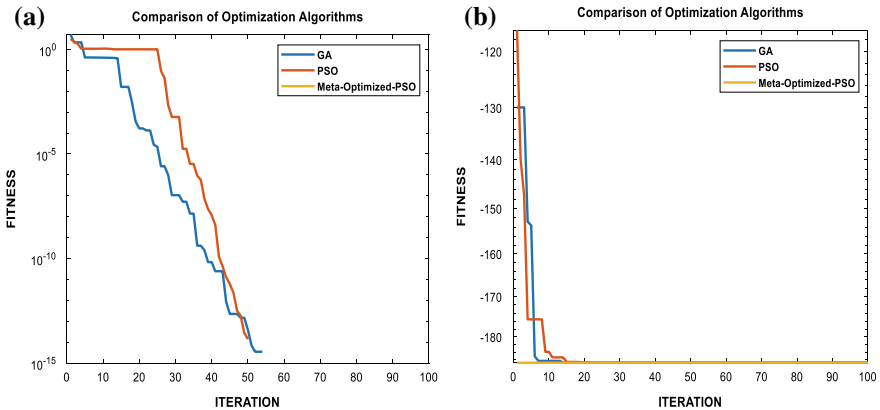
Fig. 8 Convergence curve for Rastrigin and Shubert benchmark functions

References

1. Łapa K (2019) Meta-optimization of multi-objective population-based algorithms using multi-objective performance metrics. Inf Sci 489:193–204. https://doi.org/10.1016/j.ins.2019.03.054
2. Smit SK, Eiben AE (2009) Comparing parameter tuning methods for evolutionary algorithms. In: 2009 IEEE congress on evolutionary computation. IEEE, pp 399–406. https://doi.org/10.1109/CEC.2009.4982974
3. De Jong K (2016) Evolutionary computation: a unified approach. In: Proceedings of the 2016 on genetic and evolutionary computation conference companion, pp 185–199
4. Eiben AE, Smith JE (2003) Introduction to evolutionary computation. Natural Computing Series Springer. https://doi.org/10.1007/978-3-662-05094-1
5. Eiben ÁE, Hinterding R, Michalewicz Z (1999) Parameter control in evolutionary algorithms. IEEE Trans Evol Comput 3(2):124–141. https://doi.org/10.1109/4235.771166
6. Pedersen MEH, Chipperfield AJ (2010) Simplifying particle swarm optimization. Appl Soft Comput 10(2):618–628. https://doi.org/10.1016/j.asoc.2009.08.029
7. Shi Y, Eberhart RC (1998) Parameter selection in particle swarm optimization. In: Porto VW, Saravanan N, Waagen D, Eiben AE (eds) Evolutionary Programming VII. EP 1998. Lecture Notes in Computer Science, vol 1447. Springer, Berlin, Heidelberg. https://doi.org/10.1007/BFb0040810
8. Eberhart RC, Shi Y (2000) Comparing inertia weights and constriction factors in particle swarm optimization. In: Proceedings of the 2000 congress on evolutionary computation. CEC00 (Cat. No. 00TH8512), vol 1. IEEE, pp 84–88. https://doi.org/10.1109/CEC.2000.870279
9. Clerc M, Kennedy J (2002) The particle swarm-explosion, stability, and convergence in a multidimensional complex space. IEEE Trans Evol Comput 6(1):58–73. https://doi.org/10.1109/4235.985692
10. Trelea IC (2003) The particle swarm optimization algorithm: convergence analysis and parameter selection. Inf Process Lett 85(6):317–325. https://doi.org/10.1016/S0020-0190(02)00447-7
11. Keane AJ (1995) Genetic algorithm optimization of multi-peak problems: studies in convergence and robustness. Artif Intell Eng 9(2):75–83. https://doi.org/10.1016/0954-1810(95)95751-Q
12. Lovbjerg M, Krink T (2002) Extending particle swarm optimisers with self-organized criticality. In: Proceedings of the 2002 congress on evolutionary computation. CEC'02 (Cat. No. 02TH8600), vol 2. IEEE, pp 1588–1593. https://doi.org/10.1109/CEC.2002.1004479

13. Krink T, Løvbjerg M (2002) The lifecycle model: combining particle swarm optimisation, genetic algorithms and hillclimbers. In: International conference on parallel problem solving from nature. Springer, Berlin, Heidelberg, pp 621–630. https://doi.org/10.1007/3-540-45712-7_60
14. Atashnezhad A, Wood DA, Fereidounpour A, Khosravanian R (2014) Designing and optimizing deviated wellbore trajectories using novel particle swarm algorithms. J Nat Gas Sci Eng 21:1184–1204. https://doi.org/10.1016/j.jngse.2014.05.029
15. Xu X, Tang Y, Li J, Hua C, Guan X (2015) Dynamic multi-swarm particle swarm optimizer with cooperative learning strategy. Appl Soft Comput 29:169–183. https://doi.org/10.1016/j.asoc.2014.12.026

HCShare: Blockchain Technology in Healthcare System

R. Sangeetha and M. Krishnamoorthi

Abstract On this generous growing world, the medical healthcare system is expediting by driving new modernizations in the healthcare system by providing contemporaneous varieties for an inmate's management. The enhancement of the medical system is done by collecting enormous preceding data, accommodating those records, scrutinizing of the data, and interchanging them with the different medical system layers. In the headway to enhance the medical system with further preservation of the health records, the Blockchain Technology takes the part in the system. The prospective system likewise supports to transfer or interchange the data of the inmates between different healthcare system, which is in the guaranteed way. HCShare is a confidential, protected, reliable and translucent data interchanging scheme between the different healthcare organizations in a scattered decentralized mechanism. In which some of the cryptographic algorithms are used to ensure the protection of the inmate's confidential records stored in the ledger and to enhance the data integrity of the records.

Keywords Healthcare system (HCShare) · Blockchain technology · SHA-256. · Data integrity

1 Introduction

As the Blockchain technology has a power to change dramatically as it is transformational, it is considered by the most healthcare organizations. As the generous growth in the blockchain system, extensive quantity of endeavor has done [1]. The health records of the patients are kept and maintained in the different entities such as their local, centralized servers of the hospice, pharmaceutical places, and sometimes in the physician's places according to the need and storage facility of the records over cloud platform or maybe on local paper based records, which are scattered all around the world depends on the intention of the data [2, 3]. Hence the improvement in the

R. Sangeetha (✉) · M. Krishnamoorthi
Dr. N.G.P. Institute of Technology, Coimbatore, India
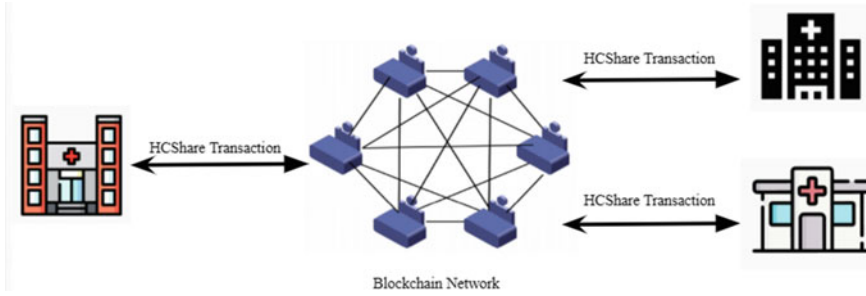
Fig. 1 Blockchain network structure

development of the healthcare system has delivered by the most looming sciences like Blockchain Technology, Internet of Things, and Edge Computing techniques. These are the major technologies drives the healthcare system in the smartest way with better security and integration among the other healthcare organization's data [4]. HCShare Healthcare system is a confidential, protected, reliable and translucent data interchanging scheme between the different healthcare organizations in a scattered decentralized mechanism [5]. HCShare system helps consistently to distribute and digitization [6] of the records between multiple medical organizations [6]. The transaction of the records of the inmate are transferred from one of the hospitals to other hospitals can be done by the blockchain network. As shown in the Fig. 1, every data that are stored in the medical organizations are shared in the scattered distributed blockchain network, in which all the data are stored in the secured mechanism by encrypting the data of the inmates [2].

The blockchain is referred to as the amalgam of couple of decrepit automation equally like cryptography [6] and peer-peer communication mechanisms [6]. It is a kind of scattered storage of records in a distributed-ledger-technology [6] in which the ledger can be increased generously within a massive collection of data directory. In the HCShare System model, the blockchain technology helps in the transaction of the medical records of the patients over the different medical institutions or over the different clinics depends on the patient's need about the records. For example, if the patient has been treated from hospital 1 for over the one years of duration, then in the critical situation, they have to claim another hospice for a quick treatment. In that situation the patient can't able to get all their previous medical reports again by doing all the test cases of the hospital to ensure his problem, or they can't go the hospital A for records of their medical reports. This HCShare system model helps the patients to get their medical data reports whenever they need through the blockchain system.

2 Related Works

In the HS-BC [7] for the electronic medical record [8], that provides the permission to approach the patients records in a secured manner, by implementing the KUNodes algorithm [7]. Similarly, in the [9], for securing the data of the inmates the algorithms likely SHA-256 [9] which is to calculate the hash key value, ECC (Elliptic Curve Cryptography) algorithm [9] for creating irregularity hash key values and for the encryption of the records of the patients, the AES (Advanced Encryption Standard) Algorithm is used. In [9] there have been some of the inadequacy such as absence of integrity of the records, and the system can't be able to perform massive requests simultaneously [2]. When the system uses cloud transaction via a blockchain technology, it is found that there is no security for the confidential data of the inmate's records. In which there is lack of the data integrity such as problems in the correctness of the records, lack of accuracy in the data, lack of consistency of the records. In some of the data sharing system, also used cloud environment, in which the data can be presented in the different layers, along with the accessibility to the data [2]. In such cases, the third party can filch the data, and can misuse the confidential data of the inmate's records. The SHealth works on the different layers such as government, user, and the IoT layers [2]. In the ChainSDI [3], the home-based medical service sharing programming introduced to manage the confidential data of the inmates [3]. In the Sshealth [4] system, massive records are analyzed also the difficulties in the management to provide the secured storage for the inmate data preservation [4]. In the [5], the objections in the blockchain scalability are well focused and found the issues like size and number of transactions of a block, and the solutions to sort it out was classified into storage optimization and redesigning blockchain [5].

3 Proposed Model

The proposed research work on the healthcare system is to yield the security of the inmate data, from the third-party attacks. Figure 2 provides the structure of the blocks [6] hash values in the blockchain system [10], in which the hash values are ordered in the Merkle tree [6]. In the HCShare the physician depot the medical reports in the corresponding servers of the specific hospice, in which the only authorized administrator who were allowed creates and depot the private passcode to ensure the safety of every dossier, in the permissioned [8] blockchain system to accomplish secured scattered ledger of hospice records. So that, every data within the ledger can be available for the hospice, physician and for the inmate of the hospice can access them in anytime from anywhere.

Fig. 2 Merkle tree block structure [6]
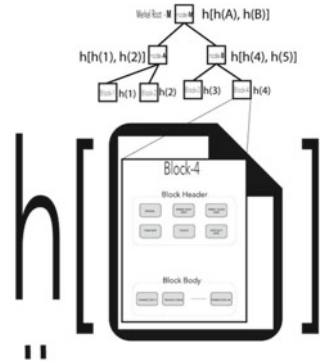
3.1 Blocks

In the HCShare System model, blocks in the blockchain transaction consists of the couple of things, which are called as block header [11] and block body [11], in which it contains the version numbers represented in the blocks, and hash content from the Merkle tree as shown in Fig. 2, timestamp value and the parental hash key values. These are shown in Fig. 3.

In the Merkle tree, the hash key generation starts from the nodes in the leaf level, in which the parental node hash value is calculated from its combination of leaf nodes, in which manner all the node's hash value is calculated. For example, assume that the leaf node of the Merkle tree as H_0, H_1, H_2, H_3, and H_4, in which the parental node hash value is calculated as follows,

HV_{01} = Hash_VALUE [HV_0 + HV_1].

HV_{23} = Hash_VALUE [HV_2 + HV_3].

HV_{0123} = Hash_VALUE [HV_{01} + HV_{23}].

In which the each and individual hash values of the leaf node contains the hash value of the transactions. As it contains every transaction details of the blocks, the complete Merkle tree contains the overall information of all transactions.
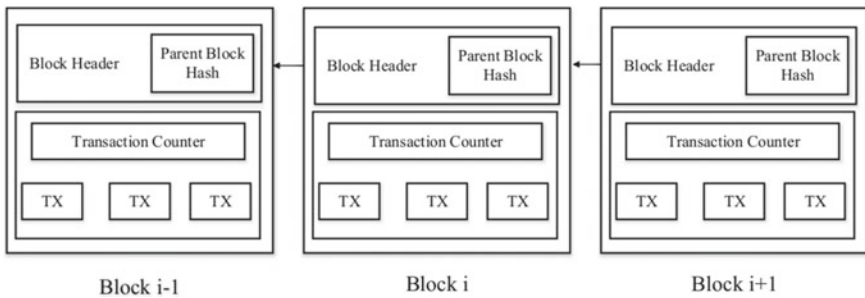
Fig. 3 Blocks in the blockchain system [11]

Fig. 4 KUNodes algorithm
[7]

$$KUNodes(BT, rl, t):$$
$$X, Y \leftarrow \emptyset.$$
for all $(\eta_i, t_i) \in rl,$
    if $t_i \leq t$, add $Path(\eta_i)$ to X.
for all $x \in X,$
    if $x_l \notin X$, add x_l to Y;
    if $x_r \notin X$, add x_r to Y.
if $Y = \emptyset$, add $root$ to Y.
return Y.

3.2 Encryption Algorithms

For the better secured data transaction among the multiple blockchain system in the healthcare system, every blocks of the transaction contains the confidential private key along with the hash values and the transaction details of the blocks. In the transaction details the inmate's records are stored according the specific hospice. For the encryption of the inmate's records, algorithms such as SHA-256, AES, KUNodes algorithms are used to ensure the safety and security of the inmate records from the malicious attacks.

In the encryption process of the HCShare System, pair of different keys has been introduced to improve the security and safety of the ledger, they are as Priority hash address key values and Secondary updated hash value. Where, with the help of priority hash key, one can only view their records, which were updated in the system by the reputed authorized hospice. So that the inmate and their family can view their data, when it is needed in the emergency situations. Another secondary updated hash value will help the hospice authorized admin or the physician to upload the records the inmates, where only with the help of the secondary hash key value, one can access completely of the ledger of particular inmate's data.

In the HCShare system model, we propose SHA-256 for generating the priority hash address key values, and to generate the secondary key values, KUNodes (Fig. 4), AES algorithm is used to ensure the data integrity of the patient's records.

3.3 Ethereum Smart Contracts

Every block in the blockchain system has its own unique key, in which each node of the blockchain system is connected to the prior blocks of the system via a individual hash key data [12]. Ethereum [12] is most widely used blockchain technology to give a better transaction over the blockchain system. In which the smart contracts [12] are used to secure the transaction details of the inmates. Smart contracts are used to execute set of programs in which the program is executed automatically when it is needed. These smart contracts manage the transaction of the ledger and the cryptocurrencies [12] of the blockchain system. These programs can be encountered without depending any other data, which performs and control overall transaction

```
1  pragma solidity ^0.4.21;
2  contract ipfs
3  {
4      struct patient_data_base
5      {
6          string file_name;
7          string data_hash;
8      }
9      mapping(uint256 =>mapping(uint256 => patient_data_base)) public patient_data_base_map;// key1: patient id  key2: file number
10
11     function get_data(uint256 _patient_id, uint256 _file_number, string memory _file_name, string memory _ipfs_hash) public payable returns(bool)
12     {
13         patient_data_base_map[_patient_id][_file_number].file_name = _file_name;
14         patient_data_base_map[_patient_id][_file_number].data_hash = _ipfs_hash;
15         return true;
16     }
17
18     function show_datas(uint256 _patient_id, uint256 _file_number) public constant returns(string,string)
19     {
20         return (patient_data_base_map[_patient_id][_file_number].file_name, patient_data_base_map[_patient_id][_file_number].data_hash);
21     }
22
23  }
24
25  // 0xc35od5bdb1f12a21d4c21888c04787f43db55336
```

Fig. 5 Inmate record mapping with database

in the blocks. In the HCShare System model, we propose Ethereum blockchain technology for the transaction among the hospice, physicist as well as to the inmates. Figure 5 shows the mapping of the hash key values to access and retrieve the records from and to the database.

4   Existing Model Analysis

SHealth [2]	Multilayer-signature [2] based healthcare system designed with sufficient functionalities like including and eradicating of the data in the system design
Sshealth [4]	External Edge is used here to manage and maintain the efficiency of the system before every transaction occurs [4]
EHR system [1]	Concussions on the blockchain system evaluated using hybrid fuzzy-ANP TOPSIS [1] to implement the top blockchain system for the better HER maintenance
ChainSDI [3]	Home-based medical data management via smart contracts [3], and CORD [3]
Scalability challenges [5]	It focuses on the solution to the scalability challenges in size of the block, high amount of transaction, number of nodes, protocol used [5]. These objections were solved by the couple of solutions as Storage Optimization [5] and Redesigning Blockchain [5]
HS-BC [7]	Efficient data security of the inmate is maintained by using update key by KUNodes algorithm [7], which provides another step of encryption to the data [7]. From the result of the [7], it is proven that very cost effective
Decentralized Authentication [11]	Efficient transaction duration for registering data and with less access time for the data obtained and evaluated with the help of NS-2V2.35 simulator [11]

In Fig. 6, the expected authentication delay of the HCShare system will be lesser than the prior existing system, as the number of transaction increases over the communication of the inmates with the hospice, the response time is efficient as compared to other systems.

Fig. 6 Expected proposed
HCShare system
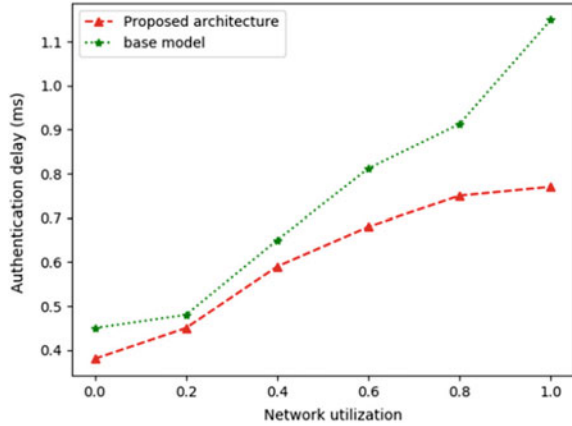
5 Conclusion

In this enormous world, the blockchain became most noteworthy technology in the internet security [13], where the it gives all kinds of facility in the smart technological world like in the fields of smart houses, smart grid [13] applications, as well as in the food safety application [13], and so on. In this proposed HCShare system model, the system gives better accessibility of the records in the blocks, by attribute-based encryption [14] technology, in which the people without having the confidential keys cannot have an access to the data of the inmates. So that the records of the inmates will be stored in secured manner in the blockchain technology. We propose HCShare system by using the Ethereum technology for the transaction of the medical reports over the different medical institutions. Our research work plans to get the better data integrity of the medical reports of the inmates in which all the records of the inmates should be transferred from and to different hospice, with the complete records in a secured manner. This HCShare is to give the better truncation efficiency over a different block to ensure the performance of the data transaction. The lack of data integrity is needs to be solved in the HCShare system, in which to overcome these kinds of issues the data are secured using pair of different addresses to ensure the safety and integrity of the data.

References

1. Zarour M et al (2020) Evaluating the impact of blockchain models for secure and trustworthy electronic healthcare records. IEEE Access 8:157959–157973. https://doi.org/10.1109/ACCESS.2020.3019829
2. Zghaibeh M, Farooq U, Hasan NU, Baig I (2020) SHealth: a blockchain-based health system with smart contracts capabilities. IEEE Access 8:70030–70043. https://doi.org/10.1109/ACCESS.2020.2986789

3. Li P et al (2020) ChainSDI: a software-defined infrastructure for regulation-compliant home-based healthcare services secured by blockchains. IEEE Syst J 14(2):2042–2053. https://doi.org/10.1109/JSYST.2019.2937930

4. Abdellatif A, Al-Marridi AZ, Mohamed A, Erbad A, Chiasserini CF, Refaey A (2020) ssHealth: toward secure, blockchain-enabled healthcare systems. IEEE Netw 34(4):312–319. https://doi.org/10.1109/MNET.011.1900553

5. Mazlan A, Mohd Daud S, Mohd Sam S, Abas H, Abdul Rasid SZ, Yusof MF (2020) Scalability challenges in healthcare blockchain system—a systematic review. IEEE Access 8:23663–23673. https://doi.org/10.1109/ACCESS.2020.2969230

6. Chukwu E, Garg L (2020) A systematic review of blockchain in healthcare: frameworks, prototypes, and implementations. IEEE Access 8:21196–21214. https://doi.org/10.1109/ACCESS.2020.2969881

7. Su Q, Zhang R, Xue R, Li P (2020) Revocable attribute-based signature for blockchain-based healthcare system. IEEE Access 8:127884–127896. https://doi.org/10.1109/ACCESS.2020.3007691

8. Niu S, Chen L, Wang J, Yu F (2020) Electronic health record sharing scheme with searchable attribute-based encryption on blockchain. IEEE Access 8:7195–7204. https://doi.org/10.1109/ACCESS.2019.2959044

9. Li H, Zhu L, Shen M, Gao F, Tao X, Liu S (2018) Blockchain-based data preservation system for medical data. J Med Syst 42(8) Art. no. 141153

10. Zhuang Y, Sheets LR, Chen Y-W, Shae Z-Y, Tsai JJP, Shyu C-R (2020) A patient-centric health information exchange framework using blockchain technology. IEEE J Biomed Health Inform 24(8):2169–2176. https://doi.org/10.1109/JBHI.2020.2993072

11. Yazdinejad GS Parizi RM, Dehghantanha A, Choo K-KR, Aledhari M (2020) Decentralized authentication of distributed patients in hospital networks using blockchain. IEEE J Biomed Health Inf 24(8):2146–2156. https://doi.org/10.1109/JBHI.2020.2969648

12. Jaiman V, Urovi V (2020) A consent model for blockchain-based health data sharing platforms. IEEE Access 8:143734–143745. https://doi.org/10.1109/ACCESS.2020.3014565

13. Li C-T, Shih D-H, Wang C-C, Chen C-L, Lee C-C (2020) A blockchain based data aggregation and group authentication scheme for electronic medical system. IEEE Access 8:173904–173917. https://doi.org/10.1109/ACCESS.2020.3025898

14. Sun J, Yao X, Wang S, Wu Y (2020) Blockchain-based secure storage and access scheme for electronic medical records in IPFS. IEEE Access 8:59389–59401. https://doi.org/10.1109/ACCESS.2020.2982964

SAZZ Converter Fed Fuzzy Logic Speed Controlled BLDC Motor Drive

A. Senthilnathan, P. Palanivel, and R. Balakrishnan

Abstract BLDC motors possess high torque than the same rated Induction motor. Due to electronic commutation and the absence of mechanic commutation and brushes, BLDC motors are widely used in many industrial applications. The main disadvantage of BLDC motor is high torque ripple. Commutation Ripple reduction of brushless direct current (BLDC) motor torque using the SAZZ converter is proposed in this work; torque ripple is proposed to carry out under both conduction and commutation intervals in this proposed work. The fuzzy logic speed controller is used for generating reference pulse in the conduction interval by processing the speed error. The performance of torque and speed based on the proposed control method was analyzed. MATLAB/Simulink platform is used for implementing as well as showing effective implementation of the proposed technique. PI and Fuzzy logic control are used for evaluating the performance. SAZZ (Snubber Assisted Zero Voltage and Zero Current Transition) topology was proposed to reduce switching losses. The main switch snubber circuit reduces the voltage during turn off, and the auxiliary circuit discharges the snubber before the next switching cycle.

Keywords BLDC motor · SAZZ converter · Fuzzy logic controller

1 Introduction

Brushless DC motors possess low electromagnetic interferences, high sensitivity, and more efficiency than a brushed motor, making BLDC motor more reliable in most industrial applications. So BLDC motors are most commonly recommended in industries for medium power applications such as electric traction, air conditioning, robotics, etc. One of the major drawbacks of a BLDC motor is torque ripples due to

A. Senthilnathan (✉) · R. Balakrishnan
EEE Department, Dr. N.G.P. Institute of Technology, Coimbatore, India
e-mail: senthilnathan@drngpit.ac.in

P. Palanivel
EEE Department, E.G.S. Pillay Engineering College, Nagapattinam, India

953
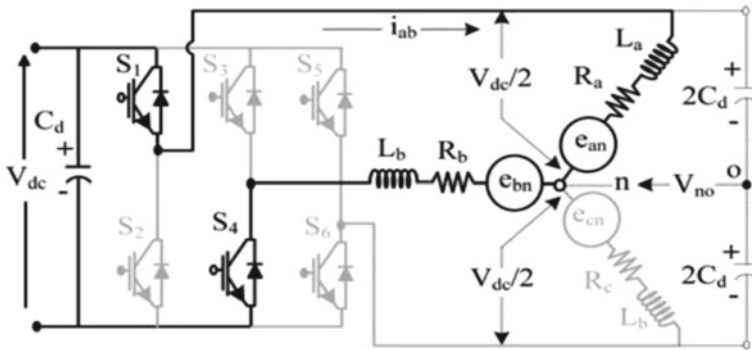
Fig. 1 Energizing of two phases in a BLDC motor

cogging, commutation, and switching power devices [1, 2]. Many research works are progressing to reduce current and torque ripples in PMBLDC motor. In the six-step inverter conduction topology, at any time, only two phases are in conduction [3–5]. Generally, the rectangular phase current is considered to be ideal for two-phase conduction mode. Due to commutation, the incoming phase current's slew rate time is not equal to the outgoing phase current, which results in current pulsation [6]. This current pulsation produced torque ripple in the PMBLDC motor. Fuzzy logic controlled SAZZ converter fed BLDC motor will create fewer torque ripples than PI- Controller.

1.1 BLDC Motor

A permanent magnet BLDC is a synchronous motor with electronic commutation rather than a mechanical one. A position sensor in a rotor such as a hall sensor is used to locate the rotor position. Using rotor position sensor firing pulses for each phase current of the Brushless DC motor is synchronized. Instead of mechanical commutation, electronic switches are provided for proper commutation. Figure 1 shows the energizing of two winding when two switches are fired one from the upper limb next from the lower limb.

1.2 SAZZ Converter

SAZZ (Snubber Assisted Zero Voltage and Zero Current Transition) converter is used to reduce switching losses. The main switch snubber circuit is used to reduce the voltage during turn off and the auxiliary circuit discharges the snubber before the next switching cycle. Figure 2 Shows the block diagram of the SAZZ converter fed
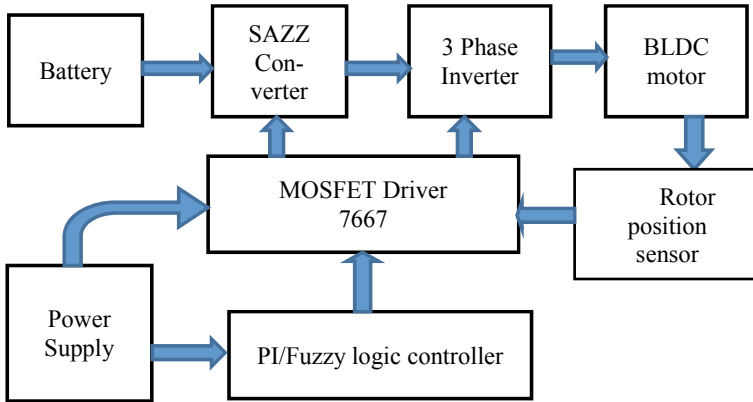
Fig. 2 Block diagram of SAZZ converter fed BLDC motor

BLDC motor. The battery is used as the input source supply and the SAZZ converter is used to step-up DC voltage and fed the supply to the inverter. Six step inverter is used to energize the BLDC motor. At any point, only two phases are energized as shown in Fig. 1. The Hall sensor is used to sense the rotor position and the PI and fuzzy controller is used to control the MOSFET driver circuit. The error and change in error are taken as two inputs to the controller and the output of the controller is used to fire switches in SAZZ converters. The block diagram of proposed work is shown in Fig. 2.

2 PI Controller Based SAZZ Converter for BLDC Motor

The reference voltage of the inverter is produced based on the speed received from the brushless DC motor. The inverter's output voltage is compared with the reference voltage, and an error voltage is generated. This error is used to control the PI controller.

The PI controller has a proportional gain and an integral gain. The proportional gain ($K_P = 0.8$) supports the analog error function and an integrator ($K_I = 1.1$) to reduce the constant level error value to zero. Based on the fault voltage, the PI controller produces pulses that regulates the MOSFET. Although the PMDC motor current follows the change in torque and change in voltage, the armature current from Fig. 3a has different deformations, and Fig. 3b and c back EMF and torque are more distorted and take longer to resolve.

a) Armature current

b) Back EMF

c) Torque
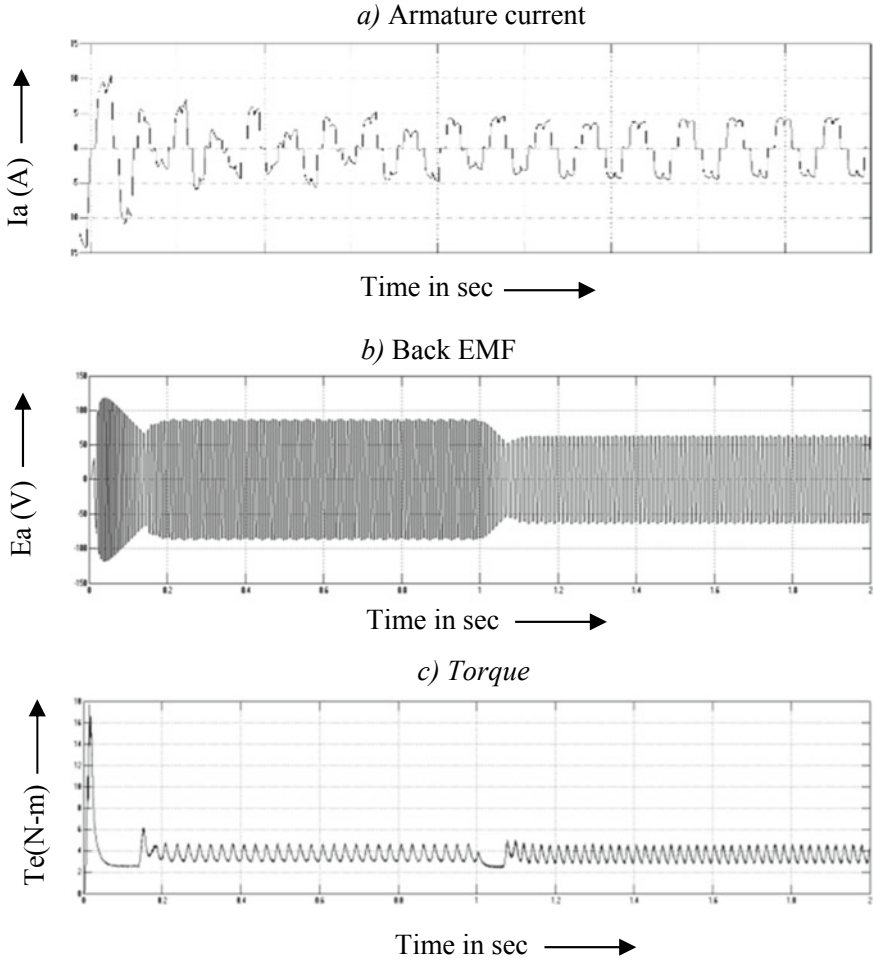
Fig. 3 Performance of PI controller based SAZZ converter fed BLDC motor

3 FLC Operation

Fuzzy logic is the usually available control approach for the above two decades. Fuzzy logic consists of fuzzification, inference engine, and defuzzification. The Mamdani fuzzy control system is used here. The input will be given to a fuzzifier [7–9]. A fuzzy rule base contains the fuzzy rule set which we have to design. This input is given to a fuzzy interface engine that compares the input with the fuzzifier inputs to give the required output to the defuzzifier. The defuzzifier gives the output after defuzzification and gives the respective output to the next system. The same input is obtained at the output but after fuzzification i.e. the character of the output does not

Fig. 4 Fuzzy partition
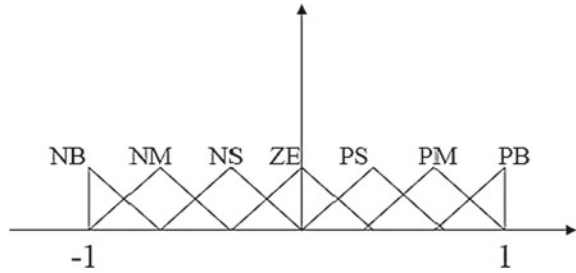
change. This output is given to the converter to boost up the voltage and obtain the required output. The Fuzzy Logic Control is designed by MATLAB R2014a.

Where,

- N B- negative Big
- N M- negative Medium
- N S- negative Small
- Z E- zero
- P S-positive Small
- P M- positive Medium
- P B- positive Big

Figure 4 shows the Fuzzy partition. Each input and output set is assigned with seven linguistic variables and forty-nine rules are enclosed in FLC. Mamdani algorithm is used to design the set of rules for the FLC. The fuzzy partition is shown in Fig. 4 which includes all the 7 sets of variables that use the IF–THEN logic to design the 49 sets of fuzzy rules.

From Fig.5a it is very clear that stator current is almost rectangular in nature. From Fig. 5c it is very clear that percentage average speed error is less than 1% and the speed follow the command as fast as possible. Moreover the step change from 200V to 150 V at t = 1 sec produce a change in speed within 0.25 s.

4 Conclusion

A Fuzzy Logic Controller based SAZZ converter for BLDC motor has been designed in this paper. This technique has been implemented as it proves to be more efficient than any other method. By employing the Fuzzy Logic Controller (FLC), the varying outputs can be brought in control to get a constant output. Here two sets of input blocks have been used where several voltage variations have been obtained on the input side in an FLC. Thus the simulation and its outputs have shown that even though voltage variations occur to a large extent, the output tends to remain the same throughout the whole process.

a) Armature current

b) Back EMF

c) Speed
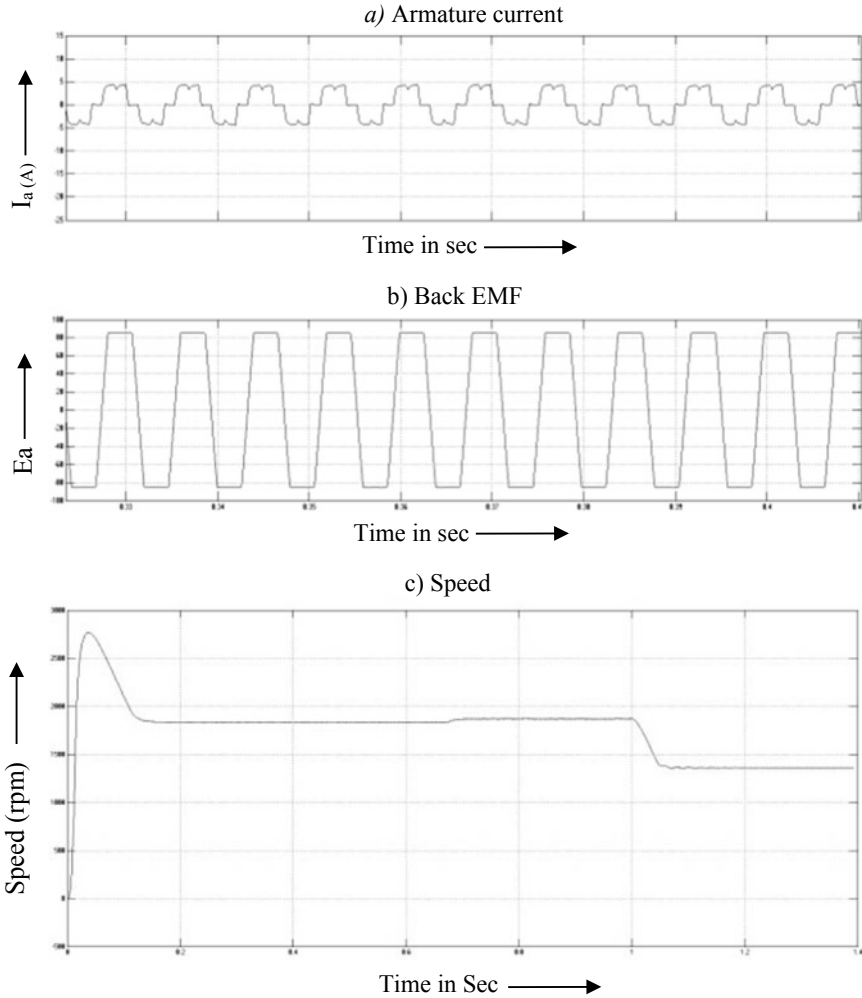
Fig. 5 Performance of Fuzzy logic controller based SAZZ Converter fed BLDC motor

References

1. Al Nabulsi A, Dhaouadi R (2012) Efficiency optimization of a DSP-based standalone PV system using fuzzy logic and dual-MPPT control. IEEE Trans Ind Inf 8(3)
2. Alajmi BN, Ahmed KH, Finney SJ, Williams BW (2013) A maximum power point tracking technique for partially shaded photovoltaic systems in Microgrids. IEEE Trans Ind Elect 60(4)
3. Chen B-Y, Lai Y-S (2012) New digital-controlled technique for battery charger with constant current and voltage control without current feedback. IEEE Trans Ind Electron 59(3)
4. Camara MB, Gualous H, Gustin F, Berthon A, Dakyo B (2010) DC/DC converter design for super capacitor and battery power management in hybrid vehicle applications—polynomial control strategy. IEEE Trans Ind Electron 57(2)

5. Vasallo Vázquez MJ, Andújar Márquez JM, Segura Manzano F (2008) A methodology for optimizing stand-alone PV-system size using parallel-connected DC/DC converters. IEEE Trans Ind Electron 55(7)
6. Senthilnathan, Palanivel P (2020) A new approach for commutation torque ripple reduction of FPGA based brushless DC motor with outgoing phase current control. Microprocess Microsyst 75
7. Veerachary M, Senjyu T, Uezato K (2003) Neural-network-based maximum-power-point tracking of coupled-inductor interleaved-boost-converter-supplied PV system using fuzzy controller. IEEE Trans Ind Electron 50(4)
8. Muyeen SM, Al-Durra A (2013) Modeling and control strategies of fuzzy logic controlled inverter system for grid interconnected variable speed wind generator. IEEE Syst J 7, No. 4, Dec. 2013.
9. Mutoh N, Ohno M, Inoue T (2006) A method for MPPT control while searching for parameters corresponding to weather conditions for PV generation systems. IEEE Trans Ind Electron 53(4)

Pedestrian Re-identification in Video Surveillance System with Improved Feature Extraction

Sina Salehian, Patrick Sebastian, and Abu Bakar Sayuti

Abstract In this work, we present a comparison between using different pedestrian re-identification (re-id) architectures. We have investigated the advantages of using more complex and deeper convolutional neural networks (CNNs) at the feature extraction stage. The re-id network is based on the summary network presented by (Ahmed and Marks 2015) which we have modified and enhanced. The comparison is done by replacing the feature extraction portion of the network. The newer improved models performed better than the baseline model and resulted in an accuracy of above 96% on our dataset and an accuracy of 92.09% on CUHK03 test dataset. The network takes 2 images as input and, outputs a confidence level indicating whether or not the 2 images depict the same person. The 2 images both go through a CNN with shared weights and the resulting 2 feature maps are used to compare and classify the 2 images as a positive or a negative match.

Keywords Re-id · Pedestrian · Re-identification · CNN · Deep convolutional

1 Introduction

1.1 A Subsection Sample

Research in the area of full body pedestrian Re-identification (re-id) has been thriving in recent years. Pedestrian re-id aims at identifying full body shots of people in non-overlapping camera views [1–7]. There are a number of challenges involved in identifying people in a new location from a different angle, such as drastic appearance, lighting and image quality changes. These types of variations can make the task of

S. Salehian (✉) · P. Sebastian · A. B. Sayuti
Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Perak, Malaysia

P. Sebastian
e-mail: patrick_sebastian@utp.edu.my

A. B. Sayuti
e-mail: sayutis@utp.edu.my

person re-id seem quite challenging. The main goal of any re-id system is to identify a subject based on a few sample images that it has seen of that subject before, these images could be from a different camera at a different location. Pedestrians re-id is of great importance for applications such as surveillance, security systems, activity analysis and tracking. Some of the other non-obvious challenges are: pose, occlusion, activity, angle and background variations in two images from two different cameras [8–12].

In this work, we are investigating the video-based re-id, by implementing the state-of-the-art image-based re-id techniques. Image based re-id is a more challenging task compare to video-based re-id due to the lack of information in the very few images available (in some cases only a single image) for each identity and also the lack of spatial and temporal information. However, an image-based re-id model would be more practical in a scenario with a small database of each ID. Moreover, the public datasets that are typically used by the image-based re-id models are much more diverse in terms of pose variation, lighting condition and angle variation, as compare to the datasets used by the video-based re-id models. Thus, alleviating biases in our model. Hence, our motivation for applying image-based re-id to perform re-id on a video [1–3, 13, 14].

The first step in re-id is to create descriptors that characterize the appearance of a person based on color and textural information of the person's clothes. The descriptor has to have high discriminative power in order to distinguish different people, but at the same time be able to handle variations in pose, angle, lighting and etc. [15].

Typically, the images have large variation of poses and are low quality; consequently, it has been proven by both, machine learning methods and also the conventional image processing methods that color and texture are the two most important features for the task of person re-id [15] and [16].

Recent work in the area of pedestrian re-id has been mainly on either creating discriminative features that are view, lighting, pose and background in variance [4, 6, 8, 11] or producing a distance metric between two positive or negative pairs of images [5, 9, 10]. In some cases, both of these methods are being implemented together, to achieve a better performance [1–3, 7, 12, 17]. Distance metric learning has proven to outperform discriminative feature learning in most cases. Hence, the focus of this work is on model utilizing both, distance metric and discriminative feature learning.

Using a single dataset that is bias to a particular pose, feature, background or appearance would not generalize well to real world scenarios. Thus, it is important to use a well-mixed batch of training data, in order to avoid bias towards a particular category. For example, most of the pedestrian re-id datasets suffer from this problem because either they are captured in an environment with a particular group of people such as CUHK1 [18] and CUHK3 [19] in which most the subjects are university students that are carrying backpacks or contains constant background information like crosswalks in PRID [20] dataset. VIPeR [21] dataset is a good example of cases where there is significant difference in resolution between the images. One solution is to use a combination of these datasets to create a more diverse dataset which in turn would help the model learn more robust features [22] the other solution is to use better and deeper models.

2   Related Work

Currently There are 2 main categories of pedestrian re-id, machine learning approaches and traditional image processing. The traditional approaches are usually carried out in two stages. Finding hand crafted features such as color histogram [23], local maximal occurrence representation (LOMO) [24] and hierarchal Gaussian descriptor (GOG) [25]. The handcrafted features are chosen in a way to contain as much discriminative features as possible. The second stage is similarity metric learning which has to determine if two images are a match or not. A number of metric learning methods have been proposed over the past few years [26–29]. The metric learning stage indicates how similar a pair of images are based on the features extracted from the first stage. It is expected that two images of the same person should have a smaller distance between them as compare to images from different people. The problem with such systems is that the two stages cannot be optimized simultaneously resulting in a sub-optimal system.

The deep learning based approaches have seen a lot of attention in recent years due to their success in various computer vision tasks. Recently, the power of deep convolutional networks in various tasks, including pedestrian re-id, has been demonstrated. In [19] a Siamese network was proposed that takes two images as inputs and compares them to one another. The visual features are extract using convolutional layers and the result will indicate the similarity score. Most common networks can be broken down into 2 main sub-network: one for feature extraction and another one for matching. The feature extraction sub-network can be either a shallow network with one or two convolutional and pooling layers or a deep network such as VGG, AlexNet and etc. The deeper network have to first be retrained over large data sets such as ImageNet dataset and then fine-tuned for person re-id. The matching sub-network has to select the features that minimize the misalignment between the learnt features and the ground truth. In short, the matching sub-network is at least one of the loss functions that the system has to minimize. Pairwise loss [30, 31] and triplet loss and its variations are the most commonly used loss functions in computer vision tasks.

In our work the feature extraction stage and the metrics learning stage are trained end to end simultaneously, which results in better optimization and higher performance.

3   Methodology

This work is aimed to be used in a distributed surveillance camera system with non-overlapping camera views. In a typical surveillance system there are multiple cameras at different locations and backgrounds. This makes it impossible to use simple tracking algorithms to follow each person around. Hence the need for a pedestrian re-id network. The first step is to select/auto select a person in one of the

cameras the subject is then detected and tracked throughout the field of view of that camera. The cropped images generated by the detector are extracted, labeled and stored in a database.

The pedestrian detector estimates the ROI (region of interest), which would be a bounding box around the area with a person in a given frame. The tracker keeps track of the identity of each box in consecutive frames in a video [32]. There are now multiple images from subject in our database.

At this stage the database contains images of the subject, these images can now be used to identify that subject in different cameras and locations. The new images are passed to the re-id network to be compared against the database to be identified.

As shown in Fig. 1, the two images are resized appropriately to the input size of the network, tabulated in Table 1. Each image is then fed to a CNN, which would result in a feature map. The two feature maps are then stacked on top of one another to form a cross-neighborhood difference layer based on the method described in the original paper [1]. The result is then fed into a fully connected layer to make a final classification. A positive match indicates that the two input images are from a same individual and a negative match indicates that they are not depicting the same person. The final stage is a SoftMax classification which would result in a probability between 0–1 for each class.

In this paper we are going to focus on the re-id block which consists of 2 parts.

- Feature extraction stage
- Difference learning stage
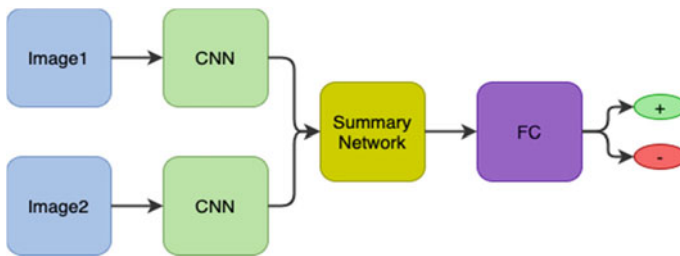
Fig. 1 Network block diagram

Table 1 Network parameters

Network	Baseline	AlexNet	DenseNet	InceptionV3
Batch size	150	150	12	12
Input size	160 × 60	227 × 227	160 × 160	160 × 160
CNN layers	2	5	121	126
CNN output feature depth	25	256	1024	2048
Total variables	2,322,192	4,180,052	8,430,356	24,400,636
Total size (bytes)	18 MB	32 MB	65 MB	186 MB

We have implemented the difference learning stage similar to the work done by [1], by computing cross-input neighborhood differences from the output of the feature extraction stage, which is then fed to a patch summary feature layer which summarizes the neighborhood difference maps by producing a holistic representation of the differences in each block. It is then followed by another convolutional layer in order to incorporate the spatial information into the model. This output is then connected to a fully connected layer, to perform the final classification of whether or not the 2 images belong to the same ID.

In this work we are going to mainly focus on the feature extraction stage, which in the case of our baseline is just 2 convolutional layers. We have experimented with 3 different convolutional neural networks (CNNs). Some of the main parameters of these networks are sown in Table 1. With the goal of achieving a much higher accuracy of re-id by learning a deeper feature map from the input.

A pair of images are fed through the network, they each go through the same CNN network and a feature map is extracted from each image. These 2 feature maps are then passed to the difference learning stage. We have replaced the CNN used in the baseline with AlexNet, DenseNet121 and InceptionV3. The networks have to first be truncated because the difference learning stage expects a feature map and all of the networks mentioned above are designed as classifiers with fully connected layers at the end. In all 3 cases the output was taken from the last layer before the fully connected layers (no top). In the case of AlexNet the original network has a max pooling layer before the fully connected layers in order to reduce the dimensions of the convolutional feature map so for AlexNet the output was taken from the max pooling layer. The network with no top is then connected to the rest of the summary network. The dimension of the feature map is different for each of the networks refer to Table 1 for input, feature map dimensions.

The database of new IDs that are saved during registration contains a 50 image for each subject. When a new image is being evaluated it has to be checked against all the images in the database. Every comparison result in a probability score which is then used to calculate an average for the 50 images that belong to each ID. The ID with the highest average probability score is chosen as the final prediction. The accuracy score is calculated based on the rate of correctly identified images in a batch.

In CUHK03, the dataset that is used for training, there are 13,164 images of 1,360 pedestrians. Therefore, the number of negative pairs is much greater than the positive pairs. If the model is trained on this dataset, it will just learn to identify every image as a mismatch which is an example of a data imbalance and overfitting. We performed a 2D translation on the dataset to produce 5 images from every image, based on [1]. However, the dataset is still imbalance and the negative pairs have to be down sampled to just twice the number of positive pairs. We randomly divide 1360 identities into non-overlapping train (1160), test (100), and validation (100) sets.

4 Results

The performance of these models was evaluated on the test data from the training dataset and also our own dataset that we created. Our dataset is divided into 2 Batches as follows: In batch number 1, all the videos have the same background with two IDs moving from side to side, recorded from different distances of 4, 6 and 8 m. There are also different conditions such as single subjects moving from left to right and vice versa, complete occlusion of either subjects in the middle of the frame by the other subject and subjects following each other. For batch number 1 all the videos are recorded 1280 × 720 pixels resolution. In batch number 2, there are multiple backgrounds and for each background there are multiple videos of two subjects (different from batch number 1) walking from side to side at different distance (3, 5 or 7 m). However, there are no special conditions. All the videos in this batch are recorded with a resolution of 320 × 240 pixels. Figure 2 depicts some sample images from our dataset.

In order to select the best Re-id dataset for the purpose of training, the baseline model was trained and on 3 different datasets and tested on our own dataset. The result from Table 2 shows that Market1501 performed well on ID0-2 but it had a very bad performance on ID3. CUHK03 had an overall higher accuracy as compare
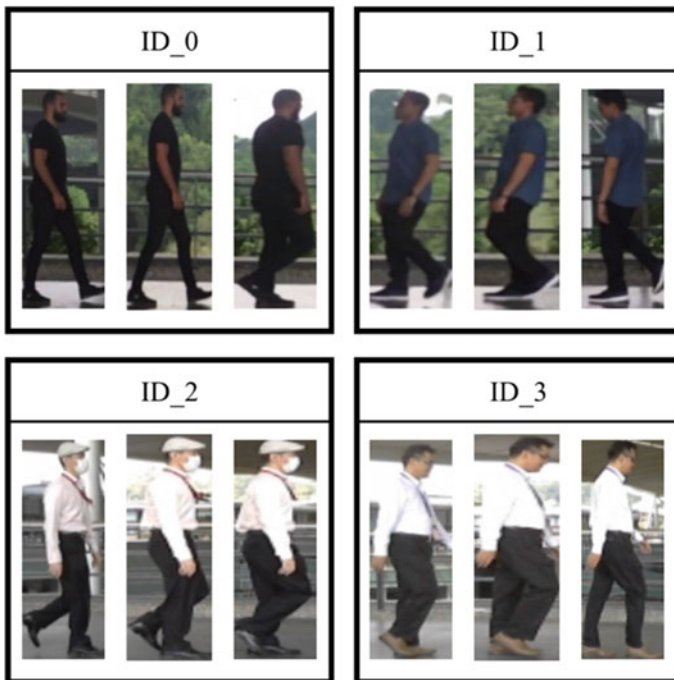
Fig. 2 Example of image from our dataset

Table 2 Training result on multiple datasets

ID	Number of images	Dataset	Accuracy (%)
0	925	CUHK01	100
		CUHK03	100
		Market1501	96.61
1	837	CUHK01	99.64
		CUHK03	93.44
		Market1501	100
2	1105	CUHK01	80.68
		CUHK03	95.55
		Market1501	99.05
3	1309	CUHK01	90.15
		CUHK03	90.12
		Market1501	54.03

to CUHK01 hence, it was chosen as the primary training dataset. Choosing a diverse dataset is an important factor in order to reduce unwanted biases, which could lead to an underperforming re-id system.

Figure 3 also shows the accuracy for each ID which is a measure of how many times the model correctly identified that ID for the given number of images.

Finding a perfect dataset with no bias is nearly impossible. Looking at the test result from Fig. 3 we can observe that selecting a better model can alleviate some of these biases. Ideally, we expect that the accuracy of the re-id be closer to a single value. However, we can see that in the case of the baseline and DenseNet121 model there is a 12% difference between the highest and lowest values. Whereas in AlexNet this difference is only 6%.

The second important criteria in evaluating an identification system is space between each user. In short, the difference in prediction probability of the model for a

Fig. 3 Result of training on multiple networks
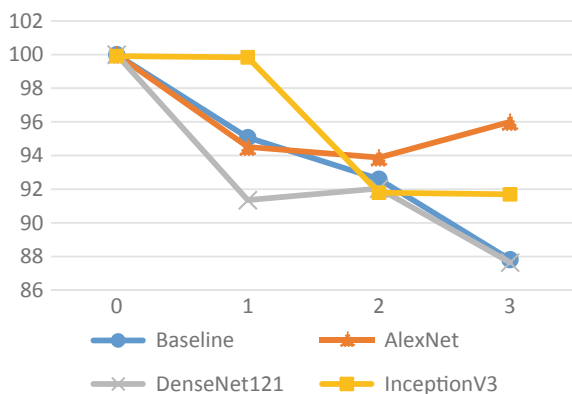
Table 3 Probabilities of different networks for each ID

ID	Network	Prob ID_0	Prob ID_1	Prob ID_2	Prob ID_3
0	Baseline	96.50	39.15	3.49	0.61
	AlexNet	95.76	22.42	1.833	0.01
	DenseNet121	99.06	14.84	1.08	0.01
	InceptionV3	92.65	5.35	0.7	0.02
1	Baseline	25.17	95.44	62.87	52.97
	AlexNet	17.8	83.29	4.87	1.41
	DenseNet121	17.83	87.99	13.39	2.43
	InceptionV3	6.14	97.17	36.37	19.7
2	Baseline	3.46	52.11	98.54	95.87
	AlexNet	0.08	1.84	89.51	73.86
	DenseNet121	0.63	7.31	99.52	82.64
	InceptionV3	0.22	19.21	94.01	84.96
3	Baseline	0.65	32.97	86.50	93.34
	AlexNet	0.00	1.70	70.48	93.38
	DenseNet121	0.13	2.52	65.17	97.49
	InceptionV3	0.04	18.06	70.40	89.44

match and non-match should be maximized. Table 3. Shows the average probability of each ID compared to all the IDs in the database.

For example, for ID 0 the average probability of all those images when they pass through the baseline model is 96.50% meanwhile, the second highest average probability for the same images is for ID_1 at 39.15% which means that the network is very confident in distinguishing between images of ID 0 and all the other 3 IDs in the database. However, the same cannot be said about ID2 where the average probability of ID_2 and ID_3 are very close 98.54% and 95.87% respectively. This is mainly caused by the bias in CUHK03 dataset that different subjects wear clothes with different colors and ID 2 and 3 have similar color clothes which increases the chance of misclassification. By optimizing the models, we can see that in all 3 cases AlexNet, DenseNet121 and InceptionV3 the differentiation between the IDs is more pronounced.

The probability column shows the average probability of those images belonging to each of the 4 IDs. Ideally the correct ID should have a probability of 100% and the rest be 0%. Therefore, a larger difference between these probabilities would ensure a lower chance of misclassification.

Typically, the datsets are recorded either in a university, a shopping mall or in the street and the subjects are wearing various color clothes and they appear in different angles and distance. This is good for training but in order to investigate what are the shortcomings of the system an engineered dataset like the one used in this paper helps with analyzing each parameter separately.
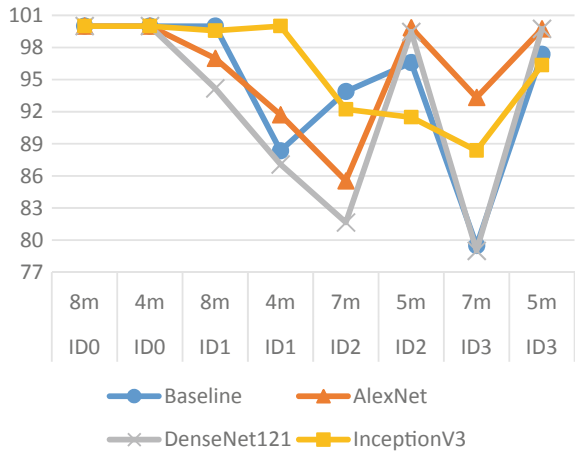
All the networks are trained on CUHK03 and the test accuracies are listed in Table 7. DenseNet121 performed the best with an accuracy of 92.09 as compare to the 88.7 which is the highest accuracy reported in the literature.

Table 6. AlexNet has the highest accuracy with InceptionV3 and baseline coming at second and third place respectively and DenseNet121 had the lowest accuracy on our dataset.

Figure 4 shows the accuracy of re-id of different models at different distances. Our dataset is recorded at different distances which enables us to investigate the effects of distance on performance of re-id networks. The database of existing IDs that every new image is compared against is consist of images from ID0 at 8 m and right side towards the camera, ID1 at 8 m and left side towards the camera, ID2 at 5 m and right side towards the camera and ID3 at 5 m and right side towards the camera

As evident by Fig. 4 and Table 4 for ID 2 and 3 the accuracy is higher at 5 m as compare to 7 m. This is mainly due to the lower quality of the images which causes the far objects to be a lot blurrier than the images at 5 m. However, for ID2 the difference between the average probabilities are much closer at 5 m than they are at 7 m. This is true for all of the models which means the models are struggling to distinguish between images of ID2 and ID3 at 5 m. The main reason for this unusual behavior is the fact that the images of ID 2 and 3 at 5 m have a lot of more similar features which results in a higher probability on both IDs.

The accuracy of the re-id for different point of view where in one case the camera sees the right side of the person and another that sees the left are shown in Fig. 5. As mentioned earlier the database contains only images from one side therefore this test can demonstrate the importance of having images from different angle in the database.

Looking at Fig. 5 we can see a pattern where the accuracy is either higher or similar when evaluating images with the same view. We can also observe that InceptionV3

Table 4 Accuracy of re-id for different architectures at different distances

ID	Range (m)	Images	Network	2nd Highest Prob	Highest Prob
0	8	447	Baseline	12.37	93.99
			AlexNet	22.52	86.34
			DenseNet121	12.02	98.96
			InceptionV3	3.33	91.96
	4	451	Baseline	12.05	93.68
			AlexNet	19.47	94.75
			DenseNet121	13.79	99.07
			InceptionV3	3.49	94.31
1	8	447	Baseline	23.93	91.74
			AlexNet	12.22	82.86
			DenseNet121	12.72	88.79
			InceptionV3	33.41	97.56
	4	451	Baseline	34.61	85.00
			AlexNet	23.55	83.56
			DenseNet121	24.1	86.73
			InceptionV3	30.86	96.4
2	7	753	Baseline	84.3	95.89
			AlexNet	62.1	84.98
			DenseNet121	76.78	99.44
			InceptionV3	80.34	91.81
	5	352	Baseline	92.59	97.07
			AlexNet	90.31	95.84
			DenseNet121	90.85	99.63
			InceptionV3	91.43	97.07
3	7	886	Baseline	63.93	82.86
			AlexNet	62.33	89.57
			DenseNet121	66.17	96.73
			InceptionV3	67.72	87.57
	5	423	Baseline	59.32	85.48
			AlexNet	81.89	98.72
			DenseNet121	63.77	98.57
			InceptionV3	74.16	92.05

Fig. 5 Result of training on multiple networks at different angles
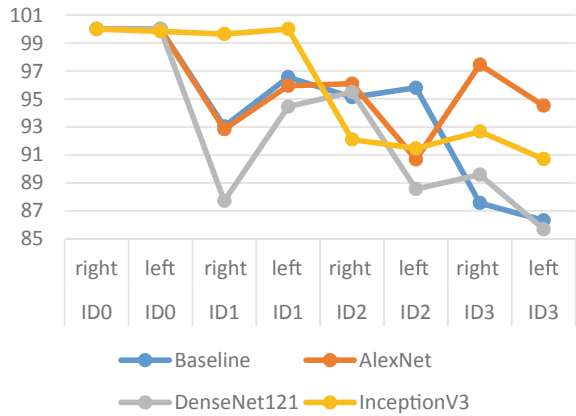
has a more uniform characteristic which results a similar accuracy for each ID, regardless of the side.

Table 5 shows that the difference between the average probabilities are above 10% for AlexNet, DenseNet121 and InceptionV3 which is an improvement over the 7% of the baseline model. The models are capable of handling variations in viewing direction which is evident by both the accuracy result and the difference in average probability values (Tables 6 and 7).

As we mentioned earlier, our dataset is consisting of 2 batches that are quite different in a few ways. Table 6 shows that all the 4 networks perform relatively good on batch number 1 with 3 of them having an accuracy of above 97% and only DenseNet121 is lower at around 95.6%. This could be either because batch 1 uses a similar background for all the images or because the 2 IDs are wearing distinctive clothing or just the fact that batch 1 has images with higher quality. As for batch 2, AlexNet performed the best with an accuracy of almost 95% which is only 3% lower than its batch 1 accuracy. InceptioV3 came second, with an accuracy around 91% in this case 8% lower than its batch 1 accuracy. The baseline and DenseNet121 had an accuracy around 90, with their accuracies dropping by 7 and 6% compare to batch 1 respectively. Because AlexNet had a high performance on both batches it has a much higher overall accuracy than the other networks.

In terms of difference in average probability all three networks outperformed the baseline based on values in Table 3.

As shown in Table 7 all three networks AlexNet, DenseNet121 and InceptionV3 have outperform previous networks in the literature. This is mainly because these designs are much deeper and more computationally expensive than the other networks. Which is a tradeoff between speed and performance. DenseNet + SumNet achieved the highest accuracy with AlexNet + SumNet in the second place and InceptionV3 + SumNet in third.

Looking at Figs. 3, 4 and 5 we can see a more uniform pattern in results from AlexNet and InceptionV3 where the results from different IDs are closer to one another.

Table 5 Accuracy of re-id for different architectures at different angles

ID	Side	Images	Network	2nd Highest prob	Highest Prob
0	Right	414	Baseline	13.17	93.86
			AlexNet	25.85	95.87
			DenseNet121	12.70	98.96
			InceptionV3	4.75	92.85
	Left	512	Baseline	12.78	93.80
			AlexNet	19.48	95.66
			DenseNet121	16.66	99.14
			InceptionV3	5.87	92.47
1	Right	360	Baseline	30.37	89.38
			AlexNet	20.07	82.32
			DenseNet121	20.36	84.38
			InceptionV3	40.46	95.77
	Left	477	Baseline	27.29	89.59
			AlexNet	15.86	84.12
			DenseNet121	15.67	91.09
			InceptionV3	32.86	98.37
2	Right	539	Baseline	88.80	96.32
			AlexNet	75.12	90.27
			DenseNet121	81.32	99.45
			InceptionV3	86.62	93.87
	left	566	Baseline	86.70	96.45
			AlexNet	72.60	88.75
			DenseNet121	83.96	99.59
			InceptionV3	83.31	94.15
3	Right	646	Baseline	59.41	80.76
			AlexNet	67.84	93.80
			DenseNet121	61.63	96.43
			InceptionV3	69.26	88.18
	Left	663	Baseline	80.17	87.14
			AlexNet	73.12	92.97
			DenseNet121	68.72	98.56
			InceptionV3	71.54	90.69

Table 6 Overall accuracy of re-id

Batch	Images	Network	Accuracy (%)
1	1762	Baseline	97.54
		AlexNet	97.26
		DenseNet121	95.67
		InceptionV3	99.88
2	2414	Baseline	90.21
		AlexNet	94.94
		DenseNet121	89.84
		InceptionV3	91.74
Overall	4176	Baseline	93.88
		AlexNet	96.1
		DenseNet121	92.76
		InceptionV3	95.81

Table 7 Overall accuracy comparison with other re-id methods

Method	Rank-1 (%)
IDLA [1]	54.74
CAN [3]	77.6
MLS [4]	87.50
PAR [5]	85.4
MSDL [6]	76.87
DCAF [8]	74.21
PersonNet [10]	64.80
LDNS [12]	62.55
Gated Siamese [14]	68.1
PDC [33]	88.7
DPFL [34]	86.70
DeepAlign [35]	85.40
Spindle [36]	88.50
JLML [37]	83.20
AlexNet + SumNet	91.89
DenseNet121 + SumNet	92.09
InceptionV3 + SumNet	90.18

The networks were trained using the following parameters, softmax cross-entropy loss function, Momentum Optimizer, learning rate 0.001 with a batch size of 12 for DenseNet121 and Inception, learning rate of 0.01 with a batch size of 150 for AlexNet. The networks were all pretrained on ImageNet which helps them converge faster. AlexNet was trained end to end for 100,000 iteration but for DenseNet121

and Inception only the top of the network was trained for the first 100,000 iterations by freezing the weights in the CNN portion of the network, it was then trained end to end for an additional 100,000 iterations.

5 Conclusion

We have investigated and demonstrated the advantages of using different datasets at the training stage. We have also looked how a deeper network with a deeper feature depth can improve the performance of pedestrian re-id systems.

Based on the result we can also conclude that a deeper network can pick up on more robust and discriminative features. The deeper networks can also help reduce the effects of biases in the data that they were trained on. This is evidence based on the improved accuracy and increased distance between the average probability values in deeper models such as AlexNet, DenseNet121 and InceptionV3.

Evaluating the network on an engineered dataset, similar to the one used in this paper, enables us to compare and test cases that do not exist on typical publicly available datasets.

During training the baseline and AlexNet models converged within 100,000 iterations however, the DenseNet121 and IceptionV3 models took much longer (more than 150,000 iterations) to converge which is expected, since they are much deeper networks.

It is important to note that the training batch size for the DenseNet121 and InceptionV3 models had to be reduced in order to reduce the overall size so that it would fit on a GPU with 11 GB of memory.

References

1. Ahmed E, Jones M, Marks T (2015) An improved deep learning architecture for person re-identification. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), Available: https://doi.org/10.1109/cvpr.2015.7299016
2. Lukezic A, Vojir T, Zajc LC, Matas J, Kristan M (2017) Discriminative correlation filter with channel and spatial reliability. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR)
3. Liu H, Feng J, Qi M, Jiang J, Yan S (Jul. 2017) End-to-end comparative attention networks for person re-identification. IEEE Trans Image Process 26(7):3492–3506
4. Guo Y, Cheung N-M (2018) Efficient and deep person re-identification using multi-level similarity. 2018 IEEE/CVF conference on computer vision and pattern recognition
5. Zhao L, Li X, Zhuang Y, Wang J (201) Deeply-learned part-aligned representations for person re-identification. In: 2017 IEEE international conference on computer vision (ICCV)
6. Qian X, Fu Y, Jiang Y-G, Xiang T, Xue X (2017) Multi- scale deep learning architectures for person re-identification. In: The IEEE international conference on computer vision (ICCV)
7. Chen W, Chen X, Zhang J, Huang K (2017) Beyond triplet loss: a deep quadruplet network for person re-identification. In: The IEEE conference on computer vision and pattern recognition (CVPR

8. Li D, Chen X, Zhang Z, Huang K (2017) Learning deep context-aware features over body and latent parts for person re-identification. In: The IEEE conference on computer Vi- sion and pattern recognition (CVPR)
9. Chen W, Chen X, Zhang J, Huang K (2017) A multi-task deep network for person re-identification. In: AAAI conference on artificial intelligence (AAAI)
10. Wu L, Shen C, van den Hengel A (2016) PersonNet: person re-identification with deep convolutional neural networks. Comput Vis Pattern Recogn
11. Matsukawa T, Okabe T, Suzuki E, Sato Y (2016) Hierarchical gaussian descriptor for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1363–1372
12. Zhang L, Xiang T, Gong S (2016) Learning a discriminative null space for person re-identification. CVPR
13. Xiao T, Li H, Ouyang W, Wang X (2016) Learning deep feature representations with domain guided dropout for person re-identification. In CVPR
14. Varior RR, Haloi M, Wang G (2016) Gated siamese convolutional neural network architecture for human re- identification. In: European conference on computer vision (ECCV)
15. Xiong F, Gou M, Camps O, Sznaier M (2014) Person re-identification using kernel-based metric learning methods. Comput Vis ECCV 1–16
16. Liao S, Hu Y, Zhu X, Li S (2015) Person re-identification by local maximal occurrence representation and metric learning. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR)
17. Li Z, Chang S, Liang F, Huang T, Cao L, Smith J (2013) Learning locally-adaptive decision functions for person verification. In: 2013 IEEE conference on computer vision and pattern recognition
18. Guillaumin M, Verbeek J, Schmid C (2009) Is that you? Metric learning approaches for face identification. In: 2009 IEEE 12th international conference on computer vision
19. Chen D, Yuan Z, Hua G, Zheng N, Wang J (2015) Similarity learning on an explicit polynomial kernel feature map for person re-identification. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR)
20. Chen J, Zhang Z, Wang Y (2014) Relevance metric learning for person re-identification by exploiting global similarities. In: 22nd international conference on pattern recognition
21. Li W, Zhao R, Xiao T, Wang X (2014) DeepReID: deep filter pairing neural network for person re-identification. In: CVPR
22. Lukežič A, Vojíř T, Čehovin Zajc L, Matas J, Kristan M (2018) Discriminative correlation filter tracker with channel and spatial reliability. Int J Comput Vis 126(7):671–688
23. Xiong F, Gou M, Camps O, Sznaier M (2014) Person re-identification using kernel-based metric learning methods. Comput Vis ECCV, pp 1–16, 2014. Available: https://doi.org/10.1007/978-3-319-10584-0_1
24. Liao S, Hu Y, Zhu X, Li SZ (2015) Person re-identification by Local Maximal Occurrence representation and metric learning. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), Boston, MA, pp 2197–2206. https://doi.org/10.1109/CVPR.2015.729 8832
25. Matsukawa T, Okabe T, Suzuki E, Sato Y (2020) Hierarchical gaussian descriptors with application to person re-identification. IEEE Trans Pattern Anal Mach Intell 42(9):2179–2194. https://doi.org/10.1109/TPAMI.2019.2914686
26. Li Z, Chang S, Liang F, Huang TS, Cao L, Smith JR (2013)Learning locally-adaptive decision functions for person verification. In: 2013 IEEE conference on computer vision and pattern recognition, Portland, OR, pp 3610–3617. https://doi.org/10.1109/CVPR.2013.463
27. Guillaumin M, Verbeek J, Schmid C (2009)Is that you? metric learning approaches for face identification. In: 2009 IEEE 12th international conference on computer vision, Kyoto, pp 498–505. https://doi.org/10.1109/ICCV.2009.5459197
28. Chen D, Yuan Z, Hua G, Zheng N, Wang J (2015) Similarity learning on an explicit polynomial kernel feature map for person re-identification. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), Boston, MA, pp 1565–1573. https://doi.org/10.1109/CVPR.2015.7298764

29. Chen J, Zhang Z, Wang Y (Dec. 2015) Relevance metric learning for person re-identification by exploiting listwise similarities. IEEE Trans Image Process 24(12):4741–4755. https://doi.org/10.1109/TIP.2015.2466117
30. Shi H et al (2016)Embedding deep metric for person re-identication a study against large variations. Available: https://arxiv.org/abs/1611.00137v1
31. Varior RR, Shuai B, Lu J, Xu D, Wang G (2016) A siamese long short-term memory architecture for human re-identification. In: Leibe B, Matas J, Sebe N, Welling M (eds) Computer vision—ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9911. Springer, Cham. https://doi.org/10.1007/978-3-319-46478-7_9
32. Salehian S, Sebastian P, Sayuti AB (2019)Framework for pedestrian detection, tracking and re-identification in video surveillance system. In: 2019 IEEE international conference on signal and image processing applications (ICSIPA), Kuala Lumpur, Malaysia, pp 192–197
33. Su C, Li J, Zhang S, Xing J, Gao W, Tian Q (2017) Pose-driven deep convolutional model for person re-identification. In: 2017 IEEE international conference on computer vision (ICCV), Venice, pp 3980–3989. https://doi.org/10.1109/ICCV.2017.427
34. Chen Y, Zhu X, Gong S (2017) Person re-identification by deep learning multi-scale representations. In: 2017 IEEE international conference on computer vision workshops (ICCVW), Venice, 2017, pp 2590–2600, https://doi.org/10.1109/ICCVW.2017.304
35. Zhao L, Li X, Wang J, Zhuang Y (2017) Deeply-learned part-aligned representations for person re-Identification. In: The IEEE international conference on computer vision (ICCV), pp.3219–3228
36. Zhao H, et al (2017) Spindle net: person re-identification with human body region guided feature decomposition and fusion. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), Honolulu, HI, pp 907–915. https://doi.org/10.1109/CVPR.2017.103
37. Li W, Zhu X, Gong S (2017) Person re-identification by deep joint learning of multi-loss classification. In: Proceedings of the 26th international joint conference on artificial intelligence, pp 2194–2200

Smart Children Management Using Data Analytics, Machine Learning and IoT

Md Jobair Hossain Faruk and Muhamad Hariz Muhamad Adnan

Abstract One of the biggest challenges faced by parents is managing the difficult or defiant behavior of children of the twenty-first century. Technology can help to provide support in managing children's activities and helping parents to be in control. Recent advances in IoT have sparked the interest of researchers to develop a smart management system for children. This paper presents a review of the smart management of children using the internet of things (IoT), machine learning, and Data Analytics. The objective is to identify potential approaches regarding smart management for children based on IoT for future researchers. The reviews were conducted from secondary sources, where conference proceedings and journals have been analyzed from Scopus and other databases. The finding indicates that IoT-based smart system has robust accuracy and performance. Thus, exploiting IoT for the development of a smart management system for children will provide significant benefits to the future generation.

Keywords Smart children management · IoT · Machine learning · Data analytics · Healthcare

1 Introduction

Children's programs are now more important to parents as they are encouraged to engage their children in as many activities as possible to help them plan for their future [1, 2]. Many studies have described various methods used by families to handle all the work they need to do around the overlapping duties of work, education, home, and enrichment events [1–7]. Managing the challenging or rebellious actions of

Md J. H. Faruk
College of Computing and Software Engineering, Kennesaw State University, Marietta, USA
e-mail: mhossa21@students.kennesaw.edu

M. H. M. Adnan (✉)
Computing Department, Faculty of Art, Computing and Creative Industry, Universiti Pendidikan Sultan Idris, Tanjung Malim, Malaysia
e-mail: mhariz@fskik.upsi.edu.my

twenty-first century children is one of the greatest challenges faced by parents [8]. As information and communication technologies (ICT) and smart devices can be an integral part of self-management plans, technology can help support the management of children's activities and help parents to be in charge [1–9].

The concept of a smart system has been proposed by many scholars where the generic term 'smart system' involves several classes of systems composed of sensors, actuators, a network, and a signal processing unit that delivers definite signals and fulfill subtasks or complete tasks [10]. Smart systems may visualize as an integration of the ideas of smart people, smart environments, and smart living and it expands into not only IoT but also data analytics and machine learning which is contextualized with an open and comprehensive network of intelligent objects that can auto-organize, share information, data and resources while reacting and acting in face of situations and changes in the environment [4–13].

In this paper, we present potential approaches towards an ideal smart management system for children that shall help future researchers to get a comprehensive knowledge of plenty of existing techniques. Figure 1. presents some potential approaches for the smart management of children.

The paper is organized as follows: In the second section, potential methods of IoT based on the smart management system for children are presented followed by presenting the discussions. The last section provides the conclusion.
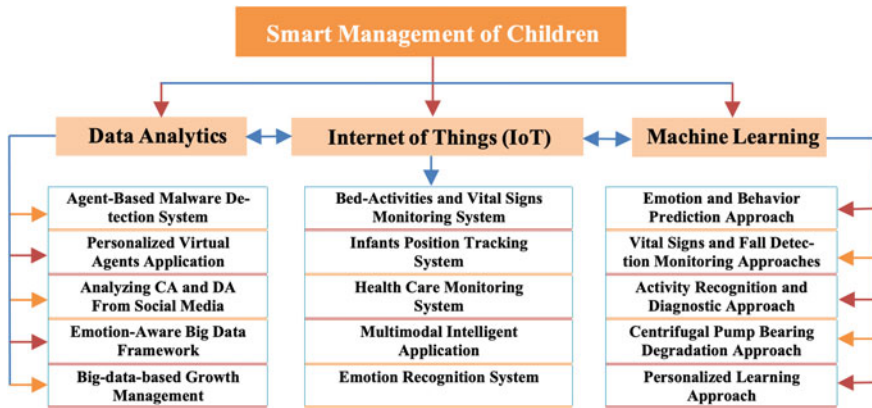
Fig. 1 Illustrations of potential techniques towards smart management of children

2 Smart Management of Children Review

2.1 *Internet of Things (IoT)*

Bed-Activities and Vital Signs Monitoring System: Monitoring sleep and physical activities are vital because of their uncertain moves and emerging technologies offer new opportunities for parents to monitor children regardless of their position and pursuit. Besides, monitoring vital signs also significant as indicators of one's health condition and assurance of proper circulatory, respiratory, neural and endocrinal functions towards keep track of the children's health condition and it refers to the parameters of the body's basic functions [14]. A group of researchers from the University of Georgia proposed vibrations based on a real-time, contactless bed-mounted monitoring system that can estimate vital signs, body movement, falls from bed, posture changes and on/off bed detection using during sleep time [15, 16].

One of its primary components is Raspberry Pi 3 that connects with a seis-mometer to get real-time measurement data, maxima statistics and an instantaneous property-based approach towards enhancing the quality of monitoring sleep condition and posture. Envelop based approach is integrated to estimate BPMh for measuring heart rate (HR) [16]. System evaluation indicates its robust accuracy of both approaches to monitoring bed activities and detecting vital signs and body movements. This approach can be one of the potential components of a smart management system because of its robust filtering techniques to estimate the heart rate (HR) and respiration rate (RR) as well as other sleep activities.

Health Care Monitoring System: According to the National Academy of Science [17], In addition to supporting this viewpoint, wellness during childhood sets the stage for adult health, but also provides an important legal, social and economic obligation to ensure that all children are as healthy as they can be. Amrita University researchers [18] present an IoT-based smart and secure health care monitor system for children to monitor behavioral and health issues and predict child disorders. C4.5 algorithm has been adopted to predict the disorder of children by analyzing collected therapeutic games and bodily parameter data. Besides. an android game has been presented that enables the device to evaluate the disorders of the child using visuospatial and strop test, sustained in attention-al blindness, and ADD/ADHD test. A case study is carried out among 500 children from schools and health centers. The result indicates its capability in predicting disorders in the child with 93.4% accuracy. Contact-free and privacy-preserving solutions for monitoring bed activities need to be emphasized towards the smart management of children.

2.2 *Machine Learning*

Activity Recognition and Diagnostic Approach: Physical activity of children recognition plays a significant role and value to measure towards analyzing, predicting,

child-development monitoring, energy-expenditure estimation, child-obesity prevention, child safety in and around the home, etc. [19, 20]. The activity recognition approach is one of the important methods for the SMART management system [21], thus a group of researchers (2020) [22] proposed a real-time and continuous activity recognition-based knowledge-driven system with aims to provide an activity recognition approach for smart homes (SH).

The proposed method adopted an ontology-based recognition model and semantic subsumption reasoning algorithms towards recognition of activities for SH. To trace the activity of daily living (ADL), the conceptual activity model has been structured for the system. The agent-based integrated system architecture was conceived for a real-time, continuous activity recognition system. The proposed system has been evaluated in various scenarios and satisfactory accuracy in activity recognition has achieved which is 94.44% while the average run time was measured 2.5 s.

2.3 Data Analytics

Emotion-Aware Big Data Framework: Recognition of feelings is an important topic in different fields of activity in particular for futuristic SMART management systems and it may be one of the influential fields for this area of research [23]. The ability to recognize and understand emotions portrayed as non-verbal cues is essential to the development of related approaches [24]. Hossain and Muhammad, [25] introduced a 5G enabled, emotion-aware connected, healthcare-based big data framework towards classifying the data to recognize the emotions of the patient; particularly whether they feel pain or not.

The work integrated Bluetooth technology for the positioning, local binary pattern (LBP), and an interlaced derivative pattern (IDP) to detect the signals. The study also adopted sensor technology towards recognizing speech and image signals of the patient as input data separately within a smart home scenario. Input data is used to estimate the final score to detect emotion where it processes the speech and signals where center-symmetric LBP (CSLBP) has been utilized to calculate the histogram, Fig. 2. The proposed approach has demonstrated by conducting experiments among fifty trained university students where the system shows the highest accuracy (99.87%) of detecting emotion.

Reading Emotional Parameters: A group of researchers [26] investigated the possibility of reading the emotional parameters of humans where they carried out an experimental study to identify six emotional parameters including engagement, excitement, focus, stress, relaxation, and interest using brain signals. Supervised learning classifiers, Naive Bayes, and Linear Regression were adopted to show the accuracy and competency of well-known electroencephalography (EEG) based Emotiv Epoc+ Neuroheadset. Experimental studies indicate 69 % and 62 % improved accuracy in reading six emotional parameters of the human subjects and such study can be a focal point for the smart children management.
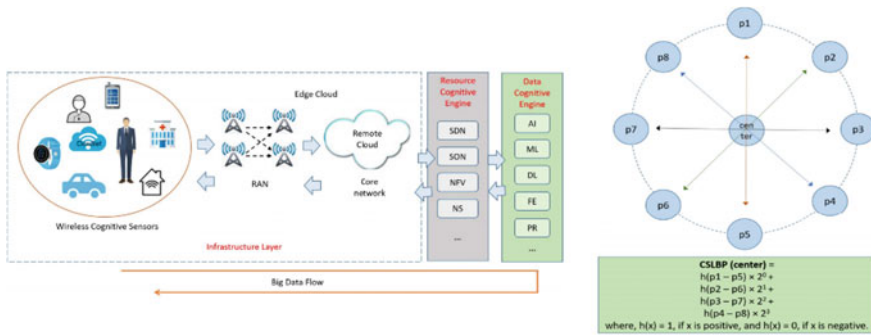
Fig. 2 Big data flow

3 Discussion

A review of smart management of children based on IoT, Machine Learning and Data Analytics has been presented where potential methods and techniques proposed by researchers have been highlighted. The summary of the findings from the literature review is illustrated in Table 1. Throughout the studies, we discovered the approaches from different areas and a combination of suitable methods that may lead us towards the development of a new smart management system for children.

Considering the challenges that the smart management applications for children faced some common approaches have been identified. Eye-tracking data analysis, emotion recognition, facial expression, infant activity monitoring and real-time interactive growth management are some of the areas that have huge potentials for the development of a smart management system for children. Similarly, the vital signs monitoring approach is an important technique that can be further investigated and integrated into a smart management system to keep track of the children's health condition. The development of such kind of SMART systems for children may initiate a huge impact on the people globally, especially among the people in Asia. We conducted a preliminary survey among 22 parents aged 35–45 from Malaysia and Bangladesh to estimate the guardian's approaches to monitoring their children. A self-administered questionnaire was utilized to collect the data from the parents and there were asked to answer and complete the questionnaire online where 48.2% are female. The study indicates that most respondents have medium knowledge of technologies, and they are more familiar and dependent on conventional techniques. A 14% of respondents utilize the existing monitoring approaches, for instance, CCTV cameras, and tracking systems. However, such a tracking system was not efficient since each component needs to monitor separately which leads the users dispassionate.

Table 1 Initial findings of selected approaches

Approaches	Objectives	Outcome
Bed-activities and vital signs monitoring system	Development of a real-time, contact-free, privacy-protective sensor-based system that shall capable to monitor not only sleeping activities, but also vital signs, and fall detection of an individual's bedtime	Utilizing verities of novel non-intrusive methods, the system demonstrated robust performance in three environments in the principle of measuring heart and respiration rate, on-bed, off-bed and sitting movements; finally, posture changes. The accuracy of detecting and alerting falls from the bed, monitoring heart and respiration rate is significantly higher than 95%
Health care monitoring system	Introduces a novel, IoT based healthcare approach to monitor behavioral and health issues and predict child disorders. Using an extended approach, the system may be facilitated by robot assistant diabetic treatment for children	Both health care and eHealth approaches indicate their potentials and can be assets for future researchers. Combinations of two novel healthcare approaches may initiate a significant impact on a smart management system for children where guardians shall be able to identify and monitor their children's health issues; particularly, a few well-known disorders. Such as ADHD, mental health, and diabetics
Emotion-aware big data framework	Introduces an emotion-aware connected, big data-based framework to recognize the emotions of the patient; particularly whether they feel pain or not	The outcome of an emotion recognition approach is interesting particularly for its accuracy that demonstrated higher than 99.87%. Although the system is evaluated among university students; however, such a system can be huge potentials for smart management of the system of children because of the demand and needs to ensure the children's health issues

4   Conclusion

In this paper, we reviewed and discussed the potentials of smart management of children. The study shows a huge expectation of parents for the SMART system towards monitoring their children in a way that is efficient, convenient, and that parallels with the expectations. By utilizing IoT, a complete and robust smart child

management system may be proposed. As similar to smart homes and smart cities, not only sensors-based IoT but also data analytics and machine learning shall be the revolutionary technique towards developing a smart management system. However, it is expected to conduct future research to design and develop a fully functional smart management system by interpreting the methods and approaches, and hundreds of thousands of people shall be benefited using the futuristic system.

References

1. Davidoff S, Lee MK, Yiu C, Zimmerman J, Dey AK (2006) Principles of smart home control. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 4206 LNCS:19–34
2. Lee MK, Davidoff S, Zimmerman J, Dey AK (2007) Smart bag: managing home and raising children. Proc 2007 Conf Des Pleasurable Prod Interfaces, DPPI'07. 434–437.
3. Lee MK, Davidof S, Zimmerman J, Dey A (2006) Smart homes, families, and controls. 2006; Available from: https://pdfs.semanticscholar.org/b751/6530ebffe5b7202f0ac201aefc7850b87 a52.pdf?_ga=2.123838666.1183646419.1597302163-2076110809.1586498474
4. Yang H, Lee W, Lee H. IoT smart home adoption: the importance of proper level automation. J Sens (2018)
5. Taylor AS, Swan L (2005) Artful systems in the home. CHI 2005 Technol Safety, Community Conf Proc - Conf Hum Factors Comput Syst 641–650
6. Tolmie P, Pycock J, Diggins T, MacLean A, Karsenty A (2002) Unremarkable computing. Conf Hum Factors Comput Syst - Proc 4(1):399–406
7. Darrah CN, English-Lueck JA, Freeman JM (2000) Living in the eye of the storm: controlling the maelstrom in silicon valley, 1–10
8. Childmind. Managing problem behavior at home: a guide to more confident, consistent and effective parenting. Child Mind Inst Inst [Internet]. Available from: https://childmind.org/art icle/managing-problem-behavior-at-home/
9. Sendra S, Parra L, Lloret J, Tomás J (2018) Smart system for children's chronic illness monitoring. Inf Fusion. 40:76–86
10. Kanoun O, Universit T, View E, Kanoun O (2014) Smart systems and devices : innovative key modules for engineering applications
11. Lanting C, Lionetto A (2015) Smart systems and cyber physical systems paradigms in an IoT and industrie/y4.0 context. S5002
12. Madakam S, Ramaswamy R, Tripathi S (2015) Internet of things (IoT): a literature review. J Comput Commun 03(05):164–173
13. Apanaviciene R, Vanagas A, Fokaides PA (2020) Smart building integration into a smart city (SBISC): development of a new evaluation framework. Energies 13(9)
14. Teixeira CC, Boaventura RP, Souza ACS, Paranaguá TT de B, Bezerra ALQ, Bachion MM et al (2015) Aferição de sinais vitais: Um indicador do cuidado seguro em idosos. Texto e Context Enferm 24(4):1071–1078
15. Clemente J, Li F, Valero M, Song WZ (2019) Demo: contactless device for monitoring on-bed activities and vital signs. Proc - 2019 IEEE Int Conf Smart Comput SMARTCOMP 2019. 472–474.

16. Clemente J, Valero M, Li F, Wang C, Song WZ (2020) Helena: real-time contact-free monitoring of sleep activities and events around the bed. In: 18th annual IEEE international conference on pervasive computing and communications
17. (2005) Children's health, the nation's wealth: assessing and improving child health. Choice Rev Online 42(10):42–5896–42–5896
18. Binu PK, Akhil V, Mohan V (2017) Smart and secure IoT based child behaviour and health monitoring system using hadoop. In: 2017 International conference on advances in computing, communications and informatics (ICACCI) 2017. 2017-Janua:418–423
19. Boughorbel S, Breebaart J, Bruekers F, Flinsenberg I, Ten Kate W (2011) Child-activity recognition from multi-sensor data. ACM Int Conf Proceeding Ser
20. Sindhu P (2018) Toddler activity recognition using machine learning. Loyola eCommons [Internet]. 2018; Available from: https://ecommons.luc.edu/cgi/viewcontent.cgi?article=4755&context=luc_theses
21. Guo J, Li Y, Hou M, Han S, Ren J (2020) Recognition of daily activities of two residents in a smart home based on time clustering. Sensors (Switzerland) 20(5)
22. Rawashdeh M, Al MG, Samarah S, Hossain MS, Muhammad G (2020) A knowledge-driven approach for activity recognition in smart homes based on activity profiling. Futur Gener Comput Syst 107:924–941
23. Dzedzickis A, Kaklauskas A, Bucinskas V (2020) Human emotion recognition: review of sensors and methods. Sensors (Switzerland) 20(3)
24. Sivasangari A, Ajitha P, Rajkumar I, Poonguzhali S (2019) Emotion recognition system for autism disordered people. J Ambient Intell Humaniz Comput
25. Hossain MS, Muhammad G (2018) Emotion-aware connected healthcare big data towards 5G. IEEE Internet Things J 5(4):2399–2406
26. Hossain Faruk MJ, Valero M, Shahriar H (2021) An investigation on non-invasive brain-computer interfaces: emotiv epoc+ neuroheadset and its effectiveness. In: 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), pp 580–589. https://doi.org/10.1109/COMPSAC51774.2021.00086

Optical Wireless Communication Based Wireless Body Area Network for Remote Patient Monitoring

Komal Masroor , **Sajid Mumtaz, Micheal Drieberg** , **and Varun Jeoti**

Abstract Internet-of-Things (IoT) based Wireless Body Area Networks (WBANs) are likely to change the traditional healthcare infrastructure by allowing remote monitoring of patients outside the conventional clinical settings e.g. in homes. However, technology remains a key enabler for such a system realization. Currently, Radio Frequency (RF-) based Zigbee networks are being employed for data transmission and reception, nevertheless, with the widespread use of IoT devices, the unlicensed Industrial, Scientific and Medical (ISM) spectrum may exhaust soon. Consequently, Visible Light Communication (VLC) technology is being considered as an alternative to RF owing to its extremely high unregulated bandwidth and low cost. Additionally, it is considered safe for human beings. Several studies in the literature have reported the effects of transceiver related parameters on a VLC link, however, in this study, we examine the impact of various parameters that are pivotal for VLC channel characterization in healthcare settings. We show that reflectivity of room surfaces, angle of irradiance of light emitiing diodes (LEDs), and on-body receiver heights have a significant impact on the received power (RXPWR). Furthermore, we validate that increase in number of transmitters (T_{xs}) reduce the fluctuations in RXPWR by 75%.

Keywords Visible light communication (VLC) · Optical body sensor network (OBSN) · Remote patient monitoring (RPM) · Uniform receiver power

K. Masroor (✉) · S. Mumtaz · M. Drieberg
Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia
e-mail: komal_17000856@utp.edu.my

K. Masroor
NED University of Engineering and Technology, Khi, Pakistan

V. Jeoti
Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia

1 Introduction

Visible Light Communication (VLC) technology is becoming increasingly popular as an alternative to Radio Frequency (RF-) based communication in many applications. Remote Patient Monitoring (RPM) is one such application where VLC enables health monitoring inside indoor spaces via the pre-installed lighting infrastructure [1]. The use of VLC in RPM can be attributed to its high unregulated bandwidth, low cost, resistance to electromagnetic interference (EMI), simplicity of the transceiver design, and its inherent features like security and safety for human body. Generally, communication in optical body sensor network (OBSN) is possible between the coordinator and on-body motes or between coordinator and the optical access point (OAP). In both the scenarios, reliable transmission and reception of patient's data is of utmost importance.

VLC enabled healthcare monitoring is in its early stage and investigations are being done to prove its viability for such systems. The initial study to demonstrate the use of optical technology for patient monitoring was presented in Torkestani et al. [2]. Following this, several studies have established the use of VLC/ infrared (IR) in monitoring applications. For instance, in Scully et al. [3], experimental demonstration of use of mobile phone camera to derive various physiological parameters from skin colour has been performed. The authors in Khalid et al. [4] used bipolar signaling and high pass filtering to avoid unwanted background ambient light interferences in a diffused mobile IR-based WBAN system along with power savings. Line-of-sight (LoS) is usually a hindrance when it comes to mobility; however, use of this type of link design to offer mobility has been investigated in Torkestani et al. [5]. The inclusion of first and second order reflections along with the major LOS component to evaluate system performance has been examined in Alyan et al. [6]. Effects of using LoS and diffuse configurations for patient monitoring in indoor environment using error coding scheme have been studied in Torkestani et al. [7]. Mostly, non-LoS (NLoS) i.e. diffuse links are employed to ensure maximum coverage indoors, nevertheless, there are several other factors that affect the OBSN link performance including room size, transceiver locations, number of transceivers, semi-angle at half power (SAHP) of the LED, field-of-view (FOV) of the receiver (R_x), reflectivity (ρ) of room surfaces and body-shadowing. In our previous study [8], we highlighted the importance of transmitter (T_x) locations on uniform distribution of RXPWR for single and four transmitter scenarios. In this study, we use a single T_x scenario to study the effects of SAHP and reflectivity on RXPWR. Subsequently, we use five T_{xs} to evaluate the impact of number of T_{xs} on uniform distribution of RXPWR. Besides, the effects of different on-body R_x heights are also invesigated (Fig. 1).
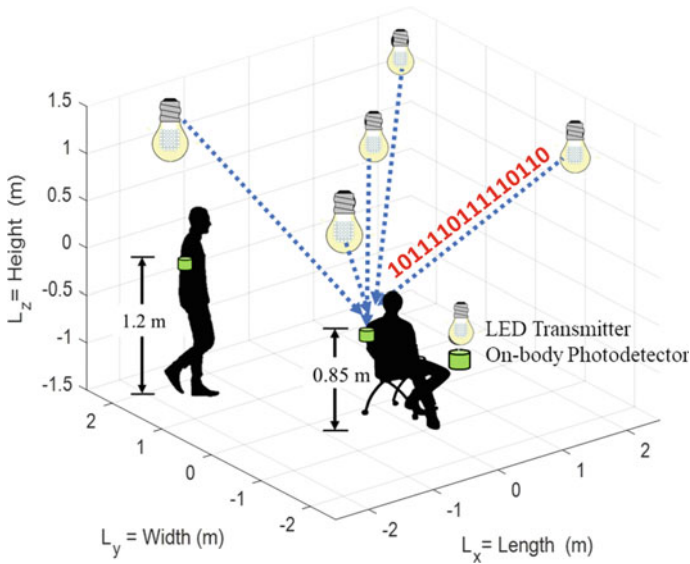
Fig. 1 System model for OBSN

2 System Description

2.1 System Model

The indoor environment is assumed to be an empty room with dimensions $5 \times 5 \times 3$ m^3 which is consistent with the literature [9, 10]. The LED transmitters, facing vertically downward, are located on the ceiling as described in Masroor et al. [8] with an additional T_x located at the centre of the room with coordinates (0, 0, 3). Also, a photodetector (PD) is assumed to be placed around the patient's arm and is oriented towards the ceiling such that the LEDs and PD are facing each other at all times. The simulations to determine the RXPWR have been performed in MATLAB using the ray tracing algorithm as described in Barry et al. [11]. The simulation parameters have been summarized in Table 1.

2.2 Channel Model and Received Power

An intensity modulation and direct detection (IM/DD) channel model has been considered for VLC-based LED transmitters, given by Ghassemlooy et al. [12]:

$$y = H\gamma x + n \tag{1}$$

Table 1 Summary of simulation parameters

Parameters	Values
For transmitter	
Center luminous intensity per LED chip	0.73 Cd
Lambert's order	1
Number of LEDs per array	3600 (60 × 60)
For receiver	
Field-of-View	85º
Physical detector area	1 cm^2
Gain of an optical filter	1.0
Refractive index of lens	1.5
Height (from the floor) → Sitting	0.85 m
Height (from the floor) → Standing	1.2 m

where y is the received signal power, H is the channel static gain, γ is the receiver responsivity, x is the transmitted signal, and n denotes the Additive White Gaussian Noise (AWGN) with double-sided noise power spectral density N_0.

The total received power ($P_{r\text{-}total}$) due to LoS ($H(0)$) and first-order NLoS (H_{nlos}^k) reflections can be calculated as [12]:

$$P_{r-total} = \left(H(0) + \sum_{k=1}^{\infty} H_{nlos}^k \right) P_t \tag{2}$$

where k is the order of reflections assumed to be 1 in this study.

3   Results and Discussion

In this study, simulations were carried out for a $5 \times 5 \times 3$ m^3 room to examine the effects of physical parameters i.e. the reflectivity of walls and SAHP of LEDs on the performance of an OBSN link. Illustrated in Fig. 2 is the RXPWR obtained in Watts for LoS configuration at different angles of irradiance. It can be observed in Fig. 2a that with single transmitter (T_x), the RXPWR is mainly concentrated at the centre when angle of irradiance is 10º while it is more evenly distributed when LED transmits at 60º (see Fig. 2b). Although the RXPWR has decreased in the latter due to the increasing distance of the PD from the T_x, it provides better coverage. Consequently, to ensure maximum coverage, further analysis will be based on LED transmissions at SAHP = 60º. Furthermore, Fig. 2c illustrates the effects of reflectivity of walls on the RXPWR via NLoS link for two values i.e. for $\rho_{\min}$ and ρ_{max} corresponding to 0.1 and 0.8, respectively. It can be observed that the reflected power reduces considerably with low reflectivity and vice versa indicating that reflectivity of surfaces contributes
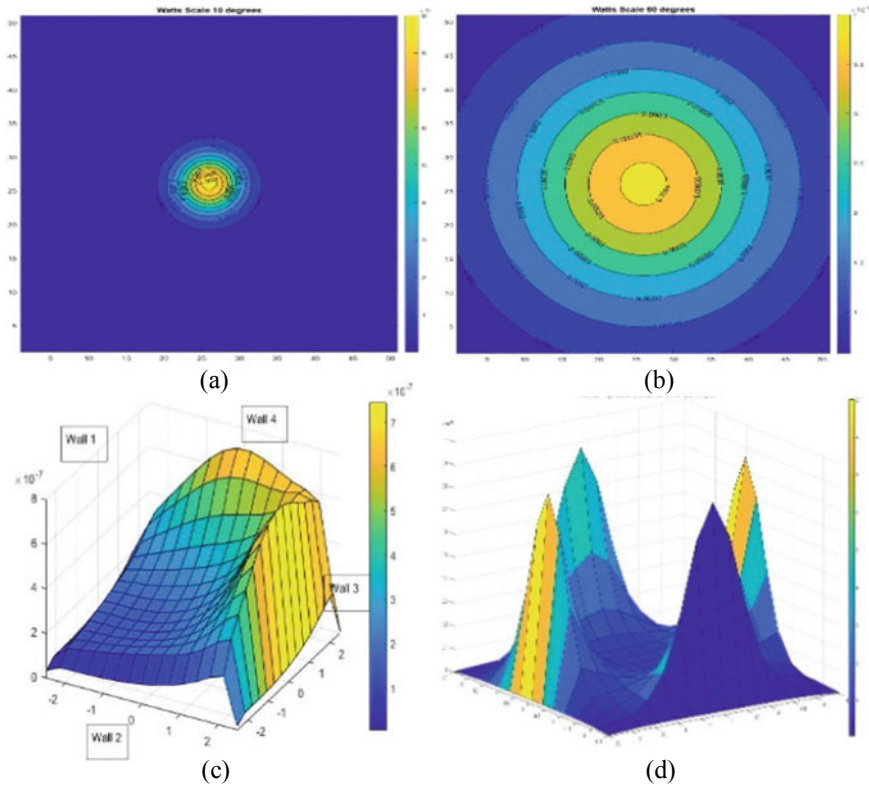
Fig. 2 Received Power for Single Transmitter with **a** SAHP 10º and **b** 60º, **c** NLoS distribution at $\rho = [0.1, 0.8]$ and **d** SAHP $= 10$º

significantly to RXPWR computations. Besides, Fig. 2d depicts how a small value of SAHP significantly reduces the RXPWR. Therefore, when designing an environment for patient monitoring using a VLC link, a high value of SAHP and ρ is desirable.

Figure 3 illustrates the impact of number of T_{xs} on RXPWR. From Fig. 3a, it is evident that a single T_x is not sufficient to provide uniform coverage to the entire area and there could be communication dips at several points in the room with RXPWR spread over ~ 20dBm range; thereby, resulting in significant fluctuations. Conversely, in Fig. 3b depicting the five transmitters' scenario, the range has reduced to ~ 5 dBm implying that the RXPWR fluctuations have reduced by 75%. Based on these results it can be said that use of multiple T_{xs} is inevitable to establish a reliable communication link with uniformly spread RXPWR. Apart from the number of T_{xs}, the R_x heights also play a key role in determining the performance of the OBSN network. As illustrated in Fig. 4, it can be seen that the values of RXPWR obtained for standing at different points in the room are higher than those obtained for sitting. This difference can be attributed to the varying distances (d_{TR}) between the T_{xs} and R_x i.e. when d_{TR} is high, the RXPWR would be less and vice versa.
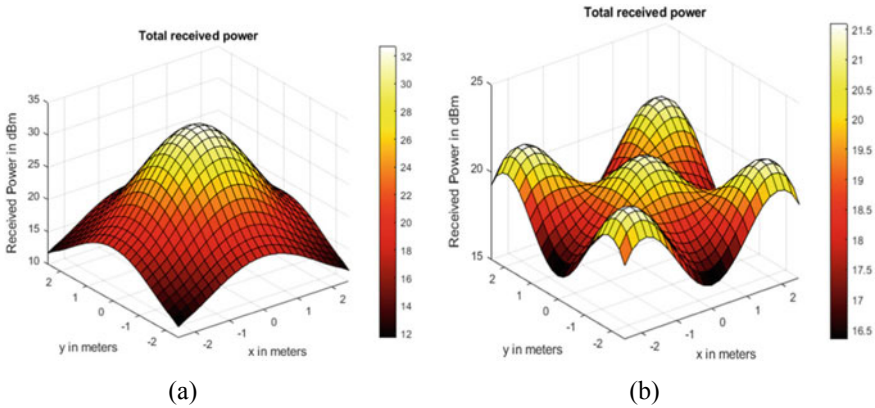
Fig. 3 Received power for **a** Single and, **b** Five transmitters

Fig. 4 Received power for varying PD heights
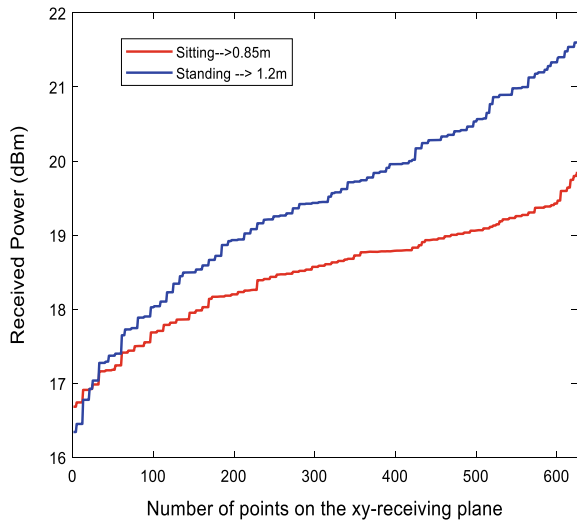
4 Conclusion

This paper emphasized the importance of different transceiver parameters that affect the performance of an OBSN link to carry out the communication effectively. It was found that reflectivity of surfaces and angle of irradiance of the LED sources play an important role in determining the power received at the PD. Also, it was shown that a single T_x leads to significant fluctuations in RXPWR, therefore, multiple T_{xs} must be used to establish uniform distribution. Moreover, it was demonstrated that as the distance between T_{xs} and R_x reduces, the RXPWR will be better and vice versa.

References

1. Julien-Vergonjanne A, Sahuguede S, Toumieux P (2015) Remote mobile patient monitoring using diffuse optical transmissions, pp 553–557
2. Torkestani SS, Julien-Vergonjanne A, Cances J-P (2010) Mobile healthcare monitoring in hospital based on diffuse optical wireless technology. In: 21st Annual IEEE international symposium on personal, indoor and mobile radio communications. IEEE
3. Scully CG et al (2012) Physiological parameter monitoring from optical recordings with a mobile phone. IEEE Trans Biomed Eng 59(2):303–306
4. Khalid A, Cossu G, Ciaramella E (2013) Diffuse IR-optical wireless system demonstration for mobile patient monitoring in hospitals. In: Transparent optical networks (ICTON), 2013 15th international conference on 2013. IEEE
5. Torkestani SS et al (2012) Indoor optical wireless system dedicated to healthcare application in a hospital. IET Commun 6(5):541–547
6. Alyan E, Aljunid S, Anuar M (2015) Mobile optical wireless system for healthcare continuous monitoring using IR technology. J Theor Appl Inf Technol 78(3):353
7. Torkestani SS, et al (2011) Performance and transmission power bound analysis for optical wireless based mobile healthcare applications. In: Personal indoor and mobile radio communications (PIMRC), 2011 IEEE 22nd international symposium on 2011. IEEE
8. Masroor K, Jeoti V, Drieberg M (2019) Analyzing the effects of LED lamp arrangements on performance of an indoor visible light communication system. In: 2019 IEEE 14th Malaysia international conference on communication (MICC) 2019. IEEE
9. Komine T, Nakagawa M (2004) Fundamental analysis for visible-light communication system using LED lights. IEEE Trans Consum Electron 50(1):100–107
10. Gismalla M, Abdullah M (2019) Performance evaluation of optical attocells configuration in an indoor visible light communication. Indones J Electr Eng Comput Sci 14(2):668–676
11. Barry JR et al (1993) Simulation of multipath impulse response for indoor wireless optical channels. IEEE J Sel Areas Commun 11(3):367–379
12. Ghassemlooy Z, Popoola W, Rajbhandari S (2019) Optical wireless communications: system and channel modelling with MATLAB® 2019. CRC Press

Development of an IoT-Enabled Stroke Rehabilitation System

Huzein Fahmi Bin Hawari and Safuan Bin Abu

Abstract Electromyography (EMG) signal plays a crucial role in rehabilitation training of the lower limb. An-IoT-enabled stoke rehabilitation system was implemented in this study what was based on microprocessor, sensor, cloud and mobile application. Some of the main concerns that we can fix for stroke rehabilitation real time monitoring for long-term. Currently, there are a lot of focus on rehabilitation for upper limb muscle as this is more important as part of the stroke patient's survival skills to use their hand for eating, drinking etc. However, the lower limb is also important skill for stroke patient to improve their mobility. Thus, this research is to help explore lower limb stroke rehabilitation. The prototype was designed by combining muscle sensor and accelerometer for real time monitoring through mobile application and server. This prototype suitable apply and use in the house for home therapy. With this prototype, doctor can monitor their patient health status in real time monitoring. Besides, can connect patient health status for the right information.

Keywords Internet of Things · Stroke · Rehabilitation · Mobile application · Cloud

1 Introduction

One of the effective methods to bounce back from stroke effectively is by taking prevention steps such as identifying and understanding the signs of a stroke, although prompt care can save lives and significantly affect the chance of full recovery. Unfortunately, in stroke survivors the risk of a second stroke experiencing increase. About 25% of stroke survivors will suffer a second stroke, according to The National Stroke Association. The probability of a second stroke is around 40% higher within the first 5 years after the first stroke. Apparently, it is predicted that about 80% of all secondary strokes are preventable with improvements in lifestyle and medical care.

By 2019, Malaysia be the 113 in the world according to stroke patients. This type of disease is very dangerous and silent killer among Malaysia citizen. Due to slowly growth disease in human body this stroke disease almost unnoticeable until disease attack at the paralyzed stage. he goal of long-term recovery is to help insure that the victim of the stroke is as healthy as possible. Theoretically this is achieved in a way that keeps the patient comfortable and inspires them to learn the new simple skills like bathing, eating, sitting and walking.

Rehabilitation usually begins in hospital following a stroke. If your diagnosis is stable, recovery will begin during two days of the stroke and proceed after discharge from hospital. It can often take several weeks, months, or even years for anyone after a stroke recovery period is dissimilar. Few people completely recover while others have long-term or lifelong disabilities. Rehabilitation may include consultation with speech therapists, physiotherapists and occupational therapists. Speech therapy supports individuals who have trouble generating expression or hearing expression. Because of the stroke, physical therapy uses exercises to help you relearn balance and agility skills you may have lost. The aim of occupational therapy is to change everyday habits, such as feeding, drinking, dressing, bathing, reading and writing. A stroke rehabilitation system can be divided to upper limb muscle and lower limb muscle [1]. The top limb or upper limb is the limb of a vertebrate extending from the deltoid area to and including the neck, including the head, axilla and shoulder Although, in the general sense of the term, the human leg is the entire lower limb for the human body, including the foot, groin, and even the hip or gluteal region. Nonetheless, in human anatomy the term only applies to the part of the lower limb extended from the knee to the ankle, also known as the crus. Legs are used for balancing, and all types of locomotion, including leisure activities such as dancing, are an important part of the mass of a individual. Figure 1, show that the people using armband upper limb stroke rehabilitation system. In past several past years, stroke rehabilitation system widely have been invent to detect the movement of upper limb muscle. A stroke rehabilitation system is a device that can detect lower limb muscle expand and contract and in the same time, able to analyses the movement of lower limb muscle.

Basically a stroke rehabilitation system consists of the 2 major parts. The first part consists of the stroke rehabilitation equipment usually in a form of machine exercise. Secondly, a wearable device that can detect the movement of lower limb muscle during the stroke rehabs. In past several years, stroke rehabilitation system widely been invent to detect the movement of upper limb muscle. From this concept, it shows that there is a lot of room of possible application of stroke rehabilitation system that can apply on lower limb of the body. In this research, we will develop an stroke rehabilitation system that will be able to detect the movement of lower limb muscle and able to provide useful information through the analysis of the stroke rehabs to diagnose the movement of lower body muscle by using muscle sensor. By looking at several criteria from lower limb muscle movement, such as how much the muscle expand and contract.

Fig. 1 Upper limb stroke
rehabilitation system

The severity of stroke complications and each person's ability to recover vary widely. Researchers have found that people who participate in a focused stroke rehabilitation program perform better than most people who don't have stroke rehabilitation. For example, these skills can include coordinating leg movements in order to walk or carrying out the steps involved in any complex activity [2]. According to Joel Stein, MD, a physician under department of Rehabilitation Machine of Columbia University's College, repetitive exercising is the key to stroke rehabilitation. Since there is no commercial device that is currently available to monitor real time stroke rehabs at home, it is important to develop a system that a user is able to monitor their health performance and in the same time, its data being monitor by medical expertise to ensure the if abnormal sign happens in data.

In IoT, health information of the user can be stored in cloud and the data accessible by third party like doctor and or medical officer to review the data and analyze the condition of the user without having a appointment between them. This exchange information of the user and the doctor will improve the medical check-up quality and able to track the rehabs activity [3, 4]. With IoT, it is hope the application of stroke rehabilitation system rehabs diagnostic will help society to boost the quality of life and the same time, provide user as alternative device to monitor their health.

2 System Design and Architecture

2.1 Central Processing Unit

NodeMCU is the controller which have an open source system for IOT devices and application which make this controller special the device equipped with ESP8266 WiFi SoC make the device ready to be used for IOT loop from devices and to the cloud based system [5]. This controller is improving version from Arduino type of controller. The functioning and software for this controller are same with Arduino and the only different is WIFI part inside the controller that makes this controller ready to connected unlike Arduino need to buy Wifi Module and write a programmer to activate the Wifi by pairing the IP Address.

2.2 Myoware Muscle Sensor

Advancer Technologies muscle module monitors the behavior of muscles by quantifying the electrical potential produced by the muscle cells. This is called electromyography (EMG). The sensor magnifies and processes a muscle's dynamic electrical action and transforms it into a simple analog signal that can be interpreted efficiently by any microcontroller with an analog-to-digital converter (ADC), such as an A-Star or Arduino-or even a Maestro servo controller as we demonstrate [6]. When the target group of muscles is flexing, the output voltage of the sensor is steadily increasing. An onboard gain potentiometer can be used to fine-tune the exact relationship between the output voltage and the muscle operation. he MyoWare Muscle Sensor is an upgraded version of the older Muscle Sensor v3 of Advancer Technologies with these upgrades, notably single-supply operation (no need for negative voltage supply) and built-in snap connectors for electrodes. Certain additional features include a raw EMG performance, safety against reverse power, a control switch and LED indicators. To be attached to the skin, the module needs three electrode (not included) which snap into the snap-style connectors of the sensor, making it simple to attach and disconnect electrodes. Two connectors are located directly on the PCB, and at the end of the connected reference electrode cable is the third one.

2.3 Accelerometer Sensor

Accelerometers seem to be instruments that evaluate acceleration, which is the rate of change of an object's velocity [6]. It calculate in meters per second squared (m/s^2) or in G-forces (g). For us here on planet Earth, a constant G-force is equal to 9.8 m/s^2, although this ranges signficantly with the height (and may be a different magnitude on different planets due to gravitational pull variations). Accelerometers are useful

for vibration detection in devices or for guidance implementations. Accelerometers can be plotted on one, two, or three axes. 3-axis platforms are now more popular as the production costs for them significantly reduce. In fact, accelerometers have internal capacitive plates. Most of these are set while some are connected to tiny springs which shift internally as the forces of acceleration operate upon the sensor. When these plates shift in relation to each other, their capacitance varies among them. The acceleration can be calculated from certain capacitance changes.

2.4 Thing Speak Cloud

In information technology (IT) sectors, cloud computing is an efficient, scalable, and costless way to employ and deploy the organization's resources that include software, platforms, and infrastructures [7–9]. ThingSpeak allows scientists and engineers to test and build IoT systems without configuring servers or designing web applications. An IoT system is one of the main parts of an IoT program. ThingSpeak is one of that kind widget toolkit that provides a variety of different of features [10, 11]. At the heart of ThingSpeak is a portal that is used to store and process information from 'things'. ThingSpeak also offers additional cloud services, other APIs and social networks collaboration apps and provides the opportunity to develop the applications as plug-ins. It is a fantastic forum with vast options to explore how the Internet of Things is embedded.

2.5 System Block Diagram

The major process flow of whole project as per Fig. 2. Myoware muscle sensor sense the muscle movement and the accelerometer give 3 axis data of the motion and send the analog signal to the microcontroller which is NodeMCU powered by battery. The data that been read by sensor will be connected to the ESP8266 which a module Wi-Fi which sends the data to the mobile application and store inside the cloud storage. Within this research, the most significant and one of the latest technologies is a mobile application, based on extracting real-time information from medical instruments attached to calculate the cardio details of patients [12].

As in Fig. 3, the wiring diagram show the wiring part and the wireless part is being described. This project base an IoT technology show that the device portable and connected to the smartphone and cloud database and smartphone via WiFi technology.
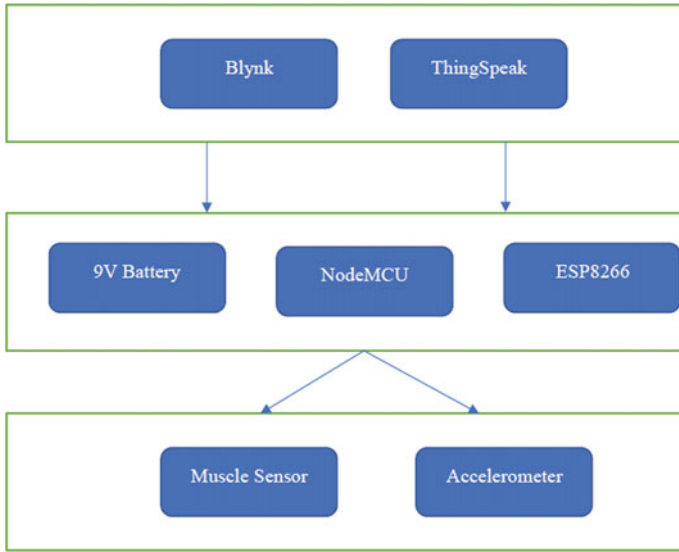
Fig. 2 Block diagram of the lower limb stroke rehabilitation system
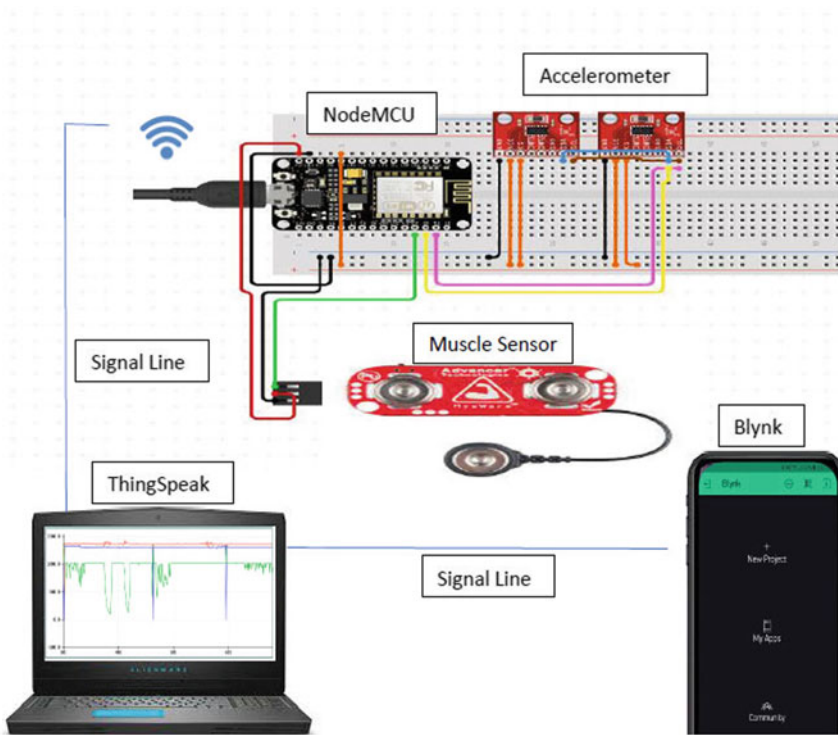
Fig. 3 Circuit diagram of the lower limb stroke rehabilitation system

2.6 System Prototype

As in Fig. 4, the description of the development of the lower limb stroke rehabilitation system is highlighted. The prototype of this project is attached with the e-box that contain microcontroller, muscle sensor, accelerometer and battery. The dimension of the system prototype will be $360 \times 290 \times 180$ mm. The user is required to place their leg on the rehabilitation platform where they are required to step down the pedal for a number of tries. The output of accelerometer and muscle sensor will be displayed real time in the Blynk mobile application and Thing Speak cloud.

3 Results and Discussion

The system prototype test is conducted at old folk's home at Rumah Seri Kenangan, Seri Iskandar Perak. The field testing is tested on random elderly at Rumah Seri Kenangan. The testing is conducted using the whole system communicate together with Thing Speak Cloud and main central system using wifi module. The data is then observed through the Thing Speak Cloud. The objective of this field testing is to study the effectiveness and the efficiency of the overall system.

Person X is a male, 42 years old person with related to asthma problem. Based on Fig. 5, the muscle sensor attached to the leg is able to provide a good response whenever the person tries to perform the steps using the system prototype where the result is posted real time on the Thing Speak cloud. For the accelerometers, it also good response on different axis, indicating good movement on the leg.

Figure 6 shows the Think Speak for Person Y. Person Y is a female, 46 years old person with related to diabetic problem and osteoporosis history. The result for the muscle sensor shows that the muscle movement is less compared to Person X. The first response is almost similar to Person X, but as the Person Y continue to perform more steps on the prototype, the movement is slightly stuttered possibly due to her health condition. As a result, the muscle sensor did not show good response for the remaining steps. Similarly, for the accelerometer the x axis indicate similarity for person X, but for Y and Z axis, the results are not different due to stuttered movement of the leg.

Person Z is an elderly male, 65 years old person with related to diabetic and gout problem. Person Z has problem with his leg and has problem to perform the steps movement on the prototype. Based on Fig. 7, the muscle sensor attached to the leg show that person Z try to perform a step. However, due to his leg problem, he was only to perform partial step and take about 10 min to perform it. From the Think Speak cloud, the accelerometer response also shows difficulty in the leg movement through indication of linear response for different x, y and z axis.
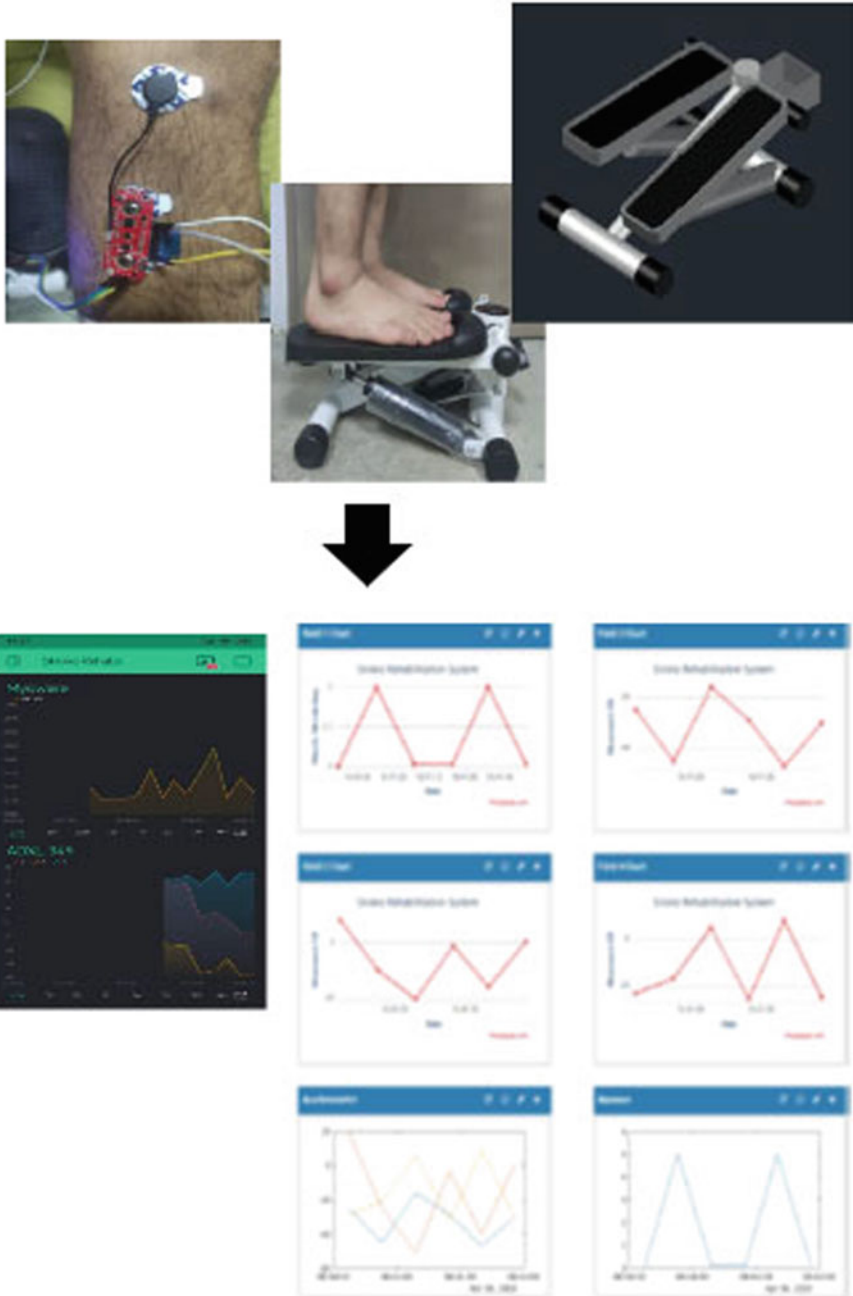
Fig. 4 Lower limb stroke system prototype
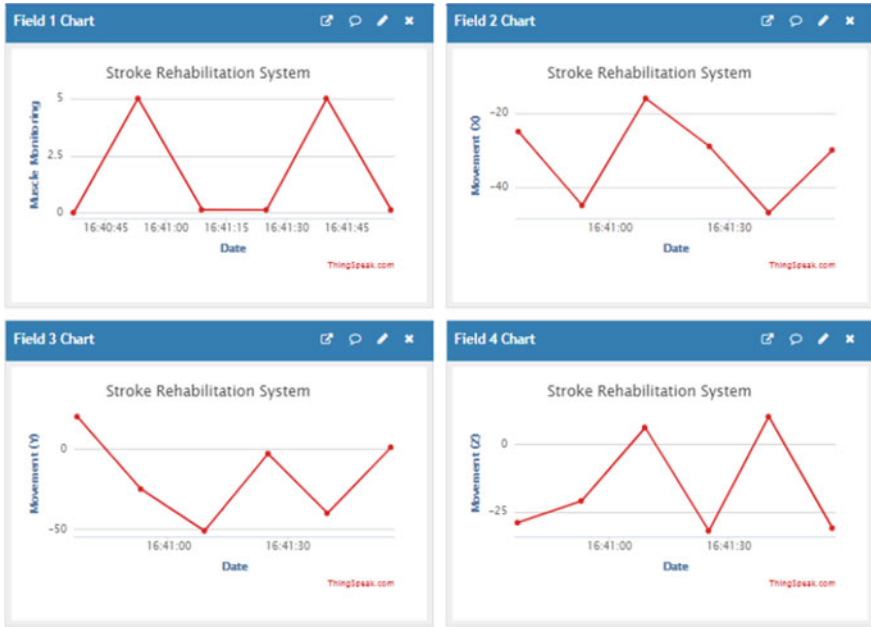
Fig. 5 Lower limb stroke system prototype think speak result for person 'X'
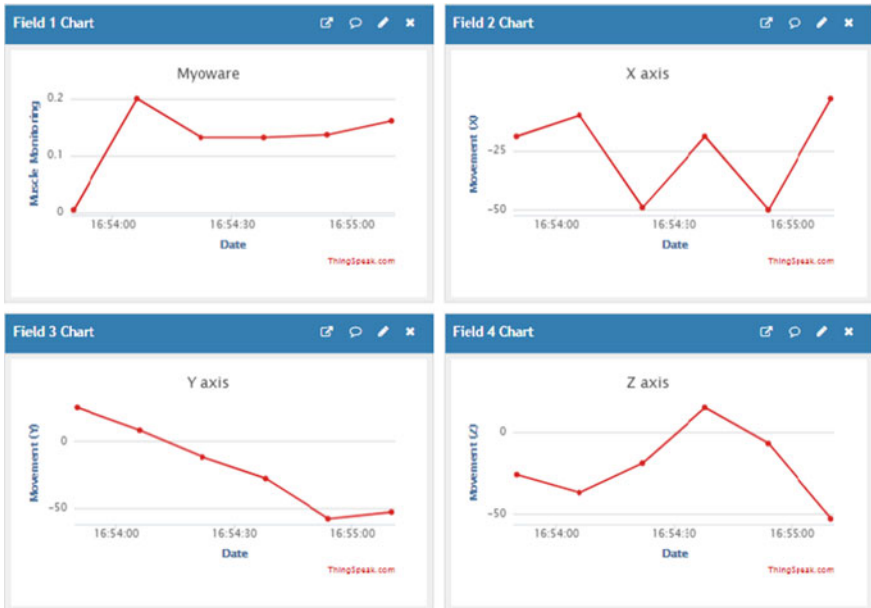
Fig. 6 Lower limb stroke system prototype think speak result for person 'Y'

Fig. 7 Lower limb stroke system prototype think speak result for person 'Z'

4 Conclusion

As a conclusion, this project is one of the important innovations needed in our rehabilitation facilities with real time monitoring technology which are very benefit for rehabilitation department to monitor citizen. The IoT based stroke rehabilitation system will gives important information about the patient's improvement for post stroke patient. Stroke patient are required to do constant rehabilitation in the hospital. In rural area there are low accessibility for rehabilitation facilities. Currently, the conventional method where it requires patient to visit the hospital. At the same time, the process is time consuming where it requires medical expert to process the data from human body and identify the illness or disease of the patient. Through the research, it will give benefits to the parents, caretakers doctors and therapist where they can monitor the post-stroke patient's online in real time. By combining the stroke rehabilitation equipment and muscle sensor with online database system will improve the efficiency in medical diagnostic where accurate data delivery which are useful for rehabilitation department to diagnose the movement of the muscle. The status of the project currently able to calibrate and measure the muscle performance with a low data error produced.

References

1. Muth CC (2016) Recovery after stroke. JAMA 316:2440
2. Islam SMR, Kwak D, Kabir MH, Hossain M, Kwak KS (2015) The Internet of things for health care: a comprehensive survey. IEEE Access 3:678–708
3. Ghosh AM, Halder D, Hossain SKA (2016) Remote health monitoring system through IoT. In: Paper presented at the 2016 5th international conference on informatics electronics and vision ICIEV 2016, pp 921–926
4. Chung Hua BD, Fahmi H, Yuhao L, Kiong CC, Harun A (2018) Internet of Things (IOT) monitoring system for elderly. In: International conference intelligent advance system ICIAS 2018, vol 201
5. Alankus G, Proffitt R, Kelleher C, Engsberg J (2010) Stroke therapy through motion-based games. In: A case study proceedings of the 12th international ACM SIGACCESS conference on computers and accessibility (2010), pp 219–226
6. Lu Z, Chen X, Zhao Z, Wang K (2011) A prototype of gesture-based interface. In: Proceedings conference human-computer interaction mobile devices services (DBLP), pp 33–36
7. Yadav N, Jin Y, Stevano LJ (2019) AR-IoMT mental health rehabilitation applications for smart cities. In: Smart cities: improving quality of life using ICT & IoT and AI (HONET-ICT) 2019 IEEE 16th international conference on, pp 166–170
8. Dohr A, Modre-Osprian R, Drobics M, Hayn D, Schreier G (2010) The Internet of Things for ambient assisted living. In: Seventh international conference on information technology, pp 804–809
9. Fan YJ, Yin YH, Da Xu L, Zeng Y, Wu F (2014) IoT-based smart rehabilitation system. IEEE Trans Ind Informat 10(2):1568–1577
10. Del Din S, Patel S, Cobelli C, Bonato P (2011) Estimating fugl-meyer clinical scores instroke survivors using wearable sensors. In: Proceeding annual international conference IEEE engineering med biolsoc EMBS
11. Hawari HF, Zainal AA, Ahmad MR (2019) Development of real time internet of things (IoT) based air quality monitoring system. Indonesian J Electr Eng CS 13(3):1039–1047
12. Zheng X, Chen W, Cui B (2011) Multi-gradient surface electromyography (SEMG) movement feature recognition based on wavelet packet analysis and support vector machine (SVM). In: Proceeding 5th International Conference IEEE Bioinformatics Biomedical Engineering, pp 1–4

MRI Brain Tumor Classification Technique Using Fuzzy C-Means Clustering and Artificial Neural Network

Angona Biswas and Md Saiful Islam

Abstract Brain tumor is a deathly disease and it is indispensable to point out the tumor immediately. Detection of brain tumor from MRI with higher accuracy has become a major research region for the medical sector. In this paper, an automatic brain tumor classification procedure applying fuzzy C-means and artificial neural network is proposed which provides higher precision. For this proposed technique, inputted MRI images are resized and then sharpening filter is used for preprocessing. After that, fuzzy C-means cluster process is chosen for image segmentation. At the next step, discreate wavelet transform is utilized for feature extraction and then features quantity are reduced by principal component analysis. Furthermore, reduced features are taken to artificial neural network for brain tumor classification. An effective training function Levenberg–Marquardt is used for neural network. This proposed method provides 99.8% accuracy, 100% sensitivity and 99.59% specificity which is comparatively better than other existing detection techniques.

Keywords Brain tumor · Fuzzy C-means · Discreate wavelet transform · Artificial neural network · Levenberg–Marquardt function

1 Introduction

Brain tumor is another disease which is playing a dominant role to cause death. ASCO published that 18,020 adults die from brain tumor in a year [1]. Cancerous brain tumors can snatch life within 5 years from the diagnosis time [1]. Brain tumors are mainly abnormal growth of cells that fill up the necessary brain space [2]. 120 types of brain tumors exist but they said all tumor types are not cancerous [2]. Glial cells are the point from where Glioma tumor arises. Pituitary tumor exists in pituitary gland [3]. A patient needs to detect the tumor type quickly to initiate treatment. It is already proved that Artificial Intelligence (AI) has enough skill to detect object more perfectly than manual process [4]. Again, without trained-experienced radiologists,

A. Biswas · M. S. Islam (✉)
Department of Electronics and Telecommunication Engineering, Chittagong University of Engineering and Technology, Chattogram, Bangladesh

MRI cannot be diagnosed perfectly [4]. Magnetic Resonance Imaging (MRI) is an efficient way for classification. In that theme, the proposed technique is based on artificial neural network (ANN) and image segmentation is done by fuzzy C-means (FCM) clustering.

Different research work was done previously to detect brain tumor types using AI and several clustering techniques were trialed. Authors N. Nandha, Dr. Karnan was found 98.87% rate for tumor detection by using FCM with Particle Swarm Optimization (PSO). They said their method provided a best result than Genetic Algorithm with FCM approach comparatively [5]. Authors Dipali M., Dr. N. Rana used histogram equalization for pre-processing and thresholding for segmentation. They used Neuro Fuzzy classifier and within eight tests they got three wrong results [6]. MRI brain tumor image segmentation analysis was done by author Heena Hooda, Om Prakash. They used K-means clustering and FCM. They claimed that FCM can detect the tumor area more perfectly than k-means [7]. FCM was used for image clustering, features were extracted by Grey level matrix (GLRLM) and the SVM was utilized for image categorization by Parveen and Amritpal Singh [8]. The authors found 91.66% accuracy by using Kernel Linear Function with 120 images [8]. Malathi and Sinthia [9] did an improved research work. They used Spatial FCM and Back Propagation Algorithm [9]. They detected area, defected cells, detection time of tumor. Two clustering algorithms were used combinedly by Rasel and Foisa [10] to reduce the error. Templated K-means and improved fuzzy C-means (TKFCM) method provided 97.1% accuracy.

This paper is proposed an effective classification technique using FCM and ANN. The novelty of the proposed method is to bring out 99.8% accuracy by utilizing proper preprocessing, clustering algorithm and effective ANN training function from an immense data. This paper arranges the remaining part as follows. Proposed framework and briefly explanation are presented in Sect. 2. Obtained results are explained using images, tables in Sect. 3. Finally, the proposed system is concluded in Sect. 4.

2 Methodology

The classification technique of proposed paper includes FCM and ANN to explore the higher accuracy. The total process of classification can be divided into four steps: pre-processing steps, image segmentation step, feature extraction and reduction and finally, classification step with ANN. The workflow of technique is exhibited in Fig. 1.

2.1 Pre-processing

This proposed technique applied resizing and sharpening filter for preprocessing:
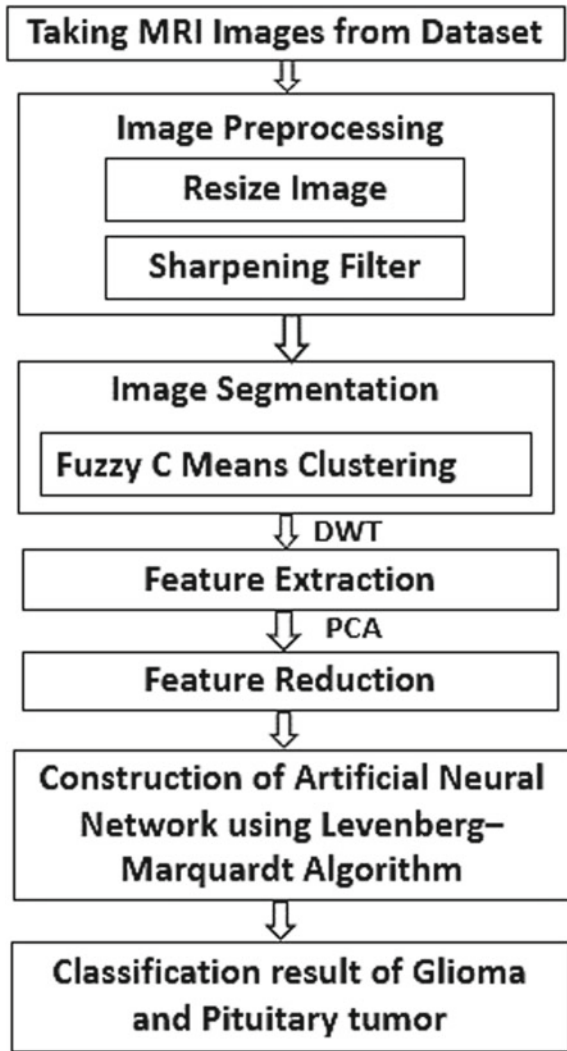
Image Resizing. Dataset consisted of different size of images. To work with all images at a time, input images were resized by 200 × 200 pixels which means to change the pixel information.

Sharpening Filter. Sharpening filter was used to strengthen the contrast shade of the bright and dark area of the input image. This also enhanced the object's edge. Sharpening filter provides better accuracy in case of image processing.

2.2 Fuzzy C-Means (FCM) Clustering

In 1981, Bezdek is the person who introduced FCM algorithm at first and FCM is an improved version than previous clustering algorithms [11] and FCM allows two or more clusters in one pixel [12]. FCM is unsupervised clustering process where data points and centers of cluster's distance create clusters [13]. It performs better than k-means (result section). FCM works based on optimizing Objective Function [11].

$$J(U, V) = \sum_{i=1}^{n} \sum_{j=1}^{c} (\mu_{ij})^m \|x_i - v_j\|^2 \tag{1}$$

where, Objective Function $= J(U, V)$, data points $= n$, fuzziness index $= m$, membership $= \mu_{ij}$, jth cluster center $= v_j$, No. of cluster center $= c$, Euclidean distance of ith particular to jth mid-point of cluster $= \|x_i - v_j\|$, k = iteration step, d_{ij} = Euclidean distance.

FCM Algorithm

Step 1: Cluster center 'c' is selected randomly
Step 2: Calculation of Fuzzy membership function 'μ_{ij}'

$$\mu_{ij} = 1 \left/ \sum_{k=1}^{c} (d_{ij} - d_{ik})^{\left(\frac{2}{m} - 1\right)} \right. \tag{2}$$

Step 3: Computation of fuzzy centers, v_j

$$v_j = \left(\sum_{i=1}^{n} (\mu_{ij})^m x_i \right) \left/ \left(\sum_{i=1}^{n} (\mu_{ij})^m \right), \quad v_j = 1, 2, \ldots, c \right. \tag{3}$$

Step 4: Repetition of step 2 and 3 before achieving the minimum Objective Function.

From taking input tumor MRI to segmentation procedure of proposed method is demonstrated in Fig. 2. Input MRI was first preprocessed and then clustered by FCM.

(a)   (b)   (c)   (d)   (e)   (f)
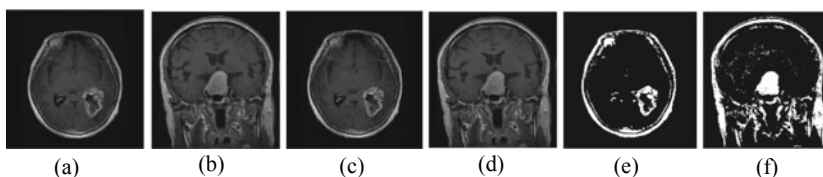
Fig. 2 **a** Glioma MRI. **b** Pituitary MRI. **c** Sharpening filtered image of Glioma MRI. **d** Sharpening filtered image of Pituitary MRI. **e** FCM clustered image. **f** FCM clustered image

Here, Fig. 2a is Glioma and Fig. 2b is Pituitary MRI images. Contrast was strengthened by sharpening filter, shown in Fig. 2c and d. Then in Fig. 2e and f, fuzzy C-means clustering was applied after filtering to segment tumor from the image background properly.

2.3 Feature Extraction and Reduction

Two-dimensional DWT was used in this proposed method and that was three-level decomposition wavelet. Discrete Wavelet Transform is employed to extract multiple features from the filtered image. Fourier transformation has a serious disadvantage as it doesn't contain any information about time domain [14]. LL, LH, HH, and HL are the four sub-bands for each scale [15]. Principal Component Analysis (PCA), an efficient way of feature reduction was used to reduce the quantity of features after feature extraction to reduce computational time and memory. Extracted multiple features were: Contrast, Energy, Kurtosis, Mean, Entropy, Correlation, RMS, Variance, Homogeneity, Smoothness, Standard Deviation, Skewness, IDM.

2.4 Proposed Artificial Neural Network Structure for Classification

Input of this ANN was collection of 13 features, hidden neurons were 10 in hidden layer and output layer had 2 neurons to indicate glioma and pituitary classes of brain tumor. This neural network was constructed with the fastest training function 'Levenberg–Marquardt'. ANN is dexterous tool for classifying brain tumors and makes flawless decisions. ANN took 70% data for training firstly, 15% sample was used for validation and then 15% was used for prediction in the testing stage.

3 Experimental Result and Analysis

3.1 Dataset

512×512 pixels dimension, T1-weighted data was obtained from [16]. Dataset consisted of total 489 MRI images where 246 glioma and 243 brain tumors. Axial and coronal plane images are used for this experiment and dataset was unbiased.
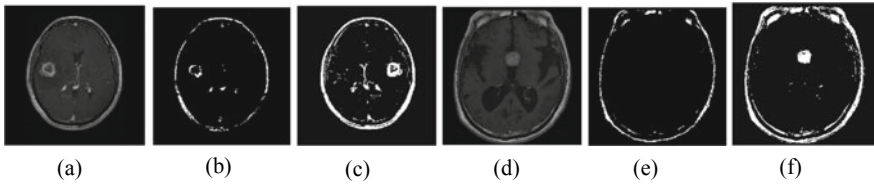
Fig. 3 **a** Glioma MRI. **b** K-means clustered of Glioma. **c** FCM clustered image of Glioma. **d** Pituitary MRI. **e** K-means clustered of Pituitary. **f** FCM clustered image of Pituitary

3.2 Selection of Perfect Clustering Algorithm

Figure 3a and d is sequentially Glioma and Pituitary MRI. Sequentially, Fig. 3b is the k-means segmented image of Fig. 3a and e is the K-means clustered image of Fig. 3d. Then, Fig. 3c and f are representing the FCM clustered image of Fig. 3a and d respectively. The result of Fig. 3 makes clear that k-means algorithm was unable to find out tumor location perfectly. But FCM had the ability to locate perfectly both of the Glioma and Pituitary tumors which is shown in Fig. 3c and f. In case of k-means clustering Fig. 3e, it was totally unable to indicate the Pituitary tumor. So, FCM was selected for methodology and features values were taken for neural network training.

3.3 Classification Accuracy of Proposed Method

The best result confusion matrix is shown in Fig. 4 where 99.8% accuracy was found for 18 iterations. Figure 4 proves that the proposed methodology can classify glioma and pituitary brain tumors flawlessly.

Fig. 4 Obtained all confusion matrix of best result

Table 1 Comparison of proposed technique with appurtenant work

Reference	Classification technique	Classification accuracy (%)
[4]	Region of interest (ROI) + DWT + gabor filter + BPNN	91.9
[8]	Image enhancement + FCM + SVM classifier	91.66
[9]	K-means and SFCM (KISFCM) + morphological filter + BPNN	93.28
[10]	Templated- K-means + improved FCM (TKFCM) + ANN	97.1
[15]	GoogLeNet (learn features) + deep CNN-SVM	98
Proposed method	Resizing + sharpening filter + FCM + 2D DWT + PCA + ANN	99.8

$$Accuracy = (TN + TP)/(TP + TN + FN + FP)100 \qquad (4)$$

$$Sensitivity = TP/(TP + FN)100 \qquad (5)$$

$$Specificity = TN/(TN + FP)100 \qquad (6)$$

where, TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative. Accuracy = 99.8%, Sensitivity = 100%, Specificity = 99.59% are elevated from the confusion matrix of Fig. 4 and Eqs. (4)–(6). Network Performance was found 0.0033 from target data and output data.

3.4 Comparison of Proposed Technique

In Table 1, different applied methods of authors and obtained result of proposed approach are compared. Table 1 makes it clear that the introduced approachment is tremendously successful. The proposed method is simple and obtain the highest accuracy of 99.8%. Reasons behind this higher precision are choosing FCM based segmentation and using 'Levenberg–Marquardt' training function.

4 Conclusion

A unique and flawless technique is proposed based on FCM clustering and ANN classifier to classify glioma tumors and pituitary tumors using MRI. This method the removes previous limitation of low accuracy by using 489 patient's data. MRI images

were pre-processed by sharpening filter to enhance the contrast. Then better clustering algorithm 'fuzzy C-means (FCM)' was selected for segmentation. DWT was utilized to extract features from clustered images and for feature reduction, PCA was applied. 13 features were collected from images and then used for artificial neural network training and testing. The fastest ANN function 'Levenberg–Marquardt' was used for proposed network training. Test result provided 99.8% accuracy at 18 iterations, 100% sensitivity, 99.59% specificity. This proposed method is able to provide 99.8% accuracy which is much better than some previous classification techniques. This result is enough to claim that proposed technique can classify glioma and pituitary tumor types magnificently. This improved accuracy is beneficial for medical sector to classify automatically.

References

1. Brain Tumor (2020) American Society of Clinical Oncology (ASCO) Publication
2. Brain Cancer Types (2020) Cancer Treatment Centers of America
3. Gliomas (2020) Johns Hopkins Medicine
4. Mustafa R, Ikhlas A (2018) Brain tumor classification via statistical features and back-propagation neural network. IEEE Xplore
5. Gopal N, Karnan S (2011) Diagnose brain tumor through MRI using image processing clustering such as FCM along with intelligent optimization techniques. In: International conference on intelligence and computing research. IEEE Xplore, pp 112–116
6. Dipali J, Rana S, Misra T (2010) Classification of brain cancer using ANN. In: 2nd international conference on electronic technology (ICECT 2010). IEEE, pp 112–116
7. Heena Hooda F, Om Verma S, Tripti T (2014) Brain tumor segmentation: a performance analysis using K-Means, FCM and region growing algorithm. In: International conference on communication and computing technologies (ICACCCT). IEEE, pp 1621–1626
8. Parveen F, Amritpal S (2015) Detection of brain tumor in MRI images, using combination of FCM and SVM. In: 2nd international conference on signal processing and integrated networks (SPIN). IEEE, pp 98–102
9. Malathi MF, Sinthia P (2020) MRI brain tumor segmentation using hybrid clustering and classification by BPNN. Asian Pac J Cancer Prev 19
10. Rasel A, Md Foisal H (2016) Tumor detection in brain MRI image using template based K-means and fuzzy C-means clustering algorithm. In: 2016 international conference on computer communication and informatics (ICCCI-2016). IEEE
11. Preetha R, Suresh FG (2014) Performance analysis of FCM in automated detection of brain tumor. In: World congress on computing and technologies. IEEE, pp 30–33
12. Shasidhar MF, Raja SSV, Kumar BVT (2011) MRI brain image segmentation using modified fuzzy C-means clustering algorithm. In: 2011 international conference on communication systems and network technologies. IEEE, pp 473–478
13. Brundha BF, Nagendra MS (2015) MR image segmentation of brain to detect brain tumor and its area calculation using K-means clustering and fuzzy C-means algorithm. Int J Technol Res Eng 2(9)
14. Zhang Y, Wu LS (2012) An MR brain images classifier via principal component analysis and SVM. Prog Electromagnet Res 130:369–388
15. Deepak S, Ameer PM (2019) Brain tumor classification using deep CNN features via transfer learning. Comput Biol Med 111
16. Brain tumor dataset into figshare (2017) https://figshare.com/articles/brain_tumor_dataset/1512427

An IoT Based Automatic Vehicle Accident Detection and Rescue System

K. Gayathri Devi ⃝ **, K. Yasoda, B. Rajesh, R. Sowmiya, and S. S. Vishalidevi**

Abstract A large variety of precious lives are lost because of road traffic accidents each day that may occur due to the mistake of driver and delayed response from the place of accident to the emergency services. The most important aspect of driver is to own a good road accident detection and data communication system in place to save injured persons. A system that sends information messages to close emergency services regarding the accident location for timely response is completely essential. The accident detection process uses raspberry pi, in detection of the data from the vibration sensor and accelerometer during accident. The process of rescue system uses GSM, in the process of sending message to the neighboring medical Centre or relatives. GPS is used to track the location of the vehicle and used in the calculation of the speed of the vehicle based on the position of (latitude, longitude) and time difference. GSM in the system is used to send the location of accident through message. Vehicle ad-hoc network (Wi-Fi) is used to transfer of the data from the controlling unit to the application server in the later retrieval of data. Hyper Text Transfer Protocol is used to transfer web page from remote server after internet connection is established. The FireBase database is used in the storage of collected data, and uploads over in the web-page and in the process informing the accident location and in alerting.

Keywords Accident detection · Human rescue system · Global positioning system (GPS) · Global system for mobile (GSM) · Vehicle ad-hoc network

K. Gayathri Devi (✉) · B. Rajesh · R. Sowmiya · S. S. Vishalidevi
Dr. N.G.P Institute of Technology, Coimbatore, Tamilnadu, India
e-mail: gayathridevi@drngpit.ac.in

K. Yasoda
SNS College of Technology, Coimbatore, Tamilnadu, India

1 Introduction

Nearly 1.25 million individuals die in road crashes annually, on average 3287 deaths on a daily basis. An additional 20–50 million are injured or disabled. Over 1/2 all road traffic deaths occur among young adults ages 15–44. There are 1012 deaths in 2019. This can be 74 deaths (+7.9%) over an equivalent amount in 2018. Within the 12 months complete October 2019 there have been 1209 road deaths. Throughout calendar 2018 there have been 1135 road deaths, 7% not up to in 2017. A recent World Health Organization (WHO) report showed that each year 1.35 million people die and 50 million place get injured. Road accidents are ranked because the eighth leading reason behinds deaths (up from ninth its previous report in 2015), with the Association for Safe International Road Travel (ASIRT) predicting that it's going to rise to the fifth leading reason behind deaths within close to future, unless forceful changes occur. Moreover, because the social damage caused by road traffic accidents, there is a big price. ASIRT estimates that between one and two percent of the annual budget of each country is spent on road accidents. Now-a-days, it has become terribly troublesome to understand that an accident has occurred and to find the position wherever it's happened. There is no system of identification and intimation relating to an accident in the past. For intimation purpose messaging services is used. GPS makes use of identification of location and GSM is built for the usage of intimation. The most intention of an accident detection is to locate the accident spot anywhere and intimating the emergency through the GPS and GSM networks. The GPS based vehicle accident identification module contains Mostly of different sensors such as accelerometer, MEMS etc., GSM module and a GPS modem connected to the microcontroller. Global System for Mobiles (GSM) technology is employed for a proper cellular mobile communication. Position of the vehicle accident location is traced down by GPS.

2 Literature Survey

The self-collision detection sensor is used to detect the accident and the information is uploaded to cloud based database server collision vehicle accident recognition, and a connected emergency notification is provided [1]. The detection of accidents supported by monitored speed and send the accident location acquired from the GPS together with the time by GSM and therefore the speed by utilizing the GPS network [2], sent to a OpenGTS server, providing a time period OpenStreetMap visualization of traffic scenario. MongoDB appraise the close mobile APPs with alert messages [3]. Auto continuous collision accident collision mechanism, exploitation physical phenomenon preview iterative algorithm, trace examination reconstruction algorithm, serial collision contact position reconstruction localization algorithm [4], self-tuning iterative hard thresholding (ST-IHT) algorithm for learning distributed spatio-temporal features and a weighted extreme learning machine (W-ELM) [5], Sudden

Lane Departure Avoidance Module and Rear-End Collision Avoidance Module [6] is used over in the process of detection of accident. The accident is detected by the in-Vehicle sensor and updates in database, verifies and compares the severity of accident supported to different accident data [7] and notifies the emergency services, uses remote cut off mechanism [8] and using GPS and GSM intimates the emergency services [9, 10] and provides over an crash path reconstruction using Kalman filter and GPS [11]. Data continuously collected from the smartphones accelerometer and analyzed using Dynamic Time Warping (DTW) [12], inputs from the accelerometer, magnetometer, gyroscope and on vehicle crash detection [13] are analyzed and intimated as e-Call with owner's information. The local prediction is exchanged from each sensor with neighboring vehicles aggregates the native predictions it receives using a weighted majority aggregation rule to get a final prediction [14], crash signals at different locations of a vehicle and their implications on crash analyzes the severity of accident [15], the system also sends a warning message. The system uses CAD-CVIS consists of varied sorts of accident types, atmospheric conditions and accident location, a deep neural network model YOLO-CA supported CAD-CVIS and deep learning algorithms to detect accidents and in alerting [16]. The system detects and assesses the severity of accidents with the assistance of an on-board unit and the TestBed analyses different accident scenarios and helps in rejecting false alert and serially detects the rollover and force [17]. Besides over these many systems, they are costly and not applicable in every cars and lacking over in the process of saving location for later use.

3   Proposed Model

This works focuses on designing and developing an accident detection using vibration sensors and wireless modules. The block diagram of the proposed system is shown in Fig. 1. In order to enhance real time accident detection and monitoring, the system has Vibration sensor SW-420, GPS and GSM module, which is connected to Raspberry pi controller, which works with Wi-Fi in transfer of data. The Vibration sensor SW-420 produces an digital output, on the detection of any collision or accident. GPS is used to track over the location in which the vehicle travels. GSM is used to send the location of accident as an URL link with the latitude and longitudinal information on the place of accident. Raspberry pi controlling unit is used to sense the information and to perform the entire controlling functions with database through Wi-Fi. For the enhancement of real time accident monitoring, the system with sensors which works on IoT basis [18, 19]. By using IoT, status on speed and the location of accident is recorded and the database manages over the data and uploads over in the webpage, in tracking of vehicle and its speed.

The connection and the flow diagram of the Vibration Sensor SW420 with Raspberry pi is given in Figs. 2 and 3. The accident detection system operates with the vibration sensor producing digital pulse output on the detection of any accident or collision. It produces over an output based on the threshold which is been set over
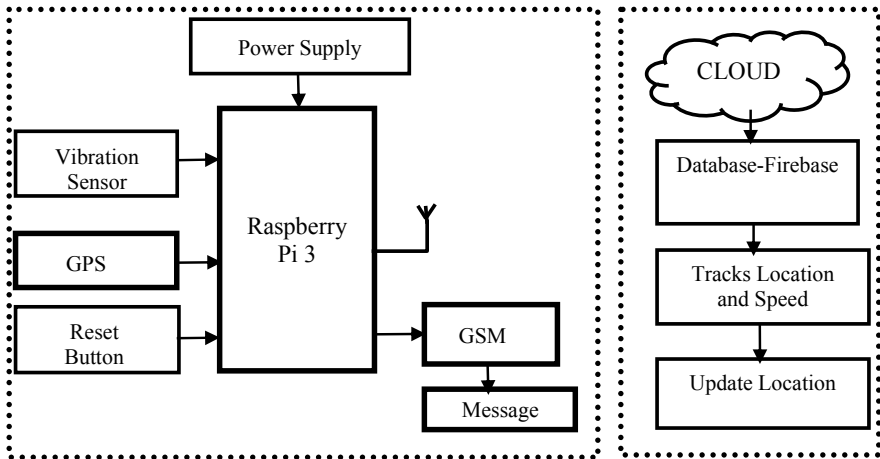
Fig. 1 Block diagram of proposed system
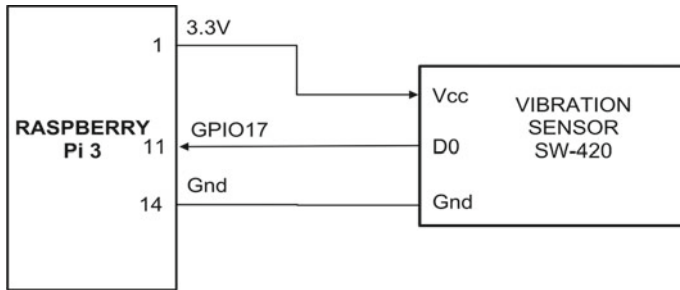
Fig. 2 Connection of the vibration sensor SW420 with Raspberry pi

Fig. 3 Flow diagram of the vibration sensor SW420 with Raspberry pi
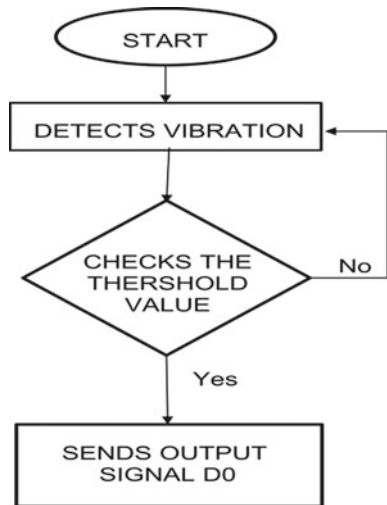
in the potentiometer. The sensor is tightly fitted over in any part of the car. The variations in the output of the sensor are trapped and those are fed as input to the micro-controller.

GPS is commonly used for both tracking and navigation in vehicles. Tracking systems helps the base station to keep track of the vehicles in the way of travel without the intervention of the driver whereas, the system also helps the driver to reach the destination by navigation. The architecture of navigation or tracking system is more or less similar. The connection and the flow diagram of the GPS with Raspberry pi is given in Fig. 4. During the occurrence of accident in any place, tracking on the position of the vehicle by GPS system and send the information by alerting person through call or SMS by using GSM.

The connection and the flow diagram of the GSM with Raspberry pi is given in Fig. 5. Now based on these variations, the micro-controller is programmed to take decision which in turn.

- Sent as an message an through the GSM to relative or nearby emergency service.
- Data is updated over in the database through wireless transmission.
- Webpage provides over the alert message and update the location in the map.
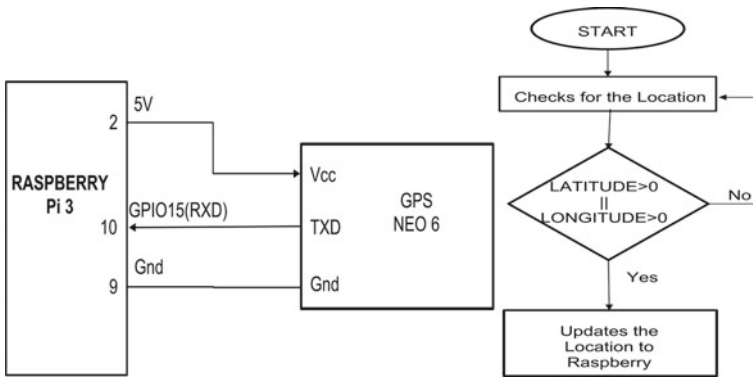
Fig. 4 Connection and flow diagram of GPS with Raspberry pi

Fig. 5 Connection diagram of GSM with Raspberry pi
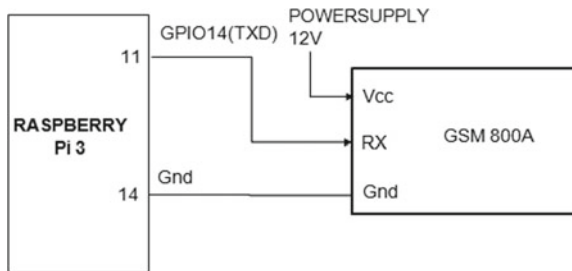
3.1 Wireless Transmission in IoT

Wireless communication system acts as a bridge for duplex communication for data collection and control message delivery and is an essential part for IoT infrastructure. Various IoT applications can be applied using Wireless communication system, which includes health care monitoring and home automation etc. and gains a lot of attention in connecting devices to internet.

3.2 Device-To-Cloud Communication

Internet cloud services like an application service provider is used to exchange data and manage the traffic of messages that will be connected directly to IoT devices in a device-to-cloud communication model. There are numerous advantages when a connection is established between the device and the IP network through cloud service when compared with the existing communications mechanisms like Ethernet or Wi-Fi connections.

The Device-to-cloud communication model that is proposed in our work is shown in Fig. 6 and maintained by more consumer IoT devices like the nest labs learning devices and many SmartTv and Smart phones. The Nest Learning Thermostat uses the devices to broadcast data to a cloud database where the data can be used to analyzed and problems can be identified. Cloud connection may help or enable user to obtain remote accessing of data through smartphones or web interface and supporting in some software updates to device.

The SmartTV technology make use of the Internet connectivity for the process of transmitting the information viewed by the user to do further analysis by the service provider and this enables the voice recognition features. The user can expand capability of device further than its original features to device-cloud module with respect to certain cases. Interoperability problem exist when there is an amalgamation of device detail from many diverse manufacturer it can be avoided if the authorized
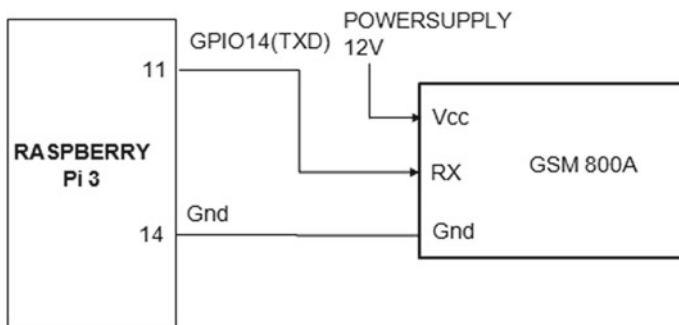
Fig. 6 Device-to-cloud communication model diagram

data protocols are used between the device and the cloud services. The device owner or user can be coupled to a precise cloud service in order to prevent them the usage of alternative service providers. This is referred to as "vendor lock-in", a term the compasses other facts of the relationship with the provider such as ownership of and access to the data. In mean time, users can have confidence that devices can be integrated based on the design for the specific platform.

IoT has made the change in the evolution today's world by connecting things over the Internet, which makes very much easier remote access. IoT is the emerging and trending technology that can be implemented over in the safety and security monitoring systems and in rescuing.

4 Results and Discussion

The Raspberry pi controls the functionality of the above mentioned sensor and modules. Programming for the controller is done using Python coding. After completion of the entire program, debugging is carried out. Finally, the program is dumped to the controller. Then, the working mechanism is verified. Figure 7 gives the snapshot of the prototype developed with major components. IoT plays a major role in collection of database, from the user. Each user can reach the server using their own e-mail id. This is achieved by means of mobile-to-cloud communication. When the user access the server, any variation in sensor detection is updated in form of database with the date and time. Therefore, real time accident detection is achieved by the implementation of IoT.

Fig. 7 Hardware implementation

Accident Occured

Location : https://www.google
.co.in/maps/place/11.0002N,76
.0094E

Accident Occured

Location : https://www.google
.co.in/maps/place/11.0002N,76
.0094E

Figure 8 gives the snapshots of the location of the accident that were transmitted through GSM. In this prototype firebase help in cloud based data storage and helps in easy retrieval of data. Firebase provides developers in the platform of mobile and web app development and which provides a plenty of tools and services which help in developing high-quality apps, grow their user base, and earn more profit without any back end server. Figure 9 gives the snapshot of the updation of location and speed in the database along with the information on accident occurrence through wi-fi using HTTP protocol. Webpage provides over the view in which location the vehicle travels,

Database   🗄 Realtime Database ▾

Data   Rules   Backups   Usage

🔗   https://silent-star-233308.firebaseio.com/Accident_Project

silent-star-233308 > Accident_Project

Accident_Project
⊟─ **RealTimeData**
      ├─ **Accident:** "Detected
      ├─ **Lat:** "11.0271
      ├─ **Lng:** "76.9413
      └─ **Speed:** 0
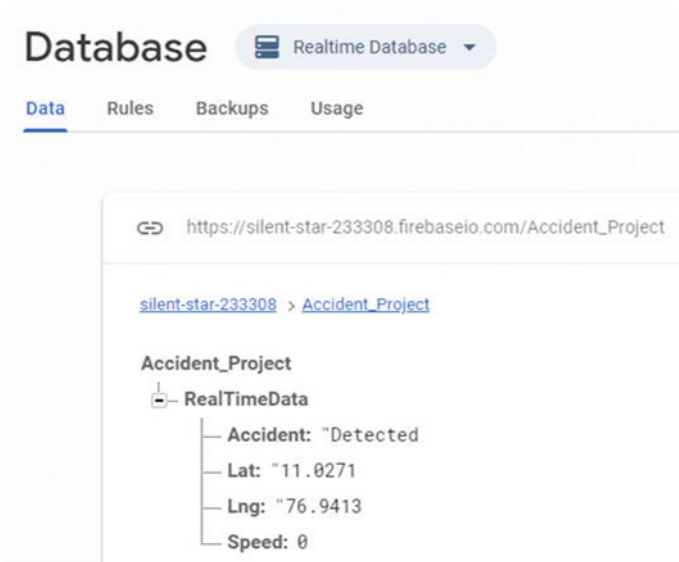
Fig. 9 Updation of accident occurrence in the database
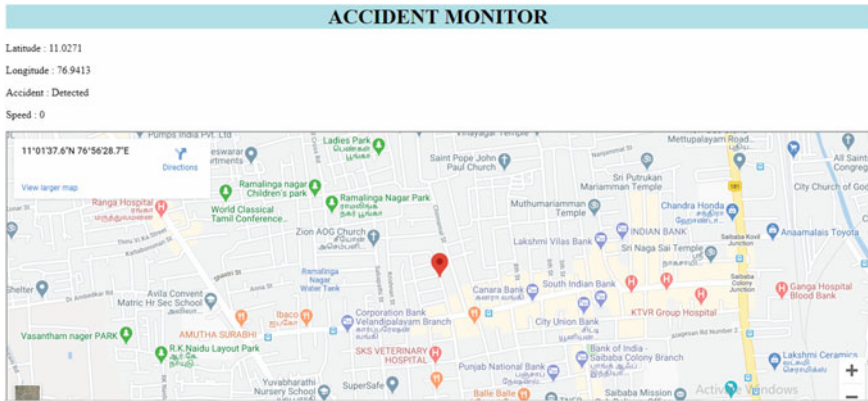
Fig. 10 Accident detection message in webpage

Table 1 Result comparison between existing and proposed system

Parameter	Existing system [20]	Proposed system
Accident detector	Li-Fi system between cars	Vibration sensor in every car
System adaption	Hard	Easy
Alert system	Implemented using connected cars	Implemented using IoT
Data collection	Not implemented	Implemented and stores the data over in database

and provides over a pop-up information on the accident occurrence. Figure 10 gives the snapshot of the webpage during accident occurrence which gathers information from the database. The information based on the location helps in easy retrieval of victim people.

The Comparison between the proposed and existing system is given in Table 1. The existing system was designed with need to transmit the accident location using IoT technology to cater the needs of people who are undergoing accident.

5   Conclusion

This proposed system with advanced features reduces manual work and allows detecting the accident and reporting to the emergency services and updating data in database. The IoT technology used in this system helps in treating the victim people instantly in case of any emergencies and the route to the location of accident.

References

1. Chang WJ, Chen LB, Su KY (2019) DeepCrash: a deep learning-based internet of vehicles system for head-on and single-vehicle accident detection with emergency notification. IEEE Access 7:148163–148175
2. Amin MS, Jalil J, Reaz MBI (2012) Accident detection and reporting system using GPS, GPRS and GSM technology. In: IEEE international conference on informatics, electronics & vision (ICIEV), pp 640–643
3. Celesti A, Galletta A, Carnevale L, Fazio M, Lay-Ekuakille A, Villari M (2017) An IoT cloud system for traffic monitoring and vehicular accidents prevention based on mobile sensor data processing. IEEE Sens J 18(12):4795–4802
4. Lang W, Biao G, Tao C (2013) Vehicle continuous collision accident reconstruction system development. Procedia-Soc Behav Sci 96:1659–1669
5. Yu Y, Xu M, Gu J (2019) Vision-based traffic accident detection using sparse spatio-temporal features and weighted extreme learning machine. IET Intel Transp Syst 13(9):1417–1428
6. Chang TH, Hsu CS, Wang C, Yang LK (2008) Onboard measurement and warning module for irregular vehicle behavior. IEEE Trans Intell Transp Syst 9(3):501–513
7. Fogue M, Garrido P, Martinez FJ, Cano JC, Calafate CT, Manzoni P (2013) A system for automatic notification and severity estimation of automotive accidents. IEEE Trans Mob Comput 13(5):948–963
8. Srabanti S, Asaduzzaman M, Mokter MKB, Anannya TT, Tumpa SN, Afroze L, Karim MM (2018) A proposed system for automatic vehicle monitoring and accident detection in Bangladesh. In: 2018 international conference on computer, communication, chemical, material and electronic engineering (IC4ME2). IEEE, pp 1–5
9. Kumar A, Jaganivasan V, Sathish T, Mohanram S (2018) Accident detection and alerting system using GPS & GSM. Int J Pure Appl Math 119(15):885–891
10. Taj FW, Masum AKM, Reza ST, Chy MKA, Mahbub I (2018) Automatic accident detection and human rescue system: assistance through communication technologies. In: 2018 international conference on innovations in science, engineering and technology (ICISET). IEEE, pp 496–500
11. Aldimirov M, Arnaudov R (2018) Method for automated reconstruction of a car's path during crash from GPS/INS data using a Kalman filter. Adv Eng Softw 115:386–390
12. Aloul F, Zualkernan I, Abu-Salma R, Al-Ali H, Al-Merri M (2015) iBump: smartphone application to detect car accidents. Comput Electr Eng 43:66–75
13. Fernandes B, Alam M, Gomes V, Ferreria J, Oliveira A (2016) Automatic accident detection with multi-modal alert system implementation for ITS. Veh Commun 3:1–11
14. Canzian L, Demiryurek U, van der Schaar M (2015) Collision detection by networked sensors. IEEE Trans Signal Inf Proc Over Networks 2(1):1–15
15. Chan CY (2002) On the detection of vehicular crashes-system characteristics and architecture. IEEE Trans Veh Technol 51(1):180–193
16. Tian D, Zhang C, Duan X, Wang X (2019) An automatic car accident detection method based on cooperative vehicle infrastructure systems. IEEE Access 7:127453–127463
17. Khaliq KA, Raza SM, Chughtai O, Qayyum A, Pannek J (2018) Experimental validation of an accident detection and management application in vehicular environment. Comput Electr Eng 71:137–150
18. Kokila J, Gayathri Devi K, Dhivya M, Haritha Jose C (2017) Design and implementation of IoT based waste management system. Middle-East J Sci Res 25(5):995–1000
19. Hou Z, Chen Y (2017) A real time vehicle collision detecting and reporting system based on internet of things technology. In: 2017 3rd IEEE international conference on computer and communications (ICCC). IEEE, pp 1135–1139
20. Krishnan P (2018) Design of collision detection system for smart car using Li-Fi and ultrasonic sensor. IEEE Trans Veh Technol 67(12):11420–11426

Automatic Firefighting System Using Unmanned Aerial Vehicle

K. Gayathri Devi ⬤, K. Yasoda, and Maria Nithin Roy

Abstract In the world of technological advancements, Internet of Things had opened a vast opportunities and gained attention in academics and in industrial automation. This paper focuses on the general representation of the design of the Unmanned Aerial Vehicle with fire extinguishing balls that can be applied for fire-fighting situation. The paper highlights the design, and the purview of the experiment conducted with respect to fire extinguishing balls that is attached to the drone. The observation of the experiments clearly signifies that the size of the fire extinguishing balls connected to Unmanned Aerial Vehicle (UAV) system should be lesser to put off the fire outbursts. The proposed work has implemented an drone of required specification to extinguish the fire using fire balls attachable to drones. The hardware consists of Navio 2 and Raspberry pi 3 as flight controller interfaced with drone (UAV). The system design is a combination of internet of things and dedicated server for continuous storage and monitoring system using Amazon web services.

Keywords Amazon web services (AWS) · Drones · Fire extinguishing balls · Remote sensing · Unmanned aerial vehicle (UAV) · Wildfires · Navio · Raspberry pi

1 Background

The implementation of the systems with drone and various remote sensing technologies is rapidly increasing to provide a mobile, and low-cost alternative for the previous systems such as ground-based systems, air controlled vehicles, and satellite-systems [1]. The images captured by the satellites with low resolution and noise are not adequate to provide effective result. It is also proves ineffective in populated areas

K. Gayathri Devi (✉) · M. N. Roy
Dr. N.G.P Institute of Technology, Coimbatore, Tamilnadu, India
e-mail: gayathridevi@drngpit.ac.in

K. Yasoda
SNS College of Technology, Coimbatore, Tamilnadu, India

[2, 3] since there exist problem to classify the region of interest and there may be misclassification. The constrained surveillance ranges affect measurement of ground equipment [4]. The early stages of fire detection UAV system with image capturing is now replaced by infrared cameras for detecting the radiation emitted by fires [5]. A direct view of the radiation is required by the infrared sensors for effective results [3]. Visual cameras, a small size high resolution camera have been used for detection of smoke produced by fire in daylight [6]. Visual cameras provide the details such as flame height, angle, location and width. These cameras produce different contrast, texture, and motion analysis [3]. Another mentioned is the usage of LIDAR devices to identify smoke particles [7].

Holman et al. [4] proposed that UAV system can be used to capture images from a camera fixed to systems in a real time environment. It was concluded that image processing quality of topographic images which are in existence and the meta data of the specific image is rectified and stored for later usage purposes. The paper discusses the quality of resolution of images and how they are processed frame by frame. He constructed a similar exiting prototype of unmanned aerial vehicle (UAV) with a quadcopter setup to collect test images for specific enhancements of low cost and flexibility. The autonomous functions in the UAV is capable of excellent ground station. The accuracy of images from ground position could be found with accuracy.

Kumar et al. [8] had developed a optimization model for the detection of fire and to extinguish the fire with fire suppressing fluid for a region where there are no habitant with multiple UAV. Two step approach is considered, the first step represents the fire by a two dimensional shape with a boundary, it requires all the UAV to work in coordination within the boundary and there should be no collision. The second step proposes a model as a function of fire and time. This approach aims at reducing the Euclidean distance between the quadcopter and the fire at any instant of time. There are few studies that considered fire suppressants other than water. They tested this theory with master model with use of 10 drones via simulation. The suppressant fluid is not practical. The undergraduate project designed as a quadcopter with a ball-dropping mechanism that can drop a single unit of fire-extinguishing ball by using compressed springs.

Wallar et al. [9] suggested a approach for the surveillance of risky areas by using three quadcopters to evaluate the coverage area of the sensors on board, the quality of the data collected with reference to various altitudes and risk involved with ground level. The two stage approach is implemented in which the movement of the quad copter to maximize the coverage area is considered in first stage and the second age optimizes the risk involved. It was intended for the examination of regions to diminish the detection risk and to ensure the data provided by the sensor is of high quality. This approach is more costly as more number of quad copters are used.

Mois et al. [10] has focused on the data transmission, security and success rate of huge amount of environmental parameter data with three the Internet of Things (IoT) based protocol, the cloud computations, and cyber-physical systems. Three types of IoT-based wireless sensors for monitoring the environment and the surroundings were analyzed in this paper. The web based communication established with User Datagram Protocol (UDP) was not reliable as the number of packets that was lost

in each minute and only the success of the transmission was 95.4% and it is mainly due to the network and is not related with the hardware. The success rate of the transmission was slightly increased, there exist a reduced battery life time when the communication was established through Wi-Fi and Hypertext Transfer Protocol (HTTP).The packets transmitted are received only when the devices is within a certain range in the case of third one using Bluetooth Smart. This paper concludes with various advantages and disadvantages of using the above three communication protocol with respect to the factors such as power consumption and life time of the devices, Internet facility, success and reliability of the transmission and revealed that all three can be used for web based application. Wi-Fi and BLE are two technologies suited for ZigBee protocol for monitoring applications that can successfully compete with the concept established.

Zhou et al. [11] proposed a tracking framework and a systematic road detection approach in video footage using UAV. Initially Graph cut algorithm is applied for the extraction of the road regions from other non road regions. After the initialization, for quick retrieval of the road region homography estimation is used. The proposed framework has a precision of 97.4% and processing 35 frames per second for 1080 × 600 videos on average. Author has published a systematic approach for the identification of road by applying the relevant algorithm for different time instant in UAV videos.

Lockheed Martin et al. conducted and demonstrated a collaborative system comprising of drone and helicopter. The helicopter is designed to drop water exactly on the hot spots identified by the UAV [12]. Companies such as Aerones, Nitrofirex, and Singular Aircraft are focusing to design and implement drone systems that make use of water to suppress fires [13–15]. Phan and Liu [16] developed a system consisting of airship, UAV and unmanned ground vehicles (UGVs). The peak level of the hierarchy the airship generates a mission plan by utilizing fire outbursts, UAV, and UGV planned models [16].

The continuous research of developing UAV system for the detection and construction of wildfire implemented with drones, application program to program the hardware and relevant algorithms for detection and tracking is in progress day by day [4]. The main focus is different from the previous establishments or papers which deliberately work in developing and resulting a remote-sensing capability which will give an alarm not only for fires, but it also in detection of the fire head and flanks' spreading. Thus remote sensing is designed to detect fires caused in buildings, firefighters, and based on that will recognize the risk of the fire spreading to any other area or a precise location parameters. The purview of this particular paper is designed for this remote-sensing system.

Howden et al. developed a system with multiple that overlooks complex areas for fires [17]. This system will split the area under surveillance into cells and the parameters such as velocity, pheromone factor and the repulsion factor will be continuously updated for every iteration. The simulation result for various parameters and giving priority to the cell was also investigated in detail. Their algorithm proved successful when implemented in real time environment as one of the UAV may have less fuel and refilling is needed suddenly, one of them may crash and one may become failure.

Pham et al. also developed a similar algorithm for multiple UAV system that identifies the fire that is spreading across the boundaries [18]. It ensures collisions are avoided between UAVs. Yuan et al. implemented a forest fire detection system for a forest fire when the fire is spreading in regions [19]. The video captured by an aircraft in real time environment and the detection proves the systems reliable and accurate. Similarly, a decentralized autonomous system of multiple UAVs to monitor and govern the perimeter of the forest fire and containment was presented by Alexis et al [20].

The detailed analysis of the implementation of various system for fire detection and methods to extinguish the fire motivated to implement the drone of required specification to extinguish the fire using fire balls attachable to drones. This system is implemented on low power chip Raspberry Pi 3 and Navio 2. The Raspberry Pi 3 is interfaced with Navio 2 module for better specific controls such as barometer, accelerometer etc. The hardware modules interfaced with raspberry pi 3 are 4 BLDC motor, 4 ESC, Power distribution board, a transceiver, and a thermal sensor. The system design is a combination of internet of things and dedicated server for continuous storage and monitoring system using Amazon web services.

2   Proposed System

Figure 1 is the representation of implemented wireless UAV system in two stages. Stage 1 is for the detection and to extinguish the fire. The stage 2 of Fig. 1 highlights the blocks used for transmitting the fire outburst information. Figure 2 shows the interconnection of the various hardware of the UAV System. The central unit of the drone is flight controller Raspberry pi3 connected to brushless motor and electronic
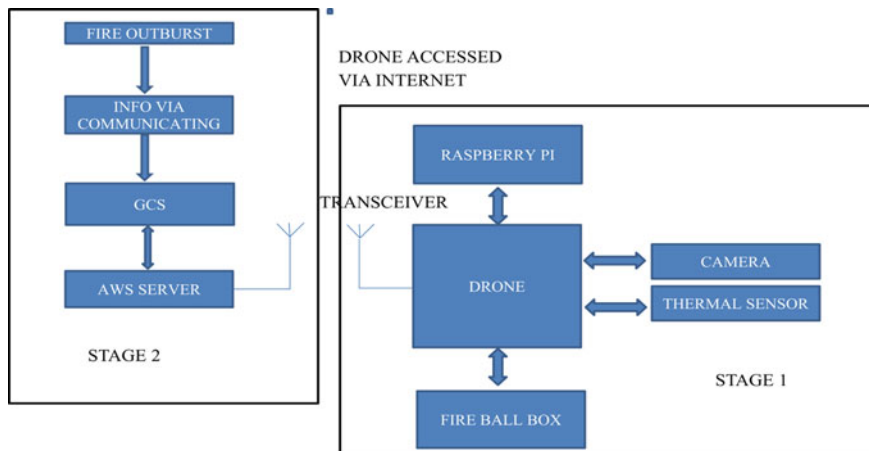
Fig. 1 Proposed UAV system
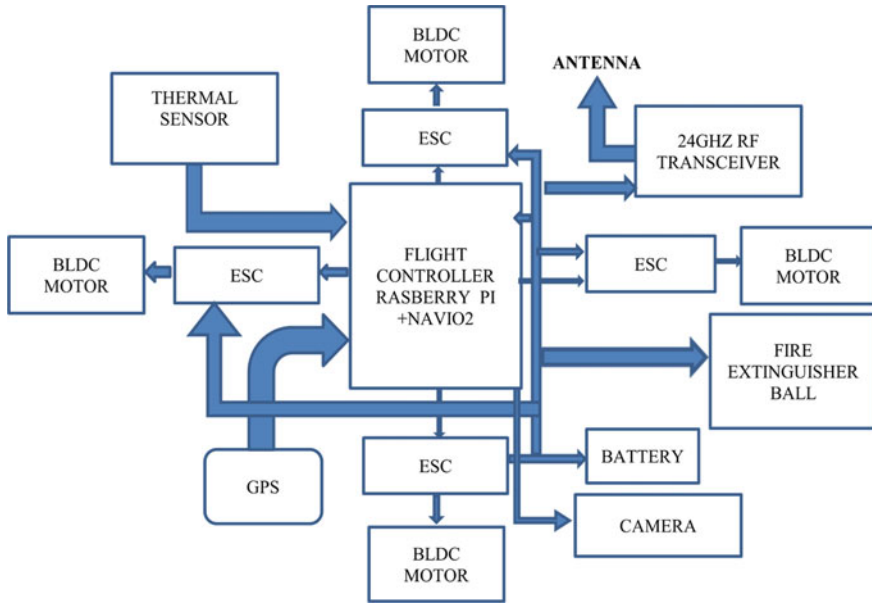
Fig. 2 Connection diagram of the hardware of the system

speed controller (ESC) for the stable movement. The purpose of using electronic speed control or ESC is for varying the speed of Brushless DC electric motor (BLDC) motor, its direction and also to act as a dynamic brake. The Raspberry Pi 3 is build with powerful processor, ten times speeder than the first generation, wireless LAN and with Bluetooth connectivity.

The drone frame used in the system is built from quality glass fiber and polyamide nylon. The main setup of the UAV is connecting the motors with the power distribution board where the power to the motors is controlled by the electronic speed controller. The power distributed board is connected to the flight controller where the input output to the ESC is connected to the flight controller. Then the LIPO battery is connected to the power distribution board and tested for the function of the motors. For wireless communications a transmitter is connected to the raspberry pi. On bottom of the UAV the fire ball drop is fixed. The balls effectiveness and robustness for building fires is maintained by conducted experiments. The fire extinguisher ball (AFO) which is approximately 0.7 kg, is used. The AFO was dropped directly into fire, whereas they were located in bottom of the drone frame which can be released the fire a few meters above.

The stage 2 focuses upon the software connection using dedicated server. Here the raspberry pi is loaded with a pre-configured operating system and connected to amazon web services (AWS) which is used to monitor and track the UAV movements. Here the operating system image configures the flight controller, so the dumping of operating system is done on a memory card. The installation takes place containing
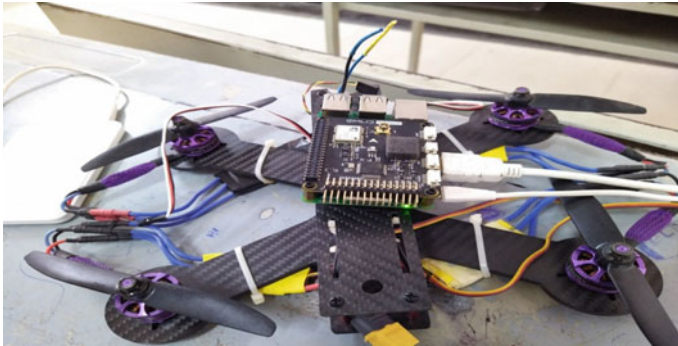
Fig. 3 Prototype of the developed drone

all services, kind of a meeting point for nodes. Hence, the nodes are discoverable and communicated to each other Raspberry Pi. There are different set of drivers within one node.

Considering fire is around two meter diameter, 0.7 kg AFO ball was dropped, the possibilities of which if the ball exploded but was not able to extinguish the fire and the fir outbursts continues or if it did not explode and the fire passed an approximate area of three meters in diameter, the fire was going to be put out manually by the firefighters These conditions were set for safety purposes. Figure 3 gives drone developed. The drone is connected to the ground station and controlled via internet. Using the AWS console a new account is created with the following services such as Lambda Service, Simple Storage Service, Cloud watch, Cloud alarm, Simple Notification Services and some additional services for external image processing. The servos are backed up by the power to so a Bit Error Correction (BEC) should be present. BEC would serve as back-up power supply to Navio + incase the main power source supply is intervened. ESCs will heat each other if many ESC central wire connected to Navio 2.

3 Results

Drone configurations are usually connected to motors and move when the downward thrust of UAV system is equal to the gravitational pull working against it. When the rotors provide an upward force greater than gravity, they automatically descend, and the UAV is directed in the desired direction using pitch, yaw, and roll. When the rotors provide an upward force greater than gravity, they automatically descend, and the UAV is directed in the desired direction using pitch, yaw, and roll. Drones are a microcosm because they are dependent to provide two-way communications for the Internet of Things. Testing of the drone must be done for safety and reliability for minimum cost and requirements. They must have fail-proof software that is certified for safety and effective in communication systems. They also should have efficient

power management and ability to operate in challenging and difficult environments. The raspberry pi acts as an interface between the Navio and AWS console. The mission planner software a GPS location is being fixed and controlled. The use of IoT core enables us to dump software into the microcontroller. The transmitter is an additional safe-switch.

The amazon web services console is a web application which enables the users or organizations to establish their application worldwide. The management of console consists of root user where certain permissions are allowed and denied. This enables us to monitor in the AWS console which will be ground control station. Initially the connection was checked using the putty window. In this window the reading that are set in the program were tested. The drone works like any other drone flying up to an extent and capturing views. But, in this case the drone flies above the fire affected areas and helps to put off the fire. This drone works with the help of controller where signals are connected through a transmitter. The drone carries fire extinguisher balls along with it and drops it from a certain height. The ball further rolls down to the center of the fire and blasts where it puts off the fire this step can be followed until the fire is completely put off, their efficient in terms of time to activate id done by AFO ball. Time taken is 280.9 s to explode. The 0.7 kg size ball extinguished a circle of 1-m diameter fire outbursts yet the 1.5 kg balls are assumed to extinguish a radius of 1.5 m. There is only need of little help from man to control the drone. AWS was connected to Raspberry Pi to capture the images from the drone.

4 Discussion

Consider the assumption that the UAVs could be re-loaded with a new method, it could be possible to prevent fire outbursts from spreading in multiple directions, or to extinguishing fires igniting the surrounding objects in large amount. The system should be increased exponentially if there is relation between increases in swarm size to the number of balls per UAV increase. The Table 1 depicts the idea of required components.

Power will be 200 amps combination of continuous current and 400 amps peak current. Motor power capability can pull up to the amps which is 25 amps continuously without any disruptions. The thrust tests, each motor provides of 3 kg with 1867 carbon propellers, at 22.3 V, 24.1 amps. A ball-release mechanism was designed and built as a mechatronic system attachable to the firefighting UAV. The system consists of electronic components such as microcontroller, power supply, sender and receiver, and motor as well as a mechanism to carry and release the ball and its connections to the fire-fighting UAV. The ball-releasing mechanism was attached with the mechatronic components.

Table 1 Table captions should be placed above the tables

Component	Description
Frame	Quadcopter frame 550 mm
Motors	2200 kV BLDC
ESCs	30 Amp
Flight controller	Navio 2 + Raspberry Pi
GPS	Navio 2
Telemetry	2.4 GHz transceiver
RC transmitter	FS-i6S 2.4 GHz transceiver
Battery	2200 mAh 3S 20C Lipo
Propellers	10 × 5 carbon fiber
Raspberry pie	Raspberry Pi 3
Thermal sensor	DS18B20 temperature sensor
Camera	Ardu-cam 5 MP 1080p sensor OV5847 mini camera

5 Conclusion and Future Work

The project can be further improved by widening the obstacle detection range with ultrasonic detectors. Further improvement can be made for this system. This system can be improved with less latency and more accuracy. Deep learning algorithms and Machine learning algorithms could be initiated. This proposed work with advanced features reduces manual work and allows people to move to that place immediately and independently. The advancements in software side can reduce the latency and produce high throughput. In the hardware side adding high resolution cameras providing high sustainable circuit boards which have high durability. Therefore, real time UAV is achieved by the implementation of web services in IOT.

These experiments were constructed and developed for a research project which the main objective is to design a system of UAVs which can be automatic connected to web server and monitored through internet and minimize the fire by extinguishing the AFO balls above them. The idea can be simple if a single system of UAV could drop them to optimal points, in optimal numbers, on time then there can be multiple systems simultaneously working at same wavelength. Remote-sensing and image capturing technology will be needed to determine points of attack and the web server interfaced should be a dedicated one which cannot be used for other services. The size of the fire-extinguishing balls depends upon the drones frame and the thrust it can lift along with its own weight is a parameter that requires further investigation for the usage of this system in real time applications. This is a promising result for further studies that will be conducted for developing the proposed drone-assisted fire-fighting system.

References

1. Rome NY (2015) Lockheed martin conducts collaborative unmanned systems demonstration; Unmanned KMAX helicopter, Stalker XE Small Unmanned Aircraft System (UAS), and UAS Traffic 25 Management (UTM) Systems Work Together to Extinguish Fire
2. Lee W, Kim S, Lee Y, Lee H, Choi M (2017) Deep neural networks for wildlife detection with unmanned aerial vehicle. In: Proceedings of the IEEE international conference on consumer electronics (ICCE). Las Vegas, NV, USA, pp 8–10
3. Martínez-de Dios JR, Merino L, Caballero F, Ollero A (2011) Automatic forest-fire measuring using ground stations and unmanned aerial systems. Sensors 11(6):6328–6353
4. Holman RA, Katherine LB, Nicholas JS (2017) Surf zone characterization using a small quadcopter: Technical issues and procedures. IEEE Trans Geosci Remote Sens 55(4)
5. Laurenti A, Neri A (1996) Remote sensing. Communications and information technologies for vegetation fire emergencies. In: Proceedings of TIEMEC 96. Montreal, QC, Canada, pp 28–31
6. Beltran C (2013) Unmanned aerial vehicle with fire extinguishing grenade release and inspection system: 100% report (Doctoral dissertation, Dissertation, Florida International University. https://mme.fiu.edu/wp-content/uploads/2013/12/F13-SR)
7. Krider EP, Noggle RC, Pifer AE, Vance DL (1980) Lightning direction-finding systems for forest fire detection. Bull Am Meteor Soc 61(9):980–986
8. Kumar M, Cohen K, HomChaudhuri B (2011) Cooperative control of multiple uninhabited aerial vehicles for monitoring and fighting wildfires. J Aerosp Comput Inf Commun 8(1):1–16
9. Wallar A, Plaku E, Sofge DA (2015) Reactive motion planning for unmanned aerial surveillance of risk-sensitive areas. IEEE Trans Autom Sci Eng 12(3):969–980
10. Mois G, Folea S, Sanislav T (2017) Analysis of three IoT-based wireless sensors for environmental monitoring. IEEE Trans Instrum Meas 66(8):2056–2064
11. Zhou H, Kong H, Wei L, Creighton D, Nahavandi S (2014) Efficient road detection and tracking for unmanned aerial vehicle. IEEE Trans Intell Transp Syst 16(1):297–309
12. Yuan C, Zhang YM, Liu ZX (2015) A survey on technologies for automatic forest fire monitoring, detection and fighting using UAVs and remote sensing techniques. Can J Forest Res 45(7):783–792
13. Lin H, Liu, Z, Zhao T, Zhang Y (2014) Early warning system of forest fire detection based on video technology. In: International symposium on linear drives for industry applications, pp 751–758
14. Sharifi F, Zhang YM, Aghdam AG (2014) A distributed deployment strategy for multi-agent systems subject to health degradation and communication delays. J Intell Rob Syst 73(1–4):623–633
15. Celik T, Demirel H (2009) Fire detection in video sequences using a generic color model. Fire Saf J 44(2):147–158
16. Phan C, Liu HH (2008) A cooperative UAV/UGV platform for wildfire detection and fighting. In: 2008 Asia simulation conference-7th international conference on system simulation and scientific computing. IEEE, Santa Clara, CA, USA, pp 494–498
17. Howden D, Hendtlass T (2008) Collective intelligence and bush fire spotting. In: Proceedings of the 10th annual conference on genetic and evolutionary computation. Atlanta, GA, USA, pp 41–48
18. Pham HX, La HM, Feil-Seifer D, Deans M (2017) A distributed control framework for a team of unmanned aerial vehicles for dynamic wildfire tracking. In: 2017 IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 6648–6653
19. Yuan C, Liu Z, Zhang Y (2017) Aerial images-based forest fire detection for firefighting using optical remote sensing techniques and unmanned aerial vehicles. J Intell Robot Syst 88(2–4):635–654
20. Alexis K, Nikolakopoulos G, Tzes A, Dritsas L (2009) Coordination of helicopter UAVs for aerial forest-fire surveillance. In: Applications of intelligent control to engineering systems. Springer, Dordrecht, The Netherlands, pp 169–19

State-of-the Art: Short Text Semantic Similarity (STSS) Techniques in Question Answering Systems (QAS)

Zaira Hassan Amur, Yewkwang Hooi, Irum Naz Sodhar, Hina Bhanbhro, and Kamran Dahri

Abstract Semantics can be used to assess responses in question answering systems (QAS). The responses are typically short sentences. Assessing short sentences for similarity with the expected answer is a challenge for Artificial Intelligence. Unlike long paragraphs, short texts lacks the adequate and accurate semantic information. Existing algorithms don't work well for short texts due to insufficient semantic information. This Paper provides the state of art on semantic similarity techniques and proposed the research framework to enhance the accuracy of short texts.

Keywords NLP · Text summarization · Semantic similarity · Short text · Information content · Information retrieval · Text assessment · Accuracy QAS · Semantic information

1 Introduction

Natural language processing the semantic similarity put its tremendous impact on question answers systems. Whereas short answers/texts observed to be most import

Z. H. Amur (✉) · Y. Hooi
Department of Management and Information Sciences, University Teknologi Petronas, Seri Iskandar, Malaysia
e-mail: Zaira_20001009@utp.edu.my

Y. Hooi
e-mail: yewkwanghooi@utp.edu.my

I. N. Sodhar · H. Bhanbhro
Department of Information Technology Shaheed, Benazir Bhutto University, Shaheed Benazir Abad, Pakistan
e-mail: irumnaz@sbbusba.edu.pk

H. Bhanbhro
e-mail: hina@sbbusba.edu.pk

K. Dahri
Department of Information Technology, University of Sindh, Jamshoro, Pakistan
e-mail: kamran.dahri@usindh.edu.pk

techniques. Which commonly recognized as STSS. These short texts include social posts, question/answers, conversations, keywords and comments etc. In STSS short-questions provide a useful understanding of specific concepts in a particular subject domain. Short texts semantic similarity (STSS) helps to measure 10–20 words long.

Due to limited information these texts consider the most challengeable for number of applications. These texts contains more than one meaning so they are generally difficult to understand. Short texts contains lots of textual errors and produce result ambiguity. These Texts generate semantic errors like grammatical and real world errors. The main reason of these errors are user's insufficient knowledge. Analyzing and fixing these errors in short-texts improves the effectiveness of learning environment. Semantic calculation of short texts influenced the text words that causes the poor accuracy [1]. Usually, understanding the hidden semantic meaning of word is difficult task.

In natural language processing the text represents sentence. Each sentence/texts has set of words, these words counted with different Parts of speech. Machines only understand the language of symbols rather than text. To recognize the particular and intended meaning of words or sentences it's highly needed to convert text into symbols. Part-of-Speech (POSs) determines that each word in the sentence contains a specific sense of information. Generally machines don't understand the dual meaning of words/sentences. So therefore, natural language processing enables machines to learn and seek the human natural language.

Text understanding plays a vital role in NLP. Most of the current methods are task oriented such as information retrieval, relationship extraction and Q&A systems. The main reason of using text in natural language processing make machines to know the human languages. The aim of this paper is to summarize STSS techniques to identify the trend of research.

1.1 Structure of the this Paper

The Paper is organized as follows: Section 2 is on literature study; Sect. 3 analyses the various similarity approaches; Sect. 4 proposed the research framework and Sect. 5. Gives the conclusion where section 6 provides the references of study.

2 Literature Review

Journal articles and proceeding papers from ISI and Scopus-indexed publications or reputable institutions on STSS research in the past years have been carefully selected. The papers are further filtered with focus on ML applications for A&Q systems. The literature review identifies interesting development and trends of the study.

Refer to Table 1.1 in Attachment I. The table characterizes the STSS methods and show the years and trends.

It is important that short texts (subjective answers) must be understandable and disambiguated so that learner can find the accurate information. However, Short text similarity also related to textual entailment (TE) and paraphrasing techniques. Which mostly use in many natural language processing tasks. These techniques differs from each other in different context. TE (Textual entailment) uses the direct relation among text fragments by using the hypothesis techniques, whereas paraphrasing used to recognize the same meaning of text. These both techniques works on yes/no decisions. STSS use these techniques to rate the semantic relatedness among groups of words/sentences [1].

There are so many traditional methods used in NLP and the most common are, Bag of words, Vector space model, and BM25 [2], helps to generate the words in text. In NLP usually traditional methods are poor methods that can't properly detect the semantics of text at conceptual levels. Because of the limited amount of text, it is not easy to calculate semantic similarity with these methods.

Several other methods like n-gram based, word-based methods, long based distance so on and so forth are used to fixing the real word errors to improve the accuracy for short answers [3]. There are some limitation in the existing intelligent subjective questions. Same similarity often repeated among the number of questions during the assessment process, increases the problem of duplication as well as low accuracy [4].

This text similarity not only because of synonyms but the similarity can also be possible because of antonyms or other similarity domains with the different learning context. When text reduced it increases the difficulty for the semantic similarity. Number of approaches have been based on traditional methods of test items and words similarity [5, 6]. But these methods are not applicable to our problem domain.

Moreover, some algorithms contain high time complexity because it uses number of semantic relations. One of study [7, 8] mentions the relationship diagram short texts known as knowledge graph and random walk method to improve the concepts of optimal word segmentation.

Another study [9] provides the attention neural network approach for reading the text. This methods distinguish the text summarization and conceptualization semantic information of text. However, the Paper [10] proposed a method that provides the input transformation to set the optimal changes to the model architecture.

Meanwhile a study [11] introduces a new Q&A system for selecting grounded and commonsense information from Concept another study [12] reviewed how to gain new facts about the world based on knowledge graphs. Most of these methods are task oriented. Although these methods have good expressiveness on particular tasks but they still lack understanding and improving the accuracy of text semantics. Some studies focuses on external knowledge to upgrade the quality of topic identification in short texts.

The textual similarity has been proposed in 2006, where only small amount of text were supposed to be included but after that the research becomes enhanced from short to long and long to individual words. [1] It also works well in many web applications like ontology generation, keyword extraction, and entity disambiguation. Like, Phan et al. [13] find out the hidden Model. According Zhang [14], suggest that when two

or more words are collocated, there must be one word that works as the center word and the other word as the collocation word. Parts of speech works well to construct the center word.

However, Kim [15] suggest that related features can also help with disambiguation. Furthermore, they worked to capture semantic relations between terms using the novel approach LDA, this approach helps to improve the accuracy of short text conceptualization by using context semantics into consideration. Whereas POSs like verbs, adjectives, and other attributes, can also help to detect instance ambiguity. Moreover, the study [16] introduces the framework for short texts that detect the errors from text. More specifically, the work has divided into three subtasks to understand the short text such as text segmentation, type detection, and concept labeling.

Wang [4] present an approach to solve the problem of semantic similarity in test papers, with the help of density entropy they selected the test papers to screen the various question papers from item bank and then applies the calculation of semantic similarity to detect the intelligent test paper from the corpus.

Furthermore, Huang [17] developed the algorithm to improve the performance of STSS with low time complexity. This algorithm incorporated the different WordNet-based word measures to address the word pairs with specific POSs that helps to enhance the evaluation of semantic similarity of unknown words in shorts texts. Hesham [1, 13] represented the work, that focusing one Semantic Textual Similarity (STS) of a question pair. In this study they finds if two questions have the same answers, then they are semantically equivalent.

Otherwise, if the answers are different to one another, or partially different then (the answer of the first question is included in the answer of the second question but it considered the incomplete answers, or the questions share a portion of their answers yet different otherwise) then these answers subsequently considered the semantically nonequivalent.

To compute semantic similarity for short texts is important in many areas. Many approaches have been proposed that use linguistic analysis. These methods determine whether the words in two short texts look alike, in terms of largest common substring [18]. These approaches usually work for trivial cases.

This kind of methods creates a major problems, because the complex features are derived from other NLP tools and components which use for other purposes. Some application depends on parse trees that are completely restricted to syntactically well-formed texts. More ever, the error from these tools also creates the problems in Text semantic similarity models that are very hard to solve individually.

Furthermore Younes [19] proposes a technique for text summarization they focused on identifying significant parts of text to create summary. To solve the problem they represented an optimized algorithm that helps to produce summary with the lexical chains and WordNet dictionary.

However, the Study [20], proposed methodology that consists on rules, based on NLP algorithms like sentence splitting, tokenization and POS tagging. These rules were applicable on user provided text which he want to summarized. In Algorithm each word is assigned a score based on its frequency, this algorithm will run sentence by sentence.

There are number of techniques that uses Part-Of-Speech (POS) as syntactic information to compute similarity between sentences. Subsequently, the work present [21], gives the concept of weighting strategy based on POS. They represent the idea that many POS and certain relation between POSs are more important and beneficial than others. In particular they combine the weighting strategy with bag-of-words approach to computer semantic (sentences) similarity in texts.

3 Techniques for Textual Similarity

In this section, this study proposes the methods that have potential for further development and the rationals. There are various approaches that used to measure the semantic similarity. This section represents the proposed method for extracting similarity in short-text questions in detail. These approaches needs to acquire similarity from semantic and syntactic information that exist in the compared natural language sentences. In a natural language processing (NLP) a sentence can be defines as a set of words with Part-of-speeches (POSs) [22] (Fig. 1).

3.1 Corpus/Statistical-Based Semantic Similarity

A large variety of the proposed approaches in words similarity or primarily corpus based. In this approach valuable data is extracted from analyzing an enormous
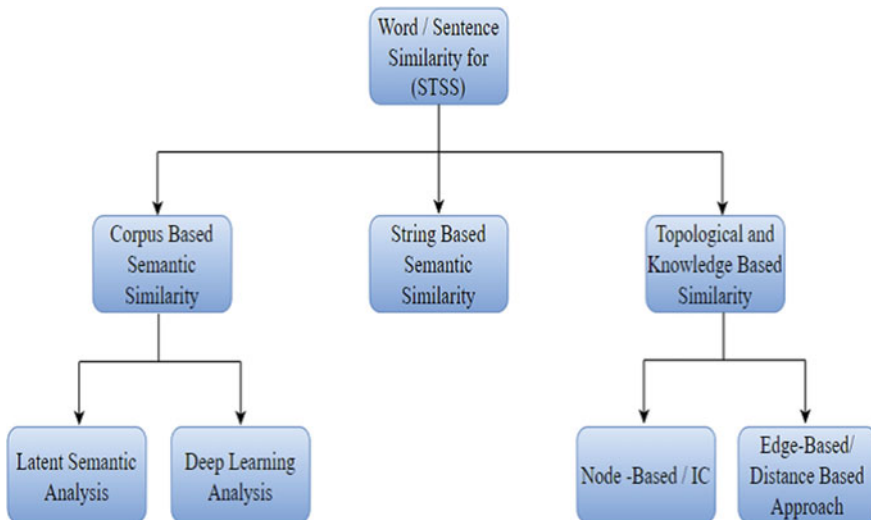
Fig. 1 Techniques for measuring the short text semantic similarity (STSS) [33]

corpora. Furthermore, analyzing words using Eq. 1 would allow the massive corpus helps to extract the words similarity accordingly. Corpus-based approaches also known as statistical-based approach. Corpus/statistical based similarities can be measured by two.

Different techniques, normal statistical analysis (LSA) and the other one is deep learning. A large corpora analyzed statistically through counting the words within the corpus and documents. Tf-idf is a vital statistical object for the corpus analysis which is used to employ as a word co-efficient [23].

3.2 Latent Semantic Analysis/Approach (LSA)

In this technique, words usually expressed as vectors. In LSA, the word matrix and vectors constructed by analyzing the large text. The word matrix represents in the form of rows and columns. In this corpus based analysis, usually cosine similarity is highly used for calculating the word similarity text, tables and figures. Substantial changes in content, e.g., new results, corrected values, title and authorship, are not possible and can-not be processed.

3.3 Deep Learning Analysis/Approach

This approach also known as word embedding technique which used to find the words semantically. In the semantic space, a very big corpus is required to extract the word representation. This technique is specifically used to guess the word from the given surrounding words, usually known as bag of words model. This model also helps to understand the vector representation. Deep learning also use the Skip gram method to understand the context meaning of word. It also helps to understand the word/vector semantically. For example if we perform some calculation on vectors semantically,

$$[\text{VECT (KING)} - \text{VECT (MAN)} + \text{VECT (Woman)}]$$

The result of this vector shows the word "queen" which is generally very close to the specific semantic calculation of word representation. Deep learning also use web based similarity.

3.4 String-Based Similarity Approach

String-based also knowns as lexical based word similarity. It uses the comparison technique between the sequences of characters. There are various

methods/approaches used in String-based similarity to calculate the similarity of words like, levenshtein distance, longest common sub string Q-gram and jaccard distance. As for as LD, it computes the distance between two strings by using different operations like, insert, delete or replace. And change the one form of string to another. Whereas, longest common substring (*LCS*) finds the longest substring of string.

Example

$$LCSubtr\ (S_1, S_2) = \max_{1 \le i \le m .1 \le j \le n} LCS_u ff(S1_1...i, S2...j),$$

3.5 Topological/Knowledge Based

These methods works well to understand the meaning of word to process computationally. It actually use the WordNet for the lexical relationships. There are three methods that helps WordNet to extract the similarity. These can be categorized as node-based/Ic, edge-based and Glass-based measure.

3.6 Node-Based/Ic (Information Content)

This approach uses concepts to extract the similarity. This measure includes the hierarchical structure. In this method every node uses IS-A taxonomy that are kept in on set. There are also other measures that considers the node/Ic based measures for example, RES, LIN and JCN [1]. Before implementing the node based model the concepts can be divide into *words*(c) and classes defined as *classes (w)* [24].

$$classes\ (w)\ =\ \{c/w\ \in\ words\ (c)\}.$$

Here in the whole hierarchy the class works as a subtree and the *class (w) contains* all sets of possible words [24].

3.7 Edge/Distance Based Measures

This is the direct approach for computing the similarity. This technique counts the edges of those nodes that are correspond to each other. Distance based measures used the hierarchal concept for WordNet where the PATH length can be used to measure the similarity There are three other approaches based on Distance measures PATH, WUP, LCH [1].

3.8   Gloss Based Measure

These measures uses the glosses between the concepts of WordNet. Vector and LESK are the type of this measure [1]. Table 1 Shows the Summary of various methods and dataset used for the semantic similarity of short texts. This table presents the latest work. Various datasets has been used by the different approaches the most common is Microsoft paraphrase corpus. This dataset contain 5801 sentence pairs that can extract from news, documents and various social network platforms.

4   Proposed Research Framework

This study proposed Research framework (Fig. 2) that has one independent variable (Text assessment) whereas the other's work as dependent variables. In this Framework the text assessment will be done by the student's essays. The evaluation will be based on examiners or expert essays that will be available to check the relatability among essays. These short texts divided into text extraction process that actually extract the core keyword from text through the machine learning process. Further the machine normalize the text through stemming and lemmatizing techniques to remove the suffixes. In summarization the machine performs the text prepressing that helps to reduce the noise from text to detect the stop words or punctuation marks from Texts. Moreover the Intent recognition check the similarity among the short texts to identify the relatable Keywords that has the highest similarity. This technique helps to give the rank to the text (answers). Whereas (Table 2) Text assessment, summarization, as well as similarity all these effects on the accuracy of short texts. The measurable variables in framework are: Text Assessment, Text extraction, intent recognition and accuracy.

In Q &A systems usually the Reponses works independently. So if the student's and examiner's responses relate with each other than machine will understand the context of word/sentence and indicates the predicate feature that has the higher similarity. If the predicate feature has high similarity then the accuracy of STSS will increase. If the predicate or candidate word failed to provide the accurate response then the accuracy of answer will not match to the specific answer.

5   Conclusion and Future Possibilites

The main consideration of this study is to give brief overview on short texts. There are various approaches like string similarity, corpus based or knowledge based. Usually these approaches provide good impact on text representation. From the literature and previous work we have identified and noticed that corpus based and knowledge based measure are most suitable measures for text summarization. Corpus based measure

Table 1 Summary of various methods and datanet used for the Semantic similarity of short texts

Research	Methods							Base method	Dataset
	WordNet	LCS	Word embedding	Structure based	Vector Based	PMI-IR	LSA		
Huang et al. [3], 2019	✓	✓			✓			String Based	MSRPar(Microsoft paraphrase corpus)
Islam et al. [25], 2009		✓				✓	✓	Combined (String & Corpus based)	MSRPar(Microsoft paraphrase corpus)
Zhao et al. [26], 2019			✓		✓			Corpus based	Gigaword & DUC-2004
Suleiman et al. [27], 2019			✓		✓			Corpus based	OSAC
Pawar et al. [28], 2018	✓				✓			Knowledge based	PILOT
Hien et al. [29], 2019 P4PIN.STS2015			✓		✓			Corpus based	MRPC
Cai et al. [30], 2020	✓	✓		✓				String & knowledge based	M&C, R&G, WS-353
Zhiguo et al. [30], 2016					✓			Corpus based	QASent, MSRP
Avani et al. [28], 2019	✓						✓	Corpus based	Kaggle
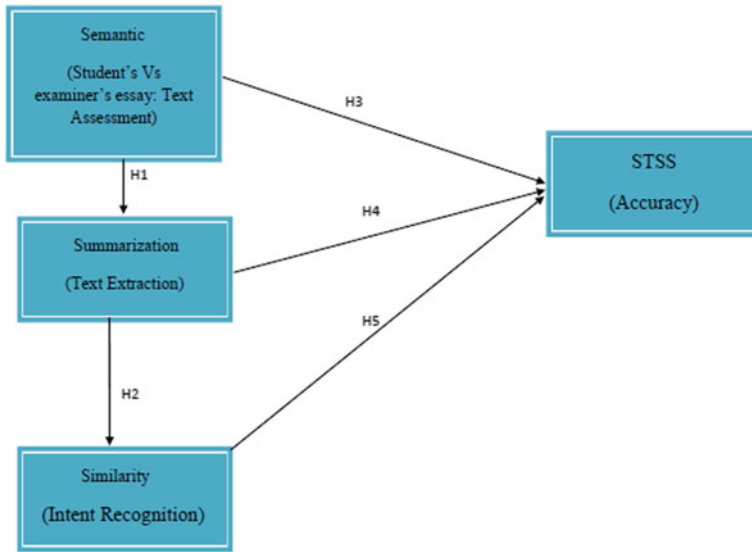
Fig. 2 Research framework

Table 2 Variables of proposed research framework

Variables	Independent variable	Dependent variable	Technique (Measurable)
Student' versus examiner's essay	✓		Text assessment
Summarization		✓	Text extraction
Similarity		✓	Intent recognition
Short Text		✓	Accuracy

provides the advantage of cosine similarity. These approaches helps to minimize the risk of errors and improve the chances of accuracy between short texts. Morever, this study proposed the research Framework that helps to find out the similarity among short texts (answers). For the future work, there is a lot of work yet to be done on unsupervised learning in semantic similarity.

References

1. Huang PS, Chiu PS, Chang JW, Huang YM, Lee MC (2019) A study of using syntactic cues in short-text similarity measure. J Internet Technol 20(3):839–850
2. Jimenez S, Cucerzan SP, Gonzalez FA, Gelbukh A, Dueñas G (2018) BM25-CTF: improving TF and IDF factors in BM25 by using collection term frequencies. J Intell Fuzzy Syst 34(5):2887–2899

3. Song X, Min YJ, Da-Xiong L, Feng WZ, Shu C (2019) Research on text error detection and repair method based on online learning community. Procedia Comput Sci 154:13–19

4. Wang H, Yang W (2019) An intelligent test paper generation method to solve semantic similarity problem. In: IOP conference series: earth and environmental science Vol 252, No 5. IOP Publishing, p 052126

5. Gashaw I, Shashirekha HL (2020) Machine learning approaches for amharic parts-of-speech tagging. arXiv preprint arXiv:2001.03324

6. Croft D, Coupland S, Shell J, Brown S (2013) A fast and efficient semantic short text similarity metric. In: 2013 13th UK workshop on computational intelligence (UKCI) (pp. 221–227). IEEE

7. Chen HH, Lin MS, Wei, YC (2006) Novel association measures using web search with double checking. In: Proceedings of the 21st international conference on computational linguistics and 44th annual meeting of the association for computational linguistics, pp 1009–1016

8. Martinez-Gil J, Pichler M (2014) Analysis of word co-occurrence in human literature for supporting semantic correspondence discovery. In: Proceedings of the 14th international conference on knowledge technologies and data-driven business, pp 1–7

9. Wang H, Zeng M, Xiong Z, Yang F (2017) Finding main causes of elevator accidents via multi-dimensional association rule in edge computing environment. China communications 14(11):39–47

10. Radford A, Narasimhan K, Salimans T, Sutskever I (2018) Improving language understanding by generative pre-training

11. Bauer L, Wang Y, Bansal M (2018) Commonsense for generative multi-hop question answering tasks. arXiv preprint arXiv:1809.06309

12. Nickel M, Murphy K, Tresp V, Gabrilovich E (2015) A review of relational machine learning for knowledge graphs. Proc IEEE 104(1):11–33

13. Phan XH, Nguyen LM, Horiguchi S (2008) Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In: Proceedings of the 17th international conference on World Wide Web, pp 91–100

14. Tachbelie MY, Abate ST, Besacier L (2011) Part-of-speech tagging for underresourced and morphologically rich languages—the case of Amharic. HLTD 2011:50–55

15. Song Y, Wang H, Wang Z, Li H, Chen W (2011) Short text conceptualization using a probabilistic knowledgebase. In: Proceedings of the twenty-second international joint conference on artificial intelligence-volume volume three, pp 2330–2336

16. Kim D, Wang H, Oh A (2013) Context-dependent conceptualization. In: Proceedings of the twenty-third international joint conference on artificial intelligence, ser. IJCAI'13. AAAI Press, pp 2654– 2661. [Online]. Available: http://dl.acm.org/citation.cfm?id=2540128 2540511

17. Hua W, Wang Z, Wang H, Zheng K, Zhou X (2015) Short text understanding through lexical-semantic analysis. In: 2015 IEEE 31st international conference on data engineering. IEEE, pp 495–506

18. Al-Bataineh H, Farhan W, Mustafa A, Seelawi H, Al-Natsheh HT (2019). Deep contextualized pairwise semantic similarity for Arabic language questions. In: 2019 IEEE 31st international conference on tools with artificial intelligence (ICTAI). IEEE, pp 1586–1591

19. Slam A, Inkpen D (2008) Semantic text similarity using corpus-based word similarity and string similarity. ACM Trans Knowled Discov Data (TKDD) 2(2):1–25

20. Jaafar Y, Bouzoubaa K (2018) Towards a new hybrid approach for abstractive summarization. Procedia comput sci 142:286–293

21. Vuong La CS (2019) Artificial intelligence versus natural stupidity: evaluating AI readiness for the vietnamese medical information system. J clin med, 8(2):168

22. Otter DW, Medina JR, Kalita, JK (2020) A survey of the usages of deep learning for natural language processing. IEEE Tran Neural Netw Learn Syst

23. Farouk M (2019) Measuring sentences similarity: a survey. arXiv preprint arXiv:1910.03940

24. Majumder G, Pakray P, Gelbukh A, Pinto D (2016) Semantic textual similarity methods, tools, and applications: a survey. Computación y Sistemas 20(4):647–665

25. Islam A, Inkpen D (2009) Semantic similarity of short texts. Recent Advances in Natural Language Processing V 309:227–236
26. Zhao F, Quan B, Yang J, Chen J, Zhang Y, Wang X (2019) Document summarization using word and part-of-speech based on attention mechanism. In: Journal of Physics: Conference Series, Vol 1168, No 3. IOP Publishing, p 032008
27. Suleiman D, Awajan AA (2019) Using part of speech tagging for improving word2vec model. In: 2019 2nd International Conference on new trends in computing sciences (ICTCS). IEEE, pp 1–7
28. Wang Z, Mi H, Ittycheriah A (2016) Sentence similarity learning by lexical decomposition and composition. arXiv preprint arXiv:1602.07019
29. Sakhapara A, Pawade D, Chaudhari B, Gada R, Mishra A, Bhanushali S (2019) Subjective answer grader system based on machine learning. In: Soft computing and signal processing. Springer, Singapore, pp 347–355
30. Cai Y, Pan S, Wang X, Chen H, Cai X, Zuo M (2020) Measuring distance-based semantic similarity using meronymy and hyponymy relations. Neural Comput Appl 32(8):3521–3534
31. Tsatsaronis G, Varlamis I, Vazirgiannis M (2010) Text relatedness based on a word thesaurus. J Artif Intell Res 37:1–39
32. Jin O, Liu NN, Zhao K, Yu Y, Yang, Q (2011) Transferring topical knowledge from auxiliary long texts for short text clustering. In: Proceedings of the 20th ACM international conference on information and knowledge management, pp 775–784
33. Han M, Zhang X, Yuan X, Jiang J, Yun W, Gao C (2021) A survey on the techniques, applications, and performance of short text semantic similarity. Concurrency Comput Pract Experience 33(5):e5971

Modeling of Wireless Power Transmission for Pacemaker Battery Charging in MATLAB Simulink Environment

G. P. Kiruthiga, Mohdfakhizan B. Romlie, K. Porkumaran, and S. Prabakar

Abstract Cardiovascular pacemaker is a little gadget that utilizes electrical driving forces conveyed to the terminal by contracting heart muscles and they can direct the thumping of the heart. It requires the energy where ordinary wires are badly arranged and dangerous to human body. When the life expectancy of the pacemaker battery is finished, continued changing of pacemaker will influence the patients. To build the lifetime of the implantable heart pacemaker. By utilizing remote force transmission for cardiovascular pacemaker depends on inductive coupling between the two curls has been proposed in this paper. The essential curl is to be put on the body, while auxiliary loop is to be set inside the body. From this, sending and getting loop can move power remotely, and beneficiary circuit will get the air conditioner sign and convert it into DC signal for the battery of the cardiovascular pacemaker. The reproduction results show that the moved productivity is about 80.01% at the distance 2 cm.

Keywords Wireless power transmission · Cardiac pacemaker battery · Inductive couplin

G. P. Kiruthiga (✉)
Dr.N.G.P Institute of Technology, Kalapatti road, Coimbatore, India

M. B. Romlie
UniversitiTeknologi Petronas, Seri Iskandar, Perak, Malaysia
e-mail: fakhizan.romlie@utp.edu.my

K. Porkumaran
Sri Sairam Engineering College, Chennai, India

S. Prabakar
Sona College of Technology, Salem, India
e-mail: srisornaprabu@gmail.com

1 Introduction

These days the world without electric force is troublesome thing to live. For instance, the day-day items like electronic vehicle, versatile, PC, clothes washer and so on, requires electric force. Like that force move through the little attachment containing numerous wires may cause fire mishaps and demise [1]. Pacemaker is only it is a little gadget that utilizes electrical motivation conveyed to the terminal by contracting heart muscles after that they can direct the pulsating of the heart [5]. In 1958, the pacemaker was embedded first with battery-powered battery made of nickel cadmium. In 1973, lithium battery was utilized to control the pacemaker. A pacemaker is put in the chest (just underneath the collarbone) to help manage moderate electrical issues with the heart.

Pacemaker leads might be situated in the chamber (upper chamber) or ventricle (lower chamber) or both, contingent upon the clinical condition [6]. Inherent coronary illness in conceived babies needs to the implantation of pacemaker from their introduction to the world to whole lifetime. And furthermore utilized in mature age individuals to aid appropriate working of heart. Boundaries of pacemaker are; capacitor, timing circuit, terminal, beat generator, battery. In this paper we zeroed in on to charge the battery.

1.1 Battery Technology

In pacemaker battery generally they are utilizing Lithium batteries in light of longer life expectancy and 5 V has been utilized, battery substitution is needed after 3.3 V. Mercury batteries likewise utilized however 3–5 cells are utilized to make one mercury battery, it makes costly. Natural force source battery is as yet in exploration not yet emerged. Plutonium based batteries were utilized in times past, and its life expectancy around 30 years however plutonium is atomic receptive and undependable, harmful subsequently not utilized.

1.2 Wireless Power Transmission

Remote Power Transfer (WPT) is a strategy for sending the electrical energy starting with one point then onto the next in a remote medium (without actual association). Two sorts in WPT,

- Near field it is non radiative and the force can be communicated to the short distance and inductive coupling is utilized.
- Far field it is radiative and the force can be communicated to the significant distance and capacitive coupling is utilized. For remote force transmission close documented is utilized on the grounds that it is non radiative (Fig. 1).

Fig. 1 Basic wireless power transmission
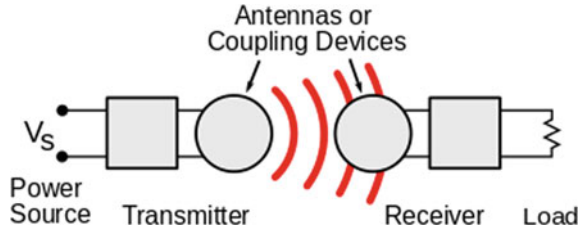
Table 1 Battery technology

Battery type	Life span
Nickel Cadmium	7 years
Mercury battery	10 years
Biological power source	Still in research
Lithium Iodide	10—12 years

From the above circuit when the power can be feed into the transmitter side then by through the coupling devices the power can be goes to the receiver side and then voltage is transferred to the battery. This is the basic wireless power transmission [7] (Table 1).

1.3 Current Issue

Once the life span of battery is over, repeated changing of pacemaker would affect the patients in these ways,

- Re-implantation of whole device cause defect to the surrounding tissues.
- Frequent surgeries for every 10 years interval for the new born babies with congenital heart disease may lead to weakness of the patient body.
- As well as senior citizens who are above 60 years cannot opt surgery method frequently [4].

To overcome the above issues, we adopted wireless power transmission to charge the cardiac pacemaker can be the alternative for surgery. In this paper we designed a simple model using MATLAB Simulink software for pacemaker wireless charger that constitutes amplifier, oscillator, inductive power transfer coupler, rectifier, LC filter and resistance load which represents pacemaker battery.

2 Proposed Work

In the wireless power transmission, the output circuit consists of a coil, a transmitter which transmits the power and receiver receives the power and send it to the battery. Whereas the transmitter side consist of an input voltage that is an AC voltage source, amplifier and the oscillator. The receiver circuit consists of rectifier, monitoring circuit and battery. All my simulation and output were generated using MATLAB (Fig. 2).

2.1 Amplifier

From the transmitter side foremost component is class A amplifier with Darlington pair. Moreover, in biomedical devices, the most commonly used type of amplifier is the class A amplifier, it is the simplest form of power amplifier during one complete cycle of input signal waveform producing minimum distortion and maximum amplitude of the output signal. Amplifier is used to increase the power of the signal. The current gain of the two-transistor individual is multiplied together along with collector currents are compared to the single transistor circuit (Fig. 3).

From the above circuit the input power 5 V and the frequency is 50 Hz then the supply goes to the capacitor C1 and value is 0.22 μF and it reaches the resistor R1, R2 and value is 2000, 7250 kΩ and then supply goes to the Darlington pair of collectors and the process continues along the Darlington pair and the amplified voltage is approximately 9 V (Fig. 4).
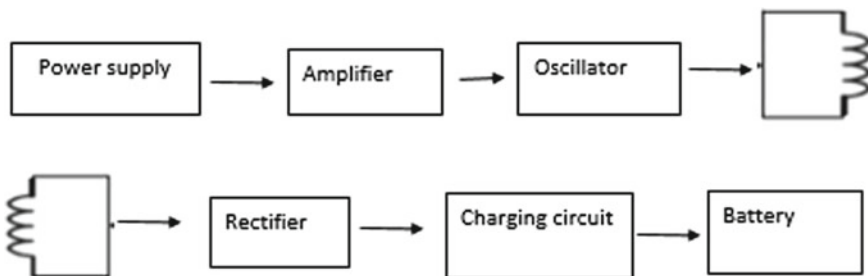
Fig. 2 Proposed work flow
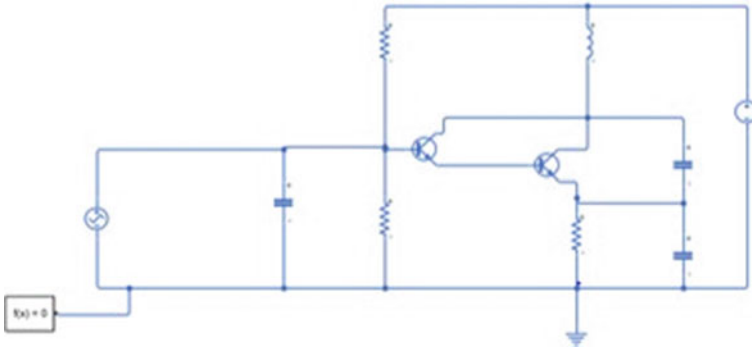
Fig. 3 Class A amplifier with Darlington pair
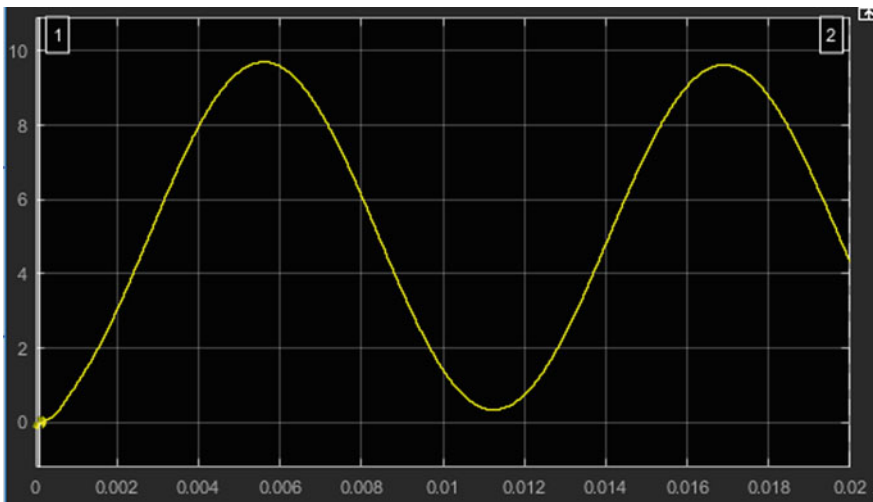
Fig. 4 Output for class A amplifier

2.2 *Oscillator*

The acceptable range of frequency for a human body is 300 kHz and 3.55 MHz. Hence the 5 V is given as the input voltage and obtaining the frequency in the range 300 kHz after oscillation (Fig. 5).

After we get the amplified voltage form the amplifier it reaches the oscillator. The oscillator can oscillate the frequency. In the above circuit 555 timer is an 8 pin IC. Pin 2 trigger is connected to the capacitor and the value is 0.1 nF, then the discharge pin is said as pin 7 and pin 4 is said as reset pin, this is connected to the input DC voltage source which is 5 V and control voltage pin is pin 5, which is grounded with capacitor value 0.1 nF. We are using the Astable mode 555 timer as an oscillator and
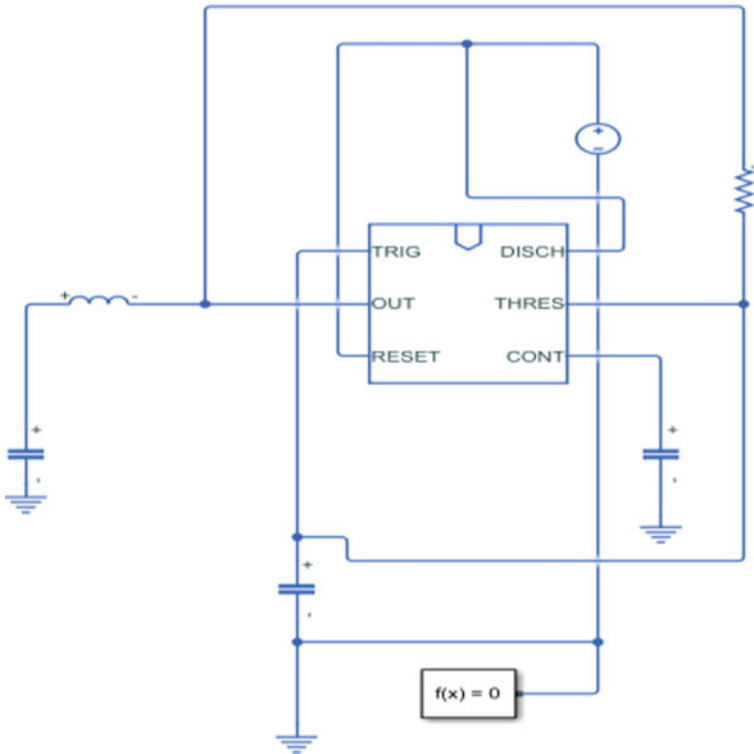
Fig. 5 Oscillator by using 555 timer

it can generate the pulse. Generally, the Astable mode generates a square wave, in order to obtain a Sine wave, we add L resonance of value 100 μH and C of value 2.5 nF along with the Astable mode 555 timer. The obtaining frequency range after simulation is about 318 kHz (Fig. 6).

$$f = 1 \div 2\pi)\overline{Lc}$$

Then substituting the values for above equation then the frequency is about, $f = 318$ kHz.

2.3 Receiver Side

The output from the oscillator goes the capacitor which is connected parallel. The AC voltage from the capacitor goes to the transmitter coil and the consists of mutual inductance L1, L2 = 95.7 nH, by through the magnetic coupling signal goes to the receiver coil. In the receiver side rectifier is present, here half rectifier is used, because
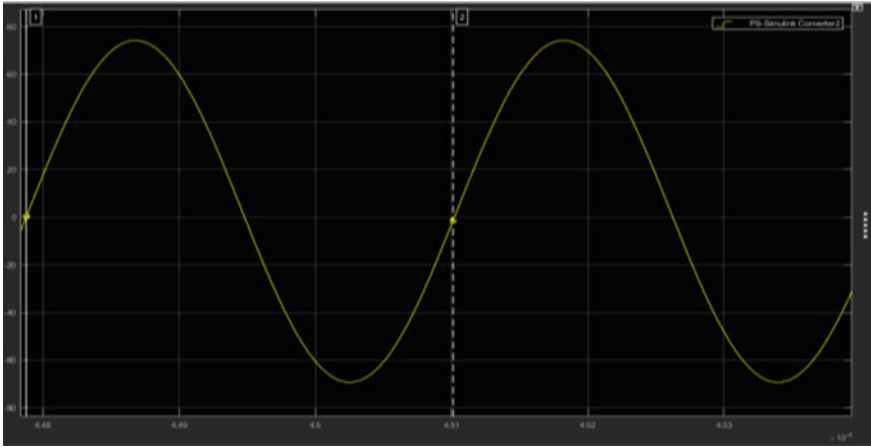
Fig. 6 Output for oscillator

the secondary coil is placed inside the heart so component should not be complex. Full wave rectifier is not used here because it is still in research. Rectifier is used to convert alternating current into direct current. To the charge the battery, the input will be in DC. The output from the rectifier is appeared as pulsated DC, for charging of the battery pulsated DC is enough. By adding LC filter pure DC is formed. This is more appropriate compared to pulsated DC (Fig. 7).
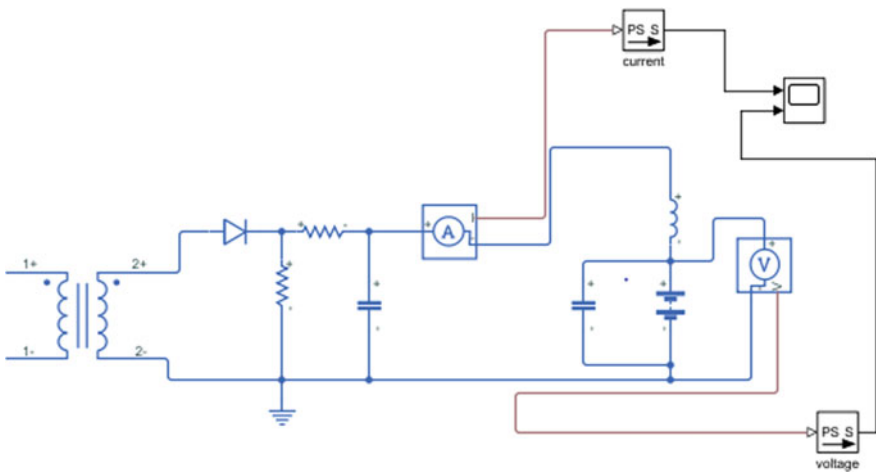
Fig. 7 Half wave rectifier with LC filter

3   Results and Discussion

The above circuit mentioned is the overall circuit. When the input here is DC of about 5 V then it goes to the amplifier circuit then it amplifies voltage about 9 V then it goes to oscillator, and it converts DC voltage into AC voltage and it shapes the frequency which is about 318 kHz and it goes to the transmitter coil through the coupling device, it goes to the receiver coil, in that rectifier converts AC to DC voltage. After rectification the output voltage is about 4 V, in spite of the loss occurred. The simulation results show that the transferred efficiency is about 80.01% at the distance 2 cm (Figs. 8 and 9).
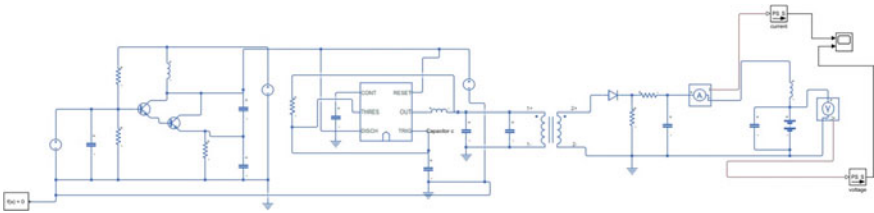
Fig. 8   Overall circuit
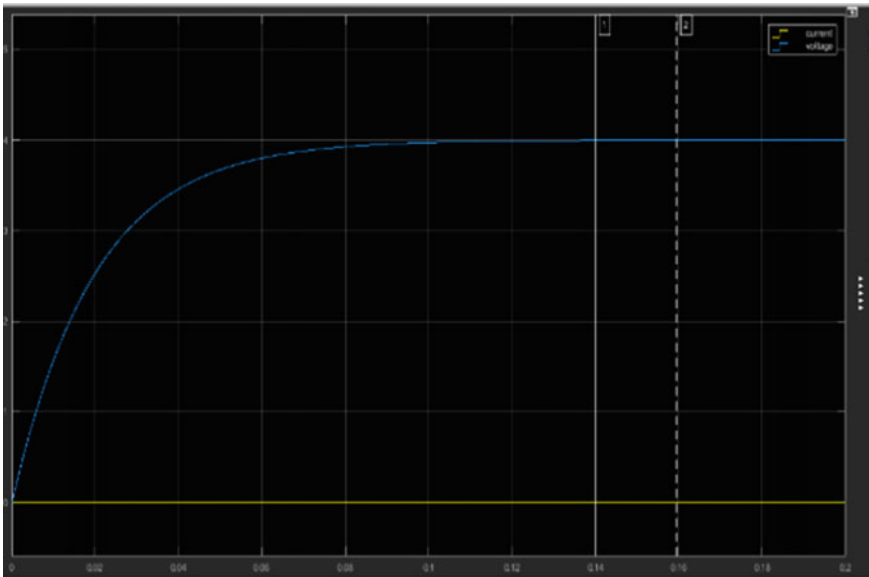
Fig. 9   Overall output for WPT

4 Future Work

More experimental analysis should be done with different background of cases in order to ensure the quality and effectiveness.

References

1. Gore VB, Gawali DH (2016) Wireless power transfer technology for medical applications
2. Peng K, Liu X, Huang P Study on the wireless energy supply system in the implantable cardiac pacemaker
3. Campi T, Cruciani S, Palandrani F, De Santis V, Hirata A, Feliziani M Wireless power transfer charging system for AIMDs and pacemakers
4. Lafrance A Who killed the rechargeable pacemaker? The Atlantic, 4 Feb 2014
5. Ali KK, Jarndal AH Remote-monitoring and wireless-powering system for pacemaker patients
6. Khan MU, Jafar A, Karimov KhS, Feroze S A proposed optimized solution for wireless power transfer using magnetic resonance coupling
7. Hu Y, Zhang Y, Xu C, Lin L,Snyder RL, Wang ZL Self-powered system with wireless data transmission
8. Vulfin V, Sayfan-Altman S, Ianconescu R Wireless power transfer for a pacemaker application
9. https://en.wikipedia.org/wiki/Wireless_power_transfer
10. Sung SH, Kim YS, Joe DJ, Mun BH, You BK, Keum DH, Hahn SK, Berggren M, Kim D, Lee KJ Flexible wireless powered drug delivery system for targeted administration on cerebral cortex
11. https://en.wikipedia.org/wiki/Artificial_cardiac_pacemaker
12. https://www.hopkinsmedicine.org/healthlibrary/test_procedures/cardiovascular/pacemaker_insertion_92, p 07980

Recent *t*-way Test Generation Strategies Based on Optimization Algorithms: An Orchestrated Survey

Ammar K. Alazzawi, Helmi Md Rais, Shuib Basri, Yazan A. Alsariera, Abdullahi Abubakar Imam, Saad Adnan Abed, Abdullateef Oluwagbemiga Balogun, and Ganesh Kumar

Abstract In software testing, test case generation is the most challenging activities and expensive tasks. Where has a considerable impact on the ability to produce a desired or intended result (i.e., quality and efficacy) of software testing. As a result, several researchers have developed a number of *t*-way test case generation strategies (where *t* points the interaction strength between parameters) due to the market demand to the various types of tests based on different approaches. This paper presents an orchestrated survey of the latest test case generation strategies such as Binary Black Hole (BBH), Sine Cosine Variable Strength (SCAVS), Combinatorial Testing Based Jaya Algorithm (CTJ), deterministic genetic multi-parameter-order (GAMIPOG) and Hybrid Artificial Bee Colony (HABC). This survey illustrates the strengths and weaknesses of each strategy, and indicates potential research studies in the field for future work.

Keywords Meta-heuristics · Combinatorial testing · Optimization algorithms · *t*-way testing · Software testing

1 Introduction

The researchers in the software testing have sophisticated a number of helpful techniques in order to avoid bugs and promote bug detection the last few decades. Although the sophisticated techniques are useful. However, the manufactured software produced has no bug-free guarantee [1]. Thus, the only procedure that can demonstrate the achieved quality and recognize the remaining problems, and risks is software testing.

A. K. Alazzawi (✉) · H. M. Rais · S. Basri · A. A. Imam · S. A. Abed · A. O. Balogun · G. Kumar
Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 32610 Perak, Malaysia
e-mail: ammar_16000020@utp.edu.my

Y. A. Alsariera
Department of Computer Science, Northern Border University, Arar 73222, Saudi Arabia

In spite of the exhaustive testing, is desirable, but often is impossible due to limitations on time and resources. As a result, a new sampling technique named *t*-way test case generation based on combinatorial testing (where *t* points the interaction strength between parameters) has begun to appear as an alternative to exhaustive testing to systematically reduce the test cases to a manageable one. For example, BTS [2–4], GS [5], IPOG families [6–8], PSTG [9] and ABCVS [10, 11]). In the last 20 years, several new *t*-way test case generation strategies have developed in order to find the most efficient strategies are able to produce optimal test cases for each configuration system (i.e., every single combination covered at most once a time).

The *t*-way test case generation strategies based on meta-heuristics have gained a huge deal of attention in software testing as part of a general interest in recent years. In spite of, is desirable, most of the current existing *t*-way test case generation strategies do not support the Input output relations, variable interaction strength, constraint, high interaction strength $t > 3$ and etc. This paper is an attempt to survey the current *t*-way test case generation strategies and collect the knowledge (i.e., the strengths and weaknesses) of some current strategies that published recently.

The rest of paper structured as follows. Sect. 2 describes the theoretical background of *t*-way testing. The *t*-way test case generation strategies based on meta-heuristics in Sect. 3. Finally, the conclusion given by Sect. 4.

2 Theoretical Background of *t*-way Testing

Every configuration system under test (CSUT) is consists of a different set of elements called parameters (*p*), and each parameter is associated with different values (*v*). Table 1 shows a simple software system to easily understand the interaction between parameters and the associated values (i.e., *t*-way testing). Where this software system consists of 3-parameters associated with 2-values each, 1-parameter associated with 3-values each and the interaction strength $t = 2$. This software system can be expressed as covering array CA (N; 2, 2^3 3^1).

The exhaustive test suite size in this software system is 24 test cases ($3 \times 2 \times 2 \times 2 = 24$) in order to test this system exhaustively as shown in Fig. 1. By executing 2-way testing, where interaction strength between parameter's $t = 2$. The possible combinations are 6 combinations between parameters (AB, AC, AD, BC, BD and CD) respectively. The optimal test suite size achieved is 6 test cases (See Fig. 1) that

Table 1 Simple software system

	Input parameters			
	A	B	C	D
System configurations	0	0	0	0
	1	1	1	1
				2
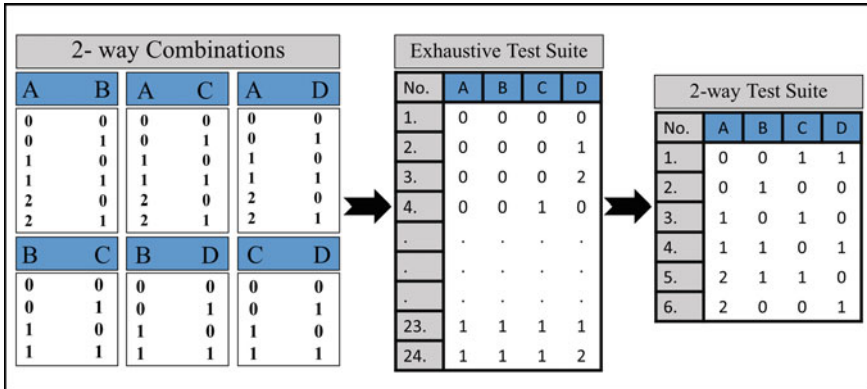
Fig. 1. 2-way test suite generation

able to cover each interaction between parameters. Therefore, *t*-way testing achieved a minimization from 24 test cases to 6 test case only (75% of total exhaustive testing).

3 Current *t*-way Test Case Generation Strategies

The generation of optimal test suite size can be considered as an optimization problem [12]. Test case generation strategies based meta-heuristics (i.e., optimization algorithm) always outstanding in this regard. Numerous of the existing meta-heuristics algorithms based strategies are using the one-test-at-a-time (OTAT) approach exclusively. In general, all the test case generation strategies based meta-heuristics generate non-deterministic solutions due to the randomization. In comparison, most studies are limited to small interaction intensity values for pure computation-based equivalents (e.g., $2 \leq t \leq 6$). In addition, there was also inadequate investigation of support for variable strength and constraints.

Binary Black Hole (BBH) is one of the most recently developed test case generation strategy based on meta-heuristic algorithms called Black Hole (BH) algorithm [13, 14]. The inspiration of BBH strategy came by the interaction of the phenomenon of the black hole and the star's behaviour together. In case of the star become close to the black hole, the black hole will swallow it. Where the algorithm will begins to generate a new solution (i.e., new star) randomly and will be searched again by the algorithm. BHH is an attractive strategy to solve the *t*-way testing problem because it is a tuning-free. BHH does not need multiple parameters for balancing between the global and local search like other generation strategies based on meta-heuristics. These features make it simple and easy for implementation. Up until now, BBH address small values of interaction strength (i.e., $t \leq 4$). In fact, BBH do not address variable strength interactions and do not provide the support for constraints.

Sine Cosine Variable Strength (SCAVS) strategy [15] based on Sine Cosine Algorithm (SCA) Algorithm for t-way variable test suite generation. SCA is meta-heuristic based on exploit the mathematical properties of both sine and cosine trigonometric functions. SCAVS is as any other algorithm starts first to produce random solutions. SCAVS supports the interaction strength up to $t = 6$ and addressing variable strength interaction. Although SCAVS addressing variable strength interaction, SCAVS does not support the constraints.

Hybrid Artificial Bee Colony (HABC) strategy [16–19] is an extended version based on Artificial Bee Colony (ABCVS) [10, 11] strategy in order to overcome the limitations of the original Artificial Bee Colony (ABC) algorithm. HABC based on hybridization of two meta-heuristic algorithms that can help the search capabilities of each by compensating for the limitations of one algorithm with the strength of others. HABC strategy based on merging the advantages of the original Artificial Bee Colony (ABC) algorithm with the advantages of an original Particle Swarm Optimization (PSO) algorithm. Unlike other strategies, HABC does support the high value of interaction strength (i.e., $t \leq 6$) and addresses the variable strength interactions. HABC also addresses the support for constraints through its variant, called PhABC [20, 21].

Unlike BBH, SCAVS and HABC that generate non-deterministic solutions, GAMIPOG is deterministic strategy. Deterministic Genetic Multi-Parameter-Order (GAMIPOG) strategy [22], based on Modified Input Parameter Order General (MIPOG) for test case generation [23]. GAMIPOG overcomes many desired challenges by combined them together in one strategy in order to take the benefits of each one such as meta-heuristics, one-parameter-at-a-time and one-test-at-a-time. In addition, GAMIPOG comparing to other existing strategies have less complexity, manageable test size and fast execution time. Where GAMIPOG works to find the global solution by breaking the problem into sub-problems merging all solution's together iterative steps. Thus far, GAMIPOG addresses high values of interaction strength (i.e., $t \leq 15$) and addresses the variable strength interactions. Furthermore, GAMIPOG does not provide the support for constraints.

Unlike all the above-mentioned strategies, CTJ is support input–output based relation (IOR) for test case generation [24]. Combinatorial Testing based on Jaya Algorithm (CTJ) strategy is one of the recently proposed strategies that support input–output based relation (IOR) for test case generation. Where CTJ adopts the Jaya algorithm as the main engine for generating the solutions. CTJ has two common control parameters: (1) maximum iteration and (2) population size. This feature makes CTJ performance to produce the solutions easily than other existing strategies. Thus far, CTJ addresses small values of interaction strength (i.e., $t \leq 3$) and addresses the input–output based relation (IOR). Furthermore, CTJ does not provide the support for constraints and variable strength interactions.

In conclusion, Table 2 gives a brief statement of the current t-way variable test generation. Referring to Table 2, all the existing strategies are non-deterministic due to the randomization (i.e., different test suite size in every implementation) except GAMIPOG strategy. There was a lack of investigation of constraints support, input output relations and as well as variable strength. GAMIPOG is the only strategy

Table 2 Current *t*-way test case generation strategies

Strategies	Non-deterministic	Deterministic	Interaction strength	Variable strength	Constraints	Input–output relations
SCAVS	✓	X	$t \leq 6$	✓	X	X
CTJ	✓	X	$t \leq 3$	X	X	✓
GAMIPOG	X	✓	$t \leq 15$	✓	X	X
BBH	✓	X	$t \leq 4$	X	X	X
HABC	✓	X	$t \leq 6$	✓	✓	X

support the addressing of high interaction strength ($t \leq 15$). Unlike GAMIPOG, SCAVS and HABC support the interaction strength up to $t \leq 6$. Unlike existing strategies, HABC only has the ability to support constraints. All strategies support the variable interaction strength, except CTJ and BBH. In contrast, CTJ addresses the input output relations unlike the other existing strategies.

4 Conclusions

This paper summarizes the current existing *t*-way test case generation strategies based on meta-heuristics and emphasises the strengths and weaknesses of each strategy. The software market is massive, and therefore, the test is needed to ensure compliances are varied; so new test ideas and test suite development are still needed. There are also numerous other meta-heuristics algorithms that still do have good potential and have not been adopted for *t*-way test suite generation.

References

1. Alsewari AA, Alazzawi AK, Rassem TH, Kabir MN, Homaid AAB, Alsariera YA, Tairan NM, Zamli KZ (2017) ABC algorithm for combinatorial testing problem. J Telecommun Electron Comput Eng (JTEC) 9:85–88
2. Alsariera YA, Zamli KZ (2015) A bat-inspired strategy for t-way interaction testing. Adv Sci Lett 21:2281–2284
3. Alsariera YA, Majid MA, Zamli KZ (2015) Adopting the bat-inspired algorithm for interaction testing. In: The 8th edition of annual conference for software testing, pp 14
4. Alsariera YA, Majid MA, Zamli KZ (2015) A bat-inspired strategy for pairwise testing. ARPN J Eng Appl Sci 10:8500–8506
5. Esfandyari S, Rafe V (2018) A tuned version of genetic algorithm for efficient test suite generation in interactive t-way testing strategy. Inf Softw Technol 94:165–185
6. Lei Y, Kacker R, Kuhn DR, Okun V, Lawrence J (2007) IPOG: a general strategy for t-way software testing. In: Engineering of computer-based systems, 2007. ECBS'07. 14th annual IEEE international conference and workshops on the, IEEE, pp 549–556
7. Lei Y, Kacker R, Kuhn DR, Okun V, Lawrence J (2008) IPOG/IPOG-D: efficient test generation for multi-way combinatorial testing. Softw Testing Verif Reliabil 18:125–148

8. Younis MI, Zamli KZ, Isa NM (2008) MIPOG-modification of the IPOG strategy for T-Way software testing. In: Proceeding of the distributed frameworks and applications (DFmA), Penang, Malaysia

9. Ahmed BS, Zamli KZ (2010) PSTG: a t-way strategy adopting particle swarm optimization. In: The fourth asia international on mathematical/analytical modelling and computer simulation (AMS), IEEE, pp 1–5

10. Alazzawi AK, Rais H, Basri S (2019) ABCVS: an artificial bee colony for generating variable t-way test sets. Int J Adv Comput Sci Appl 10:259–274

11. Alazzawi AK, Rais HM, Basri S (2018) Artificial bee colony algorithm for t-way test suite generation. In: 2018 4th international conference on computer and information sciences (ICCOINS), IEEE, pp 1–6

12. Floudas C, Pardalos P, Adjiman C, Esposito W, Gumus Z, Harding S, Klepeis J, Meyer C, Schweiger C (1999) Handbook of test problems in local and global optimization. Kluwer Academic Publishers, Dordrecht

13. Nsaif HN, Jawawi DNA (2020) Binary black hole-based optimization for t-way testing. In: IOP conference series: materials science and engineering, IOP Publishing, pp 012073

14. Al-Sammarraie HNN, Jawawi DN (2020) Multiple black hole inspired meta-heuristic searching optimization for combinatorial testing. IEEE Access 8:33406–33418

15. Altmemi JM, Othman R, Ahmad R (2021) SCAVS: implement sine cosine algorithm for generating variable t-way test suite. In: IOP conference series: materials science and engineering, IOP Publishing, pp 012011

16. Alazzawi AK, Rais HM, Basri S, Alsariera YA (2020) Pairwise test suite generation based on hybrid artificial bee colony algorithm. In: Advances in electronics engineering, Springer, pp 137–145

17. Alazzawi AK, Rais HM, Basri S (2019) HABC: hybrid artificial bee colony for generating variable T-Way test sets. J Eng Sci Technol 7:13

18. Alazzawi AK, Rais HM, Basri S (2019) Parameters tuning of hybrid artificial bee colony search based strategy for t-way testing. Int J Innov Technol Explor Eng (IJITEE) 8

19. Alazzawi AK, Rais HM, Basri S (2019) Hybrid artificial bee colony algorithm for t-way interaction test suite generation. In: Computer science on-line conference. Springer, pp 192–199

20. Alazzawi AK, Rais HM, Basri S, Alsariera YA, Balogun AO, Imam AA (2021) A hybrid artificial bee colony strategy for t-way test set generation with constraints support. J Phys: Conf Ser 042068. IOP Publishing

21. Alazzawi AK, Rais HM, Basri S, Alsariera YA (2019) PhABC: a hybrid artificial bee colony strategy for pairwise test suite generation with constraints support. In: 2019 IEEE student conference on research and development (SCOReD). IEEE, pp 106–111

22. Younis M (2020) Gamipog: a deterministic genetic multi-parameter-order strategy for the generation of variable strength covering arrays. J Eng Sci Technol 15:3142–3161

23. Younis MI, Zamli KZ (2011) MIPOG-an efficient t-way minimization strategy for combinatorial testing. Int J Comput Theory Eng 3:388

24. Younis MI, Alsewari ARA, Khang NY, Zamli KZ (2020) CTJ: input-output based relation combinatorial testing strategy using jaya algorithm. Baghdad Sci J 17

Research and Accomplishments in Applications of Non-negative Matrix Factorization

Phong Cao Nguyen, Cao Hong Nga, and Huong Hoang Luong

Abstract Since its introduction in the 1990s, non-negative matrix factorization (NMF) has captured a great amount of attention due to its capability and effectiveness in processing data in such a way that few earlier methods could perform, partly due to its non-negative constraint. This paper first briefly presents the basic NMF algorithm and concerns with the algorithm itself, then demonstrates its power with three applications in three different fields, namely face recognition in Computer Vision, distance prediction in Networking and molecular pattern discovery in Genetics. The paper ends with a quick look at other applications of NMF and recent developments that researchers have made.

Keywords Non-negative matrix factorization applications · Face recognition · Network distance prediction · Molecular pattern discovery

1 Introduction

Non-negative matrix factorization was popularized, to a large extent, by D. D. Lee and H. S. Seung when they introduced some algorithms and related proofs [1] after the idea was first described in [2] under the slightly different name *positive matrix factorization* (PMF). Since then, many applications of NMF have been found in different fields, from Informatics to Biology or Astronomy and beyond. This paper looks at some among that large collection of its uses. It starts with a scenario which reveals the need for such an efficient data analysis algorithm and touches briefly on the challenges accompanying the algorithm (Sect. 2). It then presents how NMF has been applied in Image Processing and Face Recognition, Network Distance Estimation, and Molecular Pattern Discovery by reviewing several typical works

P. C. Nguyen · H. H. Luong (✉)
FPT University, Can Tho, Vietnam
e-mail: huonghoangluong@gmail.com

C. H. Nga
National Central University, Taoyuan City, Taiwan

in the respective subjects in Sects. 3–5. Finally, it ends by describing how NMF is being applied in even more fields, and how research to enhance NMF's performance in current applications is being carried out in Sect. 6, which is followed by a brief conclusion.

2 Non-negative Matrix Factorization Overview

Methods for summarizing and analyzing mass datasets are needed in order to effectively utilize the huge amount of data available nowadays. For example, an online movie platform should be able to recommend new movies to users based on their past ratings, other users' ratings, similarities between movies and so on. Consider a simple case where users' ratings (m users on n movies) are stored in the $m \times n$ matrix X. The (i, j)-entry of X denotes the rating of user i towards movie j. Such a matrix could easily be large; one million users and 10,000 movies constitute 10 billion entries. What NMF does is, for a relatively small integer k, finding an $m \times k$ matrix W and a $k \times n$ matrix H such that every entry in W and H is non-negative and $X \approx WH$. The W and H matrices are usually called *basis* and *weight* matrices respectively because each column in X can be viewed as a linear combination of the rows in W with coefficients provided by columns in H [1] (see Fig. 1). The n movies are also said to be divided into k clusters.

The NMF problem itself is NP-hard [3], which means algorithms for exact NMF have exponential time complexity, thus too slow for real-world scenarios. Consequently, most practical algorithms are heuristic with acceptable accuracy. One of the most common algorithms is that presented by Lee and Seung in [1], on which many variants have been based. The core of the standard algorithm lies in the iterative updates of W and H [1]. The main steps of the algorithm are briefly listed below.

1. Define a function that measures the quality of the approximation, called a *cost function*. A simple measure is the Euclidean distance $E(A, B)$ between two $m \times n$ matrices A and B:
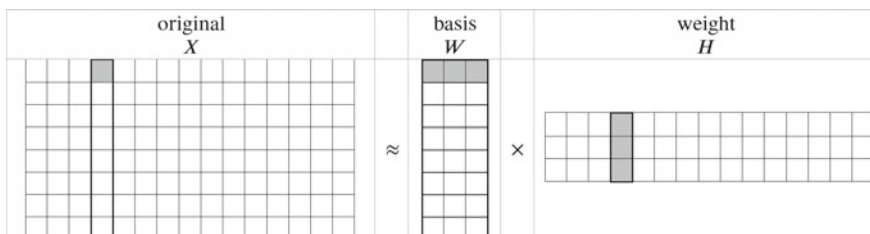
Fig. 1 Reconstructing original data using W and H

$$E(A, B) = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} \left(A_{ij} - B_{ij}\right)^2} \tag{1}$$

Alternatively, the *divergence* of A from B can be used:

$$D\left(A || B\right) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(A_{ij} \log \frac{A_{ij}}{B_{ij}} - A_{ij} + B_{ij}\right) \tag{2}$$

The goal is to minimize $E(X, WH)$ or $D(X || WH)$, depending on what function was chosen.

2. Initialize W_0 and H_0. The initialization should depend on the specific conditions of each problem.
3. At each iteration, update W and H according to the multiplicative rules.

(a) If the Euclidean distance was chosen, the (i, j)-entry of the matrices are updated as follows.

$$W_{ij} \leftarrow W_{ij} \frac{\left(XH^T\right)_{ij}}{\left(WHH^T\right)_{ij}} \tag{3}$$

$$H_{ij} \leftarrow H_{ij} \frac{\left(W^T X\right)_{ij}}{\left(W^T WH\right)_{ij}} \tag{4}$$

(b) If the divergence was chosen, the (i, j)-entry of the matrices are updated as follows.

$$W_{ij} \leftarrow W_{ij} \frac{\sum_k \frac{H_{jk} X_{ik}}{(WH)_{ik}}}{\sum_l H_{jl}} \tag{5}$$

$$H_{ij} \leftarrow H_{ij} \frac{\sum_k \frac{W_{ki} X_{kj}}{(WH)_{kj}}}{\sum_l H_{li}} \tag{6}$$

Iteration continues until the value of the chosen function falls within the pre-specified desired range.

The authors proved the convergence property of the algorithm, by which continuous iterations would converge to a locally optimal factorization [1].

The algorithm is not without its flaws, however. Certain concerns regarding the use of heuristic algorithms like this have been raised, including the non-uniqueness of W and H (multiple satisfactory pairs may exist, in which case initialization must be carefully determined) and the choice of rank k (an understanding of the underlying structure of the objects may be required). Research into these issues is still ongoing, along with the development of variants of the standard algorithm like graph regularized NMF [4] or hierachical NMF [5], just to name a few. Readers interested in such research are referred to [1, 3, 6] and the mentioned works.

3 Applications of Non-negative Matrix Factorization in Image Processing and Face Recognition

Image processing is one of the earliest applications of NMF. A collection of n images containing m pixels each is represented by the $m \times n$ matrix X. The jth image—the jth column of X—can be approximated as the linear combination of k basis images—k columns in W (see Fig. 2).

If the image is about a human face, the basis images are usually facial parts like the eyes, the nose or the mouth. An advantage of NMF becomes clear in this case: it expresses the face as a sum of facial parts. The non-negative restriction on the matrix entries is useful intuitively; it is hard to interpret what a negative weight for the mouth component means.

Lee and Seung applied this method to a collection of $n = 2429$ images containing $m = 19 \times 19$ pixels each [7]. They compared NMF with Principle Components Analysis (PCA) and Vector Quantization (VQ), which are holistic approaches; that is, in these methods, the basis images are whole faces rather than parts of a face. PCA approximates the original image with only one prototypical basis image, while VQ allows negative weights, which the authors deemed not intuitively meaningful [7]. These two algorithms were chosen for comparison primarily to highlight the effectiveness of the non-negative and part-based properties of NMF over the whole-object-based property of other existing algorithms.

The part-based property of NMF has also been utilized in face recognition. Guillamet and Vitriá compared NMF against PCA and two commercial techniques—FaceIt and Bayesian—in terms of their ability to recognize faces under different conditions [8]. FaceIt and Bayesian are considered leading techniques in computer vision, thus comparison with these two techniques aid the community in determining if NMF can replace these commercial techniques and work well in a production environment. The authors used the well-known AR face database [9]. The database consists of images of 116 people taken in two sessions two weeks apart. Features like facial expressions and external conditions were also present in the collection for each session (see Table 1) [9]. The images with neutral expression (labelled 01) were used as the training set.
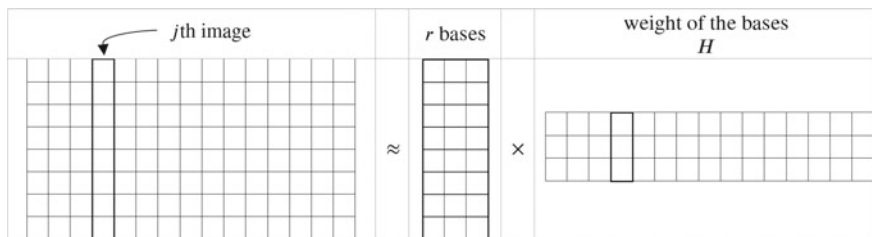
Fig. 2 NMF on a collection of images

Table 1 Features in each collection

Neutral	
01	Neutral expression
Facial expression	
02	Smile
03	Anger
04	Scream
Light	
05	Left light on
06	Right light on
07	Both lights on
Sunglasses	
08	Sunglasses
09	Sunglasses, left light on
10	Sunglasses, right light on
Scarf	
11	Scarf
12	Scarf, left light on
13	Scarf, right light on
14–26: second session	

For each of PCA and NMF, training was conducted with dimensions of 50, 100 and 150 to reveal, if any, the relationship between dimensionality and the results of face recognition. Initialization is effectively random. The experiments pointed out that NMF performed better than PCA at recognizing smiles and was comparable with FaceIt when it came to anger. However, scream proved too difficult for both NMF and PCA. For the illumination set, NMF, when used in a high dimensional space (150), even outperformed both FaceIt and Bayesian [8].

Under the effect of sunglasses, recognition rates for all approaches decreased remarkably, implying the importance of eyes in recognizing faces [8]. NMF did best when there was no light, but when lighting conditions affect the photos, it became less capable. Similar results were observed for the scarf set, suggesting NMF was good for a certain level of occlusions without another change in the scene [8].

NMF's ability to recognize face could be enhanced by adding various sets of constraints to the matrices W and H, enforcing these constraints by using different cost functions, which in turn produce variants of pure NMF. A number of such variants has been proposed, including local NMF (LNMF), Fisher NMF (FNMF) or PCA NMF (PNMF). For example, LNMF adds three constraints to pure NMF, which are (1) the number of bases should be minimized, which essentially means a basis component should not be decomposed any further, (2) bases must be made as orthogonal as possible, which is to minimize redundancy among bases and (3) only

bases with the most important information should be retained. LNMF consolidates these constraints into its divergence function (specific limits for i and j have been omitted for simplicity):

$$D(X||WH) = \sum_{i,j} \left(X_{ij} \log \frac{X_{ij}}{(WH)_{ij}} - X_{ij} + (WH)_{ij} \right) + \alpha \sum_{i,j} U_{ij} - \beta \sum_i Q_{ii} \tag{7}$$

where $\alpha, \beta > 0$ are constants, $U = W^T W$ and $Q = HH^T$.

The reader is referred to [10] for a detailed description of these variants. Other than the mentioned constraints, the core steps should remain more or less the same to what was described in Sect. 2. Experiments with different databases demonstrated the improved performance of FNMF and PNMF against NMF or LNMF in certain circumstances due to the ability of these methods to better localize basis images to the certain facial parts [10].

4 Applications of Non-negative Matrix Factorization in Network Distance Estimation

Network distances (including, but not limited to, round-trip time between hosts or network latency) have a great effect on the operation of distributed applications. A user who wants to download a file is able to choose the fastest server to fetch the file from if information about network distances is available. However, obtaining all pairwise distances for N hosts is expensive: it would take $O(N^2)$ measurements. Thus, an issue of interest is estimating unknown distances from $O(N)$ measurements and many studies have been dedicated to solving this problem. Early models are based on Euclidean distance, in which hosts are represented as coordinates in a d-dimensional systems. The distance between H_i and H_j represented by $(H_{i_1}, H_{i_2}, \ldots, H_{i_d})$ and $(H_{j_1}, H_{j_2}, \ldots, H_{j_d})$ respectively is

$$D_{ij} = \sqrt{\sum_{k=1}^{d} \left(H_{i_k} - H_{j_k} \right)^2} \tag{8}$$

Examples include Global Network Positioning (GNP), in which estimations are made by obtaining known distances from predetermined hosts called *landmarks* [11] and Vivaldi, one of the most widely-used distributed system [12]. Such systems have two common major drawbacks. First, the distances predicted are symmetric; that is, $D_{ij} = D_{ji}$. This is sometimes not true as some studies have indicated [13, 14]. Second, distances predicted this way have to observe the triangle inequality: $D_{ij} + D_{jk} \geq D_{ik}$. H. Zheng et al. have shown that triangle inequality violations are not rare in networks [15], and other studies have analyzed the effect they can have on the accuracy of Euclidean distance-based systems [16, 17].
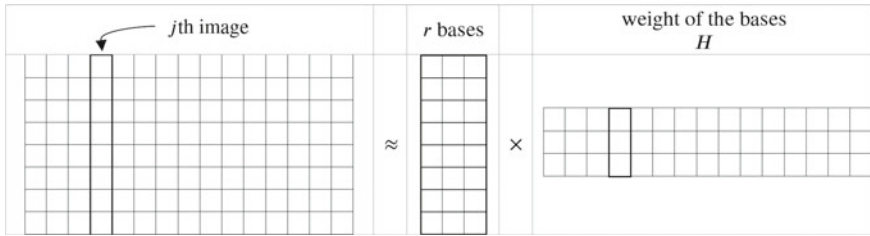
Fig. 3 Matrix factorization in estimating network distances

A system based on matrix factorization, called Internet Distance Estimation Service (IDES), was introduced in [18]. In IDES, $N \times d$ matrix W and $d \times N$ matrix H are found such that the $N \times N$ distance matrix D is approximated by WH. Then, the distance from host i to host j can be approximated as the dot product of the ith row of W and the jth column of H^T (see Fig. 3) [18]:

$$\widehat{D_{ij}} = \overrightarrow{W_i} \cdot \overrightarrow{H_j^T} \tag{9}$$

Under this model, distances do not have to observe the triangle inequality, and $\widehat{D_{ij}}$ is not necessarily equal to $\widehat{D_{ji}}$. Hence, it better predicts real-life distances compared to earlier Euclidean distance-based models. In the paper, the authors proposed two factorization algorithms—Single Value Decomposition (SVD) and NMF—and found that NMF had a major advantage over SVD: it could "fill" missing values in the distance matrix with a slight modification in the algorithm, while SVD had to "skip" hosts with missing values [18]. The predictions were closer to the actual values, but were still negatively affected by landmarks that often produced incorrect values. Chen et al. [19] introduced weight to the system—landmarks that provided correct values more often would play a larger part in calculating the prediction. They demonstrated a significant decrease in relative error on predictions of certain datasets when weight was added compared to the IDES model in [18, 19].

The NMF algorithms behind IDES and Phoenix both try to minimize the error function E, which is the Euclidean error function presented in Sect. 2 [18, 19]:

$$E = \sum_{i=1}^{N} \sum_{j=1}^{N} \left(D_{ij} - \widehat{D_{ji}} \right)^2 \tag{10}$$

Chai et al. showed that higher accuracy in predicting network distances can be achieved if the error function used is the divergence of D from $R = WH$, also mentioned in Sect. 2 [20]:

$$K\left(D || R \right) = \sum_{i=1}^{N} \sum_{j=1}^{N} \left(D_{ij} \log \frac{D_{ij}}{R_{ij}} - D_{ij} + R_{ij} \right) \tag{11}$$

This class of NMF is usually called NMF using KL divergence (KL-NMF), and it was also discussed in [1]. Apart from this difference, the steps in the algorithm of [20] are similar to those in [18], making use of landmarks or reference hosts. Chai et al. observed a 10–14.2% increase in prediction accuracy on four datasets when compared with the version of NMF used in IDES [20]. The authors attributed the improved performance to the ability of KL divergence to model network delay as having a Poisson distribution rather than the normal distribution that the Euclidean distance-based error function assumes [20].

5 Applications of Non-negative Matrix Factorization in Molecular Pattern Discovery in Genetics

Some genes in our body belong to the same group and contribute to an expression or a disease. Clustering a large number of genes into a small number of metagenes and analyzing the expression patterns of these metagenes are useful in discovering hidden or unclear biological structure or relationships between groups of genes and certain expressions. These clusters are comparable to facial parts in face recognition applications. Brunet et al. applied NMF on Leukemia, Medulloblastoma and Central Nervous System Tumors datasets to classify cancer subtypes [21]. This section describes the experiment with the Leukemia set.

N genes and their expression levels in M samples (in the Leukemia dataset [24], $M = 38$) are stored in the $N \times M$ matrix X. When an approximation $X \approx WH$ is found (W is an $N \times k$ matrix), each column in W defines a metagene that is a positive linear combination of the N genes, and each column in H represents an expression pattern of each sample as a positive linear combination of the metagenes. This factorization also divides M samples into k clusters; sample j belong to cluster i if the (i, j)-entry of H is the largest in column j (see Fig. 4) [21].

Acute leukemia is generally divided into two types—acute myelogenous leukemia (AML) and acute lymphoblastic leukemia (ALL), with the latter being further divided into two subtypes—T and B cells. The authors compared NMF with two other methods, namely hierarchical clustering (HC) and self-organizing maps (SOM). HC is essentially a process of clustering that involves merging the closest pair of clusters
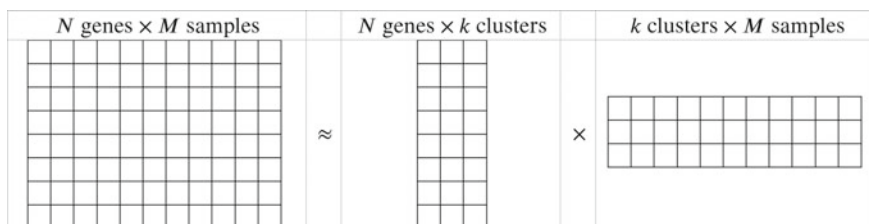

Fig. 4 NMF as a clustering method

at the previous step into a larger one until a strong cluster, which consists of all objects, are obtained [22]. The result is a tree-like structure of clustering, hence the name *hierarchical*. An SOM consists of nodes that are initially placed at random into k-dimensional space, whose positions are then adjusted by moving the points toward a random data point selected for that iteration. After the specified number of iterations (about 20,000–50,000), the points that are near each other are likely to form a related cluster [23]. Among the three, better methods would correctly classify the samples into the 2 or 3 abovementioned clusters (depending on the rank used in the algorithm) and provide more stable results (sample j belongs or does not belong to cluster j in almost all of the runs of the algorithm with different initial conditions). The findings were as follows [21].

HC

(1) HC split ALL-B samples into two subgroups (when they should have been together).

SOM

(2) While SOM could divide the samples into two clusters, such classification was not stable as the two groups could be either [AML] vs [ALL] or [AML + ALL-T] vs [ALL-B].
(3) SOM with $k = 3$ classes did not correctly divide the samples into the three known classes (ALL-T and ALL-B were mixed).

NMF

(4) NMF correctly split the samples into two clusters with remarkable stability regardless of initialization.
(5) When used with rank $k = 3$, NMF correctly distinguished ALL-T from ALL-B.
(6) Under higher ranks, NMF found a 4-cluster division quite robust while showing considerable dispersion when $k = 5$, implying that many clusters were not likely.

The authors concluded NMF had multiple strong points when compared with HC and SOM, and was better at discovering biological substructure [21]. sThis was consistent with the observation that HC imposes strict hierachical structure to objects, which may not always be the case, and at times needs subjective clustering definition, while SOM, though imposes only a partial structure on the objects, which is more appropriate, is unstable and may produce different clusterings with different initializations [21]. Y. Gao and G. Church later demonstrated the enhanced performance of sparse NMF, a variation, over classic NMF [25].

The non-uniqueness of W and H, however, may produce inconsistent results between various runs of NMF. Yang and Seoighe studied some post-processing methods to improve the stability of the algorithm [26]. They standardized W and H using 1-norm, 2-norm, 3-norm and maximum norm in addition to quantiles and sample standard deviation. Normalization maps the matrix W to a diagonal matrix D and updates the solution as follows: $W' = WD^{-1}$ and $H' = DH$. The authors also tried applying filters to the resulting matrices so that irrelevant genes were removed before running the algorithm again. Irrelevant genes were defined as those that had similar

values in the clusters, and thus it was not clear as to which metagenes they belonged to [26]. When performed on the Leukemia dataset, maximum norm remarkably made no errors even when there were two well-known outliers in the dataset [21]. In contrast, both outliers were incorrectly categorized when basic NMF was used. For most of the other datasets in the experiments, NMF with maximum norm was also singled out as the best classifier, and a filter was also proved to be able to boost performance in certain cases [26].

In medical, NMF is also used to detect new subtypes of diseases. Chalise et al. proposed combining NMF and similar network-based clustering (nNMF) across multiple data sources to recognize latent clustering structures in the data [31]. In this method, each consensus matrix constructed from each dataset represents the similarity between patients. These matrices are integrated into a robust consensus matrix which represents the network of patient similarities and the nNMF clusters. Since consensus matrices are calculated separately, they are not affected by the discrepancies in the distributions and scales of the datasets. Consensus matrices are built on a cross-validation approach that partially solves the overfitting and underfitting issues of unsupervised methods. This method is applied on both real and simulated data and proves the applicability of this method.

6   Other Applications and Recent Developments

Applications of NMF are not limited to the above three fields. Texts of hundreds of thousands of words are clustered into a few topics [27]. In astronomy, Ren et al. used NMF to extract images of astrophysical objects out of direct imaging data [28] and Berne et al. analyzed the emission of dust particles from spectro-imagery data [29]. New recommendation systems using NMF are being proposed [30]. Along with improvements in the algorithm for NMF itself, new applications in existing and new fields are certain to be discovered.

7   Conclusions and Future Works

NMF in three different fields—computer vision, networking and genetics, demonstrates the ability of NMF to both act as a predicting/estimating model and a clustering method. The results of the experiments indicate certain edges NMF has over existing approaches due to its constraint of allowing only non-negative entries in the matrices. It is certain that improved versions of these applications as well as new applications will emerge, thanks to the continuous effort researchers put into the field. As part of our future research, we will focus on a particular use of NMF, namely recommendation systems. Specifically, we will build new models by using session-based and streaming session-based in conjunction with the NMF and set up a new recommendation system using our models. In addition, we will carry out some research

to overcome the drawbacks of face recognition in some conditions like screaming faces, face with sunglasses, or face in different light conditions by combining NMF with deep neural networks.

References

1. Lee DD, Seung HS (2001) Algorithms for non-negative matrix factorization. Adv Neu Inf Proc Syst 14:556–562
2. Paatero P, Tapper U (1994) Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. Environmetrics 5(2):111–126. https://doi.org/10.1002/env.3170050203
3. Gillis N (2014) The why and how of nonnegative matrix factorization. In: Suykens JAK, Signoretto M, Argynou A (eds) Regularization, optimization, kernels, and support vector machines. Chapman and Hall/CRC, New York, NY, USA, pp 257–291
4. Cai D, He X, Han J, Huang TS (2011) Graph regularized nonnegative matrix factorization for data representation. IEEE Trans Patt Anal Mach Intell 33(8):1548–1560. https://doi.org/10.1109/TPAMI.2010.231
5. Li Y et al (2013) Hierarchical non-negative matrix factorization (hNMF): a tissue pattern differentiation method for glioblastoma multiforme diagnosis using MRSI. NMR in Biomed 26(3):307–319. https://doi.org/10.1002/nbm.2850
6. Berry MW, Browne M, Langville AN, Pauca VP, Plemmons RJ (2007) Algorithms and applications for approximate nonnegative matrix factorization. Comput Stat Data Anal 52(1):155–173. https://doi.org/10.1016/j.csda.2006.11.006
7. Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788–791. https://doi.org/10.1038/44565
8. Guillamet D, Vitria J (2002) Non-negative matrix factorization for face recognition. Catalonian Conf Artif Intell 336–344
9. Martínez AM, Benavente R (1998) The AR face database. Comput Vis Center Tech Rep #24:1998
10. Wang Y, Jia Y, Hu C, Turk M (2005) Non-negative matrix factorization framework for face recognition. Int J Patt Recogn Artif Intell 19(4):495–511. https://doi.org/10.1142/S0218001405004198
11. Ng TSE, Zhang H (2002) Predicting Internet network distance with coordinates-based approaches. In: Proceeding twenty-first annual joint conference of the IEEE computer and communications societies, New York, NY, USA, pp 170–179 vol 1. https://doi.org/10.1109/INFCOM.2002.1019258.
12. Dabek F, Cox R, Kaashoek F, Morris R (2004) Vivaldi: a decentralized network coordinate system. ACM SIGCOMM Comput Commun Rev 34(4):15–26. https://doi.org/10.1145/1030194.1015471
13. Paxson V (1997) End-to-end routing behavior in the Internet. IEEE/ACM Trans Netw 5(5):601–615. https://doi.org/10.1109/90.649563
14. Lakshminarayanan K, Padmanabhan VN (2003) Some findings on the network performance of broadband hosts. In: Proceeding 3rd ACM SIGCOMM Conference Internet Measurement, pp 45–50. https://doi.org/10.1145/948205.948212
15. Zheng H, Lua EK, Pias M, Griffin TG (2005) Internet routing policies and round-trip-times. In: International workshop on passive and active network measurement, pp 236–250. https://doi.org/10.1007/978-3-540-31966-5_19
16. Lee S, Zhang ZL, Sahu S, Saha D (2006) On suitability of euclidean embedding of Internet hosts. ACM SIGMETRICS Perform Eval Rev 34(1):157–168. https://doi.org/10.1145/1140103.1140296

17. Wang G, Zhang B, Ng TE (2007) Towards network triangle inequality violation aware distributed systems. In: Proceeding 7th ACM SIGCOMM conference internet measurement, pp 175–188. https://doi.org/10.1145/1298306.1298331

18. Mao Y, Saul LK, Smith JM (2006) IDES: an internet distance estimation service for large networks. IEEE J Select Areas Commun 24(12):2273–2284. https://doi.org/10.1109/JSAC.2006.884026

19. Chen Y et al (2011) Phoenix: a weight-based network coordinate system using matrix factorization. IEEE Trans Netw Serv Manage 8(4):334–347. https://doi.org/10.1109/TNSM.2011.110911.100079

20. Chai L, Luo X, Zhao F, Li M, Liu S (2017) Network coordinate system using non-negative matrix factorization based on KL divergence. In: 2017 19th international conference on advanced communication technology (ICACT), Bongpyeong, pp 193–198. https://doi.org/10.23919/ICACT.2017.7890082

21. Brunet JP, Tamayo P, Golub TR, Mesirov JP (2004) Metagenes and molecular pattern discovery using matrix factorization. Proc. Nat Acad Sci 101(12):4164–4169. https://doi.org/10.1073/pnas.0308531101

22. Johnson SC (1967) Hierarchical clustering schemes. Psychometrika 32(3):241–254. https://doi.org/10.1007/BF02289588

23. Tamayo P et al (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Nat Acad Sci 96(6):2907–2912. https://doi.org/10.1073/pnas.96.6.2907

24. Golub TR et al (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286(5439):531–537. https://doi.org/10.1126/science.286.5439.531

25. Gao Y, Church G (2005) Improving molecular cancer class discovery through sparse non-negative matrix factorization. Bioinformatics 21(21):3970–3975. https://doi.org/10.1093/bioinformatics/bti653

26. Yang H, Seoighe C (2016) Impact of the choice of normalization method on molecular cancer class discovery using nonnegative matrix factorization. In: PloS One 11(10):e0164880. https://doi.org/10.1371/journal.pone.0164880

27. Hassani A, Iranmanesh A, Mansouri N (2019) Text mining using nonnegative matrix factorization and latent semantic analysis. arXiv:1911.04705

28. Ren B, Pueyo L, Zhu GB, Debes J, Duchêne G (2018) Non-negative matrix factorization: robust extraction of extended structures. Astrophys J 852(2). https://doi.org/10.3847/1538-4357/aaa1f2

29. Berne O et al (2007) Analysis of the emission of very small dust particles from Spitzer spectroimagery data using blind signal separation methods. Astronomy Astrophys 469(2):575–586. https://doi.org/10.1051/0004-6361:20066282

30. Benzi K, Kalofolias V, Bresson X, Vandergheynst P (2016) Song recommendation with nonnegative matrix factorization and graph total variation. In: IEEE international conference acoustics, speech and signal processing (ICASSP). Shanghai 2439–2443. https://doi.org/10.1109/ICASSP.2016.7472115

31. Prabhakar C, Ni Y, Fridley BL (2020) Network-based integrative clustering of multiple types of genomic data using non-negative matrix factorization. Comput Biol Med. https://doi.org/10.1016/j.compbiomed.2020.103625

Feature Selection Using Correlation Matrix on Metagenomic Data with Pearson Enhancing Inflammatory Bowel Disease Prediction

Huong Hoang Luong, Trong Thanh Tran, Ngoc Van Nguyen, An Duc Le, Huyen Thi Thanh Nguyen, Khoi Dinh Nguyen, Nghi Cong Tran, and Hai Thanh Nguyen

Abstract The Fourth Industrial Revolution has brought up a vast amount of new innovative implementations. These can be used for numerous areas to make wealth and to improve human ways of living. Our point of view is to consider medical problems for enhancing prediction first. In this study, we like to rise a question about whether if we could enhancing Inflammatory Bowel Disease (IBD) prediction for early detect related sickness by feature selection on metagenomic data. Over the last few years, the prediction has been a challenge. Because of rare information and lacking data, the problem is not well considered enough. To bring back the subject, in this work, we propose a new way of enhancing Inflammatory Bowel Disease (IBD) prediction by using the Correlation Matrix with the Pearson on Metagenomic Data. Our implications have the purpose of finding out whether we could do predictions better using a specific amount of features selected by Pearson correlation coefficient. The result with the proposed method is quite promising, when we address some high correlation features out, the model can predict better comparing to randomly select features.

Keywords Feature selection · Pearson coefficient · Metagenomic data · Inflammatory bowel disease

H. T. Nguyen
College of Information and Communication Technologies (CICT),
Can Tho University, Can Tho, Vietnam
e-mail: nthai@cit.ctu.edu.vn

H. H. Luong (✉) · T. T. Tran · N. Van Nguyen · A. D. Le · H. T. T. Nguyen · K. D. Nguyen
Information Systems Department, FPT University, Can Tho, Vietnam
e-mail: huonghoangluong@gmail.com

N. C. Tran
National Central University, Taoyuan, Taiwan

1 Introduction

In recent years, the quantity of deaths related to the gastrointestinal tract is increasing. Inflammatory Bowel Disease (IBD) is not a doom disease but worth of mention because of its part to develop some deadly diseases. IBD is also one of the primary causes of colorectal cancer (CRC) [1]. IBD [2] is a term for two main categories including Crohn's disease (CD) [3] and Ulcerative colitis (UC) [4], which is characterized by chronic inflammation of the digestive tract. Digestive tract damages could be caused by prolonged inflammation which is also a major risk factor for the formation and development of malignancies. Generally, cancer [5–7] is a disease characterized by an uncontrolled nature division and the existence of abnormal cells. There are abnormal growth occurs in the colon or rectum, it is called colorectal cancer [8, 9].

There are many factors that result in the disease including lifestyle, genetics, and environmental surroundings. According to WHO statistics in 2018 [10], CRC is the second most deadly cancer and the third most commonly diagnosed cancer in the world. There are 1.80 million new cases and 862 000 deaths. In the United States, CRC is the second most common cancer diagnosis. The American Cancer Society's [11] estimate for 2020 is 147,950 new cases. It is expected to cause around 53,200 deaths by 2020. The dangers if the person has this disease: cancer cells only form and grow in the colon area in an early stage. Therefore, treatment is not too difficult. Moving into the second stage, the pathogenic cells attack other parts of the colon. This is a milestone that marks the development of cancer in the colon. If cancer cells are not controlled in time, they begin to work harder, spreading to lymph nodes. More lymph nodes that are attacked, the more severe your condition becomes, and the harder it is to control. Finally, when the disease enters a terminal stage, almost every organ of the patient is affected by the cancer cells. Treatment does not bring much hope for the patient. The risk of death is high. For prevention and early detection of the disease, an improved approach is required to cut back the number of individuals who are becoming this illness, further leading patients into a brand new healthy lifestyle.

Later to improve patients' experiences with the growth of personalized medicine. We see that patients are diagnosed with the disease in less time, assess the risks, evaluate the optimal treatment, and last but not least help them improve health while reducing expenses. On the other side, medical manufacturers also are provided a chance to develop agents that are targeted to patient groups that do not respond to medications as intended. New value assessments for personalized medicine products, together with return-on-investment models, will be required as these new strategies for pharmaceutical and diagnostic products emerge. For the direct side, hospitals, health care providers, and health plan sponsors, personalized medicine represent yet one more challenge in uncertain times. Innovation in provider and benefits management, together with clarity in regulatory and legal constructs are required, even as new national insurance reforms begin to emerge. The applications of PM are diagnosis and intervention, drug development and usage, respiratory proteomics, cancer

genomics, and population screening. And through that use metagenomic to research methods for personalized medicine. Metagenomics is the research of metagenomes that aims to obtain genetic material directly from samples within the real living environment. This field is included in environmental genetics, ecological genetics, or simply genetics generally. Due to its important role within the discovery of microbial diversity, metagenomics are often considered as a prism to assist us better understand the planet of small organisms, contributing to humanity's understanding of the entire living world. In medicine, personalized medicine could be a type of using information a couple of person's genes or proteins, from the data above wont to diagnose, prevent, or treat a disease that someone may have. For that individual field of study, aimed by the expansion of machine learning technology. We will combine medical data and machine learning by determining the concentration of bacteria that are characteristic of the chance of making tumors that cause sickness.

Some current applications of personalized medicine and machine learning technology are extremely helpful not only within the way of supporting doctors in early diagnosis but also for the sake of all stakeholders [12]. Machine learning has created an unlimited amount of applications in healthcare. These are a number of implementing machine learning in healthcare like diagnosis and identifying diseases, drug discovery, medical imaging diagnosis, etc. Personalized medicine is one in these applications which is referenced during this article [13].

During this particular article, we illustrate a way approaches for using the correlation matrix to boost Inflammatory Bowel disease prediction. Our goal is to represent a brand new efficient set of methods thanks to the trustworthy results of the prediction.

Our paper consists of 4 main sections. In the next Sect. 2, we present some highlights of related work. Afterward, the methodology Sect. 3 where we introduced to you the methods which we use on the topic later. The next Sect. 4 will be details about the Experiments. Finally, in the final Sect. 5, we conclude our paper and summarize the point of our experiments.

2 Related Work

Numerous research studies have applied Machine Learning to analyze and visualize metagenomic datasets. In particular, authors have done a study on the method to increase disease prediction based on Metagenomic using Unsupervised Binning Approaches. Their research aims to present one-dimensional (1D) using a binning and scaler approach. They also use Multi-layer Perceptron (MLP) and traditional artificial neural networks. In the paper, their purpose is to divide the bin into different which includes frequency, width, and proposed breaks. From there, the process of learning with MLP is to classify diseases with multi-fold (K-fold). The Binning classification method was implemented in the test with 10 bins. The implemented MLPs are the Rectified Linear Unit (ReLU) [14]. Additionally, the authors investigate many methods for visualizing features of metagenomic datasets relate to colorectal cancer. Their objective is to propose an efficient method to present data distribution visually

and use for classification. They proposed three methods including Image Generation and Models for Evaluation, Visualizations Based on Dimensionality Reduction Algorithms, and Visualization Based on Data Density. Image Generation and Models for Evaluation use Species Bins and Quantile Transformation Bins to generate the image, then CNN is used to evaluate the considered visualization and models. Visualizations Based on Dimensionality Reduction Algorithms apply dimensionality reduction algorithms such as t-SNE, Principal Component Analysis, Random Projection, Spectral Embedding, Isomap, Locally Linear Embedding, Multidimensional scaling and Non-negative matrix factorization. They are used to shaped high-dimensional data in 2D images for improving prediction results. Visualizations based on the Data Density illustrate the data based on the data density with Fill-up [17]. Besides, data visualization is a challenge for many areas. With measurement data, it is characterized by the size of the data that is difficult for humans to interpret. In diseases that use measurement data to predict, DL is often less effective than Classical Machine Learning in predicting CRC. Authors present an approach using Manifold Learning with t-SNE [15] and Spectral embedding to visualize digital data into images and using DL to improve performance in predicting CRC. Results come from 5 different areas used to combine visualization with Deep Learning to improve performance in diagnostics.

In the last few years, Metagenomics is a study that includes many genomes at the same time. This Metagenomics can be collected from a variety of environments. For example, Bacteria in the human intestine are estimated to have between 500 and more than 1000 species in the human intestine. However, investigating the disease remains a major challenge due to inconsistent results in disease prediction and disease complexity. High-dimensional researchers have difficulty understanding data. Many studies use 2D illustrations to interpret data and find patterns in the data. To explain metagenomic data, many studies devise methods to visualize data. Krona [16] allows to display data in order and use the zoom feature of multi-layered pie charts. The survey study shows that: Manifold Learning-based approach shows the abundance of bacteria that cause colon cancer through data samples, from which 2 out of algorithms are selected. The best is Spectral embedding (SE) and t-SNE [15] for further investigation in preparation for the filling method and based on the mean magnitude of the feature. Visualization based on the fill approach: arrange and fill the values of the features into a matrix in order of right to left and top to bottom. The author fills in using phylogenetic ordering and random sorting. Then, the proposed visuals are loaded into a DL algorithm for prediction. In short, this article uses Manifold Learning with 2 methods of embedding SE and t-SNE to put data into a Deep Learning algorithm to predict and then Fill-up [17, 18]. Moreover, authors show how to use the CNN network on data that has no original visual structure. The first is to map the data and define it to a 1D or 2D image. Based on that description they apply the CNN network to predict the disease. The collection of data in the medical field has created a revolution in medical research, application and biotechnology. These data are called omics data, they reflect aspects of the biology of the system. More and more data sets are available and models are using this information to make decisions. Today, the most successful techniques applied to this omics data set are

mainly Random Forest (RF) and sparse regression. In this paper, authors use DL directly onto 6 measurement datasets to reflect the abundance of bacterial species, their presence in the intestines of sick and healthy patients. This technology works well in image classification we focus on using CNN. The purpose of the paper is to propose an effective, concise presentation in images and demonstrate the DL techniques as an effective tool for the metagenomics prediction tasks. Methods include the Visualization of image-based representations. Its steps are 1 color is chosen and applied according to the abundance of distribution. The division can be done on a linear or logarithmic scale. The features are displayed as images in 1 of 2 ways, namely phylogenetic-sorting (PLG) or t-SNE. In short, this paper proposes the MET2IMG method to predict the patient's disease using measurement data. Authors use two main methods to construct synthetic images which are Fill-up [17] and t-SNE [19]. The Fill-up [17] used small and simple images. The T-SNE [19] feature is embedded in two-dimensional space using classic embedding in ML [20].

3 Methodology

3.1 Feature Selection Using Correlation Matrix with Pearson

In our study, we select features using the Correlation Matrix with Pearson [21]. It is a widely used method to classify and visually. We use the Correlation Matrix with Pearson which measures linear dependence between two variables [22]. Later we will be able to pick out some of the features that have a high correlation coefficient.

The datasets as mention in 4.1 Datasets Description, consisting of 6 sets of data. Each of the datasets includes the number of rows corresponding to be called, the subjects. These next columns contain the code of the bacteria and their value Metagenomic Data corresponding to the subjects. After implementing the Correlation Matrix with Pearson, we get a matrix $n \cdot n$, with n is the number of features. Matrix values are the Correlation Coefficient values range from -1.0 to $+1.0$ with -1.0 is a total negative correlation, 0.0 is no correlation and $+1.0$ is a total positive correlation. Regardless of the sign correlation, we only take absolute values. These absolute values will be considered to be greater than n to let the features to be selected. In the study, we set n to 0.1, 0.2 and 0.3.

3.2 Feature Selection Randomly for Later Comparison

For the comparison purpose, we used randomly select features to compare the Feature Selection using the Correlation Matrix with Pearson for the final result. We used the IBD datasets mentioned in the previous Sect. 3.1 as well. With this method, we based on the number of features that using Feature Selection using Correlation Matrix with

Pearson has filtered out to random. From there we will have the Feature Selection randomly correspond to Feature Selection using Correlation Matrix with Pearson in quantity terms. The later process is the same.

3.3 Predict Using K-Fold Cross Validation Method on Both 3.1 and 3.2

For both methods, randomly selection or feature selection using Correlation Matrix with Pearson, we apply K-Fold cross-validation [23] after finding out Feature Selection using Correlation Matrix with Pearson or randomly which one predict better. K-Fold cross-validation is a statistical technique to estimate efficiency of ML model to independent datasets. The goal of this method is to test the model's skill to evaluate how accurately the model's prediction result can generate with an unknown dataset with the purpose of avoiding overfitting [24] or selection bias [25]. Due to the amount of data is not abundant, we assigned K with values 5 which means divides datasets into 5 groups. In every group of data, the train data and the test data was divided by the algorithm so that we can evaluate the effectiveness of the model, and the remaining groups for training the model. Based on the metrics mention in 4.2.

4 The Experiments

4.1 Dataset Description

Data is originally published on 3 February 2016 [26] which is well-known research about Fungal microbiota dysbiosis in IBD. In this original research, the authors extract data from 235 patients with IBD and 38 healthy subjects (HS) was determined using clinical 16S and ITS2 sequencing [26]. Later on, a paper published on 6 September 2017 provided the data with classification on IBD datasets includes 6 sets of data [27]. In this paper, we use all of them as follows:

- HS_UCr - Healthy (HS) and Ulcerative Colitis (UC) in remission
- HS_iCDr - Healthy (HS) and ileal Crohn's disease (iCD) in remission
- HS_UCf - Healthy (HS) and Ulcerative Colitis (UC) in flare
- HS_iCDf - Healthy (HS) ileal Crohn's disease (iCD) in flare
- HS_CDr - Healthy (HS) Crohn's diseasen(CD) in remission
- HS_CDf - Healthy (HS) Crohn's diseasen(CD) in flare.

As the purpose of learning these data to predict whether a subject is determined as IBD patient. We will use one set for train UCr, and tests on iCDr, UCf, iCDf, CDr and CDf. Information about these sets shows in the Table 1.

Table 1 Datasets information detail

Information	Dataset					
	UCr	iCDr	UCf	iCDf	CDr	CDf
Total features	238	258	251	248	258	260
IBD patients	44	59	41	44	77	60
Healthy subject	38	38	38	38	38	38
Total subjects	82	97	79	82	115	98

4.2 Data Division and Scoring Metrics

In the previous Sect. 4.1, we have been explored 6 informative datasets. Further to the problem, we divide these datasets into two groups: one group for train and another group for tests. In the first group, we choose UCr for feature selection later on for the train. In the other group, five others set later on for tests. For evaluating the results we used three scoring metrics:

- Accuracy (ACC): ACC metrics results in range from 0.0 to 1.0. The value 1.0 indicates perfect prediction and value 0.0 means that prediction all false. Depending on the number of correct predictions $n_{correct}$ and total predictions n_{total} the equation accuracy be:

$$ACC = \frac{n_{correct}}{n_{total}}$$

- Matthews correlation coefficient (MCC): a widely used measurement in biomedical research [28]. MCC measures result in range -1.0 to $+1.0$. The value $+1.0$ shows that prediction is true one hundred percent, the value $+0.0$ means prediction is no better than random and -1.0 indicates prediction and observation has nothing to do with each other [29].
- Area Under Curve (AUC): another widely used measurement in biomedical research. AUC measures the entire area under the curve, values result in range from 0.0 to 1.0 [30].

4.3 Using Correlation Matrix with Pearson to Select Features

We describe three scenarios for selecting features using Correlation Matrix with Pearson on the dataset UCr. When all features have the absolute correlation coefficient greater than n, these features later will be used for the trained model and give out prediction.
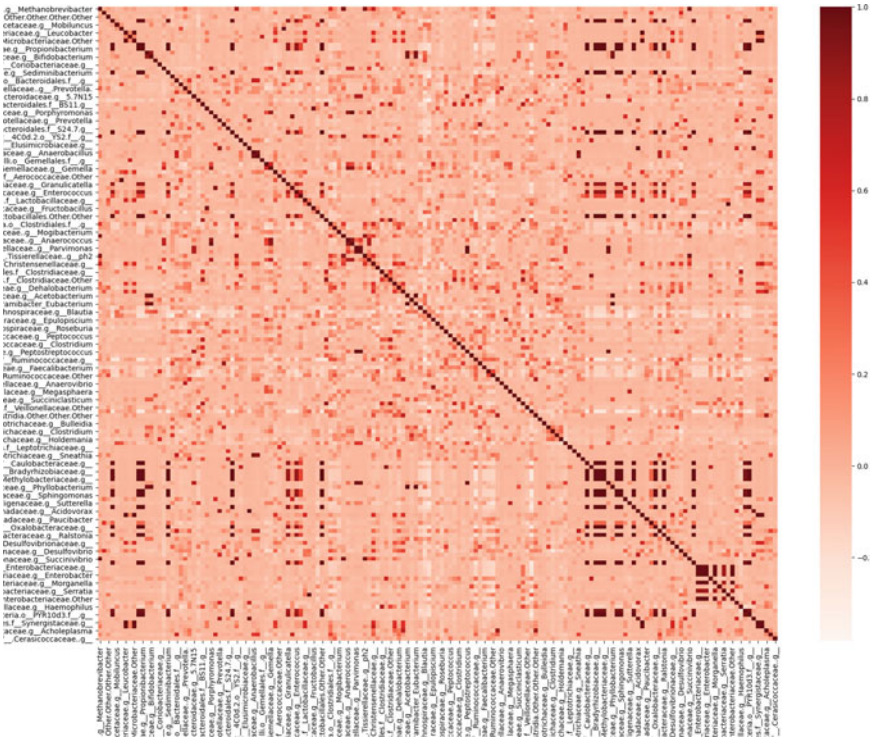
Fig. 1 Features with the absolute correlation coefficient >0.1

In the first scenario, we select all features which have $n > 0.1$ (159 features are selected). That means a feature has a positive correlation or negative correlation which is greater than 0.1 will be selected. The matrix is shown in Fig. 1

In the second scenario, we select all features which have $n > 0.2$ (30 features are selected). That means a feature has a positive correlation or negative correlation which is greater than 0.2 will be selected. The matrix is shown in Fig. 2.

In the final scenario, we select all features which have $n > 0.3$ (10 features are selected). That means a feature has a positive correlation or negative correlation which is greater than 0.3 will be selected. The matrix is shown in Fig. 3.

4.4 Prediction Result with ACC, MCC, AUC Metrics

Following Sect. 3, we got three sets of selected features from three scenarios : correlation >0.1 (159 features are selected), correlation >0.2 (30 features are selected) and correlation >0.3 (10 features are selected). We also take three sets of random features with the same size, correspondingly. Totally, we got 6 sets from UCr. These
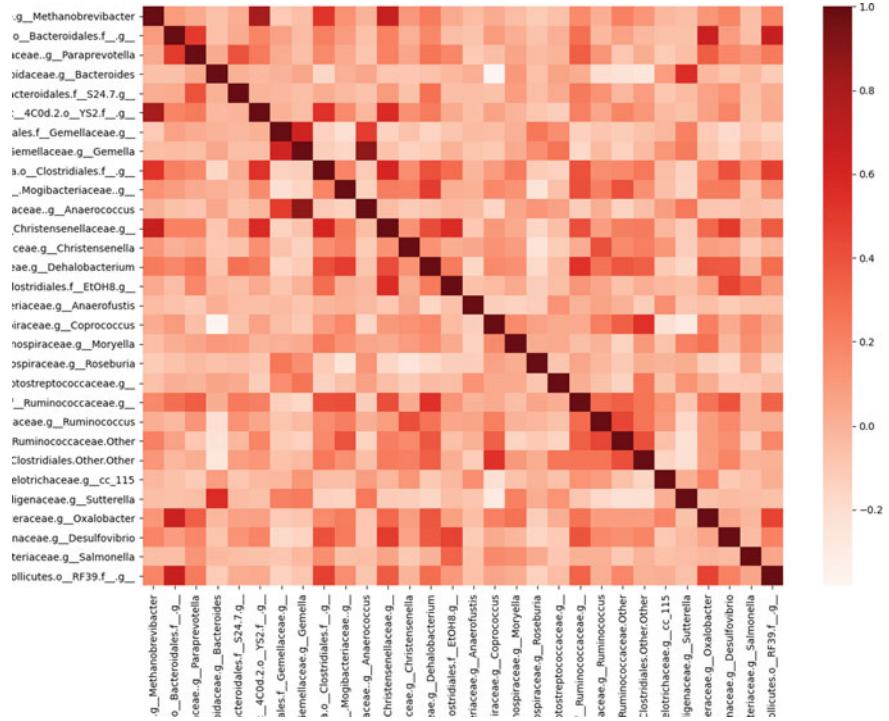
Fig. 2 Features with the absolute correlation coefficient >0.2

sets will be used for the trained model, then test on the other datasets (iCDr, UCf, iCDf, CDr and CDf).

After predict using K-fold cross-validation method we got results in three scoring metrics. These metrics include Accuracy (ACC), Matthews correlation coefficient (MCC) and Area Under Curve (AUC). The results shown in Tables 2 and 3.

5 Conclusion

As the prediction result has shown, we could state that using the Correlation Matrix on Metagenomic Data with Pearson for Feature Selection has an extraordinary improvement during predicting IBD patients compared to random features selection. This is made evident by the number that same size random features are not well suited for the train prediction model. The way of using the Correlation Matrix to take the high relate mutually feature to train is better for enhancing the prediction. Further research based on that premise could lead people to a healthy life with early detection of any kind of illness not just only IBD.
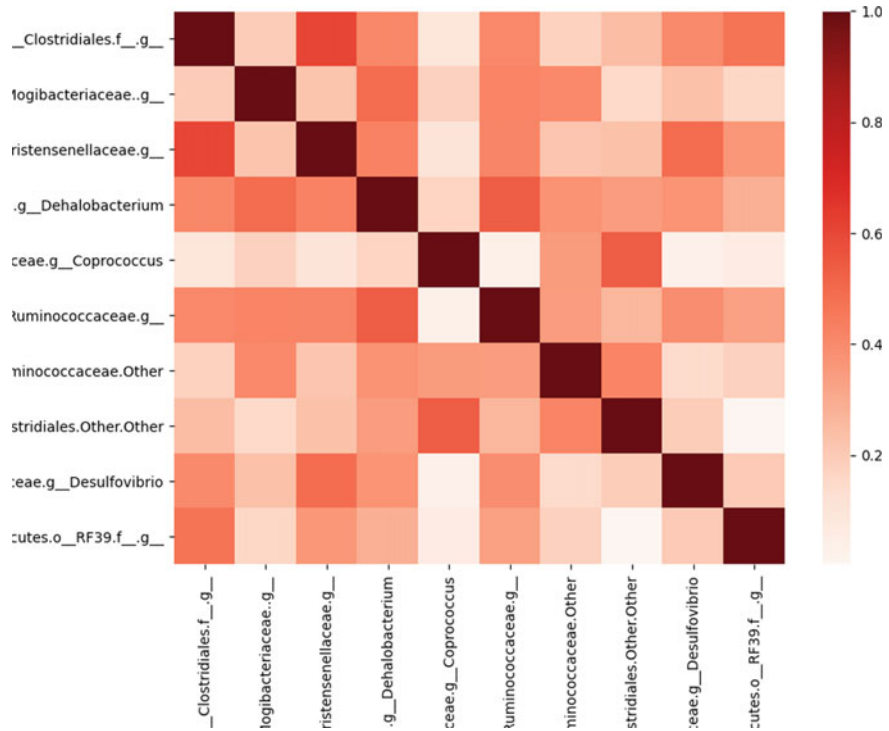
Fig. 3 Features with the absolute correlation coefficient >0.3

Table 2 Results with metrics corresponding ACC, MCC and AUC

Scenario	Features	Metrics	Dataset				
			iCDr	UCf	iCDf	CDr	CDf
corr > 0.1	159	ACC	0.95258	0.97975	0.99268	0.93750	1.00000
		MCC	0.90300	0.95957	0.98543	0.87500	0.96875
		AUC	0.95540	0.97972	0.99210	0.96875	1.00000
Random	159	ACC	0.90928	0.92911	0.94390	0.87500	0.93750
		MCC	0.82009	0.85999	0.88830	0.78125	0.87500
		AUC	0.91699	0.92978	0.94377	0.90625	0.93750
corr > 0.2	30	ACC	0.99381	0.99747	1.00000	1.00000	1.00000
		MCC	0.98734	0.99499	1.00000	1.00000	1.00000
		AUC	0.99491	0.99756	1.00000	1.00000	1.00000
Random	30	ACC	0.87422	0.90886	0.84375	0.78125	0.87500
		MCC	0.75353	0.82036	0.71875	0.59375	0.75000
		AUC	0.88443	0.90969	0.87500	0.81250	0.87500
corr > 0.3	10	ACC	1.00000	1.00000	1.00000	1.00000	1.00000
		MCC	1.00000	1.00000	1.00000	1.00000	1.00000
		AUC	1.00000	1.00000	1.00000	1.00000	1.00000
Random	10	ACC	0.72371	0.80253	0.87500	0.75000	0.87500
		MCC	0.50249	0.61778	0.78125	0.56250	0.75000
		AUC	0.75227	0.80552	0.87500	0.78125	0.87500

Table 3 Top 10 features with highest correlation score

Feature name	Score
k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__.g__	0.46782
k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__ .Mogibacteriaceae..g__	0.45053
k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__ Christensenellaceae.g__	0.38269
k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__ Dehalobacteriaceae.g__ Dehalobacterium	0.55617
k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__ Lachnospiraceae.g__Coprococcus	0.40403
k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__ Ruminococcaceae.g__	0.34405
k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__ Ruminococcaceae.Other	0.38699
k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.Other. Other	0.31592
k__Bacteria.p__Proteobacteria.c__Deltaproteobacteria.o__ Desulfovibrionales.f__ Desulfovibrionaceae.g__Desulfovibrio	0.41116
k__Bacteria.p__Tenericutes.c__Mollicutes.o__RF39.f__.g__	0.30068

References

1. Kim ER, Chang DK (2014) Colorectal cancer in inflammatory bowel disease: the risk, pathogenesis, prevention and diagnosis. World J Gastroenterol 20(29):9872–9881. https://doi.org/10.3748/wjg.v20.i29.9872
2. Centers for Disease Control and Prevention: Inflammatory bowel disease (IBD), from https://www.cdc.gov/ibd/what-is-IBD.htm
3. NIH-U.S. National Library of Medicine: Crohn's Disease, from https://medlineplus.gov/crohnsdisease.html
4. National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK): Ulcerative Colitis, from https://www.niddk.nih.gov/health-information/digestive-diseases/ulcerative-colitis
5. NHS: Overview Cancer, from https://www.nhs.uk/conditions/cancer/
6. NIH-National Cancer Institute: Understanding Cancer, from https://www.cancer.gov/about-cancer/understanding/what-is-cancer
7. American Cancer Society: What Is Cancer? from https://www.cancer.org/cancer/cancer-basics/what-is-cancer.html
8. (ASCRS) American Society of Colon & Rectal Surgeons: The Colon: What it is, What it Does and Why it is Important: Overview Cancer, from https://fascrs.org/patients/diseases-and-conditions/a-z/the-colon-what-it-is,-what-it-does
9. Innerbody: Rectum, from https://www.innerbody.com/image_digeov/dige14-new3.html
10. World Health Organization: Cancer. Retrieved September 28, 2020, from https://www.who.int/news-room/fact-sheets/detail/cancer
11. American Cancer Society: Key Statistics for Colorectal Cancer. Retrieved September 28, 2020, from https://www.cancer.org/cancer/colon-rectal-cancer/about/key-statistics.html
12. Vogenberg F, Isaacson Barash C, Pursel M (2010) Personalized medicine: Part 1: Evolution and development into theranostics. Retrieved September 27, 2020, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2957753/
13. Top 10 Applications of Machine Learning in Healthcare - FWS. (n.d.). Retrieved September 27, 2020, from https://www.flatworldsolutions.com/healthcare/articles/top-10-applications-of-machine-learning-in-healthcare.php
14. Nguyen TH, Zucker J (2019) Enhancing metagenome-based disease prediction by unsupervised binning approaches. In: 2019 11th international conference on knowledge and systems engineering (KSE), da nang, Vietnam, 2019, pp 1–5. https://doi.org/10.1109/KSE.2019.8919295

15. Nguyen TH, Nguyen TN (2019) Disease prediction using metagenomic data visualizations based on manifold learning and convolutional neural network. Lecture Notes in Computer Science, vol 11814. Springer, Cham. https://doi.org/10.1007/978-3-030-35653-8_9
16. O ndov BD, Bergman NH, Phillippy AM (2011) Interactive metagenomic visualization in a web browser. BMC Bioinform. 12:385. https://doi.org/10.1186/1471-2105-12-385. (ISSN:1471-2105)
17. Nguyen TH et al (2018) Disease classification in metagenomics with 2D embeddings and deep learning. In: The annual French conference in machine learning (CAp 2018). France: Rouen; June 2018. arXiv: 1806.09046
18. Thanh-Hai N, Thai-Nghe N (2020) Diagnosis approaches for colorectal cancer using manifold learning and deep learning. SN COMPUT. SCI. 1:281
19. Laurens van der Maaten GH (2008) Visualizing data using t-sne. J Mach Learn Res 9:8
20. Nguyen T, Chevaleyre Y, Prifti E, Sokolovska N, Zucker J (2017) Deep learning for metagenomic data: using 2D embeddings and convolutional neural networks. arXiv: 1712.00244
21. Benesty J, Chen J, Huang Y, Cohen I (2009) Pearson correlation coefficient. In: Noise reduction in speech processing. Springer Topics in Signal Processing, vol 2. Springer, Berlin, Heidelberg. from https://doi.org/10.1007/978-3-642-00296-0_5
22. Correlation Test Between Two Variables in R. (n.d.). Retrieved October 13, 2020, from http://www.sthda.com/english/wiki/correlation-test-between-two-variables-in-r
23. The 'K' in K-fold cross-validation: davide anguita, Luca Ghelardoni, Alessandro Ghio, Luca Oneto and Sandro Ridella https://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2012-62.pdf
24. Overfitting (2020) Retrieved 13 October 2020, from https://en.wikipedia.org/wiki/Overfitting
25. Selection bias (2020) Retrieved 13 October 2020, from https://en.wikipedia.org/wiki/Selection_bias
26. Sokol H, Leducq V, Aschard H et al (2017) Gut 66:1039–1048
27. Fioravanti D et al (2018) Phylogenetic convolutional neural networks in metagenomics. BMC Bioinformatics 19.S2 (2018): n. pag. Crossref. Web
28. Boughorbel S, Jarray F, El-Anbari M (2017) Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PLoS ONE 12(6): e0177678. https://doi.org/10.1371/journal.pone.0177678
29. En.wikipedia.org (2020) Matthews correlation coefficient. Retrieved October 12, 2020, https://en.wikipedia.org/wiki/Matthews_correlation_coefficient
30. Huang J, Ling CX (2005) Using AUC and accuracy in evaluating learning algorithms. IEEE Trans Knowl Data Eng 17(3):299–310. https://doi.org/10.1109/TKDE.2005.50