Lecture Notes on Data Engineering and Communications Technologies 53

V. Suma
Noureddine Bouhmala
Haoxiang Wang *Editors*

Evolutionary Computing and Mobile Sustainable Networks

Proceedings of ICECMSN 2020

Springer

Lecture Notes on Data Engineering and Communications Technologies

Volume 53

The aim of the book series is to present cutting edge engineering approaches to data technologies and communications. It will publish latest advances on the engineering task of building and deploying distributed, scalable and reliable data infrastructures and communication systems.

The series will have a prominent applied focus on data technologies and communications with aim to promote the bridging from fundamental research on data science and networking to data engineering and communications that lead to industry products, business knowledge and standardisation.

**** Indexing: The books of this series are submitted to SCOPUS, ISI Proceedings, MetaPress, Springerlink and DBLP ****

More information about this series at http://www.springer.com/series/15362

V. Suma · Noureddine Bouhmala ·
Haoxiang Wang
Editors

Evolutionary Computing and Mobile Sustainable Networks

Proceedings of ICECMSN 2020

Springer

Editors

V. Suma
Research and Industry Incubation Center,
Department of Information Science
and Engineering
Dayananda Sagar College of Engineering
Bangalore, India

Noureddine Bouhmala
Department of Technology
and Maritime Innovation
University of Southeast
Horten, Norway

Haoxiang Wang
Go Perception Laboratory
Cornell University
Ithaca, NY, USA

ISSN 2367-4512 ISSN 2367-4520 (electronic)
Lecture Notes on Data Engineering and Communications Technologies
ISBN 978-981-15-5257-1 ISBN 978-981-15-5258-8 (eBook)
https://doi.org/10.1007/978-981-15-5258-8

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

The 2020 ICECMSN conference is solely dedicated to all the editors, reviewers, and authors of the conference event.

Foreword

I extend my warm welcome in inviting you all to the proceedings of the International Conference on Evolutionary Computing and Mobile Sustainable Networks [ICECMSN 2020] organized at Sir M. Visvesvaraya Institute of Technology, on 20–21 February 2020.

The theme of the conference event is "Emerging advances in Sustainable Mobile Networks and Computational Intelligence", topics that are quickly gaining research attention from both academia and industries due to the relevance of maintaining sustainability and enhancing intelligence in smart mobile networks. The already established track record of computational intelligence models and sustainable mobile networks seems to be very functional and reliable, where it mandates the need for further exploration in this research area. This makes the ICECMSN 2020 an excellent forum for exploring innovative research ideas in the smart and intelligent networks domain.

We would like to extend our sincere gratitude to Organizing Chair Dr. V. R. Manjunath, Principal, SIR MVIT, Bangalore, India for his motivation and support to organize the conference in a successful manner. We extend our hearty thanks to Keynote Speakers Dr. Manu Malek, Editor in Chief, Elsevier CEE and Former Professor, Stevens Institute of Technology, USA and Sri B. S. Bindumadhava Scientist G & Senior Director, Centre for Development of Advanced Computing, Bengaluru, India for their valuable thoughts and discussion.

The entire success of the ICECMSN 2020 conference event depends on the research talents and efforts of the authors in the intelligent mobile networks and computer science domains, who have contributed their submissions on almost all the facets of the conference theme. An extensive appreciation is also deserved for all the conference program and review committee members who have invested their valuable time and professional expertise in assessing research papers from multiple domains by maintaining the quality standards for this conference. We extensively thank Springer for their guidance before and after the conference event.

Conference Chair
Dr. Manjula Sanjay Koti
Professor and HOD
Department of MCA
SIR MVIT, Bangalore, India

Preface

It is our pleasure to welcome you to the International Conference on Evolutionary Computing and Mobile Sustainable Networks [ICECMSN 2020] in Bangalore, India. The major goal of this conference is to bring together the academicians, researchers and industrialists under a single roof to share and exchange their research experience and results on various aspects of mobile sustainable networks and computational intelligence research and discuss about the real-time challenges and solutions adopted for it.

ICECMSN 2020 has received ample submissions of about 398 papers from both academia and industrial tracks and based on the selection of conference review committee and advisory committee members, a total of 90 papers appeared in the conference proceedings of ICECMSN 2020. It is to be noted that, all the papers regardless of their allotted tracks has extensively received at least 3 reviews from the research experts.

We hope the readers will have a productive, satisfying and informative experience from the research works gathered from all over the world. Nevertheless, this proceedings will provide a written record of a synergy of research works that exists in communication networks communities and provides significant framework for a new and futuristic research interactions. Moreover, this proceedings will pave way for the applications of computational intelligence in Mobile Sustainable Networks [MSN].

Bangalore, India — Prof. Dr. V. Suma
Horten, Norway — Dr. Noureddine Bouhmala
Ithaca, USA — Dr. Haoxiang Wang

Acknowledgments

We ICECMSN 2020 would like to extend our sincere thanks to all who have helped in making this conference event a great success. We are much pleased in thanking our educational institution Sir M. Visvesvaraya Institute of Technology, Bangalore, India for their pervasive support and effective help during the conference.

The extended support of the conference committee members before and during the conference event has helped to tackle many challenging tasks in a smooth and efficient way, where it has significantly contributed to excel the quality of the conference. Our special thanks belongs to all the conference reviewers, who played an indispensable role in providing technical and semantic reviewing assistance to all the research manuscripts received for the conference. We are thankful for their help in guiding us to select the state-of-the-art high-quality papers that deserves the publication under this conference. We also wish to thank all our faculty members and staffs for their technical and non-technical contribution for maintaining the conference participants' contentment.

The conference organizers are particularly grateful for all the authors, who have contributed their research ideas spanning over many active and emerging research domains. Our very special thanks will go exceptionally to all the conference delegates for their active participation in the conference event.

At last, the organizers would like to gladly acknowledge the local organizing committee, who ensured that all the formal steps of the conference event has been completed in an effortless way.

Contents

About the Editors

Dr. V. Suma holds a B.E. in Information Science and Technology, M.S. in Software Systems and Ph.D. in Computer Science and Engineering. Currently, she is working as Dean of the Research and Industry Incubation Centre, and a Professor at the Department of Information Science and Engineering, Dayananda Sagar College of Engineering, Bangalore, India. She has more than 17 years of teaching experience and has published over 180 papers, including research articles published in leading international journals, such as ACM, ASQ, Crosstalk, IET Software, and journals published by MIT and Dartmouth College in the USA. Her research has also been published on NASA, UNI Trier, Microsoft, CERN, IEEE, ACM and Springer portals.

Dr. Noureddine Bouhmala completed his Candidatus Magisterii in Computer Science at the University of Bergen, Norway in 1992, and completed his Ph.D. in Computer Science at the University of Neuchâtel, Switzerland in 1998 (funded by the Swiss National Science Foundation). Currently, he is working as an Associate Professor at the Faculty of Technology and Maritime Innovation at the University SouthEast, and the Faculty of Engineering and Sciences at Agder University, both in Norway. He has more than 25 years of teaching and research experience, and has published numerous research papers in international journals and international conferences. He is also an editorial board member for various respected international journals. His areas of interest include artificial intelligence, machine learning, optimization algorithms, parallel computing, data mining, autonomic and safety-critical systems.

Dr. Haoxiang Wang is currently the Director and lead executive faculty member of GoPerception Laboratory, New York, USA. His research interests include multimedia information processing, pattern recognition and machine learning, remote sensing image processing and data-driven business intelligence. He has co-authored over 60 papers in these fields in journals such as MTAP, Cluster Computing, IEEE TII, Communications Magazine, Computers & Electrical Engineering, Computers, Environment and Urban Systems, Optik, Sustainable

Computing: Informatics and Systems, Journal of Computational Science, Pattern Recognition Letters, Information Sciences, Computers in Industry, Future Generation Computer Systems, International Journal of Computers and Applications, and at conference such as IEEE SMC, ICPR, ICTAI, ICICI, CCIS, ICACI. He is a guest editor for IEEE Transactions on Industrial Informatics, IEEE Consumer Electronics Magazine, Multimedia Tools and Applications, MDPI Sustainability, International Journal of Information and Computer Security, Journal of Medical Imaging and Health Informatics, Concurrency and Computation: Practice and Experience.

Optimal Resource Sharing Amongst Device-to-Device Communication Using Particle Swarm Algorithm

H. M. Nethravathi and S. Akhila

Abstract Device-to-Device Communication (D2D) has been described as one of the important innovations in the development of 5G networks. This form of D2D networking has its own benefits of improved network capacity and reduced power usage, making it as a primary candidate for the upcoming 5G cellular networks. In this work, a single cell scenario with one base station and multiple cell users and D2D pairs are considered. As sharing causes performance degradation due to interference between CUs (cellular users) and D2D pairs, a permutation optimization strategy based on Particle swarm optimization (PSO) has been proposed to optimize resource sharing between cellular users and D2D pairs. This technique is found to maximize system performance through better resource sharing.

Keywords 5G · D2D · 3GPP · PSO

1 Introduction

Worldwide, the number of users and data traffic is increasing every day. It overloads the base stations resulting in transmission delay, latency and low data speed. In such a situation, device-to-device (D2D) connectivity is a promising candidate for fifth-generation cellular communication. In D2D communication devices transmit or receive data without assistance from the base station. This reduces the base station overload thereby increasing the overall system throughput. Optimal allocation of resources for D2D interaction has become a very important research field, with researchers showing a great deal of interest in this area. Each user will be able to operate as a conventional cellular user or a D2D user depending on their chosen

H. M. Nethravathi (✉) · S. Akhila
BMSCE, Bengaluru, India
e-mail: nethrahm@gmail.com

S. Akhila
e-mail: akhilas.ece@bmsce.ac.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_1

strategy of need and mode. Based on their model, the resources will be allocated to the associated user so as to optimize the device throughput.

D2D communication has been considered as a powerful technology to provide improved quality of wireless services. Due to the limitations in various technologies, such as power restrictions, high delays due to network congestion and adoption of new services, new techniques have been developed to replace the prevailing technology i.e., Long Term Evolution (LTE), D2D. Such approaches tackle the limitations of existing networks and meet the new needs of end-users and of the operators [1].

The advantages that these techniques would bring to an end-user would be energy savings, efficiency gains or new nearby services that can save the connection to the base station. It also provides benefits from the perspective of network operators, such as increasing the coverage area, increasing spectrum utilization or being able to meet the demand of a larger number of linked terminals in the future at the same time. 3GPP considers the use of Long Term Evolution (LTE-Direct) and IEEE 802.11 (Wi-Fi Direct) for D2D communications [2].

To achieve the expected benefits of D2D communications, the technical challenges that may arise from the difficult conditions of transmission between mobile devices needs to be addressed. In addition, D2D communications can be highly inefficient under conditions of uncertainty and low quality of connections [3]. With the use of link adaptation and Power Control techniques, this form of inefficiency can be tackled.

Aranit et al. [4] provides a review on LTE to assess its capacity to support ITS and vehicular applications. The analysis conducted qualitatively captures the main features, strengths and weaknesses of the under-development standard guidelines and solutions. In [5], the energy minimization problem for D2D communication underlying a multi-cell system has been considered to maximize throughput.

Multiantenna methods [6, 7] have also been implemented into the underlying D2D communication to eliminate the interference between D2D and cellular users. D2D communication as a cellular network underlay can share resources orthogonally or non-orthogonally with cellular users. In the orthogonal case, D2D users are allocated dedicated resources. Though simple to implement, they are unable to exploit the full potential of D2D communication to increase spectral efficiency. This has been addressed in nonorthogonal resource sharing methods. Feng et al. [8] address both D2D and cellular users' QoS specifications while optimizing the sum frequency. D2D applications should share CU's uplink resources since it is simple for the base station (BS) to manage interference problems caused by underused uplink channels.

Liu et al. [9] investigated the power control for full-duplex D2D cellular network communications. In this power control problem was formulated by maximizing the achievable sum-rate of the full-duplex D2D connection while meeting the cellular connection's minimum rate required under the cellular users and D2D users' maximum transmission power constraint.

Dynamic resource allocation is studied in [10,11, 12], where all subchannels can be used by the D2D pairs. Nevertheless, the adjacent D2D pairs will inevitably suffer extreme interference with each other.

Social and physical attributes based content sharing approach with D2MD cluster formation approach is described in [13] for 5G cellular networks, enabling traffic offloading from base stations to direct transmitting devices and enhancing power.

Interference due to resource sharing reduces network performance. In [8], authors have proposed interference management strategy by incorporating orthogonality between cellular links and D2D links.

Distance constrained based outage probability calculation is performed in [14], to justify the objective of utilizing all the possible resources in the cellular system. Deng et al. [15] proposed social aware distributed resource algorithm, this algorithm achieves convergence and stability without loss of fairness.

Khuntia et al. [16] proposed an optimal spectral allocation strategy to enhance the throughput of D2D while maintaining QoS for CUs and performance of D2D is analyzed using outage probability analysis [17].

The paper is organized in the following way. Section 2 is about the formulation of the system. Section 3 describes the methodology. The results of the simulation obtained in Sects. 4 and 5 ends with a conclusion.

2 Formulation of the System

Figure 1 shows a cellular system for sharing the uplink resources in a device to device communication framework with N number of orthogonal users and a base station (BS) [18].

The frequency band indexed by each user is expressed by $i = 1, \dots, N$.

h_i^c represents the channel between the base station and cellular user i,

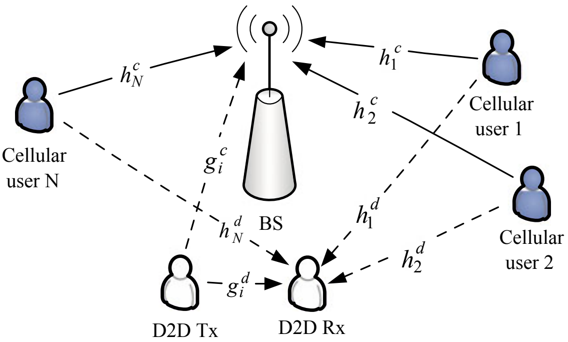
h_i^d is the channel between D2D receiver and cellular user i,

g_i^c is the channel between BS and D2D transmitter for frequency band i.

g_i^c symbolizes the channel between D2D transmitter and the receiver for frequency band i.

Let x_i^c and x_i^d be the transmitting signals for the cellular and D2D users respectively for frequency band i.

Fig. 1 Cellular system in D2D communication



Equation (1) represents the signal received at the base station by the cellular user i:

$$y_i^c = h_i^c x_i^c + g_i^c x_i^d + n_i^c \tag{1}$$

where n_i^c is the Gaussian noise with variance σ_i^c by cellular user i.

And, Eq. (2) represents the signal reception for D2D user with frequency band i:

$$y_i^d = g_i^d x_i^d + h_i^d x_i^c + n_i^d \tag{2}$$

where n_i^d is the additive zero-mean Gaussian noise with variance σ_i^d by D2D user .

Suppose that the Gaussian codes are used by both D2D and cellular users, on each frequency band i which transmits powers of

$$q_i \triangleq E\left|x_i^d\right|^2 \tag{3}$$

$$p_i \triangleq E\left|x_i^c\right|^2 \tag{4}$$

The throughputs for cellular and D2D users are given by the Eqs. (5) and (6) respectively [18].

$$R_i^c(p_i, q_i) \triangleq \log\left[1 + \frac{\left|h_i^c\right|^2 p_i}{\sigma_i^c + \left|g_i^c\right|^2 q_i}\right] = \log\left(1 + \frac{\alpha_i p_i}{1 + \theta_i q_i}\right) \tag{5}$$

$$R_i^d(p_i, q_i) \triangleq \log\left[1 + \frac{\left|g_i^d\right|^2 q_i}{\sigma_i^d + \left|h_i^d\right|^2 p_i}\right] = \log\left(1 + \frac{\gamma_i q_i}{1 + \beta_i p_i}\right) \tag{6}$$

where,$\alpha_i \triangleq \frac{|h_i^c|^2}{\sigma_i^c}$, $\beta_i \triangleq \frac{|h_i^d|^2}{\sigma_i^d}$, $\gamma_i \triangleq \frac{|g_i^d|^2}{\sigma_i^d}$ and $\theta_i \triangleq \frac{|g_i^c|^2}{\sigma_i^d}$ represents the normalized channel gains.

The resource sharing between cellular and D2D users must be designed so that the D2D can achieve maximum benefit by fulfilling the cellular user's requirements. This is accomplished through resource sharing between the D2D and cellular users.

Theoretically, this is achieved by maximizing the throughput of D2D link [18] and is represented as:

$$\begin{array}{cc} \text{Maximize} & \sum_{i=1}^{N} R_i^d(p_i, q_i) \\ \text{p, q} & R_i^c(p_i, q_i) \geq \rho_i, i = 1, \ldots, N \\ \text{subjected to :} & 0 \leq p_i \leq P_i, 0 \leq q_i \leq Q_i, i = 1, \ldots, N \\ & \sum_{i=1}^{N} q_i \leq Q \end{array} \tag{7}$$

where the QoS threshold is symbolized by ρ_i, P_i is the power budget of cellular user i, Q_i for frequency band, i is the D2D user's power limit and the overall power budget for D2D user is symbolized by Q.

It is a very challenging task to achieve optimal resource sharing. The problem in (7) is a non-convex problem as both $R_i^c(p_i, q_i)$ and $R_i^d(p_i, q_i)$ are not jointly concave in (p_i, q_i). This work aims at providing an optimized solution for resource sharing using the particle swarm optimization technique.

3 Methodology

The problem in Eq. (7) is feasible if and only if $\omega_i \triangleq 2^{\rho_i} - 1 \leq \alpha_i P_i\, fori = 1, \ldots, N$ [19].

The realization of optimal resource sharing is accomplished by assuming $\omega_i \leq \alpha_i P_i$, for $i = 1, \ldots, N$.

Let (p^*, q^*) denote the optimal solution to (7).

Define $A_i \triangleq \omega_i\beta_i\theta_i(\alpha_i\gamma_i + \omega_i\beta_i\theta_i)$,$B_i \triangleq (\alpha_i + \omega_i\beta_i)(2\omega_i\beta_i\theta_i + \alpha_i\gamma_i)$, $C_i(\lambda) \triangleq (\alpha_i + \omega_i\beta_i)\big(\alpha_i + \omega_i\beta_i - \frac{1}{\lambda}\alpha_i\gamma_i\big)$ and $D_i \triangleq \min\Big\{Q_i, \frac{1}{\omega_i\theta_i}(\alpha_i P_i - \omega_i)\Big\}$ for $i = 1, \ldots, N$.

If $\sum\limits_{i=1}^{N} D_i \leq Q$, then $p_i^* = \frac{\omega_i}{\alpha_i}(1 + \theta_i D_i)$ and $q_i^* = D_i$;

$$\text{If} \sum_{i=1}^{N} D_i > Q, \text{then} p_i^* = \frac{\omega_i}{\alpha_i}\big(1 + \theta_i q_i^*\big)q_i^* = \left[\frac{\sqrt{B_i^2 - 4A_iC_i(\lambda)} - B_i}{2A_i}\right]_0^{D_i} \tag{8}$$

Where $[D]_0^{D_i}$ symbolizes the projection onto the interval $[0, D_i]$, and $\lambda > 0$ is selected such that $\sum\limits_{i=1}^{N} q_i^* = 0$.

Substituting p_i^* into the frequency band i of D2D user denoted as $R_i^d(p_i, q_i)$ leads to:

$$R_i^d(p_i, q_i) = \log\left(1 + \frac{\alpha_i\gamma_i q_i}{\alpha_i + \omega_i\beta_i + \omega_i\beta_i\theta_i q_i}\right) \tag{9}$$

Let $h(q_i) \triangleq \frac{\alpha_i\gamma_i q_i}{(\alpha_i + \omega_i\beta_i + \omega_i\beta_i\theta_i q_i)}$, we get:

$$h''(q_i) = -\frac{2\alpha_i\gamma_i\omega_i\beta_i\theta_i(\alpha_i + \omega_i\beta_i)}{(\alpha_i + \omega_i\beta_i + \omega_i\beta_i\theta_i q_i)^3} \leq 0 \tag{10}$$

The above equation represents that $h(q_i)$ is a concave function. Therefore, Eq. (7) can be rewritten as:

$$\begin{array}{cc} \text{Maximize} & \sum_{i=1}^{N} \log\left(1+\frac{\alpha_i \gamma_i q_i}{\alpha_i+\omega_i\beta_i+\omega_i\beta_i\theta_i q_i}\right) \\ \text{q} & 0 \le q_i \le D_i, i = 1, \ldots, N \\ \text{subjected to} & \sum_{i=1}^{N} q_i \le Q \end{array} \tag{11}$$

Since the objective in Eq. (11) is increasing for each q_i, then the optimal solution will be $q_i^* = D_i$, and optimal solution to p_i will be $R_i^d(p_i, q_i)$.

3.1 Particle Swarm Optimization (PSO)

Let $f : R^n \to R$ be the function that has to be minimized and S, the number of particles that make up the swarm. Four vectors of dimension n are defined for each particle as attributes: k_i, v_i, $pbesti$ and $gbesti$. The position x_i represents a potential solution for the objective function, the velocity v_i represents the direction and intensity of the movement of the particle, $pbesti$ represents the best position found individually and $gbesti$ the best position found by the particles in their vicinity until the present moment [20].

In the PSO algorithm the steps of the canonical version of PSO are described. After the initialization of its attributes, each particle proceeds to traverse the search space by updating its speed and position. This process occurs iteratively and culminates after a predetermined number of iterations T has elapsed.

3.2 PSO Algorithm Applied for Optimal Parameter Calculation in D2D Communication

```
Define parameters constants and variables $\left(T, N, c_1, c_2, k_i^0, v_i^0\right)$
    Output:$p_i, q_i and R_i^d$ according to Eq. (11)
    for j = 1 to N do
    $pbest_i^0 \leftarrow k_i^0$
    end for
    for $j = 1$ to N do
    Update $gbest_i^0$
    end for
    for j = 1 to T do
    for i = 1 to N do
    Update $v_i^t$ and $k_i^t$
    Evaluate fitness function and update $pbest_i^t$
    end for
    for $i = 1$ to N do
```

```
Update $gbest_i^t$
end for
end for
return $gbest \leftarrow min_{gbest_i}\left\{f\left(gbest_i^T\right)\right\}$
```

The updation of velocity and positions are represented by:

$$v_{i,j}^{t+1} = v_{i,j}^{t} + c_1 r_{1_{i,j}}^{t}\left[pbest_{i,j}^{t} - k_{i,j}^{t}\right] + c_2 r_{2_{i,j}}^{t}\left[gbest_{i,j}^{t} - k_{i,j}^{t}\right] \tag{12}$$

$$k_{i,j}^{t+1} = k_{i,j}^{t} + v_{i,j}^{t+1} \tag{13}$$

The position and frequency of the movements produced by each particle in the search space, as shown in Eqs. (12) and (13), is determined by the influence of three components. The first is the impulse or impetus that represents the force that is exerted on the particle to continue the direction it leads at the current time. The second is the cognitive component that represents the force that arises from the attraction of the particle by its *pbest*, and the third is the social component that represents the force that arises from the attraction of the particle by the *gbest* of his neighbourhood [21].

4 Simulation Results

This section evaluates the performance of the proposed system under various parameters. Tables 1 and 2 provides simulation parameters and PSO parameters.

Figure 2 shows the average throughput versus SNR plot of optimal resource sharing with 8 cellular users in D2D communication. It can be observed that average throughput increases as the number of D2D SNR increase proportionally.

Figure 3 shows the average throughput versus SNR plot of PSO based resource sharing with 8 cellular users in D2D communication. It can be observed that average throughput increases as the number of D2D SNR increase proportionally.

Table 1 Simulation parameters

Parameter name	Value
Base station	1
No. of the frequency band	8
No. of cellular user	8
No. of D2D transmitter and receiver	1
Number of distance	50:50:450
SNR range	2:2:16
Dcell	300
dD2DRX	300

Table 2 PSO parameters

Parameter name	Value
Cognitive parameter	1.2
Social parameter	0.012
Swarm size	100
Inertial weight	0.0004
Number of iteration	500

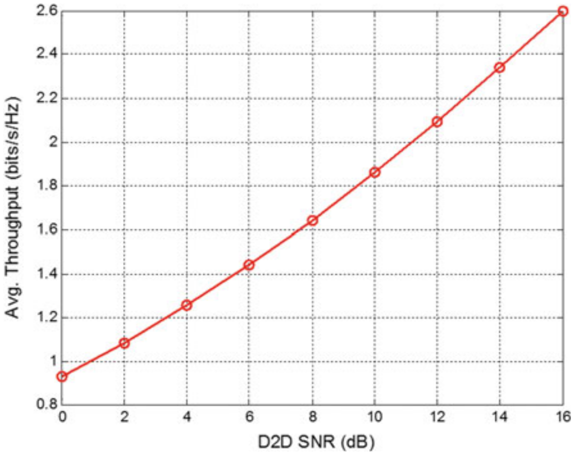


Fig. 2 Avg throughput versus SNR plot of optimal resource sharing with 8 cellular users in D2D

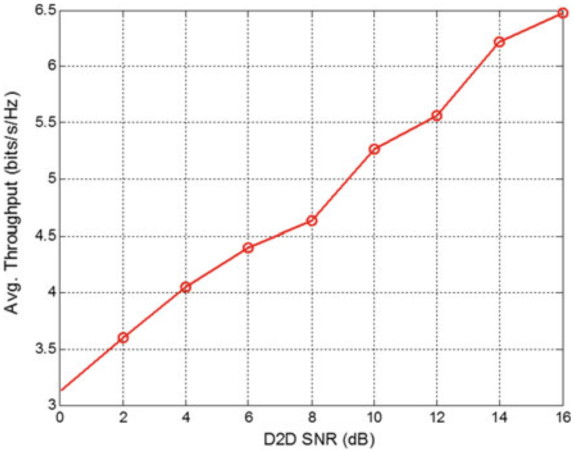


Fig. 3 Avg throughput versus SNR plot of PSO based resource sharing with 8 cellular users in D2D communication

Figure 4 shows the average throughput versus distance between D2D and BS plot of optimal resource sharing with 8 cellular users. As the distance between D2D and BS increases throughput decreases proportionally

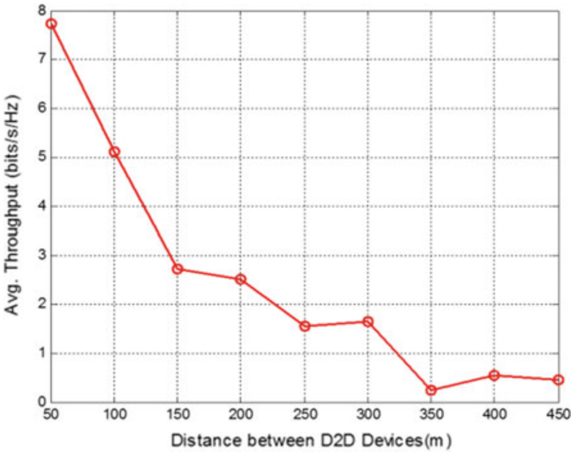


Fig. 4 Avg throughput versus distance between D2D and BS plot of optimal resource sharing with 8 cellular user

From Fig. 5 it can be observed that average throughput is high as approximately 7 bits/s/Hz at 16 dB when PSO is applied. Similarly, the optimal strategy gives average throughput is high as approximately 3bits/s/Hz at 16 dB. It can be observed that as D2D SNR increases throughput increases and can be concluded that PSO provides better performance than the optimal solution.

From Fig. 6 it can be observed that average throughput is high as 14 bits/s/Hz at a minimum distance between the base station and D2D users when PSO is applied. Further, the optimal strategy gives average throughput is high as 8 bits/s/Hz at minimum distance 50 m. The observed gain in throughput is approximate 6 bit/s/Hz at 50 m. It is observed that as distance increases between BS and D2D user throughput decreases. It can be concluded that PSO enabled resource utilization performs better than the optimal solution.

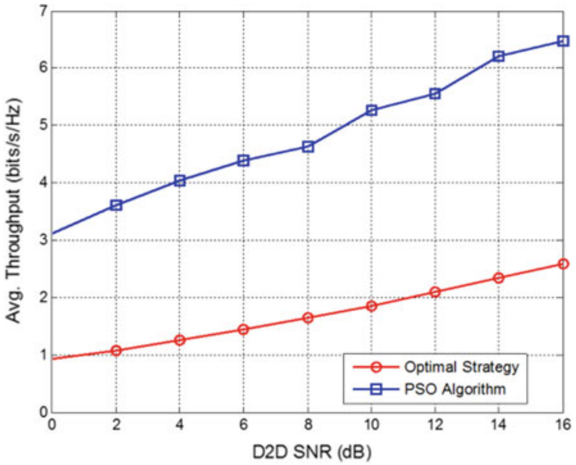
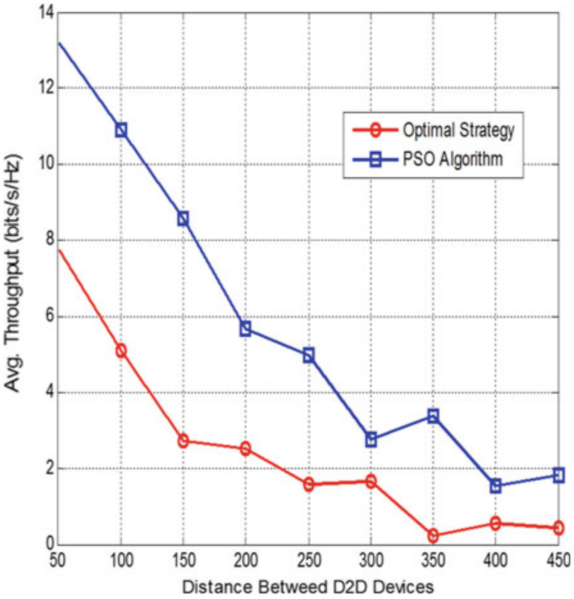


Fig. 5 Comparative result of PSO optimized and optimal strategy based average D2D throughput vs D2D SNR with 8 cellular users

Fig. 6 Comparative results of PSO optimized and optimal strategy based average D2D throughput versus the distance between the D2D link and BS



5 Conclusions

In this work, an optimized resource sharing algorithm for D2D communications using PSO has been proposed. The resource sharing problem is modelled by underlay uplink resources of multiple cellular users. It is found that the number of frequency band has been optimized. Results of the simulation show good performance of the proposed algorithm with a 6 Bit/s/Hz gain achievement in throughput at a minimum distance of D2D devices.

References

1. Karagiannis G, Altintas O, Ekici E, Heijenk G, Jarupan B, Lin K, Weil T (2011) Vehicular networking: a survey and tutorial on requirements, architectures, challenges, standards and solutions. IEEE Commun Surv Tutorials 13(4):584–616, Fourth Quarter
2. Third Generation Partnership Project, Study on enhancement of 3GPP Support for 5G V2X Services, Third Generation Partnership Project, 3GPP TR 22.886 v15.0.0, Dec. 2016. http://www.3gpp.org
3. Jiang D, Delgrossi L (2008) IEEE 802.11p: Towards an International Standard for Wireless Access in Vehicular Environments, in VTC Spring 2008—IEEE Vehicular Technology Conference, Singapore, Singapore, pp 2036–2040
4. Araniti G, Campolo C, Condoluci M, Iera A, Molinaro (2013) LTE for Vehicular Networking: A Survey, IEEE Communications Magazine, vol. 51, no. 5, pp 148–157
5. Third Generation Partnership Project, Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description;

Stage 2, Third Generation Partnership Project, 3GPP TS 36.300 v14.0.0. http://www.3gpp.org

6. European Telecommunications Standards Institute, Intelligent Transport Systems (ITS); V2X Applications, European Telecommunications Standards Institute, ETSI TS 101–539 v1.1.1. http://www.etsi.org
7. Fodor G, Dahlman E, Mildh G, Parkvall S, Reider N, Miklos G, Turanyi Z (2012) Design aspect of network assisted device-to-device communications. IEEE Commun Mag 50(3):170–177
8. Feng L, Zhao P, Zhou F, Yin M, Yu P, Li W, Qiu X (2018) Resource allocation for 5G D2D multicast content sharing in social-aware cellular networks. IEEE Commun Mag 56(3):112–118
9. Chai X, Liu T, Xing C, Xiao H, Zhang Z (2016) Throughput improvement in cellular networks via full-duplex based device-to-device communications. IEEE Access 4:7645–7657
10. Nguyen H, Duong Q, Shin OS Resource Allocation Optimization for Device-to-Device Communications in Cellular Networks, Proc. IEEE Int'l. Cnf. Information Communication Technology Convergence (ICTC'14), pp 377–378 (2014)
11. Zhao W, Wang S (2015) Resource allocation for device-to-device communication underlaying cellular networks: an alternating optimization method. IEEE Commun Lett 19(8):1398–1401
12. Gandotra P, Jha RK (2016) Device-to-device communication in cellular networks: a survey. J Netw Comput Appl 71:99–117
13. Kumar TS (2019) Efficient Resource Allocation And Qos Enhancements Of Iot With Fog Network. J ISMAC 1(02):101–110
14. Moghaddam SS, Ghavami H (2018) Joint mode selection and resource allocation in device-to-device communications. Int J Sensors Wirel Commun Control 8(3):204–216
15. Ningombam D, Shin S (2018) Distance-constrained outage probability analysis for device-to-device communications underlaying cellular networks with frequency reuse factor of 2. Computers 7(4):50
16. Deng G, Shi J, Nie G, Huang Z (2018) Time-varying Social-aware Resource Allocation for Device-to-Device Communication. IEEE Access 7:2653–2663
17. Marini F, Walczak B (2015) Particle swarm optimization (PSO): a tutorial. Chemometr Intell Lab Syst 149:153–165
18. Xu J, Guo C, Zhang H (2018) Joint channel allocation and power control based on PSO for cellular networks with D2D communications. Comput Netw 133:104–119
19. Wang J, Zhu D, Zhao C, Li JC, Lei M (2013) Resource sharing of underlaying device-to-device and uplink cellular communications. IEEE Commun Lett 17(6):1148–1151
20. Chen CY, Sung CA, Chen HH (2018) Optimal mode selection algorithms in Multiple Pair Device-to-Device Communications. IEEE Wirel Commun 25(4):82–87
21. Ke-Lin D, Swamy MNS (2016) Particle swarm optimization. In Search and optimization by metaheuristics, pp 153–173. Birkhäuser, Cham
22. Khuntia P, Hazra, R (2018) Resource sharing for device-to-device communication underlaying cellular network. In 2018 4th International Conference on Recent Advances in Information Technology (RAIT) (pp 1–5). IEEE
23. Zhang H, Song, L., Han, Z, Zhang, Y (2018) Radio Resource Allocation for Device-to-Device Underlay Communications. In *Hypergraph Theory in Wireless Communication Networks* (pp 21–39). Springer, Cham

A Systemic Method of Nesting Multiple Classifiers Using Ensemble Techniques for Telecom Churn Prediction

J. Beschi Raja, G. Mervin George, V. Roopa, and S. Sam Peter

Abstract In this contemporary world, almost every business and companies deploy machine learning methods for taking exemplary decisions. Predictive Customer analytics supports to accomplish of momentous insights from customer data. This trend is more distinct in the telecommunication industry. The most challenging issue for telecom service providers is the increased churn rate of customers. In Existing works, combining various classification algorithms to design hybrid algorithms as well as ensembles have reported best results compared to single classifiers. But, selecting classifiers for creating an effective ensemble combination is challenging and are still in investigation. This work presents a various type of ensembles such as Bagging, Boosting, Stacking and Voting to combine with different Base classifiers in a systemic way. Experiments were conducted with benchmark Telecom customer churn UCI dataset. It is inferred that; Ensemble learners outperform single classifiers due to its strong classifying ability. The models designed were affirmed using standard measures such as AUC, Recall, F-Score, TP-rate, Precision, FP-Rate and Overall Accuracy. This way of combining multiple classifiers as an ensemble achieves the highest accuracy of 97.2% from bagged and boosted-ANN.

Keywords Predictive analytics · Customer churn · Ensembles · Telecom · Classification

J. Beschi Raja (✉) · S. Sam Peter
Department of CSE, Sri Krishna College of Technology, Coimbatore, India
e-mail: beschiraja.j@skct.edu.in

S. Sam Peter
e-mail: sampeter.s@skct.edu.in

G. Mervin George
Department of IT, Sri Krishna College of Engineering and Technology, Coimbatore, India
e-mail: Mervin2me@gmail.com

V. Roopa
Department of IT, Sri Krishna College of Technology, Coimbatore, India
e-mail: v.roopa@skct.edu.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_2

1 Introduction

Prevention of churn is an important sphere of CRM in all kinds of industries. The expeditious growth of data in digital systems corresponded to information technologies furnishes immense opportunities to form customer patterns [3]. Globally, the Telecom industry has rapid growth among all businesses [1]. On the other hand, they experience furious competition in fulfilling their customers' expectations and concentrates to retain their existing valuable customers [1, 2, 4]. The customers' affinity to migrate to other feasible telecom providers varies for disparate reasons [4]. Besides, churn reflects in the outcome of the mislay of economic loss and the public image of companies. Predicting customer churn research has received apex remarkable attention all times [2, 5]. The extensive amount and divergence nature of telecom data sets are prominent impediments in acquiring the desired results for churn prediction [4]. Telecom data analysis for churn avoidance has been an imperative research domain due to its huge significance [9]. Abundant machine learning techniques were employed in Customer analytics and designing churn prediction models for precise results. The most prevailing algorithms are constructed using DT, LR, SVM and ANN. Various machine learning methods involving individual classifiers and along with the integration of multiple classifiers were applied [5–7, 10]. Churn researches exhibit that individual model-based classifiers are weak to produce desired solutions and hence researchers move to hybrid techniques, which is an integration of two or more classification algorithms [12]. The hybrid classifiers lead to better solutions compared to single classifiers and are adaptable for all environment [11]. Ensemble techniques are emerging vibrantly and got significant attention in research for increasing the accuracy of classification. Despite these evolutions, creating an effective ensemble combination of selecting algorithms is challenging and are still in investigation. It is commonly believed that the performance of ensemble techniques relay with base learners. This paper is systematized as. Session 2 reveals the discussion of existing survey works with accuracy. Session 3 presents the description of the dataset and its pre-processing process. Session 4 describes the framework proposed for ensemble classifiers while experimental results for all evaluation measures are discussed in Session 4. Finally, Session 5 concludes this article.

2 Related Works

Generally, classification methods are employed in Telecom churn prediction to analyse the customer behaviour expressed in demographics, billing data and call details for identifying customer characteristics. Initially, single classifiers are employed to predict telecom churners. For example, Xia et al. used SVM to analyse churn prediction on UCI dataset. Radial basis function reports the best result of 90% of accuracy [13]. The further improvement was performed by Sharma et al., where

Table 1 Description of dataset

Ref. no	Classifiers	Dataset	Accuracy (%)
[5]	ANN +ANN	Teradata Duke (Public)	94.32
[28]	Nave Bayesian	Teradata Duke (Public)	68
[16]	SVM with Gaussian Kernel	Teradata Duke (Public)	87.15
[25]	Decision Tree	British Telecom (Private)	82
[26]	Voted Perceptron + Logistic Regression	Asian Telecom (Private)	76

the accuracy was improved to 92.5% using feed-forward NN on UCI dataset [14]. In another study, Wouter et al. achieved an accuracy of 83.8% on the pruned dataset using various rule induction techniques [7]. Accuracy was increased by Brandusoiu et al. [15] where comparison was made between MLP and RBF. They finally realized the accuracy of 93.7% using MLP for predicting the churners. Tsai et al. [5] proposed two hybrid models with ANN (backpropagation) and SOM. The experiments reveal that ANN + ANN provides good results compared to SOM + ANN. Another hybrid model was made by Zhang et al. where KNN and Logistic Regression are combined [16]. Khashei achieves higher accuracy compared to individual methods using NN with linear regression [17]. Likewise, a work proposed by Ahmed [18] using firefly concept for optimization and increasing the accuracy. Vafeiadis et al. [19] proposed a hybrid classifier with bagging and boosting and realised a higher accuracy of 96.86% with boosted SVM. Among other techniques, Verbeke et al. performed experiments with eleven Telecom datasets. SMOTE is employed for oversampling. They realised the accuracy of 91.8 by bagging and decision Tree [20]. Likewise, De Bock et al. [10] introduced rotation ensemble techniques for churn prediction. The features were selected and extracted. The feature selection methods such as PCA, SRP and ICA are investigated for effective feature extraction. The proposed technique combines rotation Forest with Adaboost (Rot boost) which was evaluated on Telecom dataset. Stripling et al. [21] proposed a profit-driven model named as ProfLogit for churn prediction. They experimented on nine different real-time and evaluated recall and precision values (Table 1).

3 Dataset

The study has experimented with publicly available benchmark Telecom UCI dataset for designing the model and assessing the performance [22]. The UCI dataset consists of 20 features and 5000 instances. The information present is mostly connected to Call Detail Records (CDR). Out of 5000 samples, 715 samples are churners which shows the imbalance nature of the dataset.

The UCI churn data was treated for pre-processing by converting to numeric from string attributes. We excluded the variables such as 'state' and 'phone number' during training for better accuracy. Our target is design generic model that is suitable to any Telecom platform irrespective of their location and area.

4 Proposed Framework for Ensembles

The ensemble learners are composed of base and meta classifiers. Meta learners refer to the combination of various multiple classifiers. To design the ensemble framework, this phase constructs a various model by integrating base and meta learners to form multi-classifier models using boosting, stacking, bagging and voting. The dataset is initially introduced with individual classifiers and then combined with various ensemble methods such as voting, bagging, boosting and stacking. The sets formed by multiple classifiers are produced with boosting, stacking, bagging and voting and referred to as boosted sample instance, bagged sample instance, stacked sample instance and voted sample instance respectively.

4.1 Boosting

Ada boost is a technique, engage a two-step process; it first employs subclass of train data to generate set of frail rules by base classifiers and then develops an ensemble model by the rules, in turn, boosting their efficiency and performance [23] (Fig. 1).

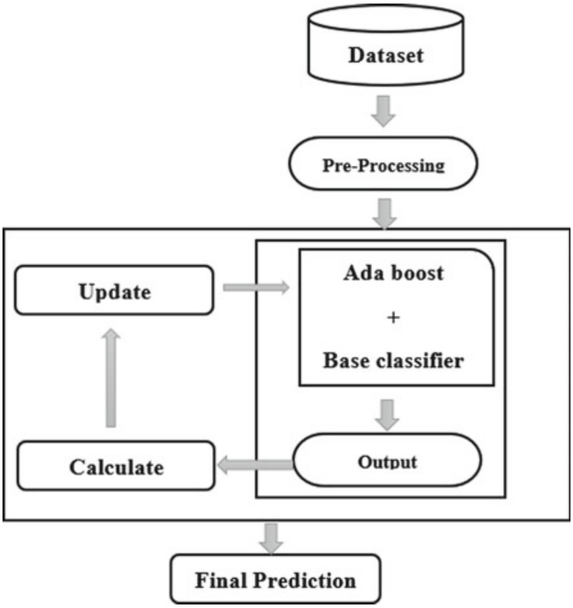


Fig. 1 Ada boost iteration and training process

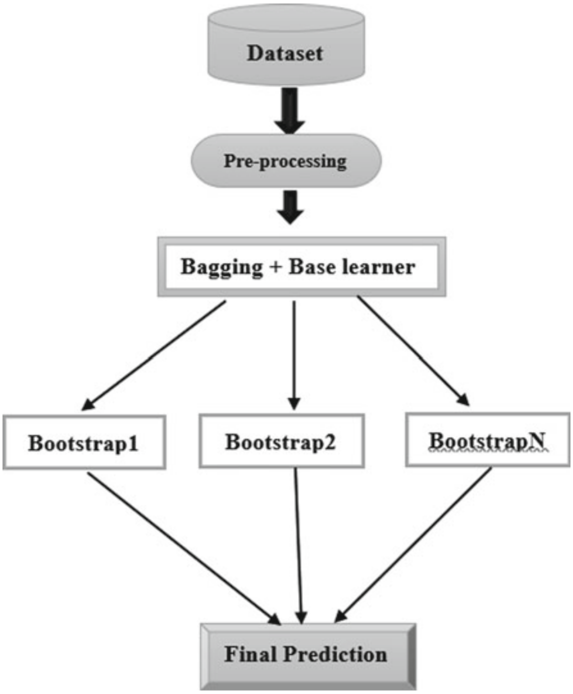
Adaptive Boost algorithm employs with labelled train data $G = \{(p\_i, q\_i)\dashv\}$, where Pi denotes samples, $i = 1\ldots N$ and q_i is connected tag per every single occurrence p_i. Proceeding each execution, $K = 1\ldots X$, a weight is allotted to all parts in train data. The frail classifier stayed to receive a frail hypothesis h_k (p_i) $= q\_i$. Learning error $\in\_k$ stands for evaluating and regenerate the weights up to the final looping is obtained. The combination of Ada boost technique with various baseline classifiers were experimented. In our work, we examined five base classifiers integrating with boosting on UCI churn Telecom dataset.

4.2 Bagging

Bagging is employed to enhance the efficiency of base classifiers. It is also called as a bootstrap aggregation that helps for overfitting [24] (Fig. 2).

Bagging creates m number of samples of bootstrap Gj ($j = 1\ldots q$) of train dataset Y with M of sample Gj. The classification results are averaged and retrieved from each bootstrap sample Gj using learning algorithm C finally. The probability of samples was selected randomly for each time $1/M$. If G as a set of bootstrap samples of size n then denoted as follows

Fig. 2 Training process of bootstrapped sets by bagging process



$$\left\{K_1^1, K_2^1, \ldots, K_G^1\right\}, \left\{K_1^2, K_2^2, \ldots, K_G^1\right\}, \ldots, \left\{K_1^G, K_2^G, \ldots K_n^G\right\}$$

$$K_n^G \equiv n\text{th observation of the } B\text{th bootstrap sample}$$

Fit G almost independent weak learners as $w_1(.), w_2(.), \ldots, w_B(.)$, and then bundle them into average to get an ensemble model with a lower variance. In our study, the investigation of various base classifiers with bagging on UCI churn Telecom dataset was carried out.

4.3 Stacking

Stacking or stacked generalization is the integration of various heterogeneous and diversity of classifier algorithms. The stacking has two levels in combining learners. First, the base learners are placed at level-0 referred to as stacked learner. Next, meta-learners are placed at level 1 referred to as meta-learner. Meta learners integrate the outcomes of base stacked learners for increasing the complete efficiency of the model. For a sample of observations $X = \left\{x_i \in R^M\right\}$ and a set of Labels $Y = \{y_i \in N\}$ and a Training Set $D = \{x_i, y_i\}$ as an input, such that learn the model M based on D, Input: $\mathrm{D} = \left\{\left(\mathrm{x_i}, \mathrm{y_i}\right)|\mathrm{x_i} \in \mathrm{X}, \mathrm{y_i} \in \mathrm{Y}\right\}$ Output: E (Ensemble Classifier). Initially, first-level classification methods are analysed and learned. Next, base methods are learned and a new sample is created using the above process. Likewise, all next classifiers are learned with newly created samples [29].

Stacking provides various heterogeneous models, integrating the diverse of base classifiers at base learners while one meta-learners placed on level one layer. Generally, 60% of the dataset is employed for training at stack learners of level 0 and rest is used evaluated with meta learners at level 1. Different classifiers with stacking for instance stacked- DT, stacked-ANN, stacked-KNN, Stacked-LR, Stacked-NB were tested (Fig. 3).

4.4 Voting

Voting Technique is mainly used for classification problems. In this method, the initial step is to design multiple classification models using train data (Fig. 4).

Here, the class label is calculated using the formulae $y = mode\{G_1(F), G_2(F) \ldots . G_x(F)\}$. Consider that if three classifiers are used for voting such as G_1, G_2 and G_3. If G_1 and G_2 predicts class 0 and G_3 predicts Class 1, then, voting classification outcome will be class 0. The base model can be designed using diverse splits of the same train data with same classifiers or with different algorithms. The voting model creates a prediction for all instances separately and derives the final outcome. If more than 50% of the votes is not given

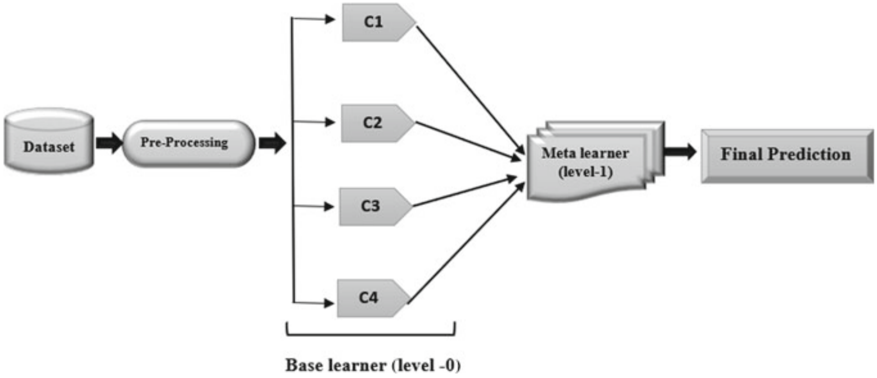


Fig. 3 Integration of base and meta learners by stacking process

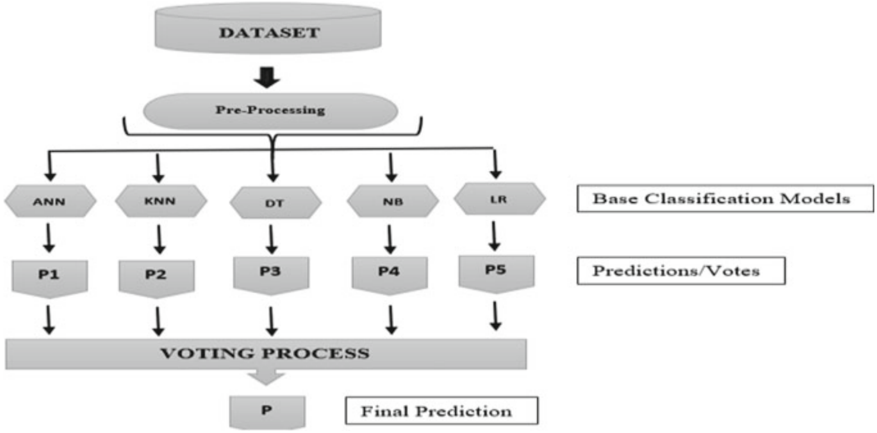


Fig. 4 Prediction from base models by majority voting process

from any of the prediction, it may say that the designed ensemble technique unable to make stable outcome [25]. The investigation on voting with five various base classifiers such as ANN, DT, KNN, NB and LR is done.

5 Experimental Outcomes and Discussion

This session presents the investigational outcomes conducted to evaluate the framework proposed. The dataset is randomly sampled 80% for the process of training and the remaining 20% for the process of testing for all phase in our analysis. The evaluation measures such as recall, ROC, AUC, Precision, TP rate, FP rate and accuracy are determined. Cross-Validation is applied for experimental analysis and testing the model. The final outcome is discussed and analysed with ensemble creation with various base learners. The experimental investigation performed using Weka and Rapidminer tool. First, experiments are carried out for base learners on UCI dataset are shown in Table 3. The highest classification accuracy is given my ANN with 95.2%. Next, the base learner algorithms are integrated with ensemble techniques such as Stacking, Boosting, Bagging and Voting to produce diverse combination. Experiments with boosting discovered that Boosted-ANN provides the highest accuracy of 97.2% which outperforms other boosting combination with other ensembles. The boosting techniques with base leaners experiments are depicted in Table 4. Analysis of bagging represents Bagged-ANN performs the best performance with 97.2.% which is the same accuracy received from Boosted-ANN. The bagging techniques with base leaners experiments are summarized in Table 5. In stacking, the highest accuracy of 96.7% was given by the combination of ANN as meta learners and other base learners are summarised in Table 7. In voting the highest accuracy was obtained was 96.1% by combination of KNN, ANN, LR and DT. The experimental analysis of voting with various combination of base learners is depicted in Table 6. It is observed that Boosted-ANN and Bagged-ANN are the prominent contributions of proposed work which outperforms all ensemble techniques used in work (Figs. 5, 6, 7, 8 and 9).

Table 3 Summary results of various base classifier for UCI dataset

Classifier	TP Rate	FP rate	Precision	Recall	F-measure	AUC	Accuracy (%)
NB	0.876	0.470	0.866	0.876	0.870	0.839	87.64
LR	0.866	0.669	0.840	0.866	0.839	0.847	86.40
ANN	0.952	0.200	0.952	0.951	0.951	0.920	95.2
DT	0.950	0.247	0.947	0.950	0.47	0.783	95
KNN	0.882	0.416	0.876	0.882	0.879	0.733	88.2

Table 4 Summary of results of boosting with various base learners

Classifier	TP Rate	FP rate	Precision	Recall	F-measure	AUC	Accuracy (%)
Boosting + NB	0.883	0.468	0.872	0.883	0.876	0.806	88.3
Boosting + LR	0.886	0.669	0.840	0.866	0.839	0.723	86.6
Boosting + ANN	0.972	0.133	0.972	0.972	0.971	0.949	97.2
Boosting + DT	0.956	0.205	0.955	0.956	0.954	0.918	95.6
Boosting + KNN	0.882	0.416	0.876	0.882	0.879	0.733	88.2

Table 5 Summary of results of bagging with various base learners

Classifier	TP rate	FP rate	Precision	Recall	F-measure	AUC	Accuracy (%)
Bagging + NB	0.950	0.545	0.913	0.950	0.931	0.870	87.9
Bagging + LR	0.865	0.663	0.838	0.865	0.840	0.847	86.5
Bagging + ANN	0.97	0.133	0.972	0.972	0.971	0.949	97.2
Bagging + DT	0.95	0.247	0.951	0.949	0.917	0.917	95.2
Bagging + KNN	0.877	0.469	0.867	0.877	0.871	0.841	87.7

Table 6 Experimented results of voting with various combination of base learners

Classifier	TP rate	FP rate	Precision	Recall	F-measure	AUC	Accuracy (%)
Voted NB-KNN-ANN-DT	0.951	0.253	0.950	0.951	0.948	0.933	95.1
Voted NB-KNN-ANN-LR	0.922	0.392	0.918	0.922	0.914	0.928	92.2
Voted KNN-ANN-LR-DT	0.960	0.222	0.960	0.960	0.958	**0.943**	**96.1**
Voted NB-ANN-DT-LR	0.95	0.271	0.90.50	0.950	0.947	0.940	95
Voted NB-KNN-DT-LR	0.928	0.367	0.925	0.928	0.921	0.922	92.8

Table 7 Experimented results of stacking with various combination of base and meta learners

Classifier	TP Rate	FP rate	Precision	Recall	F-Measure	AUC	Accuracy (%)
KNN, ANN, DT, LR	NB	0.954	0.107	0.980	0.966	0.973	**0.946**
NB, KNN, DT, LR	ANN	0.967	0.140	0.966	0.967	0.966	0.925
NB, ANN, DT, LR	KNN	0.939	0.185	0.939	0.939	0.939	0.877
NB, ANN, KNN, LR	DT	0.951	0.224	0.950	0.951	0.949	0.807
NB, ANN, KNN, DT	LR	0.949	0.242	0.948	0.949	0.947	0.910

6 Proposed Framework for Ensembles

A study on creating ensembles with heterogeneous classifiers using Booting, Bagging, Voting and Stacking with systematic experimental analysis to obtain pest performance are presented. The experiment was performed with standard benchmark Telecom UCI dataset. It is noted that Bagging and Boosting shows good results

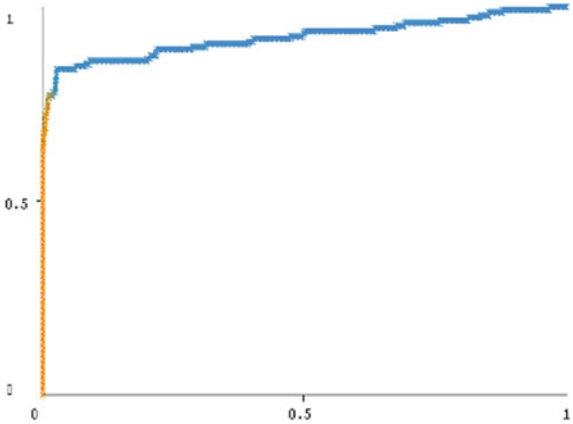


Fig. 5 ROC of ANN base learner (AUC = 0.920)

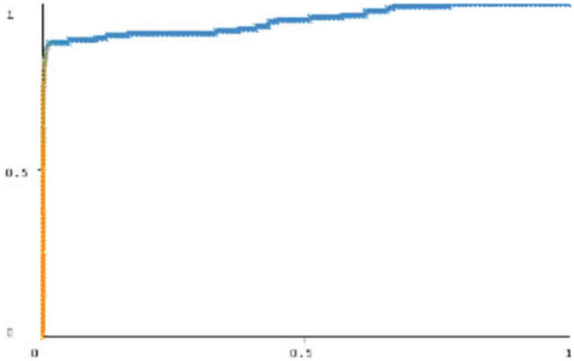


Fig. 6 ROC of boosted ANN (AUC = 0.949)

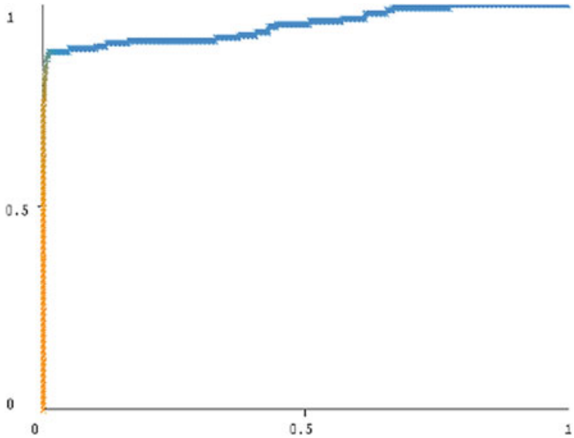


Fig. 7 ROC of bagged ANN (AUC = 0.949)

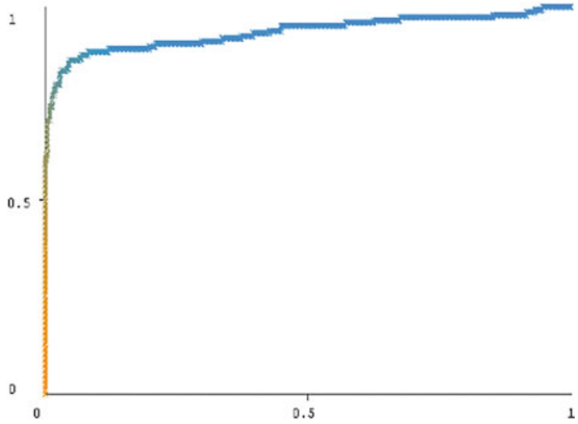


Fig. 8 ROC of voting (AUC = 0.943)

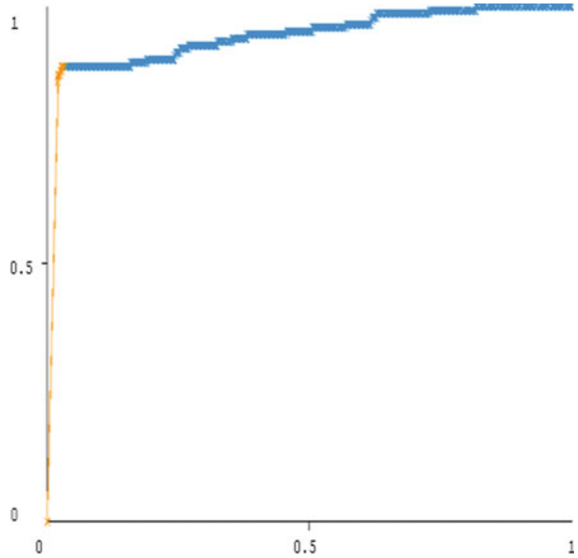


Fig. 9 ROC of stacking (AUC = 0.946)

compared to other ensemble Techniques. Despite this, it is observed that ANN outperforms other base classifiers in both ensemble model as well as in the base classifier model. Bagged-ANN and Boosted-ANN achieves the higher accuracy of 97.2%. The future work can be extended to modify ANN for acquiring further better performance and reducing time for model creation (Table 8).

Table 8 Experimented results of stacking with various combination of base and meta learners

Classifier	TP Rate	FP rate
[13]	SVM + RBF Kernel	90.9 accuracy
[26]	Neural network	92.35 accuracy
[15]	Multi-layer perceptron	93.7 accuracy
[27]	MCE + GMDH	86.9 AUC
[19]	Boosted SVM	96.86 Accuracy
	Proposed Bagged-ANN	97.2 Accuracy
	Proposed Boosted-ANN	97.2 Accuracy

References

1. Idris A, Aksam I, Zia ur Rehman (2017) Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling. Cluster Comput 1–15
2. Wenjie B et al. (2016) A big data clustering algorithm for mitigating the risk of customer churn. IEEE Trans Industr Informat 12.3: 1270–1281
3. Lu N et al (2014) A customer churn prediction model in telecom industry using boosting. IEEE Trans Industr Informat 10.2: 1659–1665
4. Idris Adnan, Rizwan Muhammad, Khan Asifullah (2012) Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies. Comput Electr Eng 38(6):1808–1819
5. Tsai Chih-Fong, Yu-Hsin Lu (2009) Customer churn prediction by hybrid neural networks. Expert Syst Appl 36(10):12547–12553
6. Vafeiadis T et al (2015) A comparison of machine learning techniques for customer churn prediction. Simulation Modelling Practice and Theory 55:1–9
7. Verbeke W et al (2011) Building comprehensible customer churn prediction models with advanced rule induction techniques. Expert Systems with Applications 38.3: 2354–2364
8. Ahmed M et al (2018) Exploring nested ensemble learners using overproduction and choose approach for churn prediction in telecom industry. Neural Comput Appl 1–15
9. Amin A, Anwar S, Adnan A, Nawaz M, Alawfi K, Hussain A, Huang K (2017) Customer churn prediction in the telecommunication sector using a rough set approach. Neurocomputing 237:242–254
10. De Bock KW, Van den Poel D (2011) An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. Expert Syst Appl 38(10):12,293–12,301
11. Sivasankar E, Vijaya J Hybrid PPFCM-ANN model: an efficient system for customer churn prediction through probabilistic possibilistic fuzzy clustering and artificial neural network. Neural Comput Appl 1–20
12. Xia Ge, Jin Wd (2008) Model of customer churn prediction on support vector machine. Syst Eng Theory Pract 28(1):71–77
13. Sharma A, Panigrahi D, Kumar P (2011) A neural network-based approach for predicting customer churn in cellular network services. Int J Comput Appl 27(11):26–31
14. Pamina J et al (2019) An Effective Classifier for Predicting Churn in Telecommunication. J Adv Res Dynam Control Syst 11 (2019)
15. Zhang Y et al (2007) A hybrid KNN-LR classifier and its application in customer churn prediction. In IEEE International Conference on Systems, Man and Cybernetics-ISIC 2007. IEEE (2007)
16. Khashei M, Hamadani AZ, Bijari M (2012) A novel hybrid classification model of artificial neural networks and multiple linear regression models. J. Expert Syst. Appl. 39(3):2606–2620
17. Ahmed, Ammar AQ, and D. Maheswari. "Churn prediction on huge telecom data using hybrid firefly based classification." Egyptian Informatics Journal 18.3 (2017): 215–220

18. Vafeiadis T, Diamantaras KI, Sarigiannidis G, Chatzisavvas KC (2015) A comparison of machine learning techniques for customer churn prediction. Simul Model Pract Theory 55:1–9
19. Verbeke W, Dejaeger K, Martens D, Hur J, Baesens B (2012) New insights into churn prediction in the telecommunication sector: a profit driven data mining approach. Eur J Oper Res 218(1):211–229
20. Stripling E, vanden Broucke S, Antonio K, Baesens B, Snoeck M, (2017) Profit maximizing logistic model for customer churn prediction using genetic algorithms. Swarm Evolut Comput 40:116–130
21. Blake CL, Merz CJ (1998) UCI Repository of machine learning databases, Irvine, University of California. http://www.ics.uci.edu/*mlearn/MLRepository.html
22. Freund Y, Schapire RE (1995) A desicion-theoretic generalization of on-line learning and an application to boosting. In: Computational learning theory. Springer, pp 23–37
23. Breiman L (1996) Bagging predictors. Mach Learn 24(2):123–140
24. Cao J et al (2012) Voting based extreme learning machine. Informat Sci 185.1: 66–77
25. Xiao J, Xiao Y, Huang A, Liu D, Wang S (2015) Feature-selection-based dynamic transfer ensemble model for customer churn prediction. Knowl Inf Syst 43(1):29–51
26. Nath SV, Behara RS (2003) Customer churn analysis in the wireless industry: a data mining approach. In: Proceedings of the Annual meeting of the decision sciences institute, pp 505–510
27. Poornaselvan KJ, Gireesh Kumar T, Vinodh PV (2008) Agent based ground flight control using type-2 fuzzy logic and hybrid ant colony optimization to a dynamic environment. 2008 First International Conference on Emerging Trends in Engineering and Technology. IEEE
28. Vijayakumari V, Suriyanarayanan N (2012) Survey on the detection methods of blood vessel in retinal images. Eur J Sci Res 68(1):83–92
29. Pamina J, Beschi Raja J, Sam Peter S, Soundarya S, Sathya Bama S, Sruthi MS (2020) Inferring Machine Learning Based Parameter Estimation for Telecom Churn Prediction. In: Smys S, Tavares J, Balas V, Iliyasu A (eds) Computational Vision and Bio-Inspired Computing. ICCVBIC 2019. Advances in Intelligent Systems and Computing, vol 1108. Springer, Cham

Hybrid Method in Identifying the Fraud Detection in the Credit Card

Pooja Tiwari, Simran Mehta, Nishtha Sakhuja, Ishu Gupta, and Ashutosh Kumar Singh

Abstract As the world is rapidly moving towards digitalization and money transactions are becoming cashless, the utilization of credit cards has rapidly heightened. The fraud activities associated with it have also been increasing which leads to a huge loss to the financial institutions. Therefore, we need to analyze and detect the fraudulent transaction and separate them from the non-fraudulent ones. The paper proffers the frame work to detect credit card frauds. A combination of the three techniques is used. These methodologies include Decision Trees, Neural Network and K-Nearest Neighbor. For each new incoming transaction, the new label is assigned by taking the majority of the labels from the output of these techniques. This model is believed to work fairly good for all sizes and kinds of dataset as it combines the advantages of the individual techniques. We conclude the paper with a comparison of our model with the existing ones.

Keywords Support vector machines (SVM) · Hidden markov model · Neural networks logistic regression · K-Nearest neighbor genetic algorithm · Random forests · Bayesian belief network · Decision trees

P. Tiwari (✉) · S. Mehta · N. Sakhuja · I. Gupta (✉) · A. K. Singh
Department of Computer Applications, National Institute of Technology, Kurukshetra, India
e-mail: pooja_51710011@nitkkr.ac.in

I. Gupta
e-mail: ishugupta23@gmail.com

S. Mehta
e-mail: simran_51710060@nitkkr.ac.in

N. Sakhuja
e-mail: nishtha_51710077@nitkkr.ac.in

A. K. Singh
e-mail: ashutosh@nitkkr.ac.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_3

1 Introduction

The utilization of the credit card has been as one of the most used financial products as cashless transactions have come into play in the 21st century [1]. Credit cards are mainly used for large transactions and hence the illegal activities relating to the cards have increased several folds with an increase in its usage. Fraud associated with the credit cards is when an illegal transaction takes places that is when unauthorized personnel tries to access the card to make certain payments. There is a large number of ways through which frauds relating to credit cards can be committed. A fraudster can steal the card, your account number, PIN and one-time passwords to commit the crime. Such frauds where the fraudster doesn't have the card physically are called as card-not-present frauds. These fraudulent transactions have incurred huge loss for the economy and hence the citizens to bear.

Causing adverse effects on business and society [2], and accounting for about billions of dollars of lost revenue each year, credit card frauds have become a major issue worldwide. While some statistics show about $400 billion cost of a year, some other figures show about 1.6 billion pounds total yearly loss of UK insurers due to such fraudulent attempts. The observed patterns in behavior of the customers regarding payments are related to the assumptions that customers use cards instead of cash (Euromonitor International, 2006) [3]. Now the loss is actually affecting everyone even if one hasn't been defrauded, paying of credit and charge debts increases the cost of goods and services. Hence the need of the hour is a fraud detection system that can distinguish an incoming transaction request as a fraudulent or non-fraudulent. and hence alarm the banks [4, 5]. Various machine learning techniques can be used for prediction.

We have come across various fraud detection techniques that exist today but none of them was competent enough to detect the fraud at the time it actually took place. All the techniques so far give accurate results only when performed on a particular dataset and sometimes with some special features only. Techniques like SVM works better than Logistic Regression when there is a class imbalance and comparatively Random Forest performs better among all three [6–8]. Bagging Ensemble Classifier is suitable for highly imbalanced dataset [9, 10]. Some techniques like Decision Tress and SVM gives better results on raw unsampled data whereas techniques like ANN and Bayesian Belief Network have high accuracy and detection rate but are expensive to train [11, 12]. Similarly, SVM and the KNN gives better results with small datasets but are not preferable with large datasets [13]. We propose a technology that works equally precisely and accurately under all circumstances and with various datasets by using a combination of existing techniques [14].

The rest of the paper is organized as:

Section 2 contains the previous study and related work in the field of credit card fraud detection. Section 3 contains a discussion about the dataset. Following which the Sects. 4 and 5 contains the actual description of model and the results obtained. And Sect. 6 is the conclusion.

2 Related Work

Fraudulent activities are the cause for huge business loss, which actuated researchers to come up with a solution that would reduce the amount of frauds by detecting and preventing them. Several methodologies that have been proposed for a fraud detection system are reviewed briefly in the following section.

In paper [15] the logfistic regression, random forest and the support vector machine was applied to the European dataset and it was found that the to hav a higher recall and precision once the dataset is balanced . In paper [16] same data set was used and the comparison was done using the models LR, DT and RF [17, 18].

According to [19] and [20], utilizes the KNN [21], and compares it with the other conventional algorithms [10].

All the techniques which were tested gave an accuracy of 80% approx. After the preprocessing of the data, algorithms showed comparatively high accuracy of over 90%. In paper [22], a neural network that included back propagation and optimized with Whale Algorithm was tested on the European dataset with test 500 samples [23, 24]. The algorithm achieved exceptional results giving 96.40% accuracy and 97.83% recall.

The deepneural networks are predominant in identifying the transactions that are fraud [25–30].

The study of Bayesian Network Classifier (HHEA) along with instance reweighing and probability analysis on some threshold shows that the Fraud Bayesian Network Classifier followed by probability threshold is more beneficial than Naïve Bayes [31], Support Vector Machines [31], Tree Augmented Naïve Bayes [32], and Decision Trees on the PagSeguro dataset when precision [33, 12], recall and economic efficiency are taken into account [34]. Bayesian Learning when taken with Dempster-Shafer Theory resulted in 98% True Positives and less than 10% False Positives when transaction history repository was implemented using the simulator [35, 36].

Although ANN detects frauds faster, Bayesian Belief is better as it is able to detect 8% more frauds as recorded on data provided by Serge Waterschoot at Europay International when only True Positives and False Positives were taken under consideration [11, 37, 38].

When dataset is highly imbalanced and independent of the rate of frauds Bagging Ensemble Classifier is one of the most stable approaches and has a high fraud catching rate. This was an observation made on the real-world credit card dataset obtained from USCO-FICO competition [39, 40], where factors were the fraud Catching frequency, no of false alarms raised, balanced classification cost and Matthews correlation coefficient [9, 41, 42, 2, 43].

The paper shows that the machine learning can be tested together on a preprocessed dataset and their average accuracies can overall give better results overcoming the limitations of each of the techniques (Table 1).

Table 1 Some previous findings from other papers

Reference no.	Technique used	Dataset used	Pre-processing	Performance metrics	Result
[11, 44]	Bayesian Belief, ANN	Provided by Serge Waterschoot at Europay International		TP, FP	Bayesian Belief, better than ANN. 8% more frauds detected. But ANN detects faster
[45, 26, 27, 28, 29, 30]	Deep Networks by training a deep network	German Credit Data		Accuracy, Variance	High Accuracy handing data granularity
[46, 33, 13]	Decision Trees, SVM	Nation banks credit card warehouse		Accuracy	Decision tree outperforms SVM
[47]	Cost-sensitive decision tree	Banks credit card data warehouses		Saved Loss Rate (SLR)	Saved much more financial resources, outperforms traditional classifier in number of frauds detected
[48, 49]	Fuzzy clustering and neural networks	Developed by Panigrahi [35].		FP, TP/Sensitivity, TN/Specificity	Up to 93.9% TP and less than 6.10% FP
[50, 10]	K-Nearest Neighbor	Real data from private bank		Recall, F-measure, Specificity, Accuracy, Precision	Performance is determined on the grounds of metrices

3 Model-Proposed

Each of the techniques had their own demerits as observed in various papers, like Neural networks have high accuracy and detection rate but are expensive to train and work well on large data sets whereas KNN gives good results with small datasets only. Therefore this paper specifies a model that works well under all circumstances with all types of datasets. This model combines the techniques of ANN, KNN and decision Trees and for each new incoming transaction the new label is assigned by

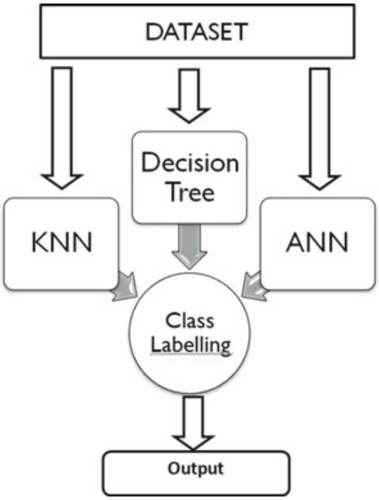


Fig. 1 Proposed model

taking the majority of the labels from the output of these techniques and the accuracy of the model is taken by taking the average of accuracies of each of the individual techniques (Fig. 1).

4 Experimental Results

4.1 Experimental Setup

Dataset. The European card holder transactions were taken as dataset in September 2013, made using credit cards. It presents transactions over the span for couple of days with average transsactions of 284,807.004 there were 492.01-fraud. 0.17201% of total transactions were highly unbalanced fraud accounts.

Preprocessing. PCA helps in acquiring principal component values, v1, v2, …v28 and contains only numerical values. The only features untouched i.e. is the time and amount. the time shows the time between the two transactions and the amount is the amount taken in each transaction and the Feature ‘Class’ that assigns the 1 for fraud and 0 for otherwise.

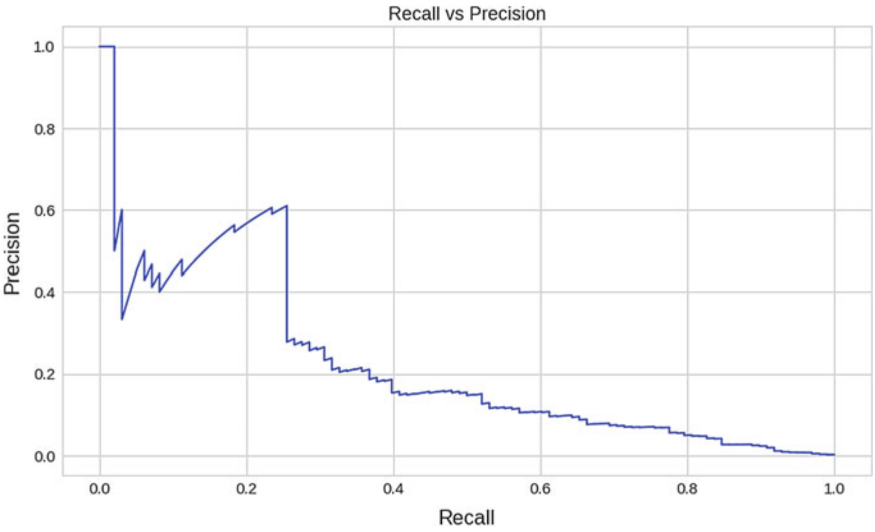


Fig. 2 Precision recall curve of the model

4.2 Result and Comparative Analysis

Figure 2 displays high precision and high recall curve which clearly shows accurate results on classification and a hghest recall. High precision is the LFPR stating the low false positive rates and High recall means LFNR stating the low false negative rate.

$$\text{Precision} = \frac{\text{True}_{\text{positve}}}{(\text{True}_{\text{ positve}} + \text{false}_{\text{ positve}})}$$

$$\text{Recall} = \frac{True_{positve}}{(True_{positve} + false_{negative})}$$

Table 2 shows the accuracies observed in various other research papers in the past wherein these individual techniques have been applied to the same European dataset for labelling transactions as fraudulent. The accuracy of our model combining these

Table 2 Comparison of accuracies

Paper Ref.	Technique	Accuracy (%)	Our model's accuracy (%)
[16]	Decision trees	94.3	93.48
[21]	KNN	95	96.96
[22]	ANN	96.40	96.55

techniques gives an accuracy of 95.66%, at the same time strengthening our model by overlooking the drawbacks of these techniques individually.

5 Conclusion

In this contribution, we developed an average accuracy-based model combining three different techniques: K-nearest neighbor, neural networks and decision trees for detecting credit card frauds. Our model is believed to works fairly good for all sizes and kinds of the dataset. This fraud detection model has been shown to provide substantial improvements in accuracy, thus overcoming the limitations of each of the models individually and making its use significant.

References

1. Statista the Statistic Portal, https://www.statista.com/topics/871/online-shopping/, March 14, 2017
2. Singh AK, Kumar J (2019) Secure and energy aware load balancing framework for cloud data centre networks. Electron Lett. https://doi.org/10.1049/el.2019.0022
3. Zheng L et al (2018) A new credit card fraud detecting method based on behavior certificate, IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), Zhuhai, pp 1–6
4. Delamaire L, Abdou HAH, Pointon J (2009) Credit card fraud and detection techniques: a review. Banks Bank Syst 4(2):57–68
5. Gupta I, Singh AK (2019) Dynamic threshold based Information leaker identification scheme. Inform Process Lett 147:69–73
6. Bhattacharyya S, Jha S, Tharakunnel K, Westland J (2011) Data mining for credit card fraud: a comparative study. Decis Support Syst 50:602–613
7. Suraj P, Varsha N, Kumar SP (2018) Predictive modelling for credit card fraud detection using data analytics. Procedia Comput Sci 132:385–395
8. Pearce J, Ferrier S (2000) Evaluating the predictive performance of habitat models developed using logistic regression. Ecol Model 133(3):225–245
9. Zareapoor M, Shamsolmoali P (2015) Application of credit card fraud detection: based on bagging ensemble classifier. Procedia Comput Sci 48:679–686
10. Sudha TRC (2017) Credit card fraud detection in internet using K nearest neighbour algorithm. IPASJ Int J Comput Sci 5(11)
11. Maes S, Tuyls K, Vanschoenwinkel B, Manderick B (2002) Credit card fraud detection using bayesian and neural networks
12. Ho TK (1995) Random decision forests, in: Proceedings of 3rd International Conference on Document Analysis and Recognition, vol 1, pp 278–282, IEEE, August-1995
13. Burges CJ (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Disc 2(2):121–167
14. Zhang ML, Zhou ZH (2007) A lazy learning approach to multi-label learning. Pattern Recogn 40(7):2038–2048
15. Mishra A, Ghorpade C (2018) Credit card fraud detection on the skewed data using various classification and ensemble techniques, in: 2018 IEEE International Students' Conference on Electrical Electronics and Computer Science (SCEECS), pp 1–5

16. Lakshmi SVSS, Kavilla SD, Machine learning for credit card fraud detection system
17. Duman E, Özçelik M (2011) Detecting credit card fraud by genetic algorithm and scatter search. Expert Syst Appl 38:13057–13063
18. Shin KS, Lee YJ (2002) A genetic algorithm application in bankruptcy prediction modeling. Expert Syst Appl 23(3):321–328
19. Malini N, Pushpa M Analysis on credit card fraud identification techniques based on KNN and outlier detection, in: 2017 Third International Conference on Advances in Electrical Electronics Information Communication and Bio-Informatics (AEEICB), pp 255–258
20. Navamani C, Krishnan S Credit card nearest neighbor based outlier detection techniques
21. Kazemi Z, Zarrabi H Using deep networks for fraud detection in the credit card transactions, in: 2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI), pp 630–633
22. Wang C, Wang Y, Ye Z, Yan L, Cai W, Pan S Credit card fraud detection based on whale algorithm optimized BP neural network, in: 2018 13th International Conference on Computer Science & Education (ICCSE), pp 1–4
23. Gupta I, Singh AK (2019) An integrated approach for data leaker detection in cloud environment. J Inf Sci Eng
24. Gupta I, Singh N, Singh AK (2019) Layer-based privacy and security architecture for cloud data sharing. J Commun Softw Syst (JCOMSS) 15(2)
25. Wang Y, Adams S, Beling P, Greenspan S, Rajagopalan S, Velez-Rojas M, ankovski S, Boker S, Brown D (2018) Privacy preserving distributed deep learning and its application in credit card fraud detection, 1070–1078
26. Reshma RS (2018) Deep learning enabled fraud detection in credit card transactions. Int J Res Sci Innov (IJRSI)
27. Pandey Y (2017) Credit card fraud detection using deep learning. Int J Adv Res Comput Sci, May–June
28. Dal Pozzolo A, Boracchi G, Caelen O, Alippi C, Bontempi G (2018) Credit card fraud detection: a realistic modeling and a novel learning Strategy. IEEE Trans Neural Netw Learn Syst 29(8):3784–3797
29. Kazemi Z, Zarrabi H (2017) Using deep networks for fraud detection in the credit card transactions, in: IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI), Tehran, 2017, pp 0630–0633
30. Lipton ZC, Berkowitz J, Elkan C (2015) A critical review of recurrent neural networks for sequence learning, arXiv preprint arXiv:1506.00019
31. Fine S, Singer Y, Tishby N (1998) The hierarchical hidden Markov model: analysis and applications. Mach Learn 32(1):41–62
32. Gupta I, Singh AK (2018) A probabilistic approach for guilty agent detection using bigraph after distribution of sample data. Procedia Comput Sci 125:662–668
33. Swain PH, Hauska H (1977) The decision tree classifier: design and potential. IEEE Trans Geosci Electron 15(3):142–147
34. De Sá A, Pereira A, Pappa G (2018) A customized classification algorithm for credit card fraud detection
35. Panigrahi S, Kundu A, Sural S, Majumdar A (2009) Credit card fraud detection: a fusion approach using dempster–shafer theory and bayesian learning. Inform Fusion 10:354–363
36. Cheng J, Greiner R (2011) Learning bayesian belief network classifiers: algorithms and system, in: Conference of the Canadian Society for Computational Studies of Intelligence, June 2011, pp 141–151. Springer, Berlin, Heidelberg
37. Kang F, Dawei C, Yi T, Liqing Z (2016) Credit card fraud detection using convolutional neural networks, 483–490
38. Kumar J, Singh AK (2018) Workload prediction in cloud using artificial neural network and adaptive differential evolution. Future Gener Comput Syst 81:41–52
39. Jurgovsky J, Granitzer M, Ziegler K, Calabretto S, Portier PE, He L, Caelen O (2018) Sequence classification for credit-card fraud detection. https://doi.org/10.1016/j.procs.2015.04.201

40. Bayer JS (2015) München, Technische Universität München. Learning sequence representations, Diss
41. Gupta I, Singh AK (2017) A probability based model for data leakage detection using bigraph, in: 7th International Conference on Communication and Network Security (ICCNS-2017), Tokyo, Japan, ACM, 2017
42. Gupta I, Singh AK (2019) A confidentiality preserving data leaker detection model for secure sharing of cloud data using integrated techniques, in: Seventh International Conference on Smart Computing and Communication Systems (ICSCC), Sarawak, Malaysia, pp 1–5
43. Kumar J, Singh AK (2019) Cloud resource demand prediction using differential evolution based learning scheme, in: 2019 Seventh International Conference on Smart Computing and Communication Systems (ICSCC), Curtin University, Miri
44. Patidar R, Sharma L (2011) Credit card fraud detection using neural network. Int J Soft Comput Eng (IJSCE) 1:32–38
45. Roy A, Sun J, Mahoney R, Alonzi L, Adams S, Beling P (2018) Deep learning detecting fraud in credit card transactions, in: Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, vol 2018, pp 129–134
46. Yusuf S, Duman E (2011) Detecting credit card fraud by decision trees and support vector machines, in: IMECS 2011 – International Multiconference of Engineers and Computer Scientists 2011, vol 1, pp 442–447
47. Sahin Y, Bulkan S, Duman E (2013) A cost-sensitive decision tree approach for fraud detection. Expert Syst Appl 40:5916–5923
48. Behera TK, Panigrahi S (2015) Credit card fraud detection: a hybrid approach using fuzzy clustering, neural network, in: Second International Conference on Advances in Computing and Communication Engineering, Dehradun, pp 494–499
49. Xie XL, Beni G (1991) A validity measure for fuzzy clustering. IEEE Trans Pattern Anal Mach Intell 8:841–847
50. Sanaz M, Mehdi S (2018) Cost-sensitive payment card fraud detection based on dynamic random forest and k-nearest neighbors
51. Srivastava A, Kundu A, Sural S, Majumdar A (2008) Credit card fraud detection using hidden markov model, in: IEEE Transactions on Dependable and Secure Computing, vol 5(1), pp 37–48
52. Onukwugha C (2018) Data mining application in credit card fraud detection system

Real-Time Human Locator and Advance Home Security Appliances

Anand Kesharwani, Animesh Nag, Abhishek Tiwari, Ishu Gupta, Bharti Sharma, and Ashutosh Kumar Singh

Abstract In this modern era, human safety has become an important issue. In the past years, crimes against girls and children have raised significantly. In this paper, a novel prototype Human Location Monitoring System is implemented using GPS and GSM. Through this device, the victim child can send their location, and also parents can check their children's location. The model is not only restricted to humans but also can be fitted in any vehicle or any device or thing which you want to keep off-track. The results show that the proposed model is far better than the existing solution which uses Bluetooth, RFID, and Wi-Fi technology. It is a desirable device for people as it has a range of the whole world and is cost-efficient.

Keywords Safety system · IoT · Arduino Uno · GPS · GSM

1 Introduction

As per the data of Indian statistics, the number of crimes in India is increasing tremendously. Figure 1 represents the number of crimes occurred in the country during the past years from 1953 to 2013 [1, 2]. This tremendous increment is noticed

A. Kesharwani · A. Nag · A. Tiwari · I. Gupta (✉) · B. Sharma · A. K. Singh
National Institute of Technology, Kurukshetra, India
e-mail: ishugupta23@gmail.com

A. Kesharwani
e-mail: anandkesharwani100@gmail.com

A. Nag
e-mail: nag.animesh56@gmail.com

A. Tiwari
e-mail: abhitiwari5796@gmail.com

B. Sharma
e-mail: bhartivijaysharma310@gmail.com

A. K. Singh
e-mail: ashutosh@nitkkr.ac.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_4

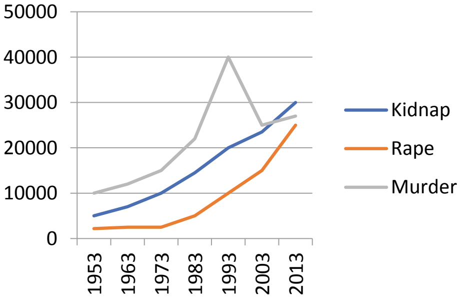


Fig. 1 Number of crimes in India from 1953 to 2013

because there is no fear of being caught [3]. Therefore, the security of children has become a major concern for their parents [4–10]. In this paper, a cost-effective and reliable Human Locator Monitoring System (HLMS) is developed. The proposed system is composed of GPS (Global Positioning System), GSM (Global System for mobile), microcontroller and it can track the location of the victim with the help of Google map application. The proposed model is a new innovative device which will provide personnel security to its client.

Also, there are three additional components along with HLMS which makes it a complete home security package. The component focuses to ease our work and prevent us from robbery or short circuit that may happen in the house accidentally anytime of which you may be completely unaware. The three components are

(1) Switching on or off lights of the home by mobile: Through this device, we can switch the lights on or off of our homes from anywhere through our mobile.
(2) Intruder Detector System: This device sends a message to the owner of the house when an intruder enters your house.
(3) Fire Detector System: This is a device which prevents your house from burning from the fire which may occur due to a short circuit or by a cylinder burst and send a message and calls the owner as soon as fire burst out.

2 Related Work

There are several children tracking system which is based on Bluetooth, RFID, Wireless LAN, and GPS. The system which is based on RFID is limited to a certain radius like a shopping mall, school, or garden [3]. Mori et al. in his paper proposed a child-tracking system which worked on Bluetooth technology and used android terminals. The Bluetooth enabled communication among different android terminals and configures a Bluetooth MANET. This system is good but if the child losses the mobile or the mobile is snatched from him then there is a problem [11]. Al-Suwaidi and Zemerly gave an idea on a mobile application which provides location information to family members using GPS and client–server approach. It also alerts the user if their friends are nearby [12]. Pham Hoang in his paper proposed a model to track a vehicle using GPS and GSM. The model is very good, and we can track our vehicle any time using this model. The disadvantage with their device is the size is not small and can easily be identified [13]. Rashed in his paper proposed a model that can track and lock the speed of a vehicle when it passes a certain speed or location [14]. Nilesh Dhawale, Mahesh Garad, and Tushar Darwatkar proposed in their papers to use a mobile phone instead of any other device. They used the GPS-enabled mobile phone and capture the location using the GPS of the mobile phone. The advantage of their proposed model is that no special device needs to be taken along with the victim. The disadvantage of their model is if the kidnappers snatch their mobile as usually the first thing the kidnappers do is to snatch all belonging of the victim. So in this case, if the victim does not have the mobile phone then their proposed model would not work [15].

3 Proposed Model

3.1 Components of the Model

This section provides a detailed description of the various hardware devices that are incorporated into this device.

(a) **Arduino Uno**: The Arduino Uno is based on ATmega328 and is a microcontroller board. It has 20 different digital input and output pins out of which 6 are used as analog input pins and 6 are used as PWM output pin. It also has a 16-MHz resonator and an in-circuit system programming (ICSP) header. The Arduino is fitted with a USB port, a power jack, and a reset button. It contains all those components needed to support the microcontroller. We just need to connect it with a USB cable and plug the other end of the USB cable to the computer or we can also provide power to it with an AC to DC adapter or a 12-v battery.
The Arduino Uno (see Fig. 2) is slightly better than the boards developed earlier as it does not use FTDI USB to serial driver chip. It uses ATmega16U2 programmed as a USB to serial converter. It has its own USB bootloader through which advanced users can reprogram it. There is an enormous support community and a large set of support libraries and hardware add ones, for instance, you can easily make the Arduino Uno from wired to wireless. It makes the Arduino a great platform for users to build embedded electronics.

(b) **SIM900**: The device is a very compact and reliable wireless module. The SIM900A (see Fig. 3) is a dual-band GSM/GPRS solution which can be embedded in applications where ever needed. It works on 1800 MHz and provides voice modulation, SMS, data exchange, and fax service. It consumes very low energy. It has a configuration of (24 * 24 * 3) mm. The small size makes it fit in a very small space that is it can be fitted where the application demands compact design.

Fig. 2 Arduino Uno

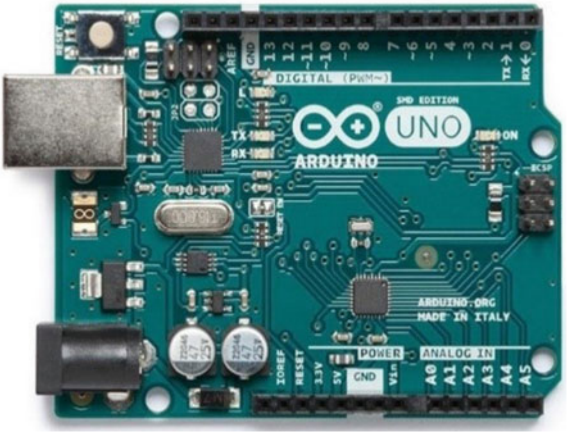
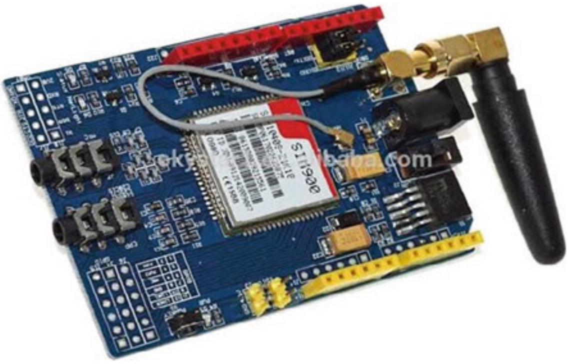


Fig. 3 SIM900A



It has a dual-band of 900/1800 MHz. It consumes very low power approximately 1.5 mA in sleep mode. It can operate in temperature ranging from -40 degrees to +85 degrees. There is a status indicator D5 which blinks continuously whenever a call comes, otherwise it does not blink. There is another LED called D6 which if constant and not blinking, then it means network connection has been established; otherwise, if it is blinking continuously, then it means it is trying to connect to the network.

(c) **NEO-6 M GPS**: The NEO-6 M (see Fig. 4) is a GPS module which receives GPS coordinates on activation. It has a built-in (25*25*4) mm ceramic antenna through which searches for satellites. There are a built power and a signal indicator through which you can check the status of your GPS module. It also provides a data backup battery which saves the coordinates data when the main power goes off accidentally.

It can be used in projects or applications for automatic returning or going to a particular spot. It has four pins: (1) RX, (2) TX, (3) GND, (4) VCC. To connect it

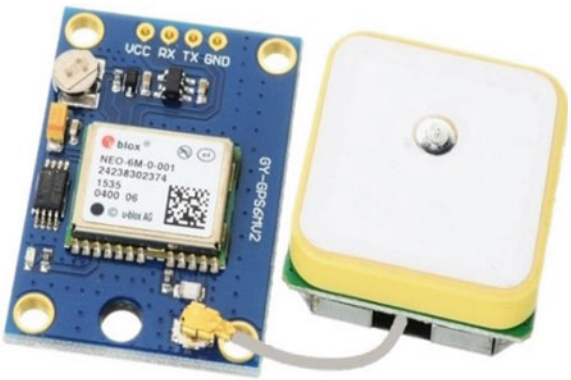


Fig. 4 NEO-6 M GPS

to Arduino, we need to connect the RX, TX, GND, VCC pins of GPS with the TX, RX, GND, VCC pins of Arduino, respectively.

3.2 Block Diagram Operational Summary

The flowchart of the model is designed in Fig. 5 which describes briefly the step-by-step procedure of the proposed model.

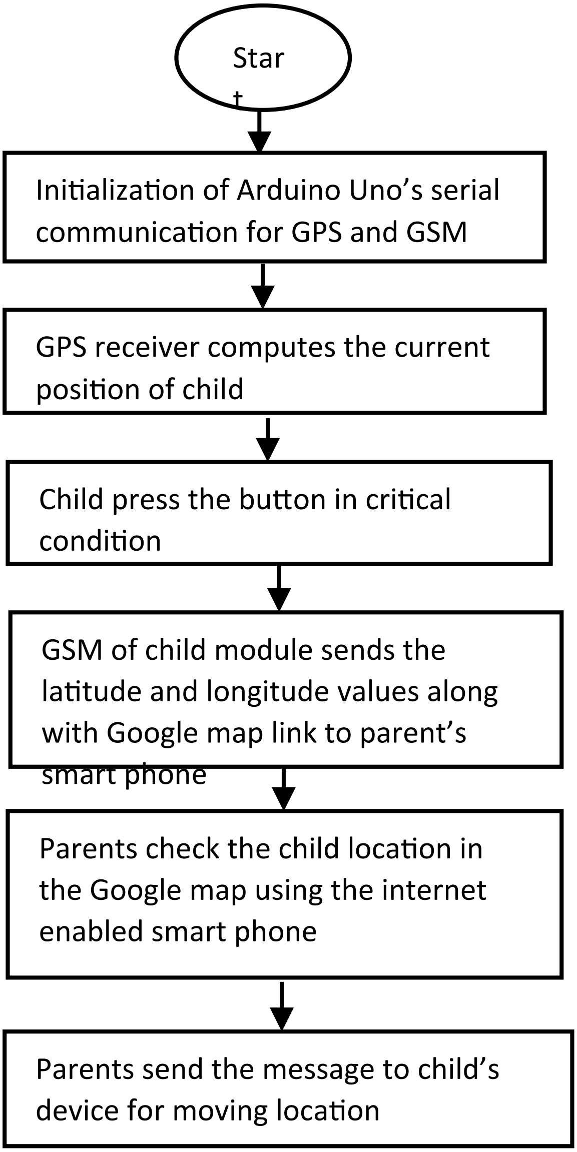
The device is simple and concise. The only thing needed is to wear this device which will be fitted inside a wristwatch. If a person is kidnapped or is struck in a flooded area, then he/she can just press a button on the wristwatch and then a message will be sent automatically to the family members and friends stating the situation along with the real-time location where the person is presently stuck. If the location is in motion then he/she can send a message again with the new location, which can be viewed directly with Google map in one click.

By sending a simple message on the device id number, the present location of the person in distress can be known that is the family member will just have to send a message “Location” and within seconds the satellite location of the device with exact coordinates, that is, latitude and longitude will be sent to the person who requested for it. With Google maps, you can view the person’s location with the exact building or place he/she is in. Here, the ID is a phone number which will be given on the time of purchase of this device and hence will be known to the family members. It can also be fitted in any vehicle, mobile device, or any equipment that we want to track so this device has a universal application and hence a desirable device for most people.

Along with this device, there are are additional components along with HLMS which makes it a complete home security package. The three components are

(1) Switching on or off lights of home by mobile.
(2) Intruder Detector System

Fig. 5 Flowchart of the proposed model


Star
t
Initialization of Arduino Uno's serial communication for GPS and GSM
GPS receiver computes the current position of child
Child press the button in critical condition
GSM of child module sends the latitude and longitude values along with Google map link to parent's smart phone
Parents check the child location in the Google map using the internet enabled smart phone
Parents send the message to child's device for moving location

(3) Fire Detector System

The first component is that we can switch the lights on or off of our homes from anywhere, for instance, if we forgot to switch the lights off of our home in the hustle to go to office or any other place, then we can switch off your home's light through a simple message sent by our phone. Our mobile number will be registered first by the device which can be done by connecting the device with a computer via a USB cable.

We can switch "on" the lights also. The message to be sent is simple, for example, "*light1 on" or "light2 on" or "all lights on" or "all lights off.*" The IoT device is attached with the main power cable and when it receives a particular message then it switches the light ON/OFF according to the message. The power cable of each fan or light is passed through a relay board which stops or lets the current flow.

The second component is the intruder detector system which sends a message to the owner of the house (that is, on the registered mobile number) when an intruder enters your house. For instance, if some thief enters your house while nobody is at house, then a message will be delivered to your registered phone number. This works on PIR sensors which if anybody comes across it will detect the movement and will send a message along with a phone call to the owner of the house immediately.

The third component prevents your house from burning from fire. For instance, if fire breaks out at your home then as soon as the device gets the sensitivity of heat, then a message will be delivered to the registered mobile number of the owner. This is a very helpful device as everything gets burnt when the building is on fire. This device works on the phenomenon that when the temperature of a thermometer rises and crosses a threshold value and then action takes place which is indicated by the microprocessor device inside it and it sends a message and a phone call to the owner of the house indicating that the house is in the fire.

3.3 System Design

The proposed system model is designed using the devices Arduino Uno, GPS, GSM module, Male-to-Female Wires, and Button. These devices are connected through various connections as shown in Fig. 6. The distinct connections that have been performed among the devices are represented as follows:

(a) **The connection between GPS and Arduino Uno**

vcc(GPS) → vcc(Arduino)
gnd(GPS) → gnd(Arduino)
tx(GPS) → pin3(Arduino)
rx(GPS) → pin4(Arduino)

(b) **Connection between SIM900 and Arduino Uno**

vcc(SIM900) → vcc(Arduino)
gnd(SIM900) → gnd(Arduino)
tx(SIM900) → rx(Arduino)
rx(SIM900) → tx(Arduino)

(c) **Connection between Button and Arduino uno**

gnd(Button) → gnd(Arduino)
vcc(Button) → vcc(Arduino)
pin3(Button) → pin8(Arduino)

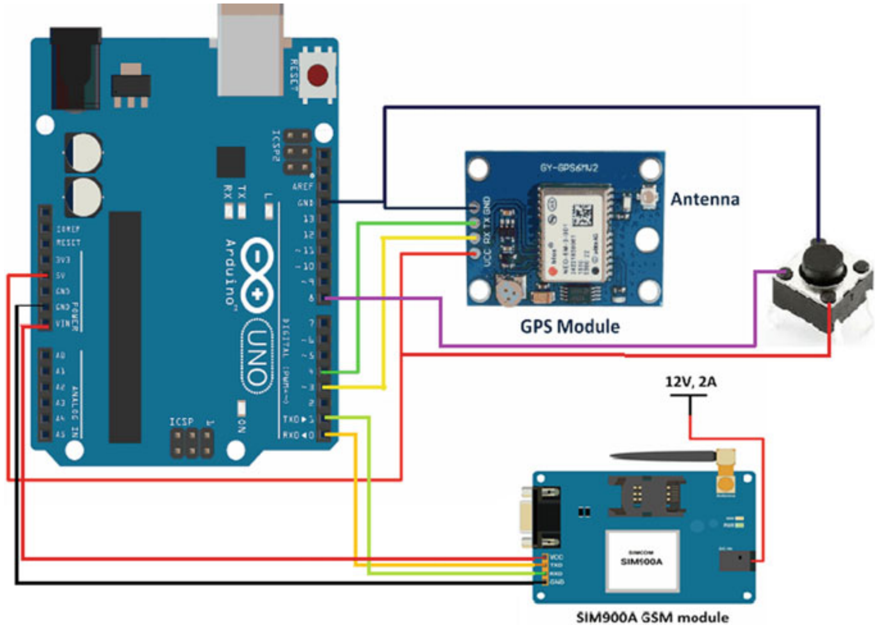
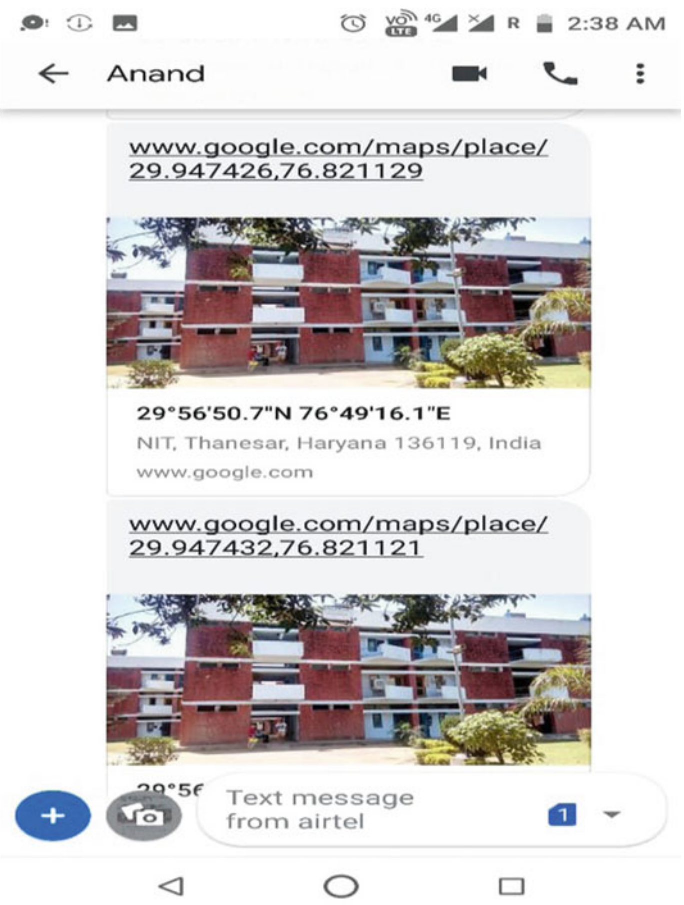


Fig. 6 Interfacing Diagram between Arduino UNO, GPS Module, and SIM900A GSM Module

4 Performance Analysis

4.1 Experimental Result

(a) **Message from the victim**: Fig. 7 is the message which is received by the receiver (which can be parents or guardians). The message contains a Google map link that is www.google.com/maps/place/29,947426,76,821129. By clicking on the link, one can get the exact location of the victim. This message is like a normal message and can be forwarded to anybody, for instance, to the police. This message depicts how easily one can get and track the exact location that is longitude and latitude of a particular area. These locations are provided by four satellites and hence are exact. One can get this location when the victim (which can be a child or a girl or anybody) presses the button on the IoT device. The image shown here is our hostel that is the victim's location or the building where the victim is kept. The line below is the longitude and latitude of a point where the victim is kept. The "NIT, Thanesar, Haryana 136119, India" represents the address of the location with the building name, city name, state name, pin code of that city, and country name. The difficulty arrives when the parents or police arrive at the location and they cannot find the exact room number or the exact floor number. The victim can be anywhere in the 10-meter radius. The building also is the front entrance that is the victim might be in another block but the



www.google.com/maps/place/29.947426,76.821129

29°56'50.7"N 76°49'16.1"E
NIT, Thanesar, Haryana 136119, India
www.google.com

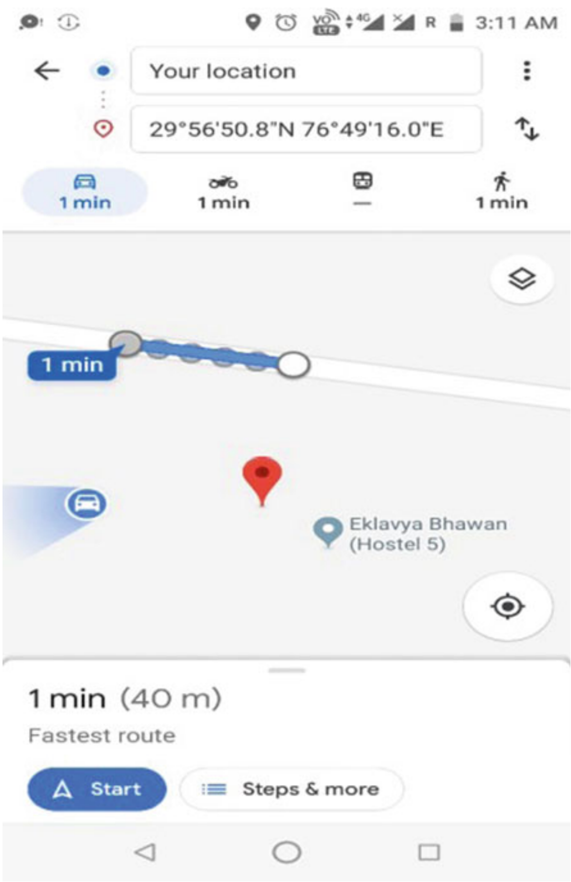
www.google.com/maps/place/29.947432,76.821121

29°56

Fig. 7 Message from victim

image will be shown of the main entrance. These are the few difficulties but these are very minor shortcomings as it is very advantageous in finding the victim in the whole city or country rather than in 10-meter circle or different floors.

(b) **Messages received by the receiver**: Fig. 8 is the window which appears after you click on the link. The window shows the distance between your location and the victim's location. It also shows the shortest route which should be taken to reach the victim's location and the time to reach between the two destinations. Distance, time, and route by car, bike, train, bus, and by walk are shown in the picture above. It is very helpful as one can reach easily and without a hustle. As the coordinates, both longitude and latitude, are exact (as they are received from four satellites), one will reach the exact building. The problem as discussed



Your location

29°56'50.8"N 76°49'16.0"E

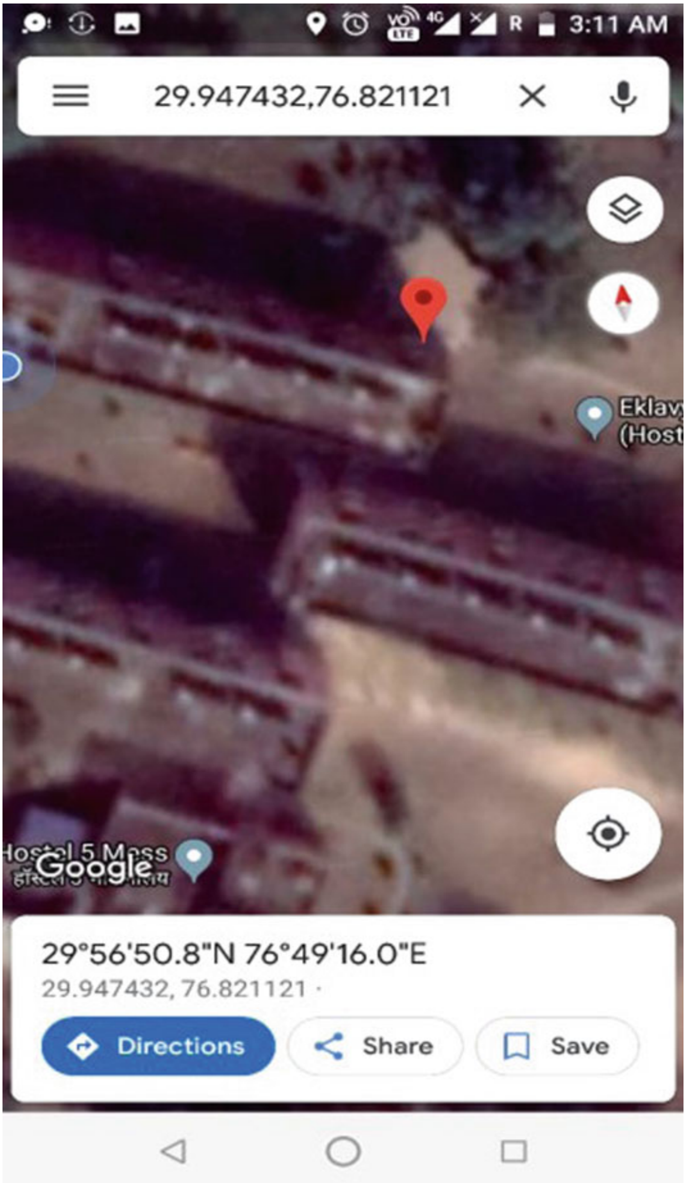
1 min (40 m)

Fastest route

Fig. 8 On clicking the message from the victim

earlier will come to find out which block in the building or which floor or in which room the victim is exactly.

(c) **Satellite mode of the message from victim**: Fig. 9 is the picture which shows the satellite image of the location of the victim. This image also shows that the victim is in the first block and the topmost floor and in the corner most room on the left side. This is the most appropriate description of the location. This type of description is available only if the device is in direct contact of the satellite without any medium in between that is the device is under open air. Also, there can be a misplacement of the location which can be in the range of 10 meters approximately.



29.947432,76.821121

29°56'50.8"N 76°49'16.0"E

29.947432, 76.821121 ·

Fig. 9 On clicking satellite mode

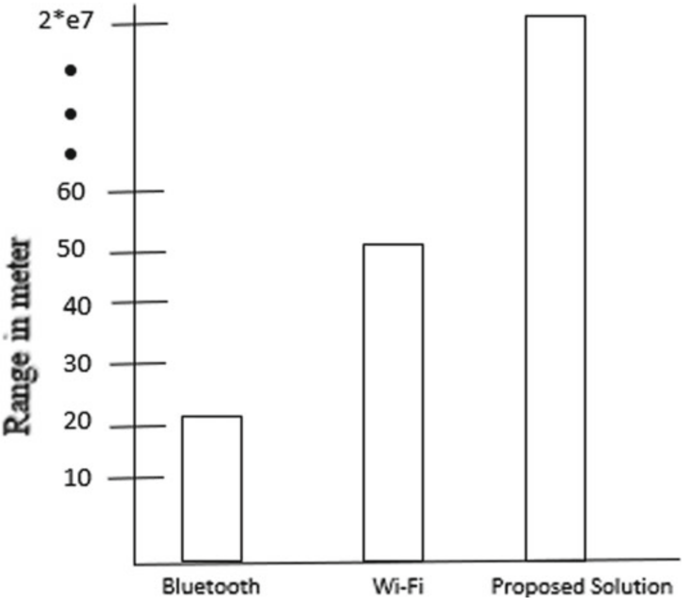


Fig. 10 The efficiency of the work

4.2 Discussion to Justify the Efficiency of the Work

The graph in Fig. 10 depicts the gigantic range (in meter) for the proposed device as well as the compared devices. It is observed that the proposed solution works on networks and hence is reachable all round the globe and is far better than previous solutions like the devices which work on Bluetooth or Wi-Fi. Also, the device is lightweight and uses very low power and hence is best suited for individuals who are kidnapped or stuck in a flooded area or stuck beneath the earth at the time of earthquake.

4.3 Applications

(1) Parents can track the location of their children, that is, if their children are kidnapped, then parents easily get the location of their children.
(2) Anyone can send a message for help if they are kidnapped or are stuck in a flooded area or struck beneath the earth when an earthquake strikes.
(3) It can be fitted in your vehicle or any device like laptop so that you can track your devices any time.

5 Conclusion

A prototype of HLMS is implemented that can monitor a human's position and give information to the parents and police using Google map application. The users can check the location of their children from anywhere through an internet-enabled smartphone. This type of device will certainly help in reducing crime against children and girls and provides means to the parents to monitor their child's whereabouts. This system can be made more efficient by making the size of the device small such that it can be fitted inside identity cards, shoes, or watches.

References

1. Sunil Yadav, Meet Timbadia, Ajit Yadav, "Crime Pattern Detection, Analysis & Prediction," University of Mumbai, Shree L.R Tiwari College of Engineering, Thane, India
2. Suhong Kim, Param Joshi, Parminder Singh Kalsi, and Pooya Taheri, "Crime Analysis Through Machine Learning," Fraser International College, Simon Fraser University British Columbia, Canada
3. Introduction to the Internet of Things security,"Standardization and research challenges"
4. Gupta I, Singh AK (2019) Dynamic Threshold based Information Leaker Identification Scheme. Information Processing Letters 147:69–73
5. I. Gupta and A. K. Singh, " An Integrated Approach for Data Leaker Detection in Cloud Environment", Journal of Information Science and Engineering, 2019
6. I. Gupta, N. Singh, and A. K. Singh, "Layer-based Privacy and Security Architecture for Cloud Data Sharing", Journal of Communication Software and Systems (JCOMSS), vol. 15, no. 2, 2019
7. Gupta I, Singh AK (2018) A Probabilistic Approach for Guilty Agent Detection using Bigraph after Distribution of Sample Data. Procedia Computer Science 125:662–668
8. I. Gupta, and A. K. Singh, "A Probability based Model for Data Leakage Detection using Bigraph", 7th Int. Conf. Commun. and Network Security (ICCNS-2017), Tokyo, Japan, ACM, 2017
9. I. Gupta and A. K. Singh, "A Confidentiality Preserving Data Leaker Detection Model for Secure Sharing of Cloud Data using Integrated Techniques", Seventh International Conference on Smart Computing and Communication Systems (ICSCC), Sarawak, Malaysia, 2019, pp. 1–5
10. Animesh Nag, Anand Kesharwani, Abhishek Tiwari, Ishu Gupta, Bharti Sharma, Ashutosh Kumar Singh, "Potential and Extention of Internet of Things", 2nd International Conference on Computer Networks and Inventive Communication Technologies, Coimbatore, India, 2019
11. Shubham Sharma, Siddhant Parashar "smart city implementation models based on IOT technology" Techno Site 2014–15
12. Ahmed M, Causevic A, Fotouhi H, Lindén M. "An Overview on the Internet of Things for Health Monitoring Systems". ResearchGate, Conference paper (2015) Malardalen University. Vasteras, Sweden
13. Mohammed ZKA, Elmustafa SAA." Internet of Things Applications, Challenges, and Related Future Technologies". World Scientific News. 2017; 67(2):126–48
14. Dhall R, Solanki VK. "An IoT based predictive connected car maintenance approach". International
15. Dhawale Nilesh, Garad Mahesh, Darwatkar Tushar (2014) GPS and GPRS Based Cost Effective Human Tracking System Using Mobile Phones. International Journal of Innovations & Advancement in Computer Science IJIACS ISSN 3(4):2347–8616

A Novel Implementation of Haptic Robotic Arm

A. Kavitha, P. Sangeetha, Aijaz Ali Khan, and K. N. Chandana

Abstract Robotics is a study of machines that are used to do different jobs. Robots are used to perform some work traditionally done by humans. Haptic technology is a growing area that will be useful for all humans, where human interactions are difficult and hazardous. The proposed system finds a range of applications in a harmful environment which can be used for medical applications and other areas, where it is difficult for humans. This method uses the technique of the master–slave concept and it is demonstrated at different stages of the performance of the glove with respect to the flex sensors.

Keywords Haptic robotic arm · Flex sensors · Microcontroller · Robotics · Embedded systems

1 Introduction

Haptic technology works on the basis of touch and motion, which can be applied in remote operations, computer simulations and it is also used to control any mechanical structure [1–4]. Robotics is a science of controlling the artificial embodies by using a processor or a controller. It is the study of robotic technology which helps mankind for its assistance. Robotics is a study that deals with the design and construction of robots to control the operation [5, 6]. The study is related to simulations of the surface on which robot moves along with the feedback from objects; there is a lot of

A. Kavitha (✉) · P. Sangeetha · A. Ali Khan (✉) · K. N. Chandana
Department of ECE, KNSIT, Bengaluru, India
e-mail: kavithaln217@gmail.com

A. Ali Khan
e-mail: aijazalikhan@gmail.com

P. Sangeetha
e-mail: psangi123@gmail.com

K. N. Chandana
e-mail: Chandu.naveen.21@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_6

mathematical and experimental models that have been done previously to develop a haptic robotic arm. This paper proposes an implementation of a haptic robotic arm that is controlled by using flex sensors mounted on the hand glove. The flex sensor captures the movement of the hand which is used to interact with the haptic robotic arm. This proposed work has a vast application in various and hazardous environments. This proposed arm can be used as a substitute for human hands and it replicates the actions. The haptic arm is controlled by the hand glove movement, whose data are collected from the flex sensors that are placed on the gloves.

2 Methodology

The flex sensors input that is mounted on the hand glove are read and, according to the input received, the DC motors will be rotated accordingly to the program that is run using the microcontroller (Figs. 1 and 2).

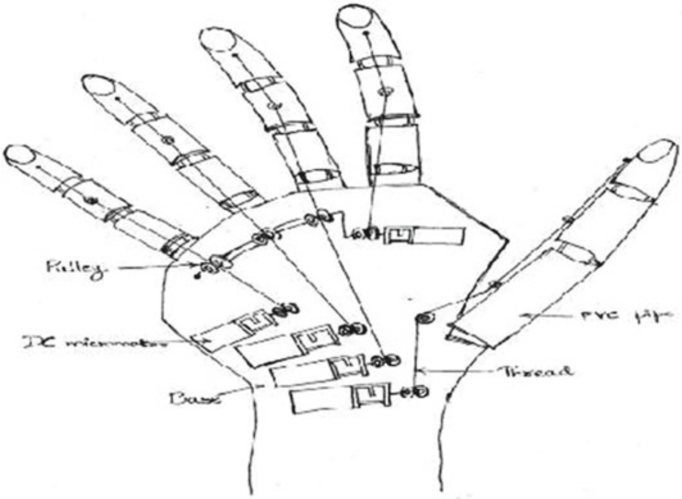


Fig. 1 Haptic robotic arm

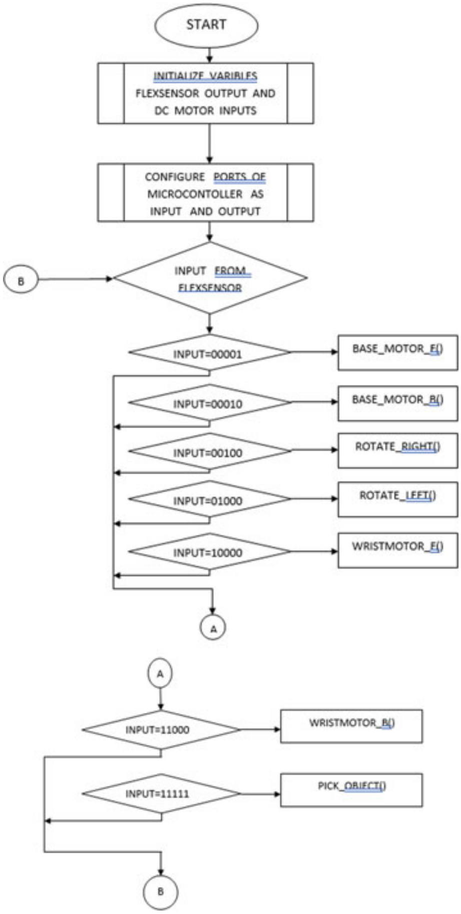


Fig. 2 Flowchart for controlling of motors

3 Block Diagram

3.1 Flex Sensor

It is a sensor which is a variable resistor and its resistance increases as the sensor is bend depending on the angle of deflection. The amount of bend varies the resistance linearly as both are directly proportional to each other.

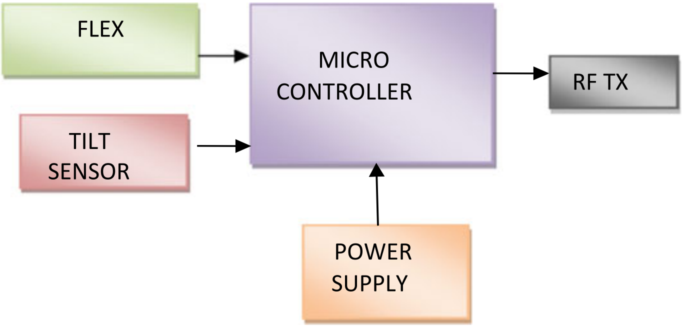


Fig. 3 Block diagram of transmitter section

3.2 *Tilt Sensor*

These are transducers which are used to produce horizontal and vertical inclinations and are interfaced with microcontroller (Fig. 3).

3.3 *Microcontroller*

Here, 8051 microcontroller is used and it is an 8-bit microcontroller where it is interfaced with encoder and decoder (Fig. 4).

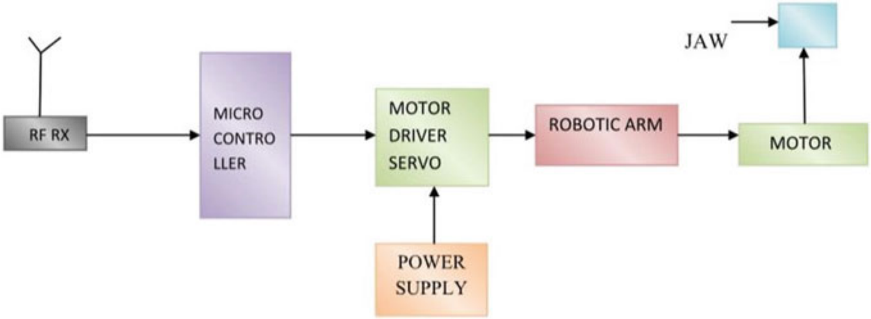


Fig. 4 Block diagram of receiver section

3.4 Power Supply

A power supply of 230 V AC RMS voltage is used which is connected to a transformer that steps down the AC voltage to the level of DC output.

3.5 RF Transmitter and Receiver

It is a wireless transceiver protocol which works with the device that supports UART.

3.6 DC Motor

It is a device that is operated to run the motor which is electrically controlled.

3.7 Gripper (Jaw)

The robotic hand is used to pick and place the object controlled by sensors. It mimics the functions performed by the human hand for holding an object, tightening the grip, etc (Fig. 5).

4 Transmitter Section

The transmitter section consists of the following components such as AT89s52 (Microcontroller), HT12E (Encoder), RF transmitter, and ADXL335 (accelerometer). The port 0 of microcontroller from P0.0 to P0.3 is the address and data lines AD0 to AD3 are connected to Y, X, Z (to measure the acceleration) and to the flex sensor to record the angle of rotation on the hand glove. The port 1 from P1.4 to P1.7 is connected to the switches through diodes to control the robot. The port 2, P2.0, is connected to the encoder (pin 13). The Port 3 from P3.5 to P3.7 is connected to the encoder Pins 10, 11, 12 for data encoding. The encoded data are sent for transmission through RF transmitter from pin 17 to data pin of RF transmitter. The inputs that are parallel in nature are first converted to the serial output and it is transmitted through RF transmitter (Tables 1 and 2).

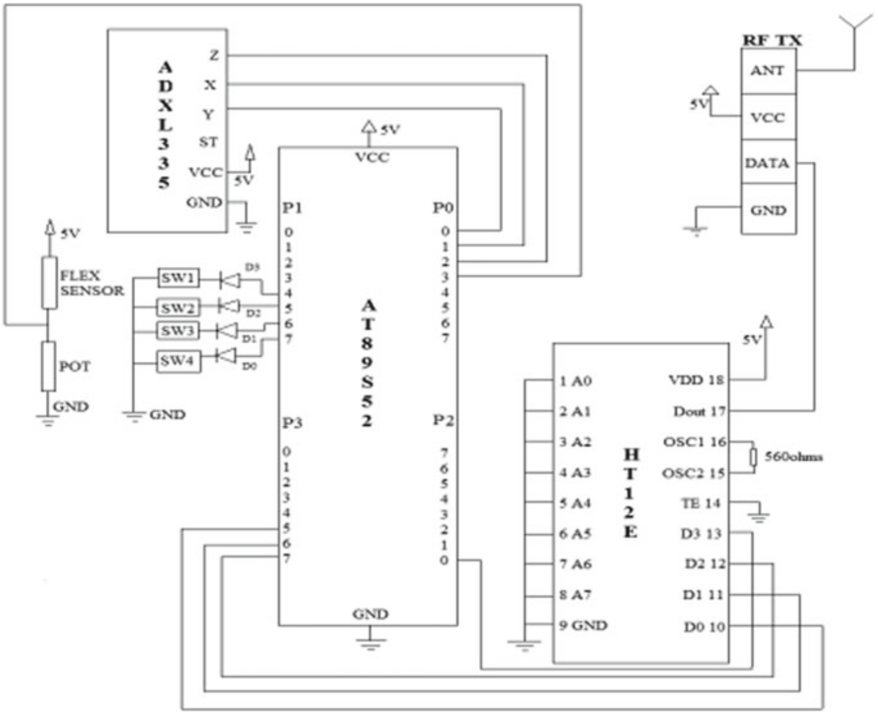


Fig. 5 Transmitter circuit diagram

Table 1 Flex Sensors Readings at different angles

Resistance (kΩ)	Voltage (V)	Angle (°)	Circuit output (V)
26.18	0.82	90	3.74
40.26	0.67	75	3.19
48.74	0.45	60	2.7
52.92	0.40	45	2.2
57.10	0.37	30	1.7
60.40	0.32	15	0.5

5 Receiver Section

RF receiver receives the signal sent from the transmitter section for processing. The data are given to decoder for decoding through Pin 14 of HT12D. The data are sent to microcontroller for processing and controlling the rotation of the motor based on the data received from the flex sensor (Fig. 6).

Table 2 Motor control through flex sensors

FLEX SENSOR INPUTS					RESULT
A F 1	A F 2	A F 3	A F 4	A F 5	
0	0	0	0	1	BASE MOTOR CLOCKWISE ROTATION
0	0	0	1	0	BASE MOTOR ANTICLOCKWISE ROTATION
0	0	1	0	0	RIGHT ROTATION
0	1	0	0	0	LEFT ROTATION
1	0	0	0	0	WRIST MOTOR CLOCKWISE ROTATION
1	1	0	0	0	WRIST MOTOR ANTI CLOCKWISE ROTATION
1	1	1	1	1	PICK OBJECT

Table 3 Flex sensor positions

Flex sensors	Fingers
AF1	Little
AF2	Ring
AF3	Middle
AF4	Index
AF5	Thumb

6 Results

Testing is done to check whether each component is working according to the design or not. Then, when the execution finishes, an outcome is checked against the expected result. The flex sensor used here gives a maximum of 75 kohms. As the flex sensor is bent the carbon resistive mounted on it expands and the surface cracks add on to the band and give more resistance. The following table shows the position of flex sensors at different angles and the resistance record at these angles is listed below (Table 1).

Table 2 shows that, if flex sensor is in the first condition, the base motor rotates clocks wise; to rotate anti clockwise, AF4 is made high, when AF3 is high it rotates in the right direction. When AF2 is high, it rotates in the left direction. When AF1 is high, wrist rotates clockwise and anti-clockwise when AF1 and AF2 are high, and when all are high it, picks the object (Figs. 7, 8 and Table 3).

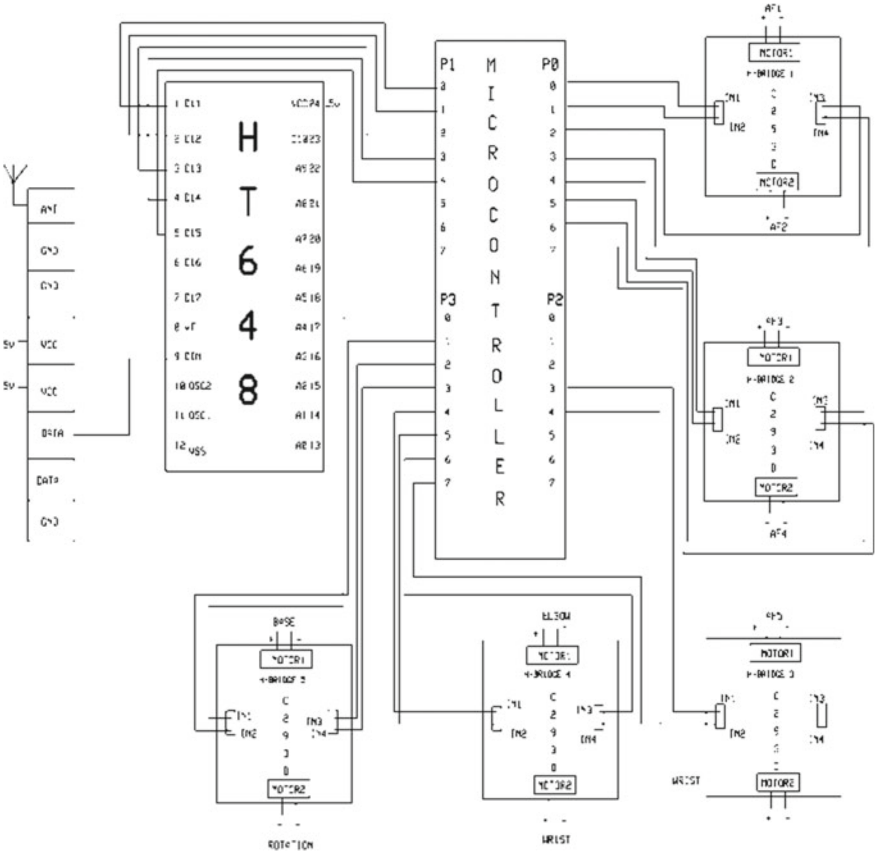
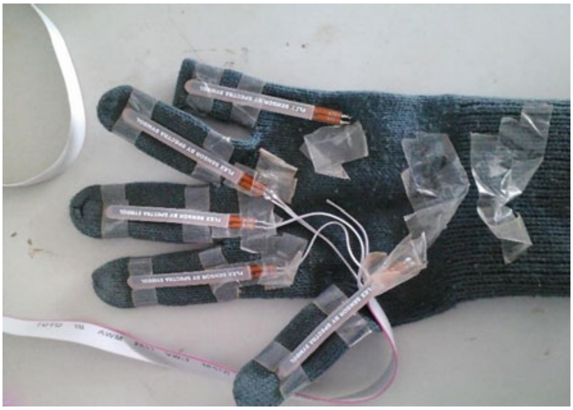


Fig. 6 Receiver circuit diagram

Fig. 7 Data glove with flex sensors



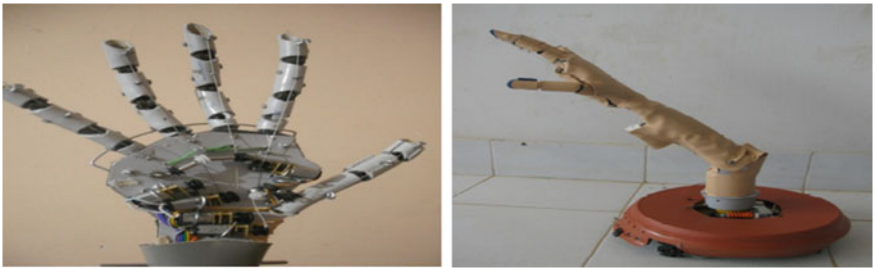


Fig. 8 Robotic arm and the model

7 Conclusion

The project model was designed keeping in mind that it should have a real-time response as the flex sensors movements. The flex sensors sense the moves and collect data from each finger, which vary with the rest depending on the amount of bend on the flex sensors, these analog values are converted to digital and given to the microcontroller for processing

References

1. Xinxing T, Hironao Y, Dingxuan Z, Tao N (2009) Haptic interaction in tele-operation control system of construction robot based on virtual reality. IEEE
2. Nisha S, Swati U, Sorabh G (2011) Based on touch: haptics technology, IEEE
3. Luca T, Rolf N, Stefania S, Amir B, Smilen D, Vincent H (2010) Audio-haptic physically-based simulation of walking on different grounds. In: Proceedings of MMSP’IO 2010 IEEE International Workshop on Multimedia Signal Processing
4. Rosenberg LB, Adel Stein BD Perceptual decomposition of virtual haptic surfaces
5. Turchet L, Nordahl R, Serafin S, Berrezag A, Dimitrov S, Hayward V (2010) Audio-haptic physically-based simulation of walking on different grounds. In: proceedings of MMSP 10 2010 IEEE International Workshop on Multimedia Signal Processing
6. Mohammed K (2010) Design of a gripper tool for robotic picking and placing. Uppsala University

A Survey on Partially Occluded Faces

Shashank M. Athreya, S. P. Shreevari, B. S. Aradhya Siddesh, Sandeep Kiran, and H. T. Chetana

Abstract Over the past decade, partially occluded face recognition has been an urgent challenge to computer visionaries due to conditions, which appear unconstrained. The main aim of the facial recognition system is to attain the ability to detect partially occluded regions of an individual's face and authenticating/verifying that face. There are existing neural networks that are proven to be perfect on analysing the patterns for constrained looks but fail to perform in analysing partially occluded faces that are common in the real world. The paper discusses the trainable Deep Learning Neural Network (DLNN) for partially occluded faces by recognizing all the possible faces in the image, either resting, posing or projecting faces and matching them across the trained datasets of DLNN and encoding the identified faces.

Keywords DL algorithm · DLNN pattern matching · Face detection · Facial analysis · Partial occlusion

1 Introduction

The face is an essential human biometrics used in everyday human communication and face detection is the process of identifying people in images or videos, which is an essential part of many biometric, security and surveillance systems. Despite the significant progress in face detection technology, it is incompatible when faced with an uncontrolled environment of occlusions, drastic illumination changes, and facial pose variations. Facial occlusion involves parts of the face hidden through particular objects/entities like sunglasses, masks, hats or scarves. Facial occlusions reduce or create inaccuracies to the performance of face detection systems. Therefore, robustness to partial occlusions is thus crucial in nowadays. This work aims to propose

S. M. Athreya · S. P. Shreevari · B. S. Aradhya Siddesh · S. Kiran · H. T. Chetana (✉)
Department of Computer Science & Engineering, Vidyavardhaka College of Engineering, Mysuru, India
e-mail: chethanaht@vvce.ac.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_7

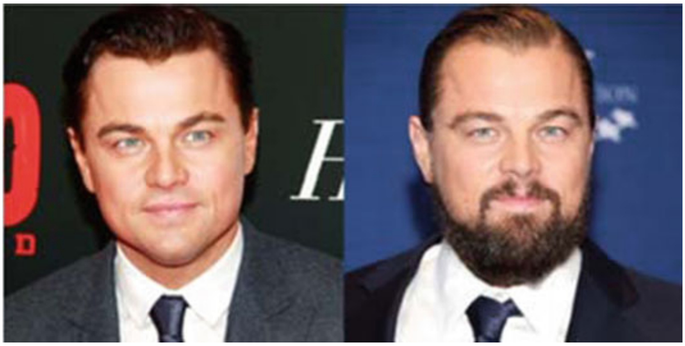


Fig. 1 Face occluded by a beard [3]

an effective detection system using Deep Learning Neural Network (DLNN) to figure out monochromatic pixel patterns based on the directional flow of luminosity (Figs. 1, 2 and 3).

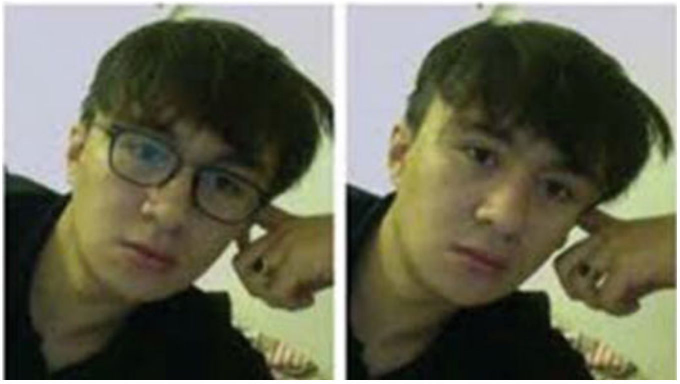


Fig. 2 Face occluded by glasses [2]

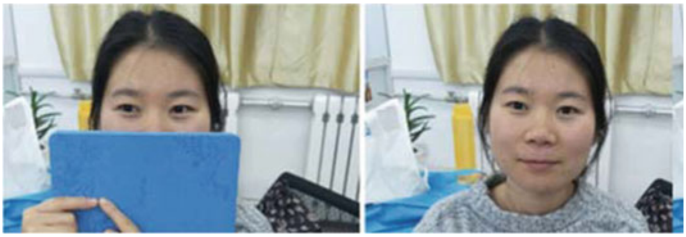


Fig. 3 Face occluded by a book [3]

2 Literature Survey

Facebook has built an exceptional capacity to recognize Facebook friends from user's posts and photos. Previously, by clicking on faces and entering a friend's username, you could tag friends in pictures. But today, Facebook magically tags all your friends in the picture for you automatically when you post a photo. Facebook algorithms can identify the faces of your Facebook friends after just a few tags. It is an incredible technology with 98% accuracy, just as good as humans are! [3] (Tables 1 and 2).

In the case of partial occlusion of the face, there are often no features identified in current discriminative or generative approaches. We interpret the reason being shortcomings of indiscriminative and generative techniques of localized facial feature detectors and appearance modelling. We propose to solve with a new method of detection that hybridizes these two methods to result in better accuracy [2]. Recognition of emotions based on facial expressions plays a vital role in various applications such as behavioural analysis, interactions between people and machines, mental healthcare services, interpersonal relationships and social surveillance. Findings from research on two public datasets showed that CENTRIST interpretation obtained strong accuracy levels for occluded and un-occluded facial expressions relative to other approaches [3].

Over the past few years, the identification of emotions based on facial expression has been attracting growing attention from the research community. Many technologies may benefit from the recognition of facial expressions, such as prediction of behaviour, personal relationships, communication between humans and computer's auto-recommenders. In this study, through the Weber Local Descriptor (WLD), we evaluate a model for emotion detection based on robust facial expressions [4].

Table 1 Comparison of different classifiers and their accuracy [13]

Features	Classifiers	Light (%)	Scarf (%)	Glasses (%)
LBPu8,2	WkNN	81.4	39.4	39.4
LBPu8,2	CBR	96.2	83.6	50.2
LBPu8,2	SVM (poly)	78.1	36.9	26.1
LBPu8,2	LR	84.8	45.0	23.4
LBPu8,2	NB	82.5	43.7	20.1

Table 2 Comparison of different methods and their accuracy [14]

Author	Year	Methods	Accuracy (%)
Seongwon	2019	Krolak	91.53
Viola	2019	REGT	92.38
Steve	2018	Geometric features	90.08
Kanade	2018	Mixture distance	89.8

3 Literature Survey Methods

There are multiple sources of work on methods to verify faces occluded partially by different objects and facial features. Face recognition and verification have seen high accuracy lately, but there is still a lot of scope for improvement with occluded faces. A comparison of the different classifiers and their accuracy is key in designing a new approach to solving the problem.

4 Applications in Partial Occlusion

Although facial recognition has historically been associated with surveillance and security, there is now a significant foray into other sectors, including retail, advertising and medicine. Shortly, the global market for facial recognition technology is expected to generate an estimated revenue of $9.6 billion with an annual compound growth rate of 21.33% [5].

1. Fraud Detection
2. Shoplifting prevention
3. Facial recognition controversy
4. Account security
5. Medication adherence
6. Target marketing

These are a few of the main applications for facial recognition systems [6].

5 Challenges in Partial Occlusion

Face Recognition has progressed considerably and has proved to be essential in many applications. Some instances are surveillance, image retrieval, access control systems, authentication and verification of personal identity to name a few. There still exist some shortcomings, which have influenced its quality of service.

Let us discuss some of the most impactful challenges to any facial recognition system [7].

5.1 Illumination

Since most image processing is done in the illumination plane of the image, this is a significant factor. For example, it has been known that a small shift in luminosity circumstances will have a huge effect on its findings. If the lighting appears to change, even if the same subject is captured with the same sensor and facial expression and posture are almost similar, the outcomes that arise may seem quite divergent [7].

5.2 Background

The subject's positioning also acts as a contributing factor to the drawbacks. A facial recognition method might not always produce the same results in outdoor conditions as it produces inside since the factors influence as soon as the positions adjust, impacting its performance variables such as personal gestures, age, etc., which make a significant contribution to these variants [8].

5.3 Pose

Facial recognition processes are extremely susceptible to changes in posture and inclination of the face. Head movements or differentiated points of view will inevitably induce facial expression differences and produce intra-class differences that make automatic face recognition a tricky game [9].

5.4 Occlusion

Face occlusions such as scars, marks, beard and moustache, apparel (glasses, gloves, mask and so on.) often interfere with a face recognition performance. The existence of such elements makes the subject complex in nature and in a real-world environment, it becomes challenging for the application to work [15].

5.5 Expressions

An important thing to consider is different expressions of the same person. Due to differences in the subject's mental state, macro- and micro-emotions find their place on the face and amid such gestures, which are many; it becomes hard to recognize them effectively.

5.6 Complexity

Established state of the art facial recognition methods rely heavily on architecture of the Convolution Neural Network (CNN) being too deep, which is very complicated and inadequate for operation on embedded real-time applications. An optimal face recognition system should be accommodating of lighting, gesture, posture and occlusion deviations. It must be accessible for vast numbers of user groups who need to capture limited images while at the same time removing complicated infrastructure during registration [10].

6 Conclusion

From the survey, we conclude that the nature of occlusion of the face can be largely varied. Hence, using a DLNN for facial verification is most promising amongst existing approaches.

Compliance with Ethical Standards.

All author states that there is no conflict of interest.

Humans/Animals are not involved in this work.

We used our own data.

References

1. Cornejo JY, Pedrini H (2016) Recognition of occluded facial expressions based on CENTRIST features. In: 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP), Shanghai, 2016, Clerk Maxwell J, A Treatise on Electricity and Magnetism, 3rd ed., vol 2. Oxford: Clarendon, 1892, pp 68–73
2. Shin J, Kim D (2014) Hybrid approach for facial feature detection and tracking under occlusion. IEEE Signal Processing Letters, 21(12):1486–1490
3. Hongxing S, Jiayi W, Peng S, Xiaoyang Z (2013) Facial area forecast and occluded face detection based on the YCbCr elliptical model. In: Proceedings 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC), Shengyang, 2013. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press
4. Ramìrez Cornejo JY, Pedrini H (2018) Emotion Recognition from Occluded Facial Expressions Using Weber Local Descriptor. In: 2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP), Maribor, 2018
5. Young M (1989) The Technical Writer’s Handbook. University Science, Mill Valley, CA
6. Aisha Azeem, Muhammad Sharif, “Survey of Face Recognition Techniques under Partial Occlusion”, The International Arab Journal of Information Technology, Vol. 11, No. 1, January 2014
7. Tanvi B. Patel, Prof. Jalpa T. Patel “Occlusion Detection and Recognizing Human Face using Neural Network” IEEE 2017
8. Rohit Tayade “Occlusion Detection Prior To Face Recognition Using Structural Feature Extraction” International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE 2017International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE2017

9. Chen YA, Chen WC, Wei CP, Wang YC (2017) Occlusion-aware face inpainting via generative adversarial networks. IEEE
10. Deng-Yuan Huang, Chao-Ho Chen, Tsong-Yi Chen, Jian-He Wu, ChienChuanKo" Real-Time Face Detection Using a Moving Camera" 201832nd International Conference on Advanced Information Networking and Applications Workshops, IEEE 2018
11. Ashwin Khadatkar, Roshni Khedgaonkar, K.S.Patnaik "Occlusion InvariantFace Recognition System", 2016 world Conference on FuturisticTrends in Research and Innovation for Social Welfare(WCFTR'16), IEEE 2016
12. Hua Wang, Xin Gu, Xiao Li, Zhe Li, Jun Ni "Occluded Face Detection Based on Adaboost Technology" 2015 Eighth International Conference on Internet Computing for Science and Engineering, IEEE2015
13. M.P. Satone, K.K. Wagh Face detection and recognition in color images" IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011
14. Foram Shah, Chandni Sharma, Shreya Patel, Abhishek More "Review of Face Detection based on Color Image and Binary Image "International Journal of Computer Applications (0975–8887) Volume 134 – No.1, January 2016
15. Asthana A, Zafeiriou S, Cheng S, Pantic M (2013) Robust discriminative response map fitting with constrained local models. In IEEE Conference on Computer Vision and Pattern Recognition, pp 3444–3451
16. Kepenekci B, Tek FB, Akar GB (2002) Occluded face recognition based on Gabor wavelets. In: Proceedings. International Conference on Image Processing, Rochester, NY, USA, pp I–I
17. Chen J, Shan S, Yang S, Chen X, Gao W (2006) Modification of the adaboost-based detector for partially occluded faces. In: 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, pp 516–519
18. Wu G, Tao J, Xu X (2019) Occluded Face Recognition Based on the Deep Learning. In: 2019 Chinese Control and Decision Conference (CCDC), Nanchang, China, 2019, pp 793–797
19. Charoenpong T, Nuthong C, Watchareeruetai U (2014) A new method for occluded face detection from single viewpoint of head. In: 2014 11th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Nakhon Ratchasima, pp 1–5

Some Effective Techniques for Recognizing a Person Across Aging

Mrudula Nimbarte, Madhuri Pal, Shrikant Sonekar, and Pranjali Ulhe

Abstract To identify a person across aging is a very challenging and interesting task. It has gained a lot of attention from the researchers as it has a wide range of real-life applications like finding missing children, renewal of passport, renewal of a driving license, finding criminals, etc. Many researchers have also proposed their own methodologies; still there is a gap to fill in. Hence, the authors have proposed some techniques for the betterment of the system performance. The first one is with GLBP as a novel feature, second with Convolutional Neural Network (CNN) and the last one is a modification of the previous method with biased face patches as inputs. CNN is found to be the perfect solution for face recognition problem over aging as there is no need for any complicated preprocessing and feature extraction steps. FGNET and MORPH II datasets are used for testing the performance of the system. All these techniques outperform available state-of-the-art methods in the Rank-1 recognition rate.

Keywords Artificial intelligence · Machine learning · Deep learning · Image processing · Aging model · AIFR

M. Nimbarte (✉) · M. Pal · S. Sonekar · P. Ulhe
JDCOEM, Nagpur, MH, India
e-mail: mrudula_nimbarte@rediffmail.com

M. Pal
e-mail: madhuridec9@gmail.com

S. Sonekar
e-mail: srikantsonekar@rediffmail.com

P. Ulhe
e-mail: pranjulhe@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_8

1 Introduction

To recognize an identity of a person from available facial images over a time span is a very challenging and. job. The process of aging drastically changes the facial shape and texture of a face, causing degradation in the performance of the system [1]. Various parameters like head position, brightness, expression make it difficult. Also, aging makes it even more difficult as it changes individually. It is mostly affected by geographical location, personal hygiene, eating habits, cosmetics, physical and mental health, etc., of a person [2]. By considering all these factors, to recognize a person across aging is difficult.

But it has very interesting real-world applications like finding missing children, renewal of passport, renewal of a driving license, finding criminals, etc. So, it needs a robust system to recognize a person across aging [3]. Two popular datasets: FGNET and MORPH are publically available for this purpose. The task becomes more difficult as there are less inter-class similarities and more intra-class variations. Hence, it has attracted researchers to provide a more robust system as it is very interesting and useful for society [4]. Many of them have proposed their own methodologies, still it is unsolved. This broad area consists of two parallel tasks: face recognition and face verification across aging. To recognize a face over aging is a multi-class problem and face verification is a two-class problem.

Bouchaffra [2] proposed a framework to identify a face over aging using α-shape-based topological features. To reduce dimensionality Kernelized radial basis function (KRBF) and for classification mixture, multinomial distributions are used. El Khiyari and Wechsler [4] introduced a robust system to recognize a face across time period with novel automatic feature extraction based on deep learning. They used a very deep CNN architecture (VGG) with 16 layers. Sajid et al. [5] attempted a novel model based on matching score space (MSS) to recognize face images separated by aging and enhanced facial asymmetry. Gong et al. [6] demonstrated a novel maximum entropy feature descriptor (MEFA) for recognizing age-variant facial images. A new feature-matching framework is also presented as Identity Factor Analysis (IFA) to improve recognition accuracy. Many studies are being presented for recognizing age separated face images over a decade using different methods [7, 8], but less are focused using CNN [9].

The aim of the paper is to focus on face recognition problem with some of the proposed techniques to improve the performance of the system. For this purpose, we have done three different experiments. The first one with novel feature descriptor (GLBP) and second with CNN, still we tried for the betterment of the performance. The last experiment involves the use of CNN with a change in traditional inputs. Instead of using the complete face as input, we used two different face patches as inputs.

The rest of the paper is organized as follows. Next, Sect. 2 provides proposed techniques to recognize a person over aging. Then Sect. 3 presents experimental details on FGNET and MORPH II standard datasets. Lastly, Sect. 4 includes conclusion and future scope.

2 Techniques for Recognizing a Person Across Aging

This section describes proposed techniques to recognize a person over aging using different approaches. The task is to recognize a person from available database images having variations in aging. It includes some conventional steps as Preprocessing, Extraction of Feature, and Image Classification. Image preprocessing helps to get better results for the system. Feature extraction is required to extract the desired feature descriptors. To recognize the identity of the person classification is required. This task belongs to the multi-class classification problem. The overall process to recognize a person across aging is shown in Fig. 1. In this paper, we have proposed three different approaches to solve the problem. The first approach involves the use of GLBP as a novel feature descriptor. The second one replaces the traditional method by using CNN. Lastly, it proposes the concept of a biased face patching approach in combination with CNN.

2.1 Using Novel Feature Descriptor

In this approach, as per the traditional methods, the input image is applied for preprocessing. As standard aging datasets contain images of varying size and brightness, it may create some problems while recognition. To bring a standard dataset in normalized form, image preprocessing is used. It needs face detection and cropping from the given input facial image. Popular Viola–Jones algorithm is generally used to detect a face. To convert the RGB image to grayscale image is the next required step. Then images are resized to 32 × 32. Histogram normalization and head pose correction are complicated preprocessing steps which are not required in this work. Figure 2 illustrates the working of the proposed model as per the first approach with GLBP feature descriptor.

The next step is to feature extraction from the normalized image. We used a combination of Gabor wavelets and LBP features for this purpose and call them as GLBP features. It gives advantages to both the features. This feature vector is then applied to reduce dimensions using principal component analysis. Lastly, k-Nearest Neighbor (k-NN) algorithm is used for classification.

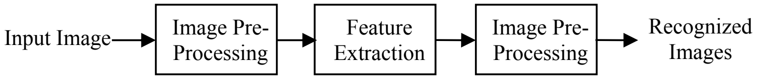


Fig. 1 Overall process for proposed methodologies

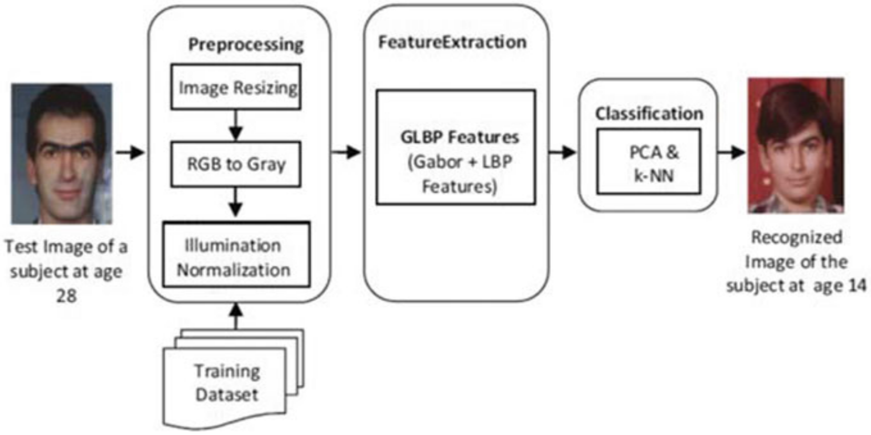

Test Image of a subject at age 28
Preprocessing
Image Resizing
RGB to Gray
Illumination Normalization
Training Dataset
FeatureExtraction
GLBP Features (Gabor + LBP Features)
Classification
PCA & k-NN
Recognized Image of the subject at age 14

Fig. 2 Proposed model for approach 1 using GLBP features [3]

2.2 Using Convolutional Neural Network (CNN)

As our aim is to improve the performance of the system, we investigate new methods to look for better solutions. Recently, the concept of deep learning is getting more popular as there is no need to apply for feature extraction and classification separately. It provides both of them a single structure. So, we tried the next approach by using convolutional neural network for solving the recognition problems over aging.

Again, we used some preprocessing steps like face detection, cropping, RGB to gray conversion, and resizing. Then for extracting features, we used seven-layer CNN architecture with three convolutional (C1, C3, C5), two sub-sampling (S2, S4), and two fully connected layers (F6, F7). It is one of the simplest networks. For convolution layer, a 5 × 5 filter is used, which is a linear operation that performs element-wise multiplication and then addition. Then sub-sampling as summing is used with 2 × 2 that results in the reduction of feature map by a factor of 2 in both dimensions. The last layers are fully connected layers where each output is connected to all inputs. An image of 32 × 32 size is used as input to this CNN architecture. Finally, the support vector machine (SVM) is used for classification as shown in Fig. 3.

2.3 Using CNN on Face Patches

Although the method proposed in the second approach is better than the first one, we tried some refinements in this method. Instead of providing complete face as input, now we provided two face patches as inputs. Periocular region (a patch with both eyes and eyebrows) is found as the most stable region of the face over aging. So, keeping this fact in mind, we have used the portion as the first patch. The second

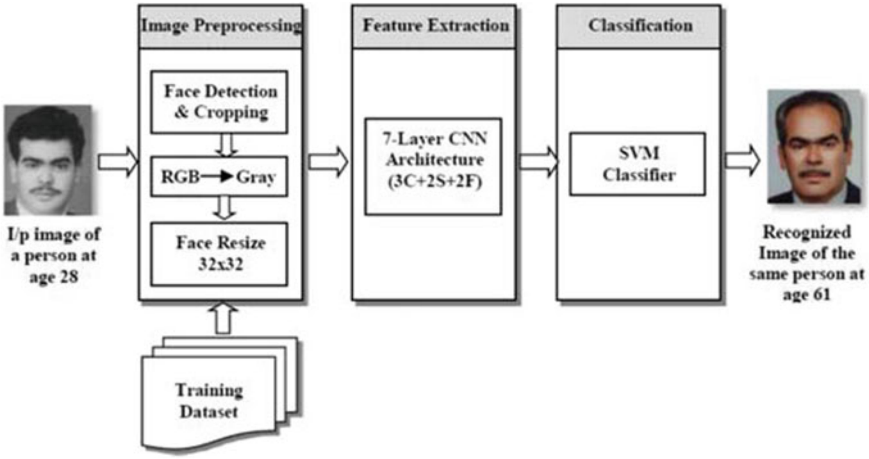


Image Preprocessing
Face Detection & Cropping
RGB→Gray
Face Resize 32x32
Training Dataset
Feature Extraction
7-Layer CNN Architecture (3C+2S+2F)
Classification
SVM Classifier
I/p image of a person at age 28
Recognized Image of the same person at age 61

Fig. 3 Proposed model for approach 2 using CNN [7]

patch includes a nose with the mouth. In this approach, two methods are followed: first with a combination of both feature vectors and second with a weighted average of both of them. For computing a weighted average of both the feature vectors, we have given higher weight to the periocular region as compared to the nose–mouth region. Again, as per the second approach, SVM is used for classification, as it is compatible with CNN. The detailed architecture is shown in Fig. 4. For this approach, we keep the same CNN architecture as discussed in Sect. 2.2.

3 Experimental Details

3.1 Experimental Setup

We proposed three different techniques to recognize a person over aging. For all these experimentations, a system with MATLAB 2015a (64 bit), 2.60-GHz Intel(R) Core(TM), i-5 CPU with 8 GB of RAM is used. We also used MatConvNet open-source library version 1.0 beta 20 for experimenting with CNN in MATLAB [10].

For all these experiments, FGNET and MORPH II popular aging datasets are used. FGNET dataset contains 1002 images of 82 subjects [11] and MORPH II contains more than 55,000 images of about 13,000 subjects [12]. For testing the performance of the system Leave-One-Person-Out (LOPO) scheme is used and can evaluate the rank recognition for performance evaluation.

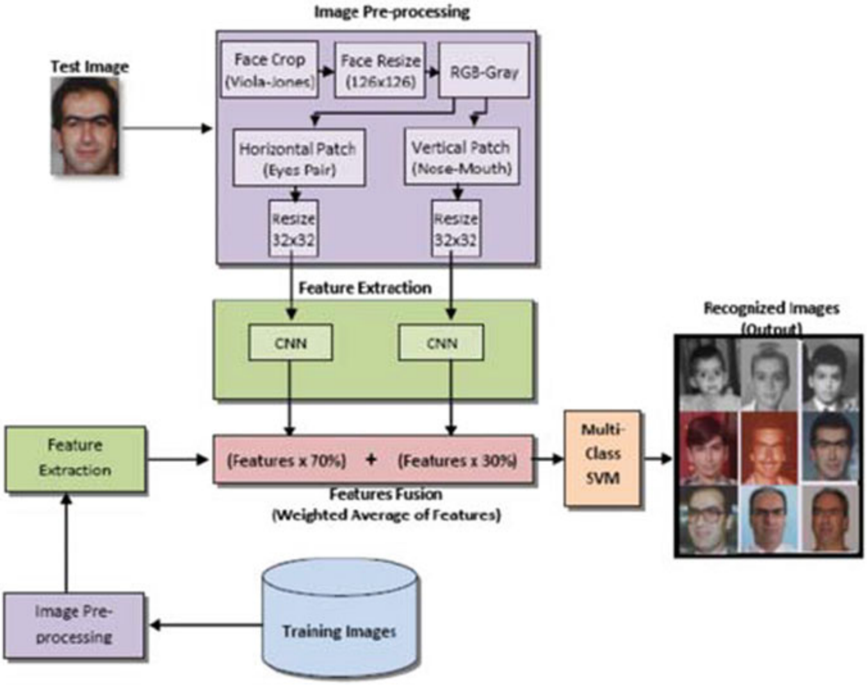


Fig. 4 Proposed model for approach 3 using CNN on biased face patches [8]

3.2 Results and Discussion

In our experiments, from FGNET dataset, a total of 980 images from 82 subjects are used. For training and testing, 852 and 128 images are used, respectively. Similarly, from MORPH II dataset, in all 1005 images of 255 subjects are used. Among these, 750 images are utilized for training process and 255 images are utilized for testing process.

Rank-1 recognition using the first approach on FGNET dataset is 76.5% and for MORPH II is 90%. The proposed system analyzed the second approach as Rank-1 recognition is 86.6 % on FGNET and for MORPH II is 92.5%. Similarly, Rank-1 recognition using the last approach is 91.4% on FGNET and 98.4% on MORPH II dataset as shown in Table 1. It is found that MORPH II dataset gives better performance as compared to FGNET dataset as there are less intra-class variations. Figure 5 shows a performance analysis of the comparison between the approaches over FGNET and MORPH II datasets.

All these proposed methods outperform over available state-of-the-art methods. The comparative analysis of proposed methods over available state-of-the-art methods is given in Table 2.

Table 1 Comparative rank-1 recognition of our experiments using both datasets

Proposed methods	Rank-1 recognition (%)	
	FGNET	MORPH II
Approach 1 (using GLBP features)	76.5	90
Approach 2 (using CNN)	86.6	92.5
Approach 3 (using CNN + Biased Face Patches)	91.4	98.4

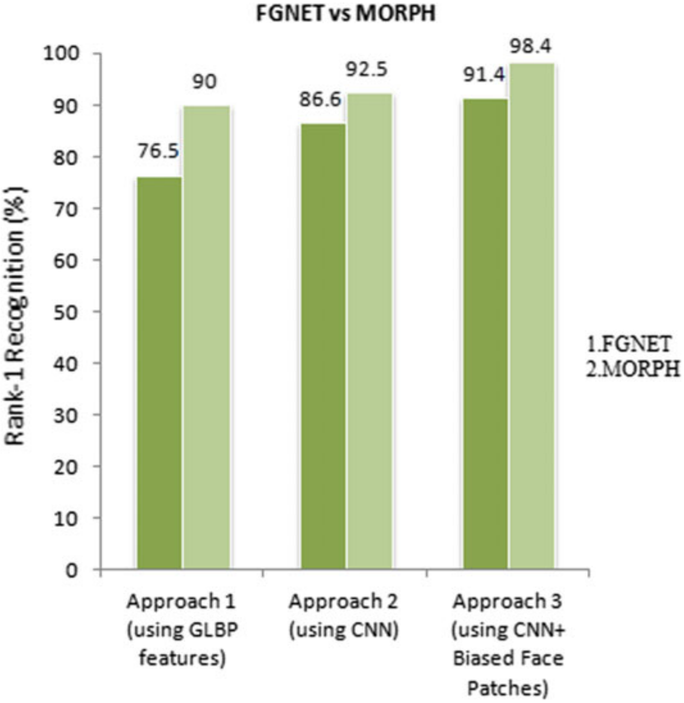


Fig. 5 Recognition rank analysis

Finally, from all of the above-used approaches, experiments and results obtained on both the datasets, the paper is summarized as,

1. The performance of the system is improved by performing preprocessing steps. In our approach, complicated steps are not performed. It makes the program simpler. Moreover, there is no degradation in performance.
2. Novel biased approach on both face patches works well in combination with CNN and SVM to improve the system performance as rank-1 recognition.
3. Our CNN architecture contains only seven layers; it is smaller as compared to architectures proposed in state-of-arts. Hence, it is simpler in nature and needs less execution time.

Table 2 Comparative analysis of proposed methods with state-of-the-arts on both datasets

Methods	Rank-1 Recognition (%)	
	FGNET	MORPH
Facial asymmetry [5]	69.5	69.4
NTCA [2]	48.9	83.8
MDL [1]	65.2	91.8
CNN [4]	80.6	92.2
MEFA [6]	76.2	93.8
LF-CNN [9]	88.1	97.5
Approach 1 (using GLBP features) [3]	**76.5**	**90.0**
Approach 2 (using CNN) [7]	**86.6**	**92.5**
Approach 3 (using CNN + Biased Face Patches)[8]	**91.4**	**98.4**

4. Overall, in all the experiments, MORPH II dataset gives a better performance over FGNET dataset. As there are more age variations in the images of FGNET dataset, FGNET contains more intra-class differences as compared to MORPH dataset.

4 Conclusion and Future Scope

To recognize a person over a span of years is a very interesting and challenging task. The authors have summarized some of the proposed methodologies for recognizing a person over aging. All experimentations were performed on two standard datasets: FGNET and MORPH II. The aim of this paper is to introduce the methods with simpler techniques and less preprocessing steps. Traditional methods need preprocessing, feature extraction, and classification separately. The use of CNN eliminates complicated preprocessing step as head pose correction and separate feature extraction algorithms. The results of these methods demonstrate that CNN with biased face patches outperform other proposed methods on both datasets as a periocular region is the most stable region. It also improves the results over available state-of-the-art methods with the advantage of CNN. The work can be extended by a varying number of layers of CNN to check the effect on the performance of the system.

References

1. Sungatullina (2013) Multiview discriminative learning for age-invariant face recognition. In: 10th IEEE international conference and workshops on automatic face and gesture recognition (FG)
2. Bouchaffra D (2014) Nonlinear topological component analysis: application to age-invariant face recognition. IEEE Trans Neural Netw Learning Syst

3. Nimbarte M, Bhoyar K (2017) Face recognition across aging using GLBP features, Springer Book Series, Smart Innovations, Systems and Technologies, Chapter 30, Vol 2, pp 275–283. https://doi.org/10.1007/978-3-319-63645-0_30
4. El Khiyari H, Wechsler H (2016) Face recognition across time lapse using convolutional neural networks. J Informat Security 7(3):141–151
5. Sajid M (2016) The role of facial asymmetry in recognizing age-separated face images. J Comput Electr Eng pp 1–12. http://dx.doi.org/10.1016/j.compeleceng.2016.01.001
6. Gong D (2015) A maximum entropy feature descriptor for age invariant face recognition. In: IEEE conference on computer vision and pattern recognition (CVPR)
7. Nimbarte M, Bhoyar K (2018) Age Invariant Face Recognition using Convolutional Neural Network. Int J Electr Comput Eng (IJECE) 8(4):2126–2138. https://doi.org/10.11591/ijece.v8i4
8. Nimbarte M, Bhoyar K (2020) Biased face patching approach for age invariant face recognition using convolutional neural network. Int J Intell Syst Technol Appl 19(2):103–124. https://doi.org/10.1504/IJISTA.2020.107216. Online only
9. Wen Y, Li Z, Qiao Y (2016) Latent factor guided convolutional neural networks for age-invariant face recognition. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 4893–4901
10. Phung S, Bouzerdoum A (2009) MATLAB library for convolutional neural networks, Technical Report, Visual and Audio Signal Processing Lab University of Wollongong
11. The FG-NET Aging Database. http://www.fgnet.rsunit.com
12. MORPH Non-commercial Release Whitepaper. http://www.faceaginggroup.com

A Comprehensive Survey on Federated Cloud Computing and its Future Research Directions

S. R. Shishira and A. Kandasamy

Abstract The cloud computing paradigm is popular due to its pay-as-you-go model. Due to its increasing demand for service, the user has a huge advantage in paying for the service currently needed. In a federated cloud environment, there is one or more number of cloud service providers who share their servers to service the user request. It improves minimizing cost, utilization of services and improves performance. Clients will get benefited as there is a Service Level Agreement between both. In the present paper, survey is provided on the benefits of the federated environment, its architecture, provision of resources and future research directions. Paper also gives the comparative study on the above aspects.

Keywords Federated cloud · Optimization methods · Cloud architecture

1 Introduction

Cloud computing is becoming more popular due to its pay per demand service. As lots of consumers hosting their application on cloud in different platforms, cloud service is in huge demand. Federated cloud refers to the greater number of service providers that are geographically distributed, who share their servers to serve the client request. Few cloud service providers are Amazon, Microsoft, Google, etc. Federated cloud concept comes under Infrastructure as a Service model, where consumer requests for the Infrastructure and it is served in the form of virtual machines [10]. These are loaded in the form of different operating systems including their software.

S. R. Shishira (✉) · A. Kandasamy
Department of MACS, NITK Surathkal, Mangalore, India
e-mail: shishirasr@gmail.com

A. Kandasamy
e-mail: kandy@nitk.ac.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_9

Cloud service providers have a huge number of resources for serving the client. In a single deployment cloud, when there exists natural disaster or any network security attacks, there is huge data loss for consumers [4]. Hence, multiple cloud deployment overcomes this disadvantage by sharing the resources and serving the request.

The cloud service broker operates as an intermediate between the provider and a consumer. The broker helps in sending the client request to the provider and getting the best service provider to service the client request [13, 18]. A broker helps in monitoring, managing, accessing the cloud service provider irrespective of the geographically situated servers. Once the broker chose the best cloud service provider to serve a client request, the Service level agreement is signed between to access the service [12]. Hence, in this type of environment quality of service is high comparing to a single deployment cloud.

2 Entities

Following are the important entities involved in a Federated cloud environment which consists of

- *Consumer*: Cloud consumer is referred to as a user who requests a particular service. Eg. Storage, network or infrastructure. Without having the knowledge of the backend process, the user or the consumer submits the task by demanding the service by paying for that particular service.
- *Public Cloud*: Resources available publicly for free or pay as per usage.
- *Private Cloud*: Cloud is owned by a single entity and the resources are not available for public. It is only for the private business.
- *Hybrid Cloud*: A consolidation of private and public cloud is termed as a hybrid cloud. One or more corporate organizations are involved in this.
- *CSP*: Cloud service provider provides resources based on the request by the user. For example, Amazon Ec2, Google and Microsoft Azure [9].
- *Workloads*: Cloud workloads are loads produced by a variety of applications and services employed on cloud infrastructures. It is a combination of jobs and tasks submitted by the user/consumer in the cloud.
- *SLAs*: Service Level Agreements are the agreed commitments between the consumer and the provider for maintaining the quality of service.

3 Federated Cloud Environment

Federated cloud comprises many service providers which are bound by standard SLAs. Each service provider serves the client based on the end resources.

Multiple clouds are used to serve different applications [2]. When the private resources are unable to service the request, these public clouds can help in overcoming the issue. Service broker can allocate different resources from geographically situated service providers to serve the consumer request.

3.1 Advantages

Over a single deployment model, the federated environment has lots of benefits which are as follows:

- Scalability: When the demand is high, cloud service provider share their resources and provide services via service broker.
- Multi-cloud deployment: It helps in reducing the cost required to serve the client request by aggregating the shared resources from different providers.
- Fault-tolerance: There will be data duplication, and hence no data loss exists during the natural disaster or downtime of servers [6].
- Performance: Due to the multi-cloud deployment, the cost is minimized, a response time of serving the request will be low, which indirectly increases the performance measure.

3.2 Coupling Levels

Federated cloud consists of different levels of coupling including resource cooperation, monitoring, remote controlling, etc. Vozmediano et al. [3] explained various levels of coupling in a centre. As given in (Table 1).Coupling levels are classified as loosely coupled, partially coupled and tightly coupled federative cloud instances.

- Loosely coupling: In this type of coupling, very less of inter-operation is done. Basic operations such as monitoring and controlling are done in this level. Advanced operations like migration are not included.

Table 1 Comparison on the types of coupling level

Level	Operations	Security
Loose	Basic functions on virtual machines	Single cloud in the organizations
Partial	Controlling and maintaining VMs	Agreements of framework
Tight	Scheduling to a particular resource, live migration	Consumer region sharing

- Partially coupling: More than one number of cloud providers partner and share their resources with respect to their terms and conditions.
- Tightly coupling: These types of coupling are done inter-organization with the same architecture and OS type. Advanced features like remote monitoring, VM migration is done here.

3.3 Architectures

We have classified Federated architecture into four types (Table 2).

- Cloud bursting architecture: When the consumer runs out of the resources in the internal cloud, he can demand the service from the public cloud by paying for it [1]. During this process, the consumer bursts his/her data to the public cloud. If the internal cloud contains a cloud VM switch, then the public cloud can be used as his own internal clouds without any network change (Fig. 1).
- Cloud broker architecture: Cloud broker is an intermediate between the provider and the consumer. It is very difficult for the consumer to directly contact the provider due to the terms and conditions. Hence there will be a broker who takes the request from consumers and search for the best provider to service the request by hiding the management difficulties (Fig. 2).
- Aggregated cloud architectures: Cloud is believed to be infinitely providing services on-demand basis to their clients. But, due to circumstances if there is no availability of hardware or specific resources, the cloud cannot serve the demands. Hence, cloud providers aggregate their resources based on the framework and agreements and individually serve the client request (Fig 3). Service broker helps in choosing the best provider for the specific client request.

Table 2 Comparison on the architecture of federated cloud

Architecture	Levels	Cloud Type	Benefits
Bursting	Loose	Bursting of workloads from private to public cloud	Helpful during resource exhaustion
Brokering	Loose	Chooses among the best public service provider	Optimization of cost, execution time
Aggregation	Partial	Hybridization of public and private clouds	Resource sharing to meet the user needs
Multi-tier	Tight	Large cloud data centres combining public, private with several data centres	Scalability

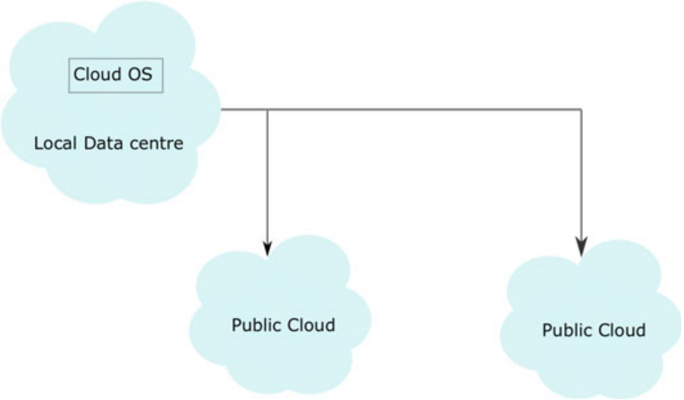


Fig. 1 Cloud bursting architecture

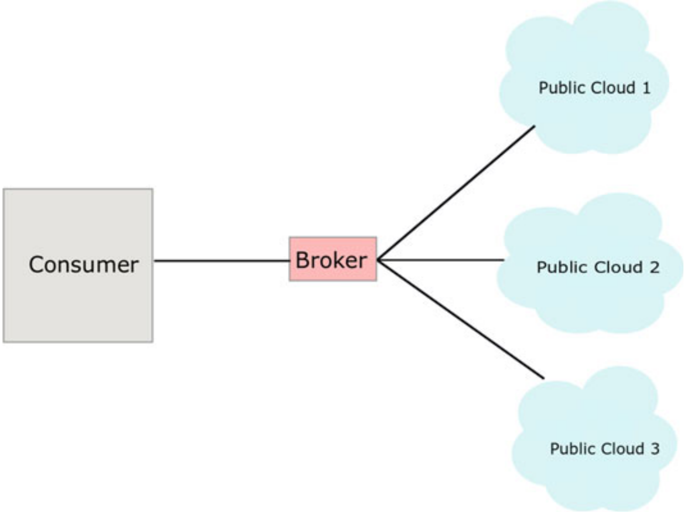


Fig. 2 Cloud broker architecture

- Multi-tier architecture: This type of architecture is a single cloud to consumers which has their cloud data centres distributed geographically (Fig. 4). The infrastructure resources can be accessed from any data centres. As it contains multiple resources distributed, the cost of the individual system is high compared to other architectures.

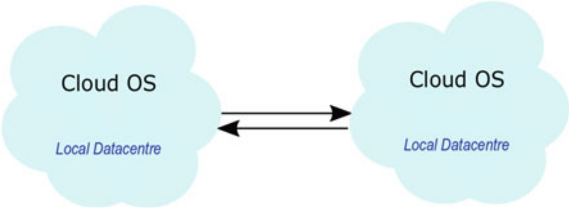


Fig. 3 Cloud aggregation architecture

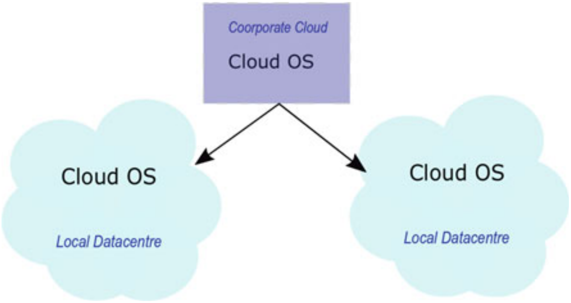


Fig. 4 Cloud multi-tier architecture

4 Resource Management

Management resources refer to selection, monitor, increase or decrease of the resources based on the demand. During off-peak hours, steady resource provisioning has no demand. Hence, automatic scaling of resources helps in auto increasing or decreasing the resources such as Infrastructure to the consumers.

4.1 Challenges in Federated Cloud

Aggregation of various cloud providers makes it a challenging task.

- Portability: It is very important to move the data safely on to different cloud data centres. Whenever there is demand from the Client, it is very necessary to serve that particular request without any delay. Also, it is important to combine two or more private and public clouds to satisfy business needs. Manage cloud services via Application Programming Interface without violating SLAs, Quality of Service, Availability and performance. A cloud service broker has to serve the request between the provider and a consumer but there are issues with interoperability. Aggregating cloud resources and satisfying the client's needs are also one of the major problems in federated cloud environment. Security is a major concern when combining two different clouds.Multiple clouds have to share their resourceswithout violating the SLAs and the audit process is differentwith respect to cloud providers.

- Deployment plan: Cloud providers offer a different set of resources based on user demand. The service broker has to choose the best provider resources based on the application requested to meet its needs. But when there is an uncertainty in user requests, maintaining the resources in the cloud provider is a tedious task. It should not go waste while in the off-peak demand.
- Quality of service: Quality of service plays a major role in serving the client's request. It is mainly based on the SLAs that are agreed between the provider and consumer. Sometimes, if the resources are exhausted due to natural disasters, it is important not to violate the SLA by compromising with the quality of service.
- Consumer specific coercion: User or a consumer can specify specific requirements to deploy their model on to the public cloud. During the process, the user can demand one more VMS which may violate the conditions of the resource present in the data centres of one of the providers. Hence, it is the service broker's job to rule out this tedious uncertainty.
- Jurisdiction: service providers have to adhere to the jurisdiction wherein the data centres placed are in specific regions. Thus, the client can deploy his/her model complying with his/her own regional laws.
- Pricing: Different service providers have different sets of pricing labels that depend on the type of services provided. Amazon EC2 generally has three types of processing reserved, on-demand and spot type. Elastic hosts allow to customize the cloud instances based on CPU, disk and memory size. Providers are charged based on network bandwidth, storage and memory.

5 Optimized Management of Resources

In the cloud both the provider and consumer try to obtain an optimal solution for resource provisioning. Provider tries to efficiently utilize the resources by not violating the QoS. While the consumer tries to minimize the service cost by deploying the applications and also get the best service during the process. Table 3 shows the existing frameworks proposed by the various authors.

- Cloud bursting architecture: Javadi et al. [19] proposed a resource provisioning model for hybrid cloud during the failure. In this paper, the author has used brokering method to make use of public resources while satisfying the users' need internally by using their private cloud. Bossche et al. [15] proposed a method to optimize the resource provisioning by bursting the cloud from private to public clouds.
- Cloud broker architecture: Chaisiri et al. [16] used stochastic programming for cost optimization. cost using broking technique. Tordsson et al. [11] considered two types of resource provisioning, the first one is to optimally place the VMS and second is to monitor and control the resources across different providers. Lucas-Simarro et al. [5] used binary integer programming to optimize the cost

Table 3 Optimized scheduling of resources in federated cloud

Authors	Framework	Parameter	Proposed model
Bossche et al. [15]	Bursting	Cost	Binary integer
Javadi et al. [19]	Bursting	Cost	Integer programming
Chaisiri et al. [16]	Brokering	Cost, response time	Stochastic method
Tordsson et al. [11]	Tight	Cost, response time	Binary integer
Lucas-Simarro et al. [5]	Brokering	Cost, execution time	Binary integer
Breitgand et al. [7]	Aggregation	Energy consumption	Greedy method
Vecchiola et al. [8]	Aggregation	Performance	Integer programming method
Wright et al. [14]	Aggregation	Cost, execution time	Binary integer
Calheiros et al. [17]	Aggregation	Cost	Deadline management method

and performance. The authors used different scheduling strategies for automatic scaling of resources during the peak time.

- Aggregated cloud architecture: Breitgand et al. [7] proposed a model for providing benefit for a cloud service provider by efficiently serving QoS for load balancing integer programming model. Vecchiola et al. [8] designed a model for the provisioning of resources from shared providers efficiently. Their proposed model helped in improving the response time of request handling. Wright et al. [14] proposed a method for efficiently searching the best infrastructure for the cloud service providers to serve the client request. A heuristic approach is used for optimizing the cost-based performance. Calheiros et al. [17] presented an architecture from intercloud to discover the best providers for the client request.

6 Conclusion

Bursting of clouds mainly focussed on cost optimizations from the service provider side, exhausting the internal cloud resources. Existing broker methods are not application-oriented but maintain QoS. Resource scheduling helps the scheduling of workloads to different resources chosen by the services broker. Hence service broker plays an important role in choosing among the number of cloud service providers in a federated environment. In this paper, a study on optimization in the management of data in the federated cloud has been done. We have presented various federated architectures and their limitations while servicing the client's request.

We have identified some of the future research directions.

- Consumer specific SLAs: A Cloud Service Provider provides SLAs depending on their resources and benefit their performance by optimizing the resources. Based on the specific applications, SLAs can be configured and benefitted as per the customer requirements.
- Consumer desired locality brokering: Based on the locality and region which is profitable to consumer, cloud broker can optimize to choose the resources from specific providers.
- Resource scheduling in the federated environment: Workloads can be predicted before going into the broker phase and the broker can choose the best CSP before handling a specific user request. Hence he can optimize parameters such as delay, response and execution time.

References

1. Selvanathan N, Jayakody D, Damjanovic-Behrendt V (2019) Federated identity management and interoperability for heterogeneous cloud platform ecosystems. In: Proceedings of the 14th international conference on availability, reliability and security. ACM
2. Buyya R, Ranjan R, Calheiros RN (2010) Intercloud: utility-oriented federation of cloud computing environments for scaling of application services. In: International conference on algorithms and architectures for parallel processing. Springer, Berlin, Heidelberg
3. Moreno-Vozmediano R, Montero R, Llorente IM (2012) Iaas cloud architecture: from virtualized datacenters to federated cloud infrastructures. Computer 45(12):65–72
4. Shishira SR, Kandasamy A, Chandrasekaran K (2017) Workload scheduling in cloud: a comprehensive survey and future research directions. In: 7th international conference on cloud computing. data science & engineering-confluence. IEEE, p 2017
5. Lucas-Simarro JL et al (2013) Scheduling strategies for optimal service deployment across multiple clouds. Future Gener Comput Syst 29(6):1431–1441
6. Rosa MJF, Aletéia PFA, Felipe LSM (2018) Cost and time prediction for efficient execution of bioinformatics workflows in federated cloud. In: 2018 IEEE international conference on bioinformatics and biomedicine (BIBM). IEEE
7. , Breitgand D, Marashini A, Tordsson J (2011) Policy-driven service placement optimization in federated clouds. IBM Res Div Tech Rep 9:11–15
8. Vecchiola C et al (2012) Deadline-driven provisioning of resources for scientific applications in hybrid clouds with Aneka. Future Gener Comput Syst 28(1):58–65
9. Habibi M, Fazli MA, Movaghar A (2019) Efficient distribution of requests in federated cloud computing environments utilizing statistical multiplexing. Future Gener Comput Syst 90:451–460
10. Bohn RB et al (2011) NIST cloud computing reference architecture. In: 2011 IEEE world congress on services. IEEE
11. Tordsson J et al (2012) Cloud brokering mechanisms for optimized placement of virtual machines across multiple providers. Future Gener Comput Syst 28(2):358–367
12. Badger L et al (2012) Cloud computing synopsis and recommendations. NIST special publication 800:146
13. Rochwerger B et al (2009) The reservoir model and architecture for open federated cloud computing. IBM J Res Dev 53(4):4–1
14. Wright P et al (2012) A constraints-based resource discovery model for multi-provider cloud environments. J Cloud Comput: Adv Syst Appl 1(1):6

15. Van den Bossche R, Vanmechelen K, Broeckhove J (2010) Cost-optimal scheduling in hybrid IAAS clouds for deadline constrained workloads. In: 2010 IEEE 3rd international conference on cloud computing. IEEE
16. Chaisiri S, Lee B-S, Niyato D (2011) Optimization of resource provisioning cost in cloud computing. IEEE Trans Serv Comput 5(2):164–177
17. Calheiros RN et al (2012) A coordinator for scaling elastic applications across multiple clouds. Future Gener Comput Syst 28(8):1350–1362
18. Najm M, Tamarapalli V (2019) A cost-aware algorithm for placement of enterprise applications in federated cloud data center. In: Proceedings of the 20th international conference on distributed computing and networking. ACM
19. Javadi B, Abawajy J, Buyya R (2012) Failure-aware resource provisioning for hybrid cloud infrastructure. J Parallel Distrib Comput 72(10):1318–1331

Decoy Technique for Preserving the Privacy in Fog Computing

K. P. Bindu Madavi and DR. P. Vijayakarthick

Abstract Fog computing is a process which computes and stores the data, which facilities network services between computing information centers and end devices. Fog computing is additionally fogging or edge computing. Fog computing focuses on increasing the efficiency and to quicken the data computing to cloud for storage and processing. This paper is mainly focus on securing personal information within the cloud employing a fog computing facility and decoy techniques. In the proposed system primary stage each licensed and unauthorized user can refer to the decoy information in fog computing. Using user profiling technique, the authorized user will be identified. If it's an unauthorized user, they cannot access the original data. If it is an authorized user, then it proceeds to the second stage by verifying the challenges. The challenge can be secured with verification code or else it can be a private question. Once authorized users clear the security challenge, they can access the original data.

Keywords FOG computing · Decoy technology · Security · User behavior profiling

1 Introduction

Cloud computing technology is a shared pool of resources; every organization is using the cloud to secure its data. We can access data stored in the cloud from anywhere and anytime. With new computing techniques and wide use of the intelligent device, new security challenges are arising as the security of cloud computing services is crucial. There are many kinds of possible threats, such as a Malware Injection, Wrapping, Browser, and Flooding [1].

K. P. Bindu Madavi (✉) · DR. P. Vijayakarthick
Department of Computer Science and Engineering, Dayananda Sagar University, Bengaluru, Karnataka, India
e-mail: madavi20@gmail.com

DR. P. Vijayakarthick
e-mail: vijay691988@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_10

The principle issue in cloud computing is to secure the user's information in such a way that specific verified users can access the data and nobody else can access that information [2]. The cloud security issues are composed of a few classes like information insurance, engineering, programming data protection, architecture, software isolation, trust, identity management, and availability [3]. The number of researches in distributed computing security has focused on anticipating ill-conceived and unauthorized access to information by creating encryptions and refined access control. Anyway, these instruments have not had the option to counteract information [2]. In the web, a username–password-based method is used but password-based authentication is vulnerable to attackers. Existing data secure mechanisms like encoding techniques are failing in protecting and preventing the information from the aggressors.

The encryption mechanism is primarily centered around the key given by the users at the time of getting to the information yet does not check the identity of the interlopers. To decrease the harm done by the aggressors, we are utilizing the strategy of decoy technology and user behavior profile.

This paper will utilize the decoy method to verify the information of the user in the cloud. This method provides fake information to the attackers replacing the original data. As soon as the attacker access the system, a fake file is presented in place of the original file; the proposed method creates a decoy file. These files are created from the beginning to guarantee enhanced security. Irrespective of any user, the system keeps the original data hidden and projects the decoy file by default. The original data are accessible to only verify users.

Structure of the paper: Sect. 2 and its sub-sections depict the foundation and preliminaries of the proposed method. Section 3 proposed framework and technique depicted, and Sect. 4 describes the conclusion.

2 Securing Clouds Using Fog

There are various ways to utilize cloud services to save data, reports, and media in remote administrations that can be recovered at whatever point users associate with the Internet. The issue of providing security to confidential data is a fundamental security issue. There are numerous ways to secure remote data in the cloud using standard access control and encryption methods [4].

Fog computing system is work against malicious users. A malicious user can access the confidential data which are stored in the cloud. Malicious attackers can easily obtain cryptographic keys and passwords to access the file. Providing security of confidential data remains a noteworthy security issue [5].

Fog computing extends the cloud computing paradigm to the edge of the network to deal with applications and services that do not match the paradigm of the cloud because of technical and infrastructure limitation including:

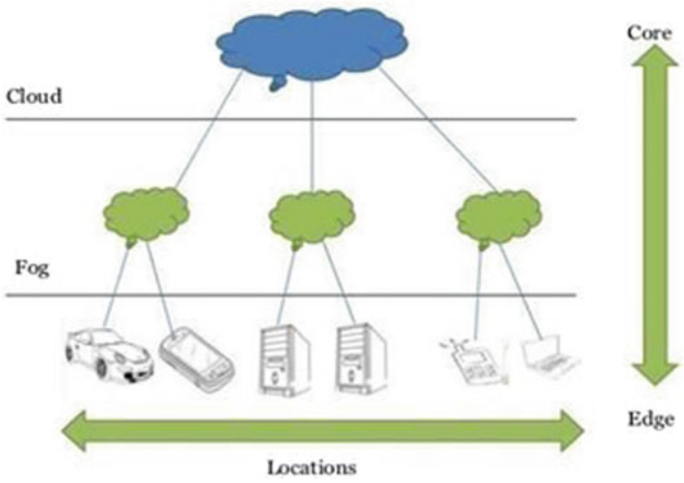


Fig. 1 Fog computing architecture

- Latency geographically circulated application
- Fast versatile applications
- Large-scale disseminated control frameworks (Fig. 1).

Creating decoy information and locating it beside the real information within the cloud to cover the important information of the user is also known as fog computing [6]. Fog Computing is a technique that helps in predicting and observance of the behavior of the user and providing security to the user's information.

Applications of fog computing:

Dependent on the data like traffic, driving conditions, atmosphere, and so on.

Smart grids, Smart Home, and smart cities: In smart utility services, fog computing is used for improving energy generation, delivery, and billing.

Real-time analytics: Fog computing is employed for time period analytics which transfers the data from manufacturing systems to financial institutions that use real-time data.

Connected cars: Self-autonomous cars are now available in the market and they produce a lot of information. The information should be dissected and handled immediately.

Health Data Management: Fog computing may be helpful in healthcare, within which real-time operation and event response area unit important.

3 Proposed Method

We propose a security model, a unique approach to preserve the private data in fog computing by decoy technique. The proposed method is divided into two main stages

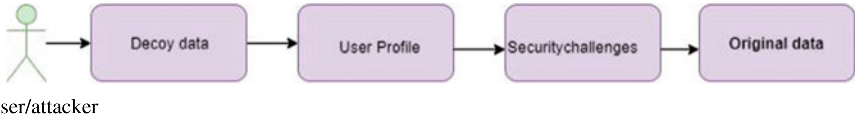


Fig. 2 Proposed system architecture

- At the first stage, both authorized and unauthorized users will refer decoy data or information by default decoy data
- At the second stage, the legitimated user will access the original data stored in the cloud bypassing all the security challenges (Fig. 2).

3.1 Methodology

There are three main modules

- User Behavior Profile
- Decoy technology
- File Generation

3.2 User Behavior Profile

The user behavior profile algorithm is used in order to detect abnormal access to the data in the cloud. It is used to determine whether a user who is accessing the data over the cloud is legitimated users or not by considering certain parameters like search behavior, amount of the information access, amount of data downloaded, etc. This method continuously monitored the behavior of the user to decide any strange access to the user's information in the cloud. This behavior-based strategy is utilized to found out the unauthorized user.

The unauthorized user most of the time includes the following:

- Trial and Error key for accessing the account
- Multiple tries for login
- Number of times a document read/write
- Visited URLs and time spent on each website

3.3 Decoy Technology

We used a unique approach for protecting the information in the cloud by decoy technique. Decoy data are honeypots and other bogus information which is used to secure the original data from an unauthorized user who is trying to access the private data. If any abnormal access in the cloud is noticed, the decoy file is generated in the cloud using the decoy technique which is sent to the unauthorized user in a manner that looks like original data. Decoy file contains fake data or bogus amount data which confusing the attackers. This decoy file is severed when the behavior of the user is been detected as an illegal user [7]. The authorized user can easily identify the vogues data instead of a legitimate file. If the user is a legitimated user, then they will pass in security challenges.

Two security features are implanted to secure the user data from unauthorized user

- Validation—checking whether or not the information access by the authorized or unauthorized user once abnormal behavior is detected
- Providing a fake amount of the decoy information's to confuse the attackers.

3.4 File Generation

When anonymous is identified using the user behavior profile algorithm rule, the file ought to be generated automatically. The file which is generated contains bogus information or fake data but it looks like the original file. The unauthorized user believes that the file is the original file and accesses the data. It is difficult to identify between original and decoy files. The original will be accessed by the only authorized user.

4 Conclusion

With the increase in malicious attackers, the security of user private information is turning into a significant issue. For which, we present a secure way to deal with secure the individual information over the cloud. The primary spotlight on verifying user's information inside the cloud by utilizing fog computing. Using the user behavior profile will determine the malicious user. In this approach, at the first stage, both authorized and unauthorized users will access the decoy data by default. But the authorized user knows the data which is accessed is not an original data. The authorized user proceeds to the second stage with security challenges; if the user passes all the security challenges, then the user can access the original data.

References

1. Zunnurhain K, Vrbsky S (2010) Security attacks and solutions in clouds. In: 2nd IEEE international conference on cloud computing technology and science, Indianapolis, December 2010
2. Kumar TS (2019) Efficient resource allocation and Qos enhancements of IOT with fog network. J ISMAC 1(02):101–110
3. Sonali K, Dhanashree B (2014) Fog computing: a new concept to minimize the attacks and to provide security in cloud computing environment. IJRET ISSN:2319–1163 |ISSN: 2321-7308
4. Reena KM , Sunil Kumar Yadav, Nikhil Kumar Bajaj and Vinay Singh (2017) Security implementation in cloud computing using user behavior profiling and decoy technology. In: International conference on inventive communication and computational technologies (ICICCT)
5. Abdullah A (2018) Fog computing and security issues: a review. In: 2018 7th international conference on computers communications and control (ICCCC)
6. Dr. Jordan S (2014) Extending the cloud with fog: security challenges and opportunities. In: 20th Americas conference on information system, Savannah, 2014
7. Jamil D, Zaki H (2011) Security issues in cloud computing and countermeasures. Int J Eng Sci Technol 3(4):2672–2676
8. Shanhe Y, Zijiang H, Zhengrui Q, Qun L (2015) Fog computing: platform and applications. In: 2015 Third IEEE workshop on hot topics in web systems and technologies
9. Ramesh K, Balaji T (2017) Alleviation of data attacks in cloud computing using offensive decoy technology. IJETT, Special Issue
10. Hadeal Abdulaziz Al Hamid, Sk Md Mizanur Rahman, M. Shamim Hossain, Ahmad Almogren, Atif Alamri (2017) A security model for preserving the privacy of medical big data in a healthcare cloud Using a fog computing facility with pairing-based cryptography. IEEE Access XXX(XX)
11. Muqtyar Ahmed S, Namratha P, Nagesh C (2013) Prevention of malicious insider in the cloud using decoy documents. IJERT 2(4) ISSN: 2278-0181
12. Farhad Foroughi and Peter Luksch (2018) Observation measures to profile user security behavior. In: International conference on cyber security and protection of digital services (cyber security 2018)

Design of Book Recommendation System Using Sentiment Analysis

Addanki Mounika and Dr. S. Saraswathi

Abstract In this paper, we propose four-level process to recommend the best book to the users. The levels are named as grouping of similar sentences by the semantic network, sentiment analysis (SA), clustering of reviewers and recommendation system. In the first level, grouping of similar sentences by the semantic network is done taking pre-processed data using parts of speech (POS) tagger from the datasets of reviewers and books. In the second level, SA is done in two phases which are training phase and testing phase by using deep learning methodology like convolutional neural networks (CNN) with n-gram method. The outcome of this level is given as input to the third level (clustering) which clusters the reviewers based on their age, locality and gender using K-nearest neighbor (KNN) algorithm. In the last level, a recommendation of books is done based on top-n interesting books using collaborative filtering (CF) algorithm. The system of book recommendation is to be done to get the best accuracy within less elapsing time.

Keywords Sentiment analysis · Document-level · Semantic network · CNN · N-gram · Doc2vec · Clustering · KNN · CF · Recommendation system

1 Introduction

Document-level sentiment analysis (SA) [1] of known lengthy texts is a difficult job, which point to a lot of words and opinions. This analysis is mainly useful for recommending books to the users. If the users get clear bifurcation about book reviews, it will be easier for them to take the decision. The below-mentioned methods in each level are used to do document-level sentiment analysis.

A. Mounika (✉)
Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry, India
e-mail: mounikaaddanki1704@pec.edu

Dr. S. Saraswathi
Department of Information Technology, Pondicherry Engineering College, Puducherry, India
e-mail: swathi@pec.edu

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_11

1.1 Semantic Network

A semantic network is a knowledge base that represents semantic or similar relations between notions in data. This is generally used as a form of knowledge representation. A semantic network is used as a tool to group similar sentences in this context.

1.2 Sentiment Analysis

Opinion aids a decision that is a personal factor of view of the reviewer. Every sentiment contains either positive, negative or neutral. Here the sentiment classification is recognized as an application recently used in the online reputation controlling systems.

Sentiment Analysis Levels. SA can be enforced on distinct levels, specifically: Word level (WL) in SA is used to determine the polarity of a word, that is either a positive, negative or neutral word. Sentence level (SL) in SA is used to determine the polarity of a sentence and is generally used to determine the analysis of opinion of the user. Particularly in social networks, SL is a series of words which addresses at deciding an opinion on a subject. Document level (DL) in SA is used to determine the polarity of a document and it is too difficult level to get the polarity compared to other levels, because, if the number of words, sentences increases, lot of noisy data will occur. So finding the polarity is a difficult task in this level.

Online book reviews are treated as one of the most fundamental sources of client opinion. Such a review may assess the book on the basis of personal taste. So, from the above-discussed levels, document level is the best level used to perform sentiment analysis on book reviews compared to other levels from the survey.

1.3 Deep Learning

In general, users can make decisions about books using online review resources. So, the grouping of similar sentences by the semantic network is done taking preprocessed data using POS tagger from the datasets of reviewers and books. Those are fed into the semantic network to form two categories of similar sentences with manually selected data in the document reviews. From the semantic network, the collectively identified words of each document with similar meaning are added to the list for every iteration and finally, the appended list is produced which will be used in deep learning methodology which shows the performance of sentiment analysis for classification of book reviews into positive, negative and neutral using CNN with n-gram method after feature extraction using word embedding to compare training phase and testing phase to get the accuracy. A clustering algorithm KNN is used to group the users into clusters of user's interest based on their age, locality and gender and recommend specific books to the user using CF method.

2 Related Works

SA is a developing research subject matter using the machine and deep learning methods. For deep learning methods, this classification calls for an essential cleaning and pre-processing step (consisting of tokenization, stop-word removal, punctuation, Html tags elimination, stemming and lemmatization) since the quality of data has a powerful impact on the overall performance of the deep learning method.

Bolanle et al. [2] worked on extracts feature and opinion pairs from reviews determine polarity and strength of evolutions classifies reviews expressed at the features of products as recommended or not recommended. In this paper, semantic orientation is used to calculate similar features. Maryem et al. [3] worked on a CNN-BiLSTM model. They proposed a solution for convolutional neural networks (CNN) and bidirectional long short-term memory (BiLSTM) models, using document to vector (Doc2vec) embedding that is appropriate for SA of lengthy texts. The CNN-BiLSTM model is compared with the other individual and combination models such as LSTM, BiLSTM, CNN and CNN-LSTM models using word2vec/doc2vec embedding. The CNN-BiLSTM model using doc2vec was performed on articles of French newspapers and obtained 90.66% accuracy with the other models. Kim [4] implemented a simple CNN with only one convolutional layer with an unsupervised model, on the dissimilar database, which includes movie reviews, subjectivity dataset, STS, TREC query dataset and customer reviews. He used a small hyperparameter, which produced extremely effective outcomes. He extensively utilized distinctive filter sizes and examined several CNN models to take out significant data. Finally, he concluded that CNN-static model gave the best overall performance outcomes. The recommender algorithms is termed model-based and makes use of matrix factorization techniques to guess on the values of latent factors, which loosely said can be an idea underlying descriptors of the domain [5]. Gao et al. [6] projected a new method, namely group-based ranking method which considers the user's reputations based on their grouping behaviors.

3 Outline of Proposed Work

In this proposal, the document-level reviews are grouped based on similar sentences by the semantic network and is done taking pre-processed data using POS tagger from the datasets of reviewers and books. From the semantic network, the collectively identified words of each document with a similar meaning are added to the list for every iteration and finally, the appended list is produced which will be used in deep learning methodology. From deep learning, convolution neural networks are used to classify the sentiments. The reviews are split into 40% of training and 60% of testing. At first, the reviews are trained using word embedding (doc2vec) which converts the text documents into vector form and then encoded with padded code sequence followed by multi-channeling for feature extraction, then max pooling is used with dropout. The n-gram method is used along with CNN to identify the negative sentences during the training phase. The k-dimension vector of softmax is used to set the range from 1 to 0. Then the testing reviews are tested with the training set to classify the reviews as positive, negative and neutral based on their polarity. The reviewers are grouped based on the type of review they are given at first and then reviewers are grouped based on their age, locality and gender depending upon their polarity using K-nearest neighbor (KNN) clustering algorithm. Finally, the recommendation system of a book selects and suggests the contents to meet user's preference automatically using data of previous users stored in the database. A CF algorithm is to be used to recommend books; it will effectively improve data scarcity and real-time problems. Based on the similarity value we predict the user's interest and recommend the top-n books.

4 Methodology

The various modules in the proposed system are grouping of similar sentences by semantic network, sentiment analysis, clustering of reviewers and recommendation system, as shown in Fig. 1. First, the data are taken from datasets of reviewers and books which are processed into grouping of similar sentences by semantic network module. Secondly, SA is done by taking data from the previous module in which pre-processing classification using CNN with n-grams is done by taking training and testing documents. After analysis, the data are moved to the next module, namely, clustering in which reviewers are clustered based on their age, locality and gender using KNN clustering algorithm. Finally, clustered data are moved to the next module, that is recommendation system in which the top-n interested books are recommended to the target user using CF algorithm.

4.1 Grouping of Similar Sentences by Semantic Network

In this module, few sentences which are manually identified as similar sentences are collected as two different types from the datasets of reviewers and books and these documents are passed to pre-processing and POS tagger. In pre-processing first the reviews are pre-processed to avoid the tags, stop words and punctuation, then stemming and lemmatization are done to identify the exact root meaning of the word and in POS tagger where the verbs, adjectives and adverbs are collected and stored in two different lists to run into the semantic network. The reviews are passed into the semantic network using pre-processing and passed to the tagger where

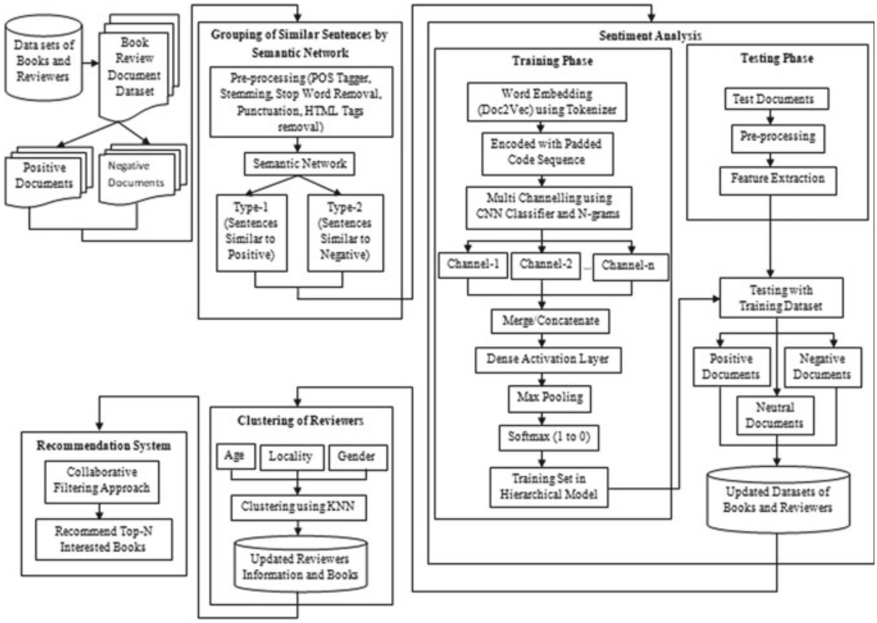

Data sets of Books and Reviewers
Book Review Document Dataset
Positive Documents
Negative Documents
Grouping of Similar Sentences by Semantic Network
Pre-processing (POS Tagger, Stemming, Stop Word Removal, Punctuation, HTML Tags removal)
Semantic Network
Type-1 (Sentences Similar to Positive)
Type-2 (Sentences Similar to Negative)
Sentiment Analysis
Training Phase
Word Embedding (Doc2Vec) using Tokenizer
Encoded with Padded Code Sequence
Multi Channelling using CNN Classifier and N-grams
Channel-1
Channel-2
Channel-n
Merge/Concatenate
Dense Activation Layer
Max Pooling
Softmax (1 to 0)
Training Set in Hierarchical Model
Testing Phase
Test Documents
Pre-processing
Feature Extraction
Testing with Training Dataset
Positive Documents
Negative Documents
Neutral Documents
Updated Datasets of Books and Reviewers
Clustering of Reviewers
Age
Locality
Gender
Clustering using KNN
Updated Reviewers Information and Books
Recommendation System
Collaborative Filtering Approach
Recommend Top-N Interested Books

Fig. 1 Module diagram

only adverbs, adjectives and verbs are extracted, which are compared with the list of already tagged words from the manual identification. The similarity is calculated using wordnet and synset. The nearby meaning of words is also taken into account to find similarity. For comparison, the mean of scores of two types is given, and the highest score and comparatively lowest score make the document to be classified into that respective type (either Type-1 or Type-2). These types (Type-1, Type-2) of the document are passed as input to the next module.

4.2 Sentiment Analysis

In this module, the process is done in two phases which are training and testing phases. In the training phase, the word embedding is done which converts the text documents into the vector form. In the word embedding, document to vector (dov2vec) model is performed to convert the data from the previous module into the vector model where similar types of words are represented by the same vector using tokenizer. These vectors are converted into digits by encoding the words from document using padded code sequence and then it is incorporated with multi-channeling using CNN classifier combined with n-gram which divides the single-input channel output from coded sequence into multiple channels, namely channel 1, channel 2 and so on up to channel n, to increase the speed and reduce the time of the function. After that, the formatted channels are merged or concatenated into a dense activation layer which does max pooling. After max pooling, the k-dimension vector of softmax is used to set the range from 1 to 0. The above same process is saved as a hierarchical model which makes it easy to compare with other sets of data. All the above-mentioned process comes under feature extraction which is also done in testing phase taking pre-processed data from the test documents. After completion of both the phases, they are compared to classify the data into positive, negative and neutral documents. The final data is stored in the database in the update process. The comparison is represented by calculating the accuracy and its performance versus with time.

4.3 Clustering of Reviewers

The input to this module is taken from the previous module in which clustering is performed. It is used in unsupervised learning where similar instances are grouped, based on their features or properties. In this module, the reviewers are grouped based on the type of review they are given at first, and then based on the reviewers' age, gender and locality they are grouped using the MySql. The queries and the groups may be of more than two groups. For clustering, the data which is going to be processed will be fully in the.csv format so that clustering is going to take place with various parameters like locality, gender, age, and most importantly based on the type of review (positive, negative or neutral). Then this clustering is visualized in python using the matplot library, and it is going to be executed by the K-nearest neighbor algorithm (using Euclidean distance formula).

4.4 Recommendation System

The clustered data from the previous module are sent as input to this module named personalized recommendation system (RS). In this system, recommendations are made using collaborative filtering (CF), which suggests the books to people with similar preferences like in the past. At the last, top-n interested books are recommended to the end-user so that the dynamic reviews are also taken into the account and the recommendation will be also dynamic according to the positive reviews of the new customers too.

5 Dataset Description

The book details were collected from the Kaggle website; about nearly 50,000 books were collected and stored in the database. The reviewer details were collected from the fake name generator, and repetition of the names is avoided. The reviews collected are on the document level. These reviews are collected from the Amazon website. There are in total 10,000 documents of reviews collected. The reviews are pre-processed to do stemming which find the root word. Those reviews are processed via POS tagger.

6 Conclusion

In this paper, we present the research work in the recommendation system using SA, by taking clustered data of reviewers, which is used to recommend books to the target user. The performance and accuracy are calculated to provide the best books to the users. This proposed work can solve the limitation in the already existing recommendation systems and provide improved accuracy. In our future work, a hybrid approach would consist of both model-based and memory-based algorithms in order to increase the performance of the system. A similar recommendation system can be incorporated for book recommendation systems.

References

1. Pang B, Lee L (2008) Opinion mining and sentiment analysis. Foundat Trends Informat Ret 2(1–2):1–135
2. Bolanle A, Ojokoh O (2012) A feature–opinion extraction approach to opinion mining. J Web Eng 11(1):051–063 © Rinton Press
3. Rhanoui Maryem, Mikram Mounia, Yousfi Siham, Barzali Soukaina (2019) A CNN-BiLSTM Model for Document-Level Sentiment Analysis: Machine Learning and Knowledge Extraction (MDPI) 1:832–847
4. Kim. Y:Convolutional Neural Networks for Sentence Classification: In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); Association for Computational Linguistics: Doha, Qatar, pp. 1746–1751, (2014)
5. H. Koohi and K. Kiani: A new method to find neighbor users that improves the performance of collaborative filtering: Expert Systems with Applications, vol. 83 (C), pp. 30–39, (2017)
6. J. Gao, Y.W. Dong, M. Shang, S.M. Cai and T. Zhou: Group-based ranking method for online rating systems with spamming attacks, Europhysics Letters and Applications, vol. 110, no. 2, pp. 28003 p1-p6, ©EPLA, (2015)
7. C. Balasubramanian, J. Raja Sekar and M. Shenbaga Devi: A Personalized User Recommendation Based on Attributes Clustering and Score Matrix: International Journal of Pure and Applied Mathematics (IJPAM), vol. 119, no.12, pp. 13751–13757, (2018)
8. C.X. Zhang, Z.K. Zhang, L. Yu, C. Liu, H. Liu and X.Y. Yan: Information filtering via collaborative user clustering modeling: Physica A: Statistical Mechanics and its Applications, vol. 396, pp. 195–203 © Elsevier, (2014)
9. https://archive.ics.uci.edu/ml/datasets/Amazon+book+reviews
10. Abdi A, Shamsuddin AM, Hasan S, Piran J (2019) Deep learning based sentiment classification of evaluative text based on multi feature fusion. Inf Process Manage. Springer. 54(4):1245–1259
11. Chakravarthy A, Deshmukh S, Desai P, Gawande S, Saha I (2018) Hybrid architecture for sentiment anlaysis using deep learning. Int J Adv Res Comput Sci 9:735–738
12. Santosh K, Varsha (2018) Survey on personalized web recommender system. J Inf Eng Electron Bus (IJIEEB) 4:33–40

Review of Python for Solar Photovoltaic Systems

R. Sivapriyan, D. Elangovan, and Kavyashri S. N. Lekhana

Abstract In recent years, the usage of solar energy as a source to produce power has increased exponentially as it provides a clean and efficient alternative to depleting non-renewable resources. The normal working period of a photovoltaic (PV) panel is 20 years, but due to defects in manufacturing or atmospheric condition changes, the efficiency and the lifespan of the panel decrease each year. The objective of this review article is to present and analyze the different methods that can be used to reduce the degradation rate of the PV cells in an economically viable way. Open-source frameworks are important to make any solution affordable; hence we explore the usage of python language in developments relating to improvement in the performance of PV cells. Based on this review a practically employable solution to improve working conditions for PV cells can be obtained.

Keywords Solar energy · PV panels · Python · Modeling · Monitoring · Analysis · Fault detection

1 Introduction

Developed in 1989 by Guido Von Rossum, the python language was originally conceived as a project to provide emphasis on code readability and advanced developer productivity. It was released in 1991 as a general-purpose, object-oriented, high-level language with an extensive number of standard libraries and modules. Being a free and open-source language with a vast network of community development and support, it quickly rose through the ranks to become one of the most widely used programming languages. Python is praised for its ease of usage, learning, and portability. It has a wide array of versatile applications involving web frameworks,

R. Sivapriyan (✉) · K. S. N. Lekhana
Sir MVIT, Bengaluru, India
e-mail: sivapriyan@gmail.com

D. Elangovan
SELECT, VIT, Vellore, India

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_12

GUI-based applications, prototyping, the ability to perform complex numeric and scientific calculations, along with advanced data analysis and visualization. It is used by many major tech conglomerates such as Firefox, YouTube, Dropbox, Microsoft, and by social sharing platforms like Quora and Reddit. Python has been adapted over the years to function beyond the scope of computer applications. An abundance of easy-to-use libraries is being customized to support electrical applications involving analysis of power systems [1], design of electronic circuits, and optimization of electrical machines [2]. The tools in python provide good support to students to obtain various power electronic waveforms of voltage and current without doing complex calculations. It simplifies and solves the compound calculations in Gauss–Siedel, Newton–Raphson, and other power analysis problems with minimal effort. Researchers and engineers are often required to develop their own software design to simulate electrical machines [3]; they are more inclined to use python for these purposes due to its affordability, versatility, and ease of use. Python is an interpreted language and hence allows for interactive testing of snippets of code, where the commands are executed instantaneously. This assists experimentation as there is no hassle of compilation delays [1]. Libraries such as Scipy, Numpy, and Pandas provide analysis of large datasets with precise and concise reports for required parameters. A real-time, quality 2D and 3D graphical representation of data can be obtained from libraries such as Matplotlib and Plotly. With increasing concerns over the usage of non-renewable resources, we have turned to solar energy to produce power, as sunlight is an inexhaustible resource. We use solar PV cells that absorb sunlight, using it as a source to produce direct current electricity. The demand for PV systems has increased with more people shifting toward clean energy. But, the performance of a PV module degrades with aging effect and atmospheric conditions [4]. As the industry expands, more research is being done to improve the functioning of the system, to achieve maximum productivity, and prolong its lifespan. Solar PV helps to get water for the off-grid agriculture field [5]. Python plays an important role in these developments as it is used in a large range of solar applications. It is chosen to be a good software [6] used to make the systems more affordable while developing tools for accurate solar power forecasting, to simulate modeling, design fault detection, and implement monitoring and analysis of PV parameters.

2 Python Libraries and Tools in the Solar PV System

Python has many libraries for solar PV analysis [7, 8], as shown in Fig. 1. Out of many libraries PVLIB Python, Solpy, Pandapower, Pyleecan, Scipy, Numpy, and Matplotlib are used by the researchers and teaching faculties for the analysis of solar PV systems.

Fig. 1 Python libraries

2.1 PVLIB Python

It is a freely available python package that has a group of classes and functions to model the PV systems. Using the implementation of models relevant to solar energy, PVLIB Python provides the algorithms to find the solar angle, clearness of the irradiance, transposition of irradiance, power, and the conversion of DC to AC voltage. The system fault detection can be obtained from measured and modeled current using PVLIB Python package [9].

2.2 Solpy

Solpy is an open-source python library which is used for performance modeling of solar power systems. Solpy will predict system performance by accessing the old weather data. The GHI, DNI, DHI implementation models can be used to calculate the irradiance and also the solar insolation. Different resource terms like ephemeris of sun, irradiance components, direct normal beam, sky diff, PV module temperature, historic weather data, and total irradiance and PV module model were described and were given with formula for their calculation.

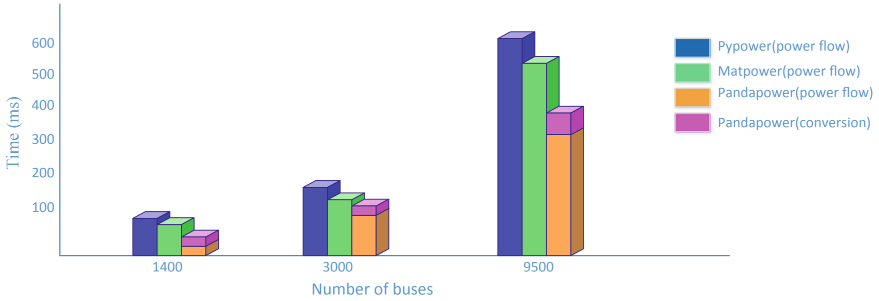


Fig. 2 Pandapower analysis time comparison

2.3 Pandapower

It is an open-source python tool used to analyze the power system. This tool mainly aims at the static analysis of balanced power systems. This tool permits the analysis of symmetrically operated transmission and sub-transmission systems and also distributed systems [10]. As shown in Fig. 2, using Pandapower, complicated mathematical calculations like Gauss–Seidal and Newton–Raphson power flow can be solved easily. Pandapower is very convenient to use and also it gives extension with libraries. Pandapower is widely used for various grid studies and for educational purposes.

2.4 Pyleecan

Pyleecan is one of the python libraries that is widely used for electrical engineering computational study [3]. Pyleecan aims at providing a user-friendly and flexible simulation structure for the upgradation of electrical drives and machines.

2.5 SciPy

It is a python-based free and open-source library, which is used in mathematical calculations, technical and scientific computing [1]. It is capable of operating on an array of NumPy libraries. SciPy includes modules for integration, image processing, linear algebra, interpolation, FFT, and signal processing.

2.6 NumPy

It is a python-based primary package for scientific computation. It is used as a proficient multi-dimensional container of generic data [1]. It is used for computing linear algebra, matrix computations, and numerical analysis.

2.7 MatplotlibNumPy

It is an open-source python library used for plotting the graphs and data visualization. It can be used in web server applications, python scripts, and user graphical interface toolkits. This library supports LATEX formatted texts and labels. It is most suitable for generating figures, where the figure can be supervised programmatically.

3 Implementation of Python in Solar PV System

3.1 Modeling

The proper functioning of a solar PV cell depends on different variable factors, such as temperature, irradiance, voltage, and current. To develop an ideal solar cell it is important to understand how a change in each factor affects the functioning of the system, as shown in Fig. 3. Modeling PV cells plays a vital role in designing and optimizing the structure of PV cells [4–8]. It helps us understand the behavior and characteristics of the cells under various conditions. Over the years different types of methods (analytical, numerical, linearized, etc.) have been used for modeling PV cells to obtain informative results [4]. Using of computational models is an accurate and highly valuable method to test and evaluate the performance of PV cells [8]. The data so collected in labs can be compared with actual results and used to develop new designs that can function at higher efficiency with fewer faults. Although several tools exist for the sole purpose of modeling solar cells, they are not user-friendly or open-source and are difficult to modify or reuse. Developing modeling platforms using an open-source programming language such as python is seen to be beneficial, as it allows collaboration and peer-reviewing which leads to the rapid development of the software [8]. One of the most accessible libraries available for solar PV modeling at present is PVLIB [9], originally developed in MATLAB. It was optimized to function in python environment so that it could be used by more people. PVLIB python is being used as a base to develop specialized platforms to deal with the mathematical side of modeling PV cells.

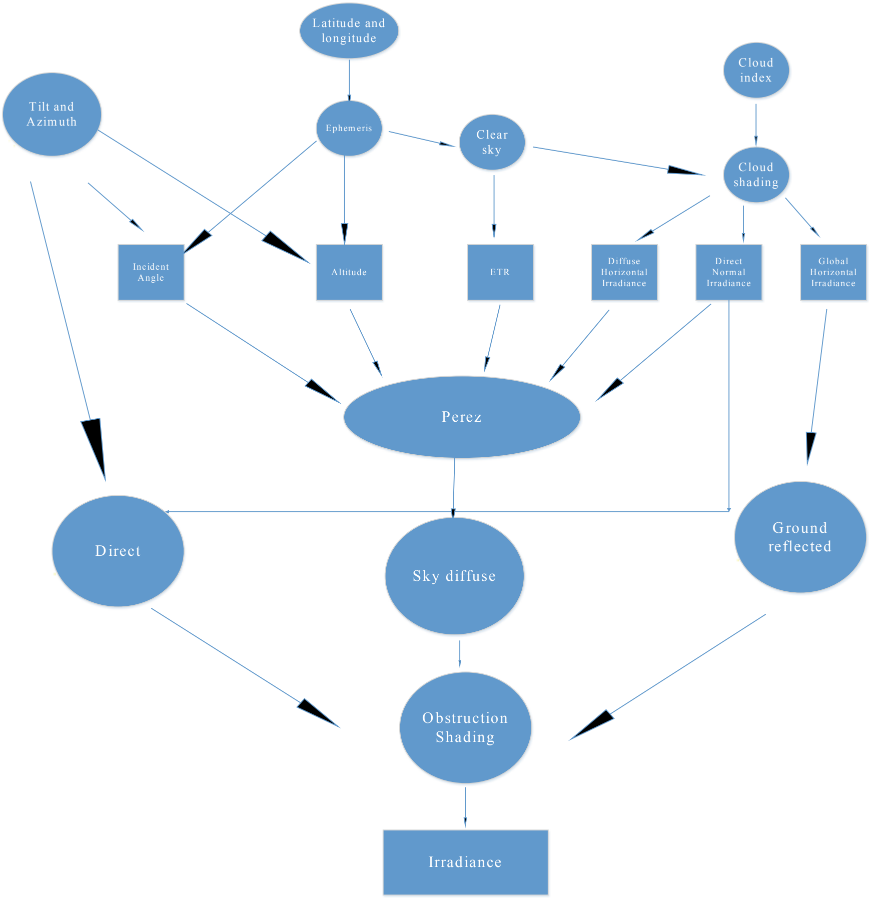


Fig. 3 Parameters in modeling PV cell

3.2 Analysis

The analysis of power system study is primarily to find the strength and the weak instances of the solar photovoltaic system. For this, the analysis of the solar photovoltaic system is required to maintain and decrease the unfavorable points that increase the dependability in the system [11]. Analysis of a system is a way of reaching the source of a problem. The analysis identifies what the system should do. Python in the analysis of the power system has many available open-source tools [11]. For the analysis, modeling, and optimization of power systems, python-based open-source tool called Pandapower is used. It gives optimal power flow, state estimations, and short-circuit calculations. This tool also includes calculations of the Newton–Raphson power flow.

3.3 Monitoring of PV Systems

The monitoring of photovoltaic systems becomes a very important task for the reliable functioning of PV systems. This system makes certain that the PV arrays are functioning correctly by tracing the output of the solar panel. A monitoring system offers us information about energy production and consumption, any damage to the solar system, optimization of energy use, and more. The parameters which require monitoring are $V_{\text{dc}}, I_{\text{dc}}, I_{\text{ac}}, P_{\text{dc}}, P_{\text{ac}}$, irradiance, and ambient temperature. To monitor these parameters collection of data is mandatory. The data can be collected from PV inverter through RTU or MODBUS/TCP. Python language which is an open-source programming language is used as a software programming language for this proposed system and pyModbusTCP which is a python package can be used [12–15]. Pymodbustcp gives access to the Modbus TCP server through the Modbus client object which is defined in the client module, as shown in Fig. 4.

The data then collected will be sent to a remote server or cloud-based monitoring system which can be made accessible to the user. Since most monitoring systems are expensive, smart monitoring of PV systems can also be performed. With a smart remote monitoring system, the monitoring, controlling, and maintenance of solar plants can be done easily without any human interface. Monitoring tasks can be performed by remote control and their communication between inverters can be done with wireless communication through RS232 interface. Different ways of communication for remote monitoring are GSM, Ethernet, and the internet. This particular smart monitoring system includes solar panels, IoT, Raspberry PI, different sensors, Arduino, LDR, and relay. This proposed system is used for improving the performance, efficiency, and also for real-time monitoring. Solar power generation can be monitored using the IoT platform. It can be used for maintenance like photovoltaic array cleaning. Here Raspberry PI is used as a processor that enables us to compute and browse from the internet. Raspberry PI microcontroller is also used in a small-scale standalone PV system in real time and at a low cost.

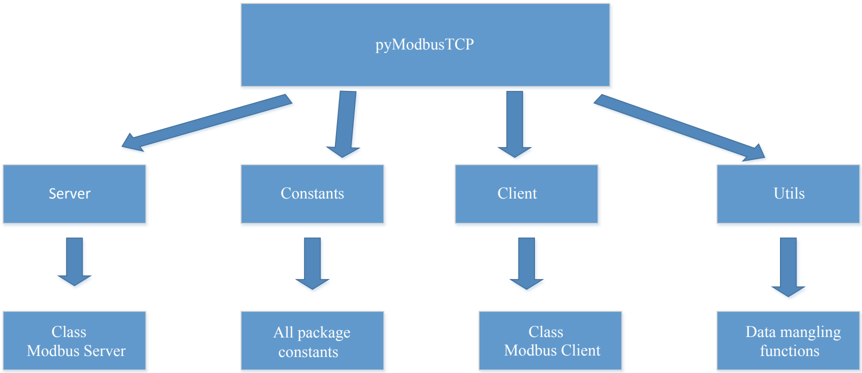


Fig. 4 Solar PV monitoring architecture

This monitoring system is used to store, measure, and display PV parameters like current, voltage, and temperature. The monitoring of outdoor PV systems is possible for real-time performance analysis and for the detection of a fault. In situ I–V measurements allow PV systems performance monitoring. Using this method power loss on different loss mechanisms of the PV module can be calculated. Python algorithms and programs are used for processing and data collection. Hence smart, simple, and cost-efficient methods of monitoring can be adopted.

3.4 Fault Detection

Renewable energy resources play a major role in energy management. The renewable energy sector is growing rapidly day-by-day all over the world, as it provides clean and adequate power. Even though the PV systems do not require a supervisory mechanism, they still might have internal or external faults reducing the efficiency. As the solar PV power adoption increases, it arises a demand for the availability and reliability of the power generated from PV systems. Any failures and damages that occur in PV systems may cause energy loss and unnecessary shutdown of the systems. The normal working period of a photovoltaic array is 20 years, but due to various factors like the changes in atmospheric conditions and manufacturing defects, dust, hotspots, crackers reduce the performance and efficiency of PV modules and also the panel yield rate and the lifespan of the panel decrease. The damage to the PV panels is caused by non-optimal cleaning which leads to improper maintenance. There is a need to detect and prevent these faults. This calls for monitoring of the panels continuously in order to decrease the rate of degradation. The model of fault detection takes the power, temperature, and irradiance as input data and precisely gives the fault type in the PV system as an output. It is very important to investigate the reliability of the PV systems. If the system is more reliable, then the cost of the photovoltaic system will be reduced and the service life of the PV modules increases. The parameters to be monitored are voltage, current, maximum power, efficiency, temperature, irradiance, and also the parameters that cause the failure of panels like dust, hotspot, cracks, and so on. The monitored parameters are uploaded to the cloud using Raspberry [16]. When the monitored parameters exceed the prescribed limits, an alert signal should be displayed on the webpage and also the alert message is directly sent to the phone, so the panel is protected by taking the protecting actions.

The fault detection in the grid-connected PV plants can be done by using detection algorithms. For simulating the PV systems, PVLib can be used. This package consists of a class called PV system that has the function to find the various model parameters. In order to check the behavior of the particular PV system and then comparing it with the original output, it uses the measured temperature, irradiance, current, voltage, and power. The dissimilarity measure is obtained by using mathematical models. Data preprocessing is used to get clean data by avoiding all the possible noise signals before it is used in the algorithms. The data will be collected by using sensors in the plant which measures the weather conditions, string level currents, performance level, gird power data, and so on. A SCADA system gathers all the signals and it allows control of the subsystems of the plant remotely. The expected and measured current are compared in order to detect the anomalous behavior of the system, and also the faults are analyzed. Also, the defects within the solar panel can be detected using thermal images. A framework is created for the detection of the defects automatically using the thermal imaging in order to create a precise alert system of risky conditions. The detected results give the alert signal, as in which part the solar panel is not working under expected conditions. So, by detecting the faults and taking the necessary actions, the average working period of the PV array can be increased.

4 Conclusion

In this paper, the application of python in modeling, analysis, monitoring, and fault detection of the photovoltaic system is discussed. Python language is used as a platform for its versatility, low cost, and ease of use. Python and its libraries and tools are accurate for modeling, analysis, monitoring, and fault detection of PV systems. The use of PVLIB python is to develop specialized platforms to deal with the mathematical side of modeling PV cells and also for simulating PV systems. A python-based open-source tool, Pandapower is a tool aiming for the optimization and analysis of solar PV systems. PYPSA is also a python power system analysis toolbox. The collected data from inverter are passed onto the server or cloud through pyMODBUS/TCP protocol. Python algorithm is used for processing these data. The use of python is to assist in the monitoring of complex systems and also alert the system operators when the behavior of the system is changed. The fault detection is based on comparing the expected and measured values of parameters. So, by continuous monitoring of system conditions and detecting the faults, stable output power delivery of the solar panel is ensured in remote areas. Hence the best, smart, and cost-efficient methodologies can be obtained through python.

Acknowledgements The authors would like to thank Punitharaj Research and Training Institute (OPC) Private Limited, Bangalore (https://prti.co.in) and Sir MVIT, Bangalore for their technical and financial support for this work.

References

1. Fernandes TR, Fernandes LR, Ricciardi TR, Ugarte LF, de Almeida MC (2018) Python Programming Language for Power System Analysis Education and Research. In: IEEE PES Transmission & Distribution Conference and Exhibition-Latin America (T&D-LA), pp. 1–5. IEEE
2. Grout I (2018) Electronic Circuit and System Design using Python and VHDL. In: 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, pp. 13–16. IEEE
3. Bonneel P, Le Besnerais J, Pile R, Devillers E (2018) Pyleecan: an open-source Python object-oriented software for the multiphysic design optimization of electrical machines. In: XIII International Conference on Electrical Machines, pp 948–954. IEEE
4. Umahsnakar S, Arunkumar G, Sivapriyan R (2017) Comparative analysis of solar photovoltaic cell models. In: 2017 International Conference On Smart Technologies for Smart Nation, pp 46–51. IEEE
5. Sivapriyan R, S. Umashankar, P. Sanjeevikumar, and Atif Iqbal.: Direct-Coupled Permanent Magnet DC Motor-Driven Solar Photovoltaic Water Pumping System—A Literature Review. In: Advances in Smart Grid and Renewable Energy, pp. 307–314. Springer, Singapore (2018)
6. White, Jeremy T., Michael N. Fienen, and John E. Doherty.: A python framework for environmental model uncertainty analysis. In: Environmental Modelling & Software 85. pp. 217–228. (2016)
7. Alonso-Álvarez D, Wilson T, Pearce P, Führer M, Farrell D, Ekins-Daukes N (2018) Solcore: a multi-scale, Python-based library for modelling solar cells and semiconductor materials. J Comput Electron 17(3):1099–1123
8. Charles, Nathan, Mahmoud Kabalan, and Pritpal Singh.: Open source photovoltaic system performance modeling with python. In: IEEE Canada International Humanitarian Technology Conference, pp. 1–4. IEEE (2015)
9. Stein, Joshua S., William F. Holmgren, Jessica Forbess, and Clifford W. Hansen.: PVLIB: Open source photovoltaic performance modeling functions for Matlab and Python. In: ieee 43rd photovoltaic specialists conference, pp. 3425–3430. IEEE (2016)
10. Thurner, Leon, Alexander Scheidler, Florian Schäfer, Jan-Hendrik Menke, Julian Dollichon, Friederike Meier, Steffen Meinecke, and Martin Braun.: Pandapower—An open-source python tool for convenient modeling, analysis, and optimization of electric power systems. In: Transactions on Power Systems 33, no. 6, pp. 6510–6521 (2018)
11. Phongtrakul, Tipthacha, Yuttana Kongjeen, and Krischonme Bhumkittipich.: Analysis of Power Load Flow for Power Distribution System based on PyPSA Toolbox. In: 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, pp. 764–767. IEEE (2018)
12. Kang, Shinyuk, and Ilwoo Lee.: Implementation of PV Monitoring System Using Python. In: 21st International Conference on Advanced Communication Technology, pp. 453–455. IEEE, (2019)
13. Deshmukh, Neha S., and D. L. Bhuyar.: A Smart Solar Photovoltaic Remote Monitoring and Controlling. In: Second International Conference on Intelligent Computing and Control Systems, pp. 67–71. IEEE (2018)
14. Walters, Joseph, Siyu Guo, Eric Schneller, Hubert Seigneur, and Matthew Boyd.: PV module loss analysis using system in-situ monitoring data. In: IEEE 7th World Conference on Photovoltaic Energy Conversion, pp. 2204–2208. IEEE (2018)
15. Prasanna, J. Laxmi, D. Lavanya, and T. Anil Kumar.: Condition monitoring of a virtual solar system using IoT. In: 2nd International Conference on Communication and Electronics Systems, pp. 286–290. IEEE (2017)
16. Dyamond WP, Rix AJ (2019) Detecting Anomalous Events for a Grid Connected PV Power Plant Using Sensor Data. In: Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa, pp. 287–292. IEEE

Data Exploratory Analysis for Classification in Machine Learning Algorithms

Jesintha Bala Chandrasekar, Shivakumar Murugesh, and Vasudeva Rao Prasadula

Abstract Availability of big data transformed the way machine learning works and the way data is used in machine learning. In real time the data gathered from various sources might be unstructured, incomplete, unrealistic and incorrect in nature. Transforming the data with the above-mentioned qualities and making it ready for analysis is a challenging task. As the quality of data have direct impact on the efficiency of the trained model, data exploratory analysis (DEA) plays a major role in understanding the data and forms the quality training dataset for the machine learning algorithms. This paper emphasizes the importance of DEA in the selection of the significant attributes and filling of missing values to form the quality training dataset. The dataset considered for experimentation is a binary classification problem "Survival prediction of Titanic Passengers". Experimental results show that training the model with the quality dataset has improved the accuracy as compared to the case when the model was trained with a raw data.

Keywords Data exploratory analysis · Data visualization · Big data analytics · Logistic regression

1 Introduction

Machine learning revolutionized the way the data is used in making a decision and thriving the technology toward automation. A computer software can work according to the way it was programmed or designed; that is, it will work with a fixed set of rules and instructions. Machine learning differs from the ordinary computer software by instilling the learning ability into the algorithm by training data, testing the learning ability by test data and the effectiveness of learning is measured by the error measurements. Extraction of knowledge or meaningful information from the big data is an important aspect of data science [1]. Data available from a different source in different formats at different times are gathered together to extract insights and

J. B. Chandrasekar (✉) · S. Murugesh · V. R. Prasadula
Central Research Laboraotry, Bharat Electronics Limited, Bengaluru, India
e-mail: jesintha@bel.co.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_13

knowledge. The availability of the data for learning and the ability of the machine learning to learn automatically from the big data is the basic foundation of artificial intelligence. However, preparing the data which is readily available for analysis or learning is a tedious job as most of the real-time data is of unstructured, unclean, and incomplete in nature [2]. As the effectiveness of learning is completely dependent on the training data, the preparation of data (data pre-processing) is very important.

Making efficient and well-informed decisions by acquiring insights from the massive amount of data is the primary aim of big data analytics. Data should be proactively collected, instantly processed and constantly monitored. Exploratory data analytics is the first step in the big data analytics that helps to understand the data patterns and the correlation, market trends and so on [3]. Data visualization—the graphical representation of the data—helps the users to understand the nature of data and its relationship irrespective of the data size. Traditional methods to visualize the data are quite difficult because of the big data's characteristics—volume, variety and velocity. Breaking the big data into manageable and logical portion is the basics for data visualization. It leads the preparation of data by discovering, structuring, cleaning, enriching and validating data [4] so that it accelerates the analysis and learning process. This can be done by selecting the significant attributes, eliminating the unwanted observation, filling the missing data and maintaining the balance within the dataset.

2 Related Works

Dimensionality reduction is a method of reducing the number of features in a dataset by using different methods. Generally, it is subdivided into two components: feature selection and feature extraction. Feature selection is a dimension reduction technique to select the subset of the relevant/significant features to create the model. As feature selection doesn't alter the original features, and it keeps the subset of the original features, it results in maintaining the physical meaning of the original feature sets, better model readability and interpretability. Owing to these advantages, it is used in many real-world applications. Feature selection retrieves original features by removal of redundant and insignificant features. As insignificant and redundant features are removed from the dataset, it results in reduction of the computational, storage costs without significant loss of actual information or negative degradation of the learning performance. To summarize feature selection is a very important step in the data pre-processing module to find the correlated features and delete the redundant or uncorrelated features from the dataset. Noisy features often result in increasing the complexity of the classifier without any contribution of effective information to the classifier. Filter, wrapper and embedded approaches are the different types of feature selection methods.

Filter method: It is generally used in the pre-processing step. Generally the subset of features is selected based on the scores obtained in different statistical test for their correlation with the output variable.

Wrapper method: In wrapper methods, the model is trained by selecting a subset of feature. The new features will be added or deleted based on the inferences obtained in the preceding module. Finally, the problem is shortened to a search problem. Generally, wrapper methods are very expensive in computation.

Embedded method: It merges the characteristics of the filter and wrapper methods. Implementation of this method is in such a way that the algorithms have their own intrinsic feature selection methods.

Importance of feature selection is summarized below:

- It supports faster learning in machine learning algorithm.
- It helps in reducing the complexity of the model
- It helps in improving the efficiency of the model if the right features are selected
- It reduces the over fitting.

Compared to the conventional data, some important points have to be considered on extracting valuable information from the big data. Traditional feature selection methods have challenges when big data characteristics are into consideration.

1. Generally, traditional methods require large amount of learning time; as a result, processing speed finds hard in catching up with the transition in big data.
2. As real-time big data not only includes redundant features, it also includes wrong information, missing values, outlier and so on, so selecting the features from the data by following the conventional methods is not easy.
3. As the data are acquired from different sources in different means, always the received data are not reliable. It might be altered/forged; as a result, it builds up in the complexity of the feature selection.

3 Exploratory Data Analysis

An art of looking at the data to understand the same of its fullest potential is called data exploratory analysis. Data exploratory analysis as the name indicates explores the data to understand the way it is generated, how it is distributed, how it is related with other attributes and how far it is useful [5]. Visualization and transformation are the building blocks of exploratory data analysis. Understanding the data which is of small size will not be tedious as the users can go through the data manually or the traditional approach can help in plotting when the data is well structured and organized. Incomplete and unorganized data and massive amount of data cannot be visualized so easily as small datasets. That's where data visualization comes into picture to understand the huge datasets. Selecting an effective way of presenting the data for understanding is very important as most of the data generated today are incomplete and unstructured in nature. Visualizing the data of afore-mentioned characteristics is possible only by transforming it. Visualization/graphical representation of data starts with the formation of questions. Generating/forming of different set of questions will help in viewing the data, its distribution and relationship between various parameters. Visualizing the data by generating questions is an iterative process by asking questions, finding out answers for that questions and again it continues with new questions. The art of asking quality questions generates large quantity of questions and it aids in better understanding of the data [6].Creating different types of plots based on different types of questions leads to another different set of questions, different sets of graphs and finally helps in deep understanding of the data [7]. The fundamental aims of data visualization are mentioned in the following:

- To understand the relationship between attributes with respect to the context (e.g. number of hours of usage of mobile vs. battery life)
- To find out the data which are very far from the centre/median (outlier detection)
- To understand the underlying structure of the data (hidden pattern)
- To find the most important attribute (feature selection) [8].
- To understand the distribution of data, numeric summaries, aggregation
- To establish a model that explain the data using minimum attributes (dimensionality reduction) [9].

Some of the types of visualization are mentioned as follows:

Interactive visualization: Visualization tool should not be static. User interaction of selecting, zooming and filtering is required for effective visualization. This is very much required especially in the case of viewing spatial temporal data [5].

Streaming visualization: Visualizing the data which is generated in the real time. It should have the capability of dynamic generation of data visualizations from the streaming data [10].

4 Architecture

Considering the importance of data exploratory analysis in identifying the significant attributes, the proposed method used exploratory analysis to select significant attributes and filling of missing values. The flow of the work starting from data collection till the building of the model is depicted in Fig. 1. The steps followed in the process are as follows:

Step 1: Collection of data; data collection might be either online (real-time data or offline data).

Step 2: Identification significant attributes through data exploratory analysis. It also includes elimination of irrelevant variables.

Step 3: Once the significant attributes are identified, the next step is to fill the missing values. Various methods are available to fill the missing values; identifying the appropriate method to fill the missing value is done through exploratory analysis.

Step 4: Once data pre-processing is done, the next step is to split the dataset into training dataset (to train the model) and testing dataset (to evaluate the model).

Step 5: Training dataset will be given to the model for learning and it is an iterative process.

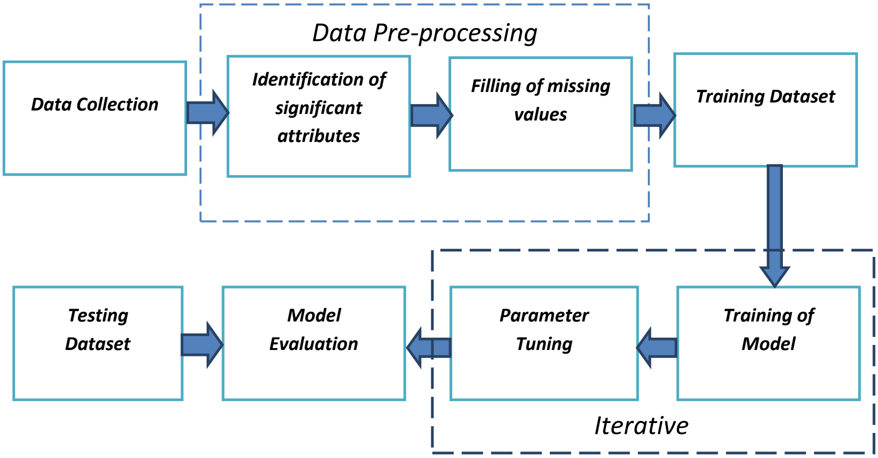


Fig. 1 DEA for ML architecture

Step 6: Tuning of parameter to optimize the model.

Step 7: Once the model is ready, then its performance is evaluated by using the testing dataset.

Logistic regression classification algorithm [11] is used in the proposed methodology. Logistic regression is one of the machine learning classification algorithms that is used to forecast the probability of a categorical-dependent parameter. In logistic regression, the dependent parameter is a binary variable that contains data coded as 1 (True, Yes, etc.) or 0 (False, No, etc.). The basic equation of generalized linear model is:

$$g(E(y)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \tag{1}$$

Here, $g()$ is the link function, $E(y)$ is the expectation of target variable and $\beta_0 + \beta_1 x_1 + \beta_2 x_2$ is the linear predictor (β_0, $\beta_{1,}\beta_2$ to be predicted). The objective of the link function is to 'link' the expectation of y to linear predictor.

Prediction of the Survival of the Titanic Passengers is the problem statement considered for this paper. The model has to predict whether the passenger has survived or not.

The general linear regression equation with the identified significant parameters such as Gender, PClass, Age (dependent variables are identified using data exploratory analysis, which is mentioned in the experimental results) from the dataset is mentioned in the following:

$$g(E(y)) = \beta_0 + \beta_1(\text{Gender}) + \beta_2(\text{PClass}) + \beta_3(\text{Age}) \tag{2}$$

Gender, PClass, Age and IsAlone are the independent variables and survival (y) is the dependent variable that depends on the independent variables.

In logistic regression, the probability of outcome dependent variable (success or failure) is the fundamental thing. As mentioned above, $g()$ is the link function. Link function is based on two things: probability of success (p) and probability of failure $(1 - p)$.

Probability 'p' should satisfy the following conditions:

1. The value should always be positive (since $p >= 0$)
2. The value should be less than or equal to 1 (since $p \leq 1$)

Probability 'p' has to satisfy the above-mentioned conditions which are the basis of logistic regression. Since probability must always be positive, the linear equation has to be represented in exponential form.

$$p = \exp(\beta_0 + \beta_1(\text{Gender}) + \beta_2(\text{PClass}) + \beta_3(\text{Age}) \tag{3}$$

Equation (3) can be rewritten as

$$p = e(\beta_1 + \beta_1(\text{Gender}) + \beta_2(\text{PClass}) + \beta_3(\text{Age}) \tag{4}$$

In order to formulate the probability less than 1, divide p by a number larger than p. This can be simply done by:

$$p = \frac{\exp(\beta_0 + \beta_1(\text{Gender}) + \beta_2(\text{PClass}) + \beta_3(\text{Age})}{\exp(\beta_0 + \beta_1(\text{Gender}) + \beta_2(\text{PClass}) + \beta_3(\text{Age}) + 1} \tag{5}$$

$$p = \frac{\exp(\beta_0 + \beta_1(\text{Gender}) + \beta_2(\text{PClass}) + \beta_3(\text{Age})}{\exp(\beta_0 + \beta_1(\text{Gender}) + \beta_2(\text{PClass}) + \beta_3(\text{Age}) + 1} \tag{6}$$

Using Eqs. (1), (4) and (6), the probability can be redefined as:

$$p = \frac{\mathrm{e}^{\mathrm{y}}}{e^{y} + 1} \tag{7}$$

where p is the probability of success. Equation (7) is the logit function.

If p is the probability of success, then $1 - p$ will be the probability of failure which can be written as

$$\mathrm{q} \ = 1 - \mathrm{p} \tag{8}$$

$$q = 1 - \frac{\mathrm{e}^{\mathrm{y}}}{e^{y} + 1} \tag{9}$$

where q is the probability of failure.

On dividing, Eq. (7)/Eq. (9), the resultant equation is

$$\frac{p}{1-p} = \mathrm{e}^{\mathrm{y}} \tag{10}$$

After taking *log* on both sides,

$$\log\left(\frac{p}{1-p}\right) = y \tag{11}$$

$\log(p/1-p)$ is the link function. Logarithmic transformation on the outcome variable results in modeling a nonlinear association in a linear way.

The resultant equation after replacing the value of y is

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1(\text{Gender}) + \beta_2(\text{PClass}) + \beta_3(\text{Age}) \tag{12}$$

This is the equation of linear regression.

5 Experimental Results

Data visualization as part of data exploratory analysis can be used to identify the relationship between variables (correlated features) and distribution of the variable. With this property of visualization, it can be used in identifying the significant attributes for the training dataset. Dataset which is used in this paper for analysis is Titanic Dataset and the details are mentioned in Table 1. Imbalance factor (IF) in the dataset is calculated by the formula mentioned in Eq. (13).

$$IF = 1 - \frac{n}{\text{Total Samples of all Class}} * min_{j=1,2\ldots n}\left(L_j\right) \tag{13}$$

Table 1 Dataset details

S. no.	Parameter	Details
1	Number of class	2
2	No. of parameters	12 + 1 (output class)
3	No. of records in training dataset	891 (68.07%)
4	Imbalance factor in testing dataset	23.23%
5	No. of records in testing dataset	418 (31.93%)
6	Imbalance factor in testing dataset	27.3%

Table 2 Passengers age representation

S. No.	Age range	Class	Age type
1	0–12	Children	C
2	13–30	Youth	Y
3	30–50	Middle	M
4	Above 50	Senior	S

Table 3 Passenger traveled alone or not

S. No.	Attribute name	Class	IsAlone
1	Sibsp	=0	Yes
2	Sibsp	$! = 0$	No
3	Parch	=0	Yes
4	Parch	$! = 0$	No

where n is the number of class and L_j is the number of samples belonging to the jth class.

Visualization techniques were used in identifying the relationship between the features and the output class for the prediction of the Survival of the Titanic Passengers. It is a structured dataset having 12 attributes (11 attributes + 1 output class). The attributes are Passenger ID, Passenger Class, Passenger Name, Gender, Age, Sibsp, Parch, Ticket ID, Fare, Cabin, Embarked and Survived (Output Class). Attributes such as Passenger Name, Passenger ID, Ticket ID, Fare and Embarked can be eliminated as it will not have any effect on the output class.

Some of the original attributes such as Age, Sibsp and Parch are transformed into another format for having a better understanding. Age values are transformed (categorized) into four different classes as mentioned in Table 2.

Similarly, the Sibsp (sibling/spouse) and Parch (parents/children) parameters are combined together and transformed into another variable called IsAlone (whether the passenger is traveling alone or along with children/parents/spouse/siblings). Transformed variable and its representation are mentioned in Table 3.

Relationship between the input features such as Gender, Passenger class type and Age and the output class is visualized through the bar plots. The data plotted on the plot clearly indicates how these features influence the output class (survival). From Fig. 2, it is known that the percentage of survival rate is more for females (74.2%) than compared to male passengers (18.9%).

Survival rate (SR) is calculated by the formula which is mentioned in the Eq. (14).

$$\mathrm{SR} = \frac{\text{Number of Passengers Survived}}{\text{Number of Passengers Travelled}} \tag{14}$$

Similarly, the next attribute which impacts the output class is the class in which the passengers traveled. Percentage of survival rate of the passengers who has traveled

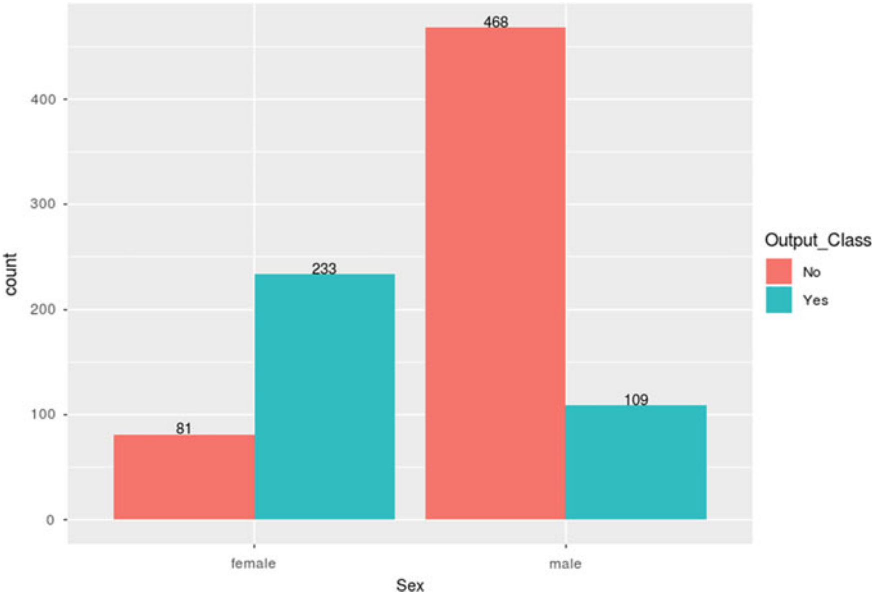


Fig. 2 Gender-wise survival

in the first class (62.9%) is more compared to passengers who have traveled in the second (47.2%) and third class (24.23%). The plot is depicted in Fig. 3.

Another attribute which impacts the output class is the Age. As the age is a continuous value, it is transformed into categorical variable for better analysis. Transformed variable Age_Type was analyzed. It was found that the percentage of survival rate of the children (57.9%) is more compared to other age groups (Middle—42.1%; Senior—34.3%; Youth—33.8%) and the result is depicted in Fig. 4.

Another interesting attribute which is to be considered is IsAlone parameter. This is a derived attribute based on the other two attributes Sibsp and Parch. Analysis results of the same are plotted in Fig. 5. Survival rate of the passengers who traveled along with their family (50.5%) is more than the passengers who has traveled alone (30.3%).

Among the identified significant attributes, attribute Age had missing values. From the analysis it was found that the Passenger Class and the Gender highly correlated with the Passenger's age. So the average age of passengers was calculated passenger's class-wise and gender-wise. The missing values in the age column were filled with the values mentioned in Table 4.

From the data exploratory analysis for the titanic dataset, the identified significant parameters are Gender, Passenger class, Age and IsAlone. Thus, the significant parameters and the filling of the missing values formed a quality training dataset. Once the quality training dataset was formed, the identified significant parameters were given as input to the logistic regression algorithm for training. Once the model is trained using the training dataset, it was evaluated using the test dataset. From the

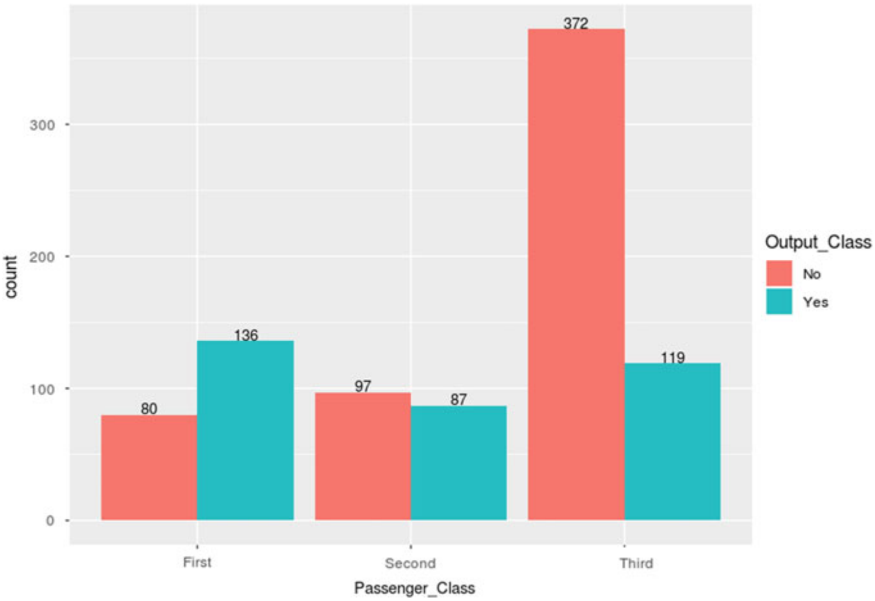


Fig. 3 Class-wise survival

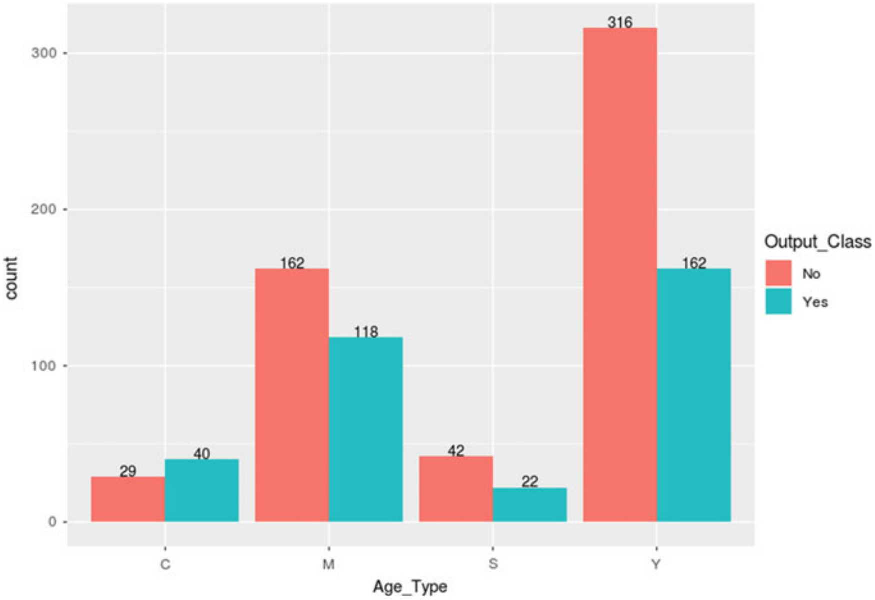


Fig. 4 Age-wise survival

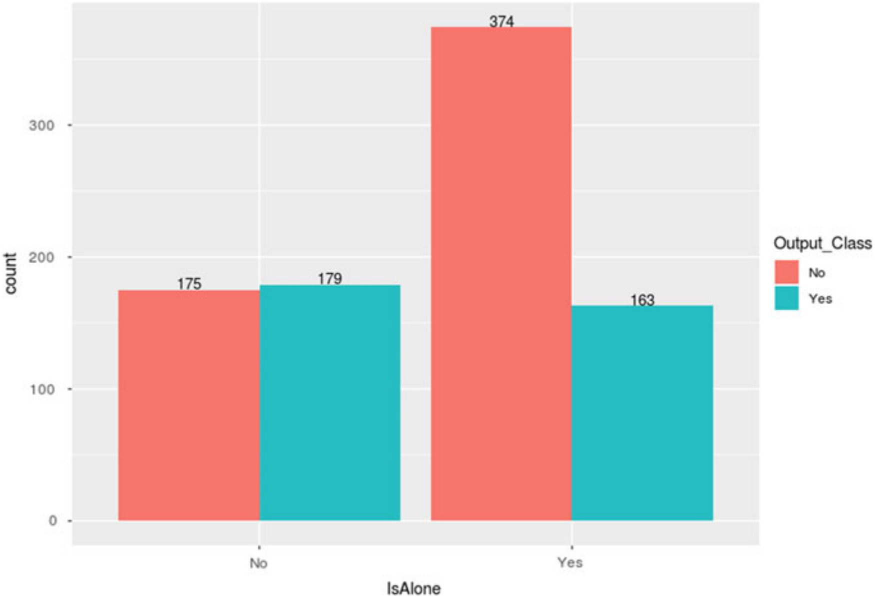


Fig. 5 Survival analysis w.r.t. IsAlone

Table 4 Filling of missing values

S. No.	Class	Gender	Age
1	First	Female	35
2	First	Male	42
3	Second	Female	29
4	Second	Male	31
5	Third	Female	22
6	Third	Male	27

results it was found that the algorithm which is trained with the significant parameter (95.22%) exceeds (by 23.21%) the algorithm which is trained with all the parameters (72.01%).

Further it was analyzed with the IsAlone parameter that though the survival rate is higher for passengers who has traveled not alone than the passengers who has traveled alone, the percentage value is around 50%. So, the algorithm was trained again with the chosen parameters (Gender, PClass, Age) by eliminating IsAlone parameter. The trained model was able to achieve the same efficiency (95.22%). The efficiency of the model trained with all the parameters and the model trained with chosen parameters is depicted in Fig. 6.

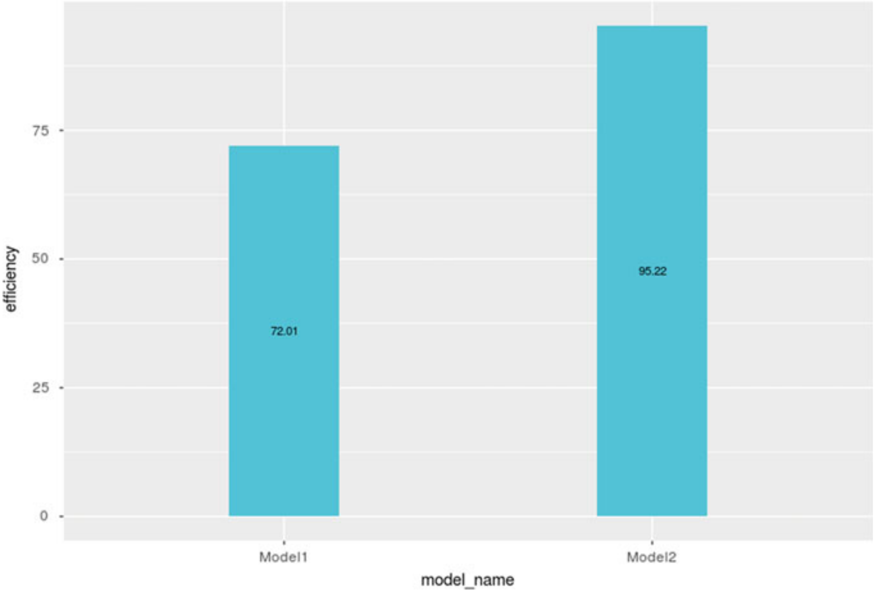


Fig. 6 Efficiency of the two models

6 Conclusion

In this paper we explored the importance of big data and the data exploratory analysis. Data visualization part of exploratory analysis is very important in understanding the big data, its deep structure and its distribution. Indeed, this is achieved by finding out the relationship between variables in the data, outliers, hidden pattern and so on. Mining valuable information from big data is indeed difficult and challenging. As an important data pre-processing technique, feature selection can greatly improve the efficiency of utilizing data. Data visualization technique to find out the relevant attributes for the prediction of survival of the Titanic Passengers was discussed and the results were plotted. Also, the logistic regression model was trained with the quality training dataset and results were discussed and it was found that the model which is trained with the significant parameters performs better than the model which is trained with all the parameters. For future review work, the challenges in the visualization of higher dimensional dataset can be explored and also reducing the processing time of the huge data and the possibilities of parallel visualization can be explored.

Acknowledgments The authors would like to thank the Group Head, Cloud and Data Analytics Department and Chief Scientist of Central Research Laboratory, Bharat Electronics Limited, Bengaluru for their encouragement and support.

References

1. Tsai CW, Lai CF, Chao HC, Vasilakos AV (2015) Big data analytics: a survey. Big data analytics: a survey. J Big Data Springer Open J 2(21):1–32. https://doi.org/10.1186/s40537-015-0030-3
2. Marjani M, Nasaruddin F, Gani A, Karim A, Hashem IA, Siddiqa A, Yaqoob I (2017) Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges. IEEE Access, vol. 5, pp. 5247–5261. https://doi.org/10.1109/access.2017.2689040
3. Samuel Soma Ajibade, Anthonia Adediran.: An Overview of Big Data Visualization Techniques in Data Mining. International Journal of Computer Science and Information Technology Research, 4(3):pp 105–113. (2016)
4. ZhihanLv, Houbing Song, Basanta-Val, Anthony Steed, MinhoJo.: Next-Generation Big Data Analytics: State of the Art, Challenges, and Future Research Topics. IEEE Transactions on Industrial Informatics, 13(4), pp. 1891–1899. https://doi.org/10.1109/tii.2017.2650204 (2017)
5. Andrea Batch, Niklas Elmqvist.: The Interactive Visualization Gap in Initial Exploratory Data Analysis. IEEE Transactions on Visualization and Computer Graphics, 24(1): pp. 278—287. https://doi.org/10.1109/tvcg.2017.2743990. (2017)
6. Sun Guo-Dao, Ying-Cai Wu, Liang Rong-Hua, Liu Shi-Xia (2013) A Survey of Visual Analytics Techniques and Applications: State-of-the-Art Research and Future Challenges. Journal of Computer Science and Technology 28(5):852–867. https://doi.org/10.1007/s11390-013-1383-8
7. Banu onanc Uyan Dur.: Analysis of data visualizations in daily newspapers in terms of graphics design. Procedia - Social and Behavioral Sciences, Elsevier Science Direct, 51: pp. 278–283. https://doi.org/10.1016/j.sbspro.2012.08.159. (2012)
8. Li Jundong, Liu Huan (2017) Challenges of Feature Selection for Big Data Analytics. IEEE Intell Syst 32:9–15. https://doi.org/10.1109/mis.2017.38
9. de Leeuw Wim, Verschure Pernette J, van Liere Robert (2006) Visualization and Analysis of Large Data Collections: a Case Study Applied to Confocal Microscopy Data. IEEE Trans Visual Comput Graphics 12(5):1251–1258. https://doi.org/10.1109/tvcg.2006.195
10. Satyanarayan A, Russell R, Hoffswell J, Heer J (2015) Reactive vega: A streaming dataflow architecture for declarative interactive visualization. IEEE Trans Visualizat Comput Graph 22(1):659–668.https://doi.org/10.1109/tvcg.2015.2467091
11. https://medium.com/greyatom/logistic-regression-89e496433063

Keystroke Dynamics for User Verification

Ashwini Sridhar and H. R. Mamatha

Abstract With the evolution of internet, the dependency of humans on them has increased. This has led to an increase in attacks, forgery, impersonation and so on, which require that a user and his privacy be maintained. Thus the need to protect a user has increased intensifying protection, authentication and verification methods of a user. There are many methods of authenticating a user, which include traditional methods of authentication such as passwords, personal identification numbers and so on, However, these methods have their drawbacks and hence biometrics have replaced these methods in some cases and in some cases biometrics has turned out be an additional layer of security, therefore providing better security. In this paper we propose one of the behavioral methods of biometric authentication called keystroke dynamics which uses a user's typing rhythm to verify a user. One of the most common examples of this method is the verification of user using CAPTCHA, where the user is asked to type the letters to be verified as a genuine user and thus the user's typing rhythm is captured based on which a match is generated and the user is verified. This method is most commonly used in applications such as online banking, email verifications and other such areas. This method acts as an additional layer of security to an existing system and helps protect the sensitive information of the user.

Keywords Biometrics · User verification · Keystroke dynamics · Authentication · Security · Typing rhythm

A. Sridhar (✉) · H. R. Mamatha
PES University, Bengaluru, India
e-mail: asashwini38@gmail.com

H. R. Mamatha
e-mail: mamathahr@pes.edu

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_14

1 Introduction

For over many decades, the combination of username/password has been used for protecting electronic information systems and services. Although there are variations to this, like usage of email address or user ID instead of username, the fundamental concept has remained the same.

The combination of username/password for securing information systems is nearly 50 years old. This method was at first developed in 1961 at MIT (Massachusetts Institute of Technology) and has been in use thereon for securing most of the online services that comprise email service, banking systems and so on. Figure 1 presents a traditional authentication system.

However, due to availability of modern commodity hardware systems with better processing and storage capacity, it is becoming easier for hackers to crack the password. Hence the research community in the security domain has been working on novel type of authentication and authorization system for securing the systems.

Biometric authentication has replaced the traditional authentication method. There are two types of biometrics: physical and behavioral. This work focuses on a behavioral-based biometric called keystroke dynamics. Keystroke dynamics is the analysis of a user's typing pattern based on which a user can be verified as genuine or not. The basic features usually collected are keydown-keydown time, hold time and keyup-keydown time. Figure 2 presents these features.

This work adopts the method of keystroke dynamics as a means to verify the genuity of a user. This method thus provides better protection and an additional layer of security when combined with the traditional methods. It is also proved to be a strong method of authentication when used alone. The rest of the paper is divided as follows: Related works are presented in Sects. 2; Sect. 3 outlines the methodology and implementation used; the results and discussions are depicted in Sect. 4; Sect. 5 and Sect. 6 represents the conclusion and future work respectively.

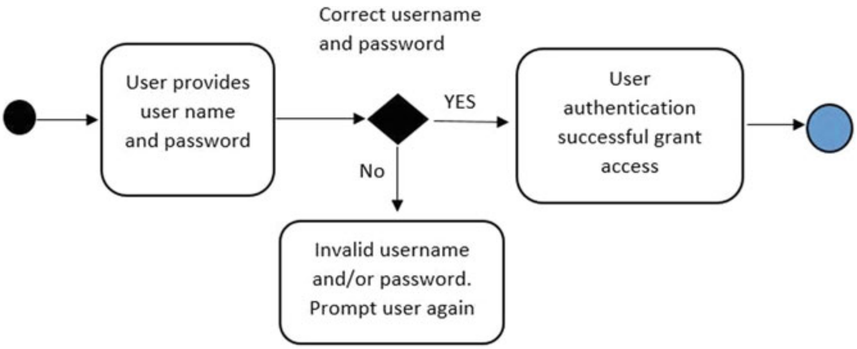


Fig. 1 Traditional authentication system

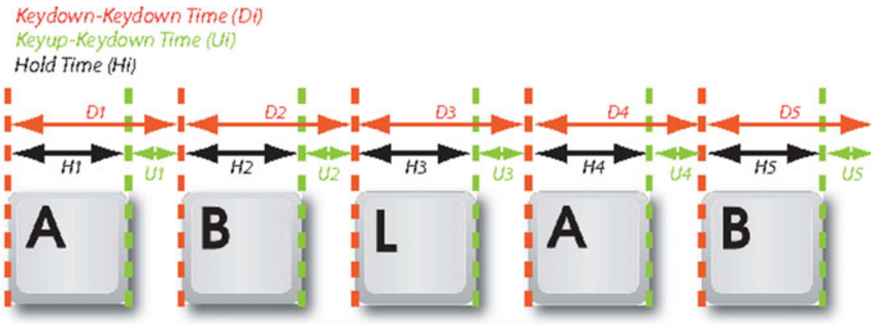


Fig. 2 Basic features of keystroke [6]

2 Related Works

Keystroke dynamics analysis has been done by different people in different ways, and each of them have arrived at their own results. This section describes in brief the work done by different people and the algorithms used to analyze the keystroke patterns of a user.

According to the work done by Killourhy and Maxion, comparing anomaly detection algorithms [1] states the best performing algorithms based on the equal error rate which was calculated on the dataset collected consisting of the user's typing patterns. The dataset comprised 51 users, which were then evaluated for a total of 14 different classifiers.

Keystroke dynamics has proved to be a wide field of research and a lot of studies have been conducted recently. There are many parameters that are taken into consideration while considering a user's typing pattern. In [2] the author talks about such parameters. This work also focuses on increasing the reliability of authentication of a user and hence makes use of keystroke dynamics as a biometric method.

The work done in the field of keystroke dynamics consists of multiple features and methods of evaluation. The work by Abdullah et al. [2] talks about an algorithm called dynamic time warping (DTW) which makes use of waveforms in order to arrive at a suitable estimation of performance.

The traditional methods of authentication make use of passwords, PINs and so on as a method of authentication. However, with the evolution of technology, it was observed that these methods of authentication alone do not provide enough security for the user data. Hence to improve the security, keystroke dynamics is used as an additional layer of security. The author in [3] includes keystroke dynamics as an additional layer of authentication to the traditional password-based authentication. The anomaly scores are calculated by using various distance-metric algorithms such as Manhattan distance and Mahalanobis.

With the increase in risk to security everyone requires a safe, quick and trustable source of communication. This requires protection of data by means of authentication. The work done by Maheshwary et al. [4] describes the method of safe, quick and trustable source of communication. The work makes use of keystroke dynamics as a method of authentication, and this is done by using the nearest neighbor algorithm.

3 Methodology

In this work of verifying a user based on their keystroke dynamics, we have studied the performance of different algorithms. The general methods followed are: data loading, data selection, training, testing and calculation of equal error rate (EER).

In the data loading phase, the data are loaded from a text file into the system. These data are then split into training and testing sets, where the first 15 vectors are for training and the rest for testing. The data are then trained where the current user and his data are taken as a genuine user data for training and the rest of the data are treated as imposters.

For the test phase, user and imposter scores are calculated. If the score is high, it is proved that the user is not genuine. EER is calculated as the total number of incorrect predictions divided by the total number of values in the dataset. 0.0 and 1.0 are considered to be the worst and the best error rates, respectively. It can be represented as follows:

$$\mathrm{EER} = \frac{\mathrm{FP} + \mathrm{FN}}{\mathrm{P} + \mathrm{N}}$$

where FP: false positive, FN: false negative, P: positive, N: negative

Accuracy (ACC) can be calculated as the number of correct predictions to the total number of values in the dataset. 0.0 and 1.0 are considered to be the worst and the best accuracies, respectively. It is calculated as

$$\mathrm{ACC} = 1 - \mathrm{EER}$$

Therefore, it can be concluded that better the EER, better the accuracy.

We have used various classification systems in order to measure the performance of the algorithms on the system and verify a user. The algorithms used are Manhattan scaled distance [1], nearest neighbor Mahalanobis [1], outlier count [1], K-nearest neighbor (KNN) [5], recurrent neural network (RNN) [6], dynamic time warping (DTW) [2], convolutional neural network (CNN) [6] and decision tree.

3.1 Implementation

The accuracy and effectiveness of the authentication system depend on the input dataset used. The dataset should comprise large data in order to successfully verify a user's identity. For the current project we have collected a dataset from 78 users,

Table 1 Data collected

Subject	H.period	DD.period.t	UD.period.t	H.t	DD.t.i	UD.t.i
Ashwini	0.078	0.531	0.453	0.078	0.281	0.203
Ashwini	0.063	0.391	0.328	0.078	0.266	0.188
Ashwini	0.079	0.547	0.468	0.078	0.188	0.11
Ashwini	0.062	0.546	0.484	0.079	0.282	0.203
Ashwini	0.063	0.5	0.437	0.094	0.157	0.063
Ashwini	0.078	0.437	0.359	0.094	0.125	0.031

each of them typing the password used in [1], “tie5roanl”, 30 times. Table 1 presents the dataset collected, which represents the timing data of each key press. The basic features include keyup-keydown time, which is the time between release of one key and the press of next; keydown-keydown is the time between continuous key presses; and the hold time which is the time between the press and release of each key. These features are collected for each letter of the password. In order to increase the efficiency of the algorithm, attributes like age, gender, trigram and bigram time are also added.

This dataset is evaluated using different detection algorithms like KNN, RNN, CNN, and the top performing algorithms used in [1] that are Mahalanobis and Manhattan scaled.

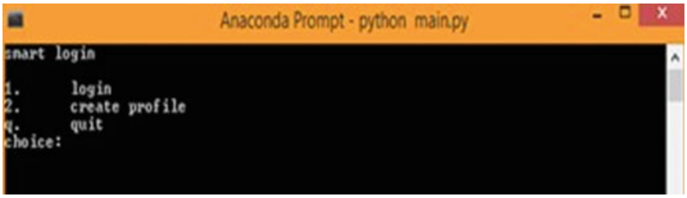
In Table 1 the first column represents the subject, that is, the user; the second column represents the hold period duration for the password typed by the user where each row represents the password typed by the user once. The third column represents the keydown-keydown period, and the fourth column represents the keyup-keydown period. The rest of the columns represents the hold time (H time), keydown-keydown time(DD time) and keyup-keydown time (UD) for each letter in the password typed by the user.

The dataset was collected with the help of a console-based application that was developed. Figure 3 presents this application. There were two options provided in the application:

1. Login: In this option the user is asked to login with his username and password which is used to authenticate the user.
2. Create profile: This option is used whenever a new user profile has to be created in order to collect the features. Once this option is selected the user is asked to type his username and the password which are stored in the text file.

This file is then used as the basis for authentication of a user.

In order to provide accurate results, it is important that we have the right functional, data and system requirements. Since it is based on machine learning algorithms used, it is important that we have enough data. Hence we collected a total of 2500 keystrokes. The general requirement includes a Windows or Linux OS with 4 or 8 GB RAM with suitable python environment and packages. We used Python 2.7 environment along with the scipy, numpy packages.



```
smart login

1.      login
2.      create profile
q.      quit
choice:
```

Fig. 3 Dataset collection application

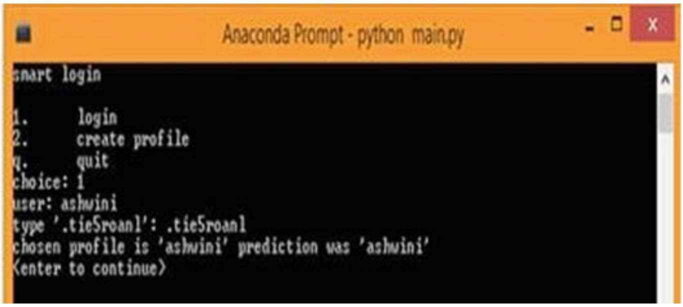
4 Results and Discussions

The data of approximately 80 users was collected. Figure 4 presents the authentication of a user.

The initial accuracy of the data seemed good. Table 2 presents the initial accuracy. As the number of users increased for the data when tested, it seemed to decrease the accuracy.

Hence the number of users was reduced with each turn and the accuracy was tested. Table 3 presents the variance in the accuracy for the dataset as the users are reduced.

In spite of reducing the users, it was seen that the maximum accuracy obtained was 50% for 40 users. Hence there was a need to re-evaluate the same data with additional features and algorithms to achieve better accuracy.



```
smart login

1.      login
2.      create profile
q.      quit
choice: 1
user: ashwini
type '.tie5roanl': .tie5roanl
chosen profile is 'ashwini' prediction was 'ashwini'
<enter to continue>
```

Fig. 4 Authentication of a user

Table 2 Initial accuracy

User Id	No. of times password entered	Correct	Wrong	Accuracy (%)
1	10	6	4	60
2	10	8	2	80
3	10	9	1	90
4	10	4	6	40

Therefore we added attributes such as age, gender and trigram time. The system performance was also measured with other algorithms, such as K-nearest neighbor (KNN), recurrent neural network (RNN), convolutional neural network (CNN), dynamic time warping (DTW) and decision tree classifiers.

After the addition of new features, we re-evaluated the algorithms. Table 4 presents the evaluation done based on the equal error rate (EER).

The first column in Table 4 represents the top performing algorithms based on the work done by Killourhy and Maxion [1]. The second column represents the EER results obtained in the benchmark dataset, that is, the evaluation done in [1]. The last column represents the evaluation based on EER for the dataset collected. Based on this EER, the top performing algorithms were established.

Similarly, the other algorithms such as K-nearest neighbors (KNN), recurrent neural networks (RNN), convolutional neural networks (CNN) and decision tree were evaluated. Table 5 summarizes the algorithms used with their accuracy.

Table 3 Variance in the accuracy of data

Total number of users in the dataset	No. of times password typed by a single user	No. of correct predictions	Accuracy
78	10	3	30
56	10	3	30
43	10	4	40
40	10	5	50

Table 4 Top performing algorithms based on EER

Algorithms	EER of benchmark dataset	EER of dataset collected
Manhattan scaled	0.17681169012847736, 0.10416561236462442	0.0674319757690701, 0.04717726419913844
Nearest neighbor (Mahalanobis)	0.3063765640274061, 0.10974436873298003	0.10893319797198889, 0.18134704000510013
Outlier count	0.13730802044159202, 0.09555389158452095	0.06919125305634136, 0.06404148467115772

Table 5 Accuracy of algorithms evaluated

S. No.	Algorithm	Accuracy (%)
1	KNN	90
2	RNN	85
3	CNN	2
4	Decision Tree	7

Dynamic time warping presents the comparison between two waveforms of a user. Figure 5 presents the peak comparison of a single user. From Fig. 5 it is observed that the peaks of a single user vary each time the user inputs the password. This is because of the key press and typing rhythm of the user which also varies with each input. Figure 6 presents the peak comparison of different users. From Fig. 6 it can be observed that peaks of each user vary due to the difference in the typing rhythm as well as the key press durations.

In order to arrive at the best performing algorithm, it is important that the factors like false positive and true positive be considered. This helps in determining the accuracy of a system. Thus it leads to an appropriate conclusion. Table 6 presents the false positive and true positive for all the algorithms used in the evaluation.

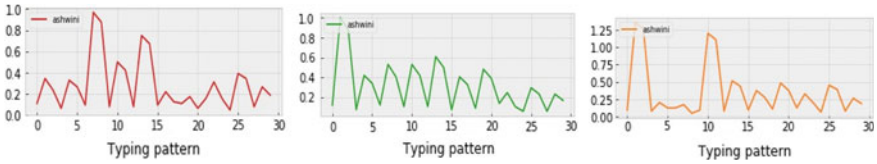


Fig. 5 Peak comparison of one user

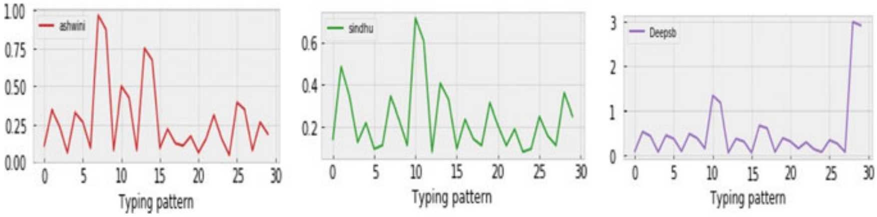


Fig. 6 Waveform Comparison of different users

Table 6 False positive and true positive for algorithms

Algorithm	Number of samples	False positive	True positive
Manhattan scaled	77	9	68
Nearest neighbor (Mahalanobis)	77	43	34
Outlier count (Z-score)	77	8	69
CNN	474	463	11
RNN	15	5	10
Decision tree	712	473	39

On the basis of our analysis, it was found that the best performing algorithm based on the equal error rate (EER) from Table 4 when compared with the benchmark dataset is Manhattan scaled algorithm. However, the outlier count and nearest neighbor (Mahalanobis) were found to be the second and third best when compared to the benchmark dataset. This may be due to slight variations in the data collected.

Of the algorithms in Table 5, KNN was found to be the most accurate algorithm, while RNN was slightly less accurate in comparison to KNN. The algorithms CNN and decision tree were found to be the least accurate algorithms with accuracy of below 10%.

Hence from Table 5 it can be concluded that CNN and decision tree algorithms are not suitable for time series data because CNN requires a large amount of multi-dimensional data collected over a long period of time for each individual in order for it to be thoroughly trained and tested. The data we have collected here are not enough. Therefore they do not produce accurate results and cannot be used.

It is therefore clear that the performance of the verification systems depends on the data collected and the features used. The performance of the algorithm, as well as the accuracy also, depends on the data and the features. Thus it can be concluded that the dataset and the features play an important role.

The proposed method of user verification using keystroke dynamics when compared to the existing techniques provide better security in terms of user privacy, verification, imposter user and other such things. In the techniques that are usually used, such as authentication through passwords, it becomes easy for an imposter to impersonate the password and the user's passwords and PINS can be hacked easily.

Keystroke dynamics acts as an additional layer of security protecting the user's privacy and user information as the typing and key press rhythm of each user is different. The difference in the typing and key press of each user makes this method better when compared to the traditional methods of security, and thus it is impossible for an imposter to impersonate the user. Hence keystroke dynamics proves to be one of the most preferred methods of user verification.

5 Conclusion

As witnessed in the design and implementation of verification of a user using keystroke dynamics, it can be concluded that the experiment is successful in achieving the targeted application feature.

The goal of authenticating a user based on the user's keystroke dynamics by building a security application has been successfully achieved. It was observed that as the number of users increased, the accuracy decreased. Hence it was necessary that different algorithms be applied and additional features be added in order to improve the efficiency of the system.

Based on Sect. 4 from Tables 4 and 5, the best performing algorithms were found to be Manhattan scaled and KNN with an accuracy of 88.3 and 90%, respectively, while the least performing algorithms were found to be decision trees and CNN with an accuracy of 7 and 2%, respectively.

The user verification method proposed in this work can be used as an additional layer of security for many applications, such as banking, various transactions and other such areas, therefore improving user authenticity, genuity and thus help preserve user security.

6 Future Work

This work is only limited to desktop applications and makes use of basic features, such as keyup-keydown time, keydown-keydown time, hold time and trigram time. It can be further extended to other computing devices such as smart phones and tablets with the addition of features suh as right handed or left handed etc.

References

1. Killourhy KS, Maxion RA (2009) Comparing anomaly-detection algorithms for keystroke dynamics. In: 2009 IEEE/IFIP International conference on dependable systems & networks. https://doi.org/10.1109/dsn.2009.5270346
2. Sulavko AE, Eremenko AV, Fedotov AA (2017) Users' identification through keystroke dynamics based on vibration parameters and keyboard pressure. In: 2017 IEEE dynamics of systems, mechanisms and machines (dynamics) (Omsk, Russia) 14 Nov–16. https://doi.org/10.1109/dynamics.2017.8239514
3. Abdullah A, Frans C, Danushka B (2016) Towards keystroke continuous authentication using time series analytics, Springer International Publishing AG 2016 M. Bramer and M. Petridis (eds.), Research and Development in Intelligent Systems XXXIII, https://doi.org/10.1007/978-3-319-47175-4_24
4. SoumenRoy,Utpal Roy, D. D. Sinha, September 2014. Enhanced Knowledge- Based User Authentication Technique via Keystroke Dynamics. International Journal of Engineering Science Invention ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726 www.ijesi.org Volume 3 Issue 9 || September 2014 || PP.41–48.33
5. Lu X, Zhang S, Yi S (2018) Continuous authentication by free-text keystroke based on CNN plus RNN. In: 2018 International Conference on Identification, Information and Knowledge in the Internet of Things, IIKI 2018, 147, pp 314–318, https://doi.org/10.1016/j.procs.2019.01.270
6. Venugopalan S, Juefei-Xu F, Cowley B, Savvides M (2015) Electromyograph and keystroke dynamics for spoof-resistant biometric authentication. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). https://doi.org/10.1109/cvprw.2015.7301326
7. Saket M, Vikram P (2016) Mining keystroke timing pattern for user authentication, Springer International Publishing AG 2017 A. Appice et al. (Eds.): NFMCP 2016, LNAI 10312, pp 213–227. https://doi.org/10.1007/978-3-319-61461-814
8. Obaidat MS, Macchairolo DT (1994) A multilayer neural network system for computer access security. IEEE Trans Syst Man Cybernet 24(5):806–813. https://doi.org/10.1109/21.293498

Activity Prediction for Elderly Using Radio-Frequency Identification Sensors

Prashant Giridhar Shambharkar, Sparsh Kansotia, Suraj Sharma, and Mohammad Nazmud Doja

Abstract In hospitals and nursing homes, older people usually fall due to weakness and disease. Standing or walking for a long time can be two of the many reasons for falling. One of the better ways for fall prevention is to monitor patient movement. A new kind of batteryless light sensors is providing us with new opportunities for activity prediction, where the inconspicuous nature of such sensors makes them very suitable for monitoring the elderly. In our study, we analyze such sensors known as radio-frequency identification (RFID) tags to predict the movements. We try to study a dataset obtained from 14 healthy old people between 66 and 86 years of age who were asked to wear RFID sensors attached with accelerometers over their clothes and were asked to perform a set of pre-specified activities. This study illustrates that the RFID sensor platform can be successfully used in activity recognition of healthy older people.

Keywords RFID sensors · Elderly · Activity · Healthcare · XGBoost · Random forest · Classification · Fall detection

1 Introduction

The count of people above the age of 60 years is increasing rapidly. India is ranked second in population and has almost 72.2 million people aged 60 and above. Falls are common among these people and recurrent falls are a major cause of morbidity. In

P. G. Shambharkar (✉) · S. Kansotia · S. Sharma
Delhi Technological University, Jamia Millia Islamia, New Delhi, India
e-mail: prashant.shambharkar@dtu.ac.in

S. Kansotia
e-mail: sparshkansotia_bt2k16@dtu.ac.in

S. Sharma
e-mail: surajsharma_bt2k16@dtu.ac.in

M. N. Doja
Jamia Millia Islamia, New Delhi, India

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_15

July 1999–April 2000, a survey was performed with 250 subjects from Chandigarh and Haryana in India. Various socio-demographic characteristics as provided by the study were recorded at baseline. A detailed and examined diagnosis was reported by a physician based on observed medical concerns, organized inspection, and cross-checking of subjects' medical records and related medications [1–6]. According to the survey, almost 103 people (that contributed to 51.5%) of the total population that participated in the survey obtained with the help of demographic study had fallen. There was a fracture reported in 21.3%, and other injuries occurred in 79.6% of those who had fallen. Falls in hospitals are responsible for deaths, hip fractures, and fractures to other parts as well. The psychological consequences of plummeting include dismay, consternation, cynicism, the feeling of negativity, fear of toppling, and decline in health. Without productive policies to counter these, the total count of and loss from plummeting in hospitals will rise further. The current research and development on fall prevention constructively introducing good measures which are somewhat helping in bringing down the rates, but nevertheless, the results still are unsatisfying. Other techniques such as installing cameras to monitor patients raise privacy concerns.

The vision-based recognition techniques have disadvantages because insufficient lighting, changing backgrounds, and the presence of multiple people in the room under focus can pose difficulties for detection using the equipment. Wearable sensors solve these problems where multiple subjects can be treated without interaction with each other [1]. Sensors capture details from body movements of individuals which are used in classifying activities of individuals as a task of classification using machine learning. Mostly sensors were heavy and battery-operated and were inconvenient to carry for patients [6]. Some other studies consisted of multiple sensors clipped to distinct parts of the patient's body to provide detailed data regarding the motion of the person. For instance, kinematic sensor data is evaluated with the help of Fourier transformations and wavelet decomposition methods as in [4] while others use environmental variables and vital signs (heart rate, respiration rate) are also used in addition to feature extracted motion data in [5]. Older people mostly prefer light-weighted and small-sized wearables that are easy to maintain and carry. These characteristics of sensors make people to really opt for such activity selection using motion sensors otherwise people look for more convenient options. A new generation of passive (batteryless) sensors which are sensor-enabled radio-frequency identification (RFID) tags solved the problems of the battery-operated wearables. In contrast to sensors that operate on batteries, passive sensors are light-weighted and smaller. In addition, passive sensors do not ask for maintenance as they do not make use of chemical energy stored in accumulators. Moreover, passive devices can be easily carried by clipping them on the outfit and thereby preventing removal of the monitoring device, especially by cognitively impaired patients. Therefore, monitoring systems with passive sensors are more common nowadays. In this study, data to be used were collected by making use of a single passive sensor-enabled radio-frequency identification tag embedded with an accelerometer for capturing the accelerations along different axes (henceforth sensor tag) for real-time activity recognition (Fig. 1).

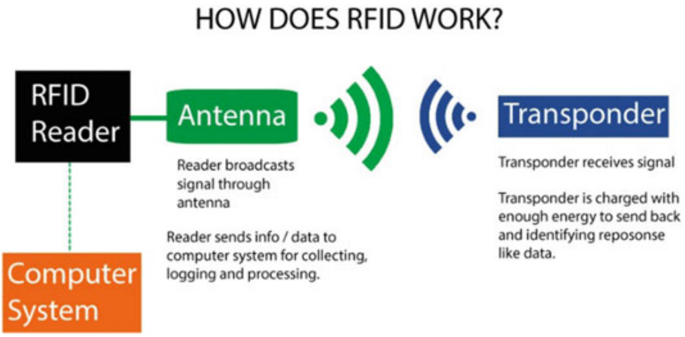
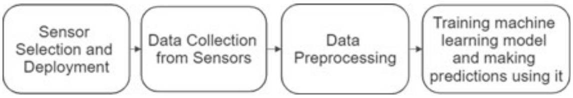

HOW DOES RFID WORK?
RFID Reader
Antenna
Transponder
Reader broadcasts signal through antenna
Reader sends info / data to computer system for collecting, logging and processing.
Transponder receives signal
Transponder is charged with enough energy to send back and identifying reposonse like data.
Computer System

Fig. 1 Working of RFID sensors. *Source* https://armourcard.com/what-is-rfid/

Apart from the acceleration signal, sensor-enabled RFID tags also provide information related to signal strength, which provides contextual information [7].

Since passive sensors are batteryless, their signals capture a fair amount of noise along with the transmitted data. The transmitted data streams have low sampling rates. Radio-frequency identification tags occupy a range of electromagnetic spectrum for broadcasting of signals to its receptors, much like most of the wireless equipment transmitter for communication. The frequency with which to transmit can be selected from the given range of frequencies from the spectrum depending on the purpose. This selection must be done such that it prevents the interference of the RFID signal with the signals from other tags or devices operating in concerned range as the interference of signals would lead to mixed signals which would distort the information from both transmissions. To ensure that the information is transmitted properly and without any distortion, RFID systems make use of a technique called time division multiple access (TDMA). As batteryless RFID tags do not utilize power from stored sources, they obtain energy from the reader. The information from the tags can be obtained from up to 20 feet of distance. Another highlight of the RFID tags is that they are produced with minimal costs due to the technology they use and therefore can have many applications in many fields. These tags can be disposed of when done with, due to their nature and low cost. The three storage types that decide the production cost of the tags are read-write, read-only, and WORM (write once, read many). With a read-write tag, information can be either overwritten or can be added to the previously contained data. Only the information written in the read-only tags can be read from it. Neither can the read-only tag be overwritten nor any data can be added to it. WORM tags cannot be overwritten, though some extra information can be added to them for one time [9] (Fig. 2).

Fig. 2 Architecture used


Sensor Selection and Deployment
Data Collection from Sensors
Data Preprocessing
Training machine learning model and making predictions using it

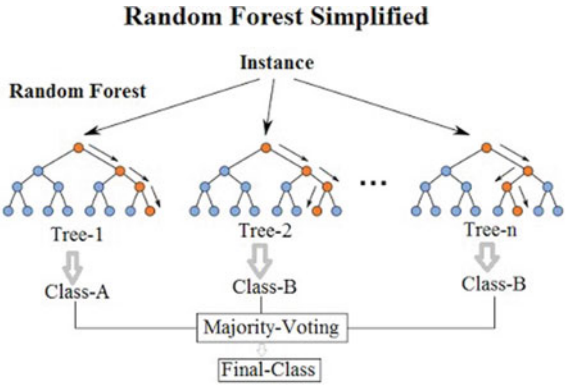


Fig. 3 Random forest. *Source* https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d

In the course of this study, we investigate whether motion data possible from interpolated acceleration signals from passive sensor-enabled RFID tags would provide a good basis for real-time activity recognition using machine learning classification technique called random forest. As a prerequisite, we have obtained two real-world datasets accumulated for ambulatory monitoring of older patients in patients' room settings [7] (Fig. 3).

2 Related Works

Table 1 provides us with the details of previous works on fall prevention. The table species the set of features and the techniques used in each of the research given in the table along with the results achieved. The features are explained in the next section.

The study taken from [2] focused on recognizing the exits from bed and chair as a part of a larger problem of preventing falls among the elderly. The concern with this paper was that the precision recognizing exits was low as the number of false exits was large. This was due to the combination of the ineffectiveness of the classifier used and the inability of the applied sensors to observe the activities

Table 1 Study of features used by previous research papers and their results

S.No	Author(s)	Year	Features Used																			Techniques Used	Results				
			α_f	α_t	α_v	θ°	α°	β°	a_x	a_y	a_z	a	RSSI	d_s	v_s	$\dot{y}$	RMS(y)	σ_y	σ^2_y	G	αID		Precision	Recall	Specificity	Fscore	Accuracy
1.	Roberto L. Shinmoto Renuka Visvanathan Derek Abbott Keith D. Hill Damith C. Ranasinghe	2017	✓	✓	✓	✓						✓	✓	✓						✓	✓	Weighted Support Vector Machines	66.82	81.44		72.48	
2.	Asanga Wickramasinghe Damith C. Ranasinghe Christophe Fumeaux Keith D. Hill Renuka Visvanathan	2016	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓					✓	✓	Bed-egress recognition proposed framework	85.00	86.00			
3.	Asanga Wickramasinghe Damith C. Ranasinghe	2015	✓	✓	✓	✓	✓	✓				✓	✓	✓	✓						✓	Support Vector Machine (linear) Support Vector Machine (rbf) Conditional Random Fields	87.87 90.39 85.97	83.44 87.42 82.35		84.96 88.45 83.73	
4.	Asanga Wickramasinghe Damith C. Ranasinghe	2015	✓	✓	✓	✓	✓						✓	✓	✓						✓	Random Forest Conditional Random Fields	94.53 87.11	93.95 92.99	91.31 90.36	94.00 89.20	92.62 91.69
5.	Óscar D. Lara Miguel A. Labrador	2012										✓				✓	✓	✓	✓		✓	Artificial Neural Network Support Vector Machine					97.90 97.51
6.	Pierre Barralon Nicolas Vuillerme Norbert Noury	2006							✓	✓	✓						✓				✓	Continuous Wavelet Transform Discrete Wavelet Transform Short Term Fourier Transform		58.20 68.70 47.10	60.1 56.7 38.0		

during transition. Non-overlapping windows method was used to reduce the false exit count which worked to some extent but some delay was added in the alarming system which indicated an exit from the bed or chair. The study of [3] considered only those features that were previously used by other activity recognition studies without performing feature learning. Also this study does not take care of the situation when the data stream obtained from the sensors is sparse. The next study from the same researchers as [3] obtained low f-score due to inadequate features used. Furthermore, the data used by this study was collected from young healthy adults and was not similar to the target group of elderly people. In the study given in [4], the researchers devised their own sequence classification algorithm that was supposed to be good for sparse data streams from the sensors but in practice it did not give better results than other studies. Further, it also incorporated some delay in its response for the falls. The techniques used in the study given in [4] gave inferior performance as compared to other studies. The study given in [5] used some extra parameters along with sensor data for activity recognition. These parameters included environmental attributes, location, and physiological signals which though helped in enhancing the performance of the system but it increased the costs incurred and reduced the chances of its practical application.

3 Proposed Work

3.1 Radio-Frequency Identification Sensor System

The data that is utilized in this study is captured by wireless sensors using a sensing and identifying platform and these can be attached on the clothes worn by the subjects during experimentation [7].

The sensor is a passive device and hence does not require batteries. In turn, the sensor tag gathers energy from the electromagnetic field spread around itself. This field is generated by the RFID reader antennas during the interrogation from the RFID reader when information that is captured by the tag needs to be transmitted for research. When enough energy is captured by the tag, the energy moves to the integrated circuit of the tag that powers the microprocessor which in turn backscatters a signal back to the field. Owing to a new signal, the nature of the electromagnetic field is changed which is then noticed by the reader antenna. Then the antenna captures the information transmitted through this signal [2]. Even though there are multiple benefits like convenience, portability, affordability, and disposability in using the wearable sensors for gathering the data, certain limitations arise in their use. First, and most importantly, due to lack of a dedicated antenna the tags can transmit to a minimal range of frequency. The tags that use low to medium frequency may transmit to a maximum range of 2 ft and if a high frequency is used then up to 7 ft may be achieved. These distances are still relatively minimal. This leads to incomplete and non-uniform information to which certain noise is added due to the nature of environment and this data is transmitted to the sensing platform for the analysis as stated in [8]. The other limitations include limited memory capacity of the tags, requirement of high-powered reader, and lack of constant powered source which is needed by most sensing application. The movements being considered for activity recognition here are typical movements for the older people that would greatly help in identifying activities.

The set of movements considered consists of only those that greatly influence the predicted results and no extra data is involved. The motion data is gathered by RFID sensing platform. Presence of some extra clothing over the sensors does not affect the sensing operation but any solid object may cause hindrance.

3.2 Data Collection

Experimental data were obtained from the UCI repository and was captured by University of Adelaide, Australia for activity recognition in aged people separately in two rooms and datasets were prepared accordingly taking essential parameters of the observations into account [7]. The activities included in the two datasets consisted of 1: sit on bed, 2: sit on chair, 3: lying, 4: ambulating. The datasets as stated earlier were captured in two well-prepared room configurations (Room1 and Room2). Room1 and Room2 were set up with four and three RFID reader antennas, respectively. While in the Room1, one of the antennas was attached to the ceiling right above the bed, there were two antennas in the same position in Room2. The remaining antennas in Room1 were positioned on the walls of the room in such a manner that they could very well capture disturbances in the area along the chair and the bed in the room. Similarly, one antenna in Room2 was positioned such that its focus is on the chair.

The wearable sensor tags from the sensing system were attached to a group of people aged between 66 and 86 years for this study. The tag is to be placed in the center position of the chest on top of the clothes that the subject is wearing. The position of placement of the tag is important as the best indication of movement of the torso or the core of the body is given in this manner. All the subjects of the research were asked to perform their routines normally while taking care that data is being captured and sensor is working correctly. The activities mostly consisted of those mentioned earlier and were performed in a sequential manner such that there are no half actions which would somehow distort collected data and cause difficulties in analysis afterwards. The activities that are taken into consideration cover most of the routine of hospitalized aged people, such as having food while sitting strolling inside a cabin [7].

Attributes used by previous researchers are:

α_F	"Frontal acceleration"
β	"Angle on transverse plane"
RSSI	"Received signal power from the sensor"

σ_y	“Standard deviation for acceleration signals”
α_L	“Lateral acceleration”
a_x	“Anteroposterior acceleration (using θ, a_F, and a_V)”
d_v	“Vertical displacement”
σ_y^2	“Variance for acceleration signals”
α_V	“Vertical acceleration”
a_y	“Mediolateral acceleration (using α, a_L, and a_V)”
v_v	“Change in vertical velocity”
G	“Gender of the subject”
θ	“Angle on sagittal plane”
a_Z	“Dorsoventral acceleration (using θ, α, a_F, a_L, and a_V)”
$\acute{y}$	“Mean value of acceleration signals”
αID	“RFID antenna receiving current tag reading”
α	“Angle on coronal plane”
a	“Magnitude of acceleration vector”
RMS(y)	“Root mean square for acceleration signals”

Attributes used by this paper are:

Attribute 1	“Time of activity (Seconds)”
Attribute 2	“Frontal acceleration”
Attribute 3	“Vertical acceleration”
Attribute 4	“Lateral acceleration”
Attribute 5	“Identification number of antenna sensor”
Attribute 6	“Received signal strength indicator (RSSI)”
Attribute 7	“Signal’s phase”
Attribute 8	“Signal’s frequency”
Attribute 9	“Labels of activities, 1: sitting on the bed, 2: sitting on the chair, 3: lying in the bed, 4: ambulating” [7]

3.3 Data Preprocessing

The complete data consist of data from the two clinical room setups (Room1 and Room2) which must be appended to one dataset and checked for null values for activity recognition using machine learning classification tasks. Then the values in the dataset are normalized or converted to a relatively similar scale without losing any pattern information and introducing any noise in the values. This is due to the better performance of machine learning models when features are normally distributed, that is all values are scaled between specified ranges.

To normalize an independent variable or feature means to scale all the values belonging to that feature for different instances to a common range of values without compromising the differences among values which contribute toward efficient analysis.

Normalization or feature scaling becomes extremely important when the data to be modeled in skewed, that is, there is a huge difference in the number of instances of each class that is to be considered a label. Both text classification and clustering require normalization as a common practice in their operations.

3.4 Random Forest Classifier

What is a Decision Tree?

Decision trees are machine learning classification and regression algorithm that can be utilized to identify the pattern present in the data being studied even if it is complicated to do so. The aim of making it a tree is that each node is divided into two or more branches based on a splitting criterion that selects on which attribute values the splitting is to be done and this would classify new instances in the best way. The tree starts with a root which grows into more nodes through the branches of nodes and leads to what we call leaf nodes that are present in the last level of a complete tree. This procedure of tree growing is done in accordance with one of the many splitting criteria defined. The chosen one is called optimal splitting criteria for the given problem and it depends on the nature of the data. Every node poses a condition for the unclassified instance which branches to the child node based on what value this instance has for the feature depicted by the node. The prediction process proceeds in such a manner that every new instance begins at the root node and progresses to children nodes through branching until it reaches to the nodes at the last level called leaves where there is no further branching possible.

3.5 What Is an Ensemble Method?

Ensemble learning methods aim to get better results by aggregating multiple results of a common problem from many predictors. Otherwise stating, it obtains a “strong learner” by utilizing the predictions of many “weak learners” which could be any machine learning classification or regression algorithm. Reduction in invariance and hence decrease in overfitting is achieved using ensemble methods.

This is because multiple predictors are taken that are either similar or can be distinct and the average of their predictions is considered in the final results. Hence greater error of one predictor is dealt with by other better predictors.

In this study we use an ensemble of decision trees which is commonly known as random forest in machine learning terminology.

3.6 How Are Random Forests Trained?

Two main types of ensemble techniques are boosting and bagging. Both these methods sample subsets of data from original dataset and train machine learning models on the subsets. These trained models are called weak learners. The results of these predictors obtained for a new instance to be predicted are aggregated to form the prediction of the strong learner. Random forest classifier makes use of the bagging ensemble method. Subsets for training are sampled with replacement which means that instances from the original dataset can be repeated in many subsets. The training of individual decision trees and eventually combining their results is called "tree bagging". More randomness is achieved in decision trees when features for nodes of tree are selected in the bagging method. It means that instead of looking for the best features to form nodes in a decision tree, we randomly select features in all of the trees which diversify the algorithm by reducing variance and increasing bias. This deals with the assumed loss in efficiency due to absence of stronger features in higher nodes of the trees. This process leads to a stronger model and is called "feature bagging". Each new instance travels through the tree, starting from the root node choosing its path at each node based on the value it possesses for the specified attribute on that node until it ends at one of the leaf nodes. On the other hand, each new instance is processed by every decision tree in the ensemble or the random forest which was obtained by training on the sampled data from the original dataset. The classification output label obtained from the strong learner which can be termed as aggregation varies according to the classification problem. For classification this is done by taking the most frequent result called mode from all of the weak learners while the average of prediction results is taken in case of regression task [11–15].

3.7 Algorithm

1. "k" number of features are chosen in a random manner from a total of m features of the dataset such that $k \ll$ m.
2. A node "d" is selected by examining the optimal split point from the pool of "k" previously selected features.
3. The node is then split into children nodes according to the optimal split criterion.

4. The steps 1 to 3 are repeated until single node has been reached.
5. "n" number of trees are to be built to obtain a forest. This is done by repeating steps 1 to 4.

3.8 Construction of the Classification Tree

The two main components are the measure of class inaccuracy and the split patterns of members of variables of class. Tree formation aims to get the nodes whose entities form a single class eventually. Such a node is called pure node. We start with the function of class relative frequencies of impure nodes [12].

$$\eta(t) = \phi\big(v_1, v_2, \ldots\ldots v_j\big) \tag{1}$$

where $\mathrm{v_j}$ denotes the relative frequencies of j different classes at the considered node. ϕ is symmetric function of (v_1, v_2, …v_j).

Different inaccuracy measures are followed by different tree implementations. The focus is on reduction of inaccuracy that the node split attains.

Split performance s at node d is calculated as

$$\Delta\eta(s, t) = \eta(t) - \kappa(l)\eta(l) - \kappa(r) \tag{2}$$

where $\kappa(l)$ and $\kappa(r)$ are left side and right side proportion of the cases by the split.

The Gini Index

Gini index is an important technique for inaccuracy measurement. It can be obtained as follows for the case of two classes:

$$\eta(t) = v(0|t)v(1|t) = v(0|t)(1 - v(0|t)) = v(1|t)(1 - v(1|t)) \tag{3}$$

where the class label frequencies are denoted by v(0lt) and v(1lt), respectively.

Gini index can be generalized for more than two classes as follows:

$$\eta(t) = \Sigma v(j|t)(1 - v(j|t)) \tag{4}$$

where the relative frequency of class j at node t is represented as v(j| t).

3.9 XGBoost Classifier

XGBoost is a systematic and expandable application of the gradient boosting machine (GBM). It has multiple features such as high prediction accuracy and easy parallelism. The execution speed and model performance are the main reasons for using

XGBoost. This algorithm implements gradient boosting decision tree algorithm. The ensemble techniques in which new models are used to correct the errors and deficiencies of the currently used models is known as boosting. New models are added until no improvement in optimization parameters is noticed. An example of boosting algorithm is AdaBoost [12].

Also one of most important thing is the same code runs on every other environment, be it Hadoop or any other environment and it can solve problems using very large number of examples [14].

We are also very positive that XGBoost can very easily be implemented to make use of graphical processing units to speed up the computation which is going to be very important while developing a working prototype for the activity prediction [13].

Assuming the dataset to be D = $\{(x_i, y_i){:}i = 1 \ldots n, x_i \in \mathrm{R}^m, y_i \in \mathrm{R}\}$, it consists of n samples with m features. Let y_i denotes the value predicted by the model:

$$y_i = \sum_{k=1}^{k} f_k(x_i),\ f_k \epsilon F \tag{5}$$

where f_k denotes an individual regression tree and $f_k(x_i)$ gives the prediction evaluation given to the ith sample by the kth tree. Decreasing the objective function lets us learn the set of functions f_k:

$$Obj = \sum_{i=1}^{n} l(y_i, y_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{6}$$

The training loss function is represented by l, and it measures the difference between values predicted by the model y and the actual value y. The term Ω applies a penalty to minimize overfitting:

$$\Omega(f_k) = \gamma T + 1/2\left(\lambda||w||^2\right) \tag{7}$$

where the regularization degrees are represented by λ and γ. w and T are the scores on each leaf and the numbers of leaves, respectively. We can train the ensemble model with trees in an additive manner. Let the prediction of the ith instance at tth iteration be represented by $y_i^{(t-1)}$, and f_t needs to be added to minimize the following objective:

$$Obj^{(t)} = \sum_{i=1}^{n} l\left(y_i, y_i^{(t-1)}| + f_t(x_i)\right) + \Omega(f_t) \tag{8}$$

We simplify Eq. (8) by making use of the second-order Taylor expansion and by removing the constant term we obtain Eq. (9):

$$Obj^{(t)} = \sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2}h_i f_t(x_i)^2\right] + \Omega(f_t) \tag{9}$$

where the first-order gradient is $\mathrm{g_i}$ and the second-order gradient is $\mathrm{h_i}$. We again write the objective as:

$$Obj^{(t)} = \sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2}h_i f_t(x_i)^2\right] + \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{t} w_j^2$$

$$\sum_{j=1}^{T}\left[\left(\sum_{i\in I_j} g_i\right)w_j + \frac{1}{2}\left(\sum_{i\in I_j} h_i + \lambda\right)w_j^2\right] + \gamma T \tag{10}$$

where the instance set of leaf j is denoted by $I_j = \{i|\ q(x_i) = j\}$. The optimal weight of leaf j denoted by $w{*}j$ and the corresponding optimal value, for a fixed tree structure q, is calculated as:

$$w_j = \frac{-G_j}{H_j + \lambda},\ Obj = \frac{-1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j + \lambda} + \lambda T \tag{11}$$

$$G_j = \sum_{i\in I_j} g_i,\ H_j = \sum_{i\in I_j} h_i \tag{12}$$

By obtaining the maximum gain partition and the minimum target function, gain formula is obtained as:

$$G = \frac{1}{2}\left[\frac{\left(\sum_{i\in I_L} g_i\right)^2}{\sum_{i\in I_L} h_i + \lambda} + \frac{\left(\sum_{i\in I_R} g_i\right)^2}{\sum_{i\in I_R} h_i + \lambda} - \frac{\left(\sum_{i\in I} g_i\right)^2}{\sum_{i\in I} h_i + \lambda}\right] - \gamma \tag{13}$$

The split candidates are recognized using this formula.

Many simple trees are generated by XGBoost model, and it then uses them to score a particular leaf node for performing the splitting.

The left, right, and original leaf are represented by the first, second, and third term of the equation, respectively.

The regularization on the additional leaf is represented by γ, and to be used during the training.

4 Result Analysis

We have used random forest classifier and XGBoost classifier to perform the classification of activities on our dataset.

Table 2 Results obtained

Technique used	Results							
	Dataset1				Dataset2			
	Precision	Recall	F1Score	Accuracy	Precision	Recall	F1Score	Accuracy
Random forest	0.99	0.97	0.98	99.36	0.93	0.9	0.91	99.02
XGBoost	0.97	0.9	0.92	98.23	0.91	0.88	0.89	98.81

We evaluate the behavior of activity recognition model in this study based on the performance measure called F1-score which is calculated as F1-score = 2*P*R/(P + R) where the well-defined metrics of machine learning, namely precision (P) and recall (R) are used. As it can be observed, the harmonic mean of precision and recall is called the F1-score. In contrast to accuracy, F1-score provides a better view of the classifier performance, especially for datasets with imbalanced class distributions, because F-score is not biased toward the majority class [2].

We will present our final results using accuracy, precision, recall for completeness (Table 2).

We achieved an F1-score of 0.98 using random forest classifier and 0.92 using XGB classifier for Dataset1 from Room1. Similarly, the scores of 0.91 using RFC and 0.89 using XGB classifier were achieved for Dataset2 from Room2, which are clear upgrades from previous researches.

5 Conclusion

In our study, we used the dataset obtained from RFID sensors to try and predict the activities of the elderly. We found out that it is possible to develop a robust system to do the activity prediction. And if that is achieved, then we can go on developing smart mobile applications to take it to another level and find a solution for the fall problems of the elderly.

If this idea is to be extended, then along with activity prediction, a lot of other things like stress level, different hormone level and so on can be monitored and analyzed to help the elderly. There can be many future improvements to this study. First, we have only taken into consideration the motion-related parameters as previously provided data without incorporating other informative features. This can be improved by thinking of other details of the subject’s movements that can further enhance the purpose of activity recognition. Secondly, the activities related to ambulatory monitoring are limited in number in the context of monitoring older people in hospital and nursing homes. Therefore, further work using sparse data streams when such datasets become available will be needed in future to further generalize the current results.

In addition to the above-stated improvements, other options can be experimenting by positioning the sensors on distinct parts on the body of the subject and improvements to the deployment of the antenna of RFID reader.

Acknowledgments The authors would like to thank the faculty members of the Computer Science & Engineering Department of Delhi Technological University for the guidance provided during the course of the research. The authors would also like to thank the University of Adelaide, Australia and UCI for giving us access to such an amazing dataset, and they feel very grateful to have got the opportunity to explore different possibilities where the dataset can help us.

References

1. Krishnaswamy DD, Usha G (2014) Falls in older people. Department of Geriatric Medicine Madras Medical College and Government General Hospital
2. Shinmoto Torres RL, Visvanathan R, Abbott D, Hill KD, Ranasinghe DC (2017) A batteryless and wireless wearable sensor system for identifying bed and chair exits in a pilot trial in hospitalized older people. PLoS ONE 12(10):e0185670
3. Wickramasinghe A, Ranasinghe D (2015) Recognising Activities in Real Time Using Body Worn Passive Sensors With Sparse Data Streams: To Interpolate or Not To Interpolate?. https://doi.org/10.4108/eai.11-8-2015.151111
4. Wickramasinghe A, Ranasinghe DC, Fumeaux C, Hill KD, Visvanathan R (2017) Sequence Learning with Passive RFID Sensors for Real-Time Bed-Egress Recognition in Older People. IEEE Journal of Biomedical and Health Informatics 21(4):917–929
5. Barralon P, Vuillerme N, Noury N (2006) Walk Detection With a Kinematic Sensor: Frequency and Wavelet Comparison, 2006 International Conference of the IEEE Engineering in Medicine and Biology Society. New York, NY, pp 1711–1714
6. Lara OD, Labrador MA (2013) A Survey on Human Activity Recognition using Wearable Sensors. IEEE Commun Surv Tutorials 15(3):1192–1209
7. Wikipedia contributors, 'Radio-frequency identification', *Wikipedia, The Free Encyclopedia.* https://en.wikipedia.org/w/index.php?title=Radiofrequency_identification&oldid=929967836 (9 September 2019)
8. Activity recognition with healthy older people using a battery less wearable sensor Data Set. https://archive.ics.uci.edu/ml/datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor
9. Bonsor K, Fenlon W (2007) How RFID Works. https://electronics.howstuffworks.com/gadgets/high-tech-gadgets/rfid.htm
10. SuzSmiley: active-rfid-vs-passive-rfid. https://blog.atlasrfidstore.com/active-rfid-vs-passive-rfid
11. Brid, R. S.: Decision trees simple way to visualize a decision tree. https://medium.com/greyatom/decision-trees-a-simple-way-to-visualize-a-decision-dc506a403aeb
12. Polamuri, S.: How random forest works in machine learning. https://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learing/
13. Reinstein I (2017) KDnuggets, Random forests explained. https://www.kdnuggets.com/2017/10/random-forests-explained.html
14. Koehrsen W (2017) Random forest simple explanation. https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d
15. Liaw, Andy & Wiener, Matthew. Classification and Regression by RandomForest. Forest. 23. (2001)
16. Jaiswal, J. K., Samikannu, R., Application of Random Forest Algorithm on Feature Subset Selection and Classification and Regression. 978-1-5090-5573-9/16 IEEE (2016)

17. Chen, M., Liu, Q., Chen, S., Liu, Y., Zhang C. H., Ciu, R.: XGBoost-Based Algorithm Interpretation and Application on Post-Fault Transient Stability Status Prediction of Power System. (2019)
18. XGBoost documentation. https://xgboost.readthedocs.io/en/latest/#xgboost-documentation
19. Mitchell R, Frank E (2017) Accelerating the XGBoost algorithm using GPU computing. PeerJ Computer Science 3:e127

The Role of Predictive Data Analytics in Retailing

Mohammed Juned Shaikh Shabbir and C. M. Mankar

Abstract Large information analytics is a new practice in business analytics today. However, later modern overviews locate that huge information investigation may neglect to meet business wants as a result of the absence of business context and cluster arranged framework. In this paper, we present an objective situated large information investigation system for better business choices, which comprises a theoretical model which associates a business side and a major information side, setting data around the information; a case-based assessment technique which empowers to center the best arrangements; a procedure on the best way to utilize the proposed structure; and an associated device which is a constant enormous information examination stage. In this structure, issues against business objectives of the present procedure and answers for the future procedure are expressly estimated in the reasonable model and approved on genuine large information utilizing huge inquiries or enormous information examination. As an exact examination, a shipment choice process is utilized to indicate how the system can bolster better business choices as far as extensive understanding both on business and information investigation, high need and quick choices.

Keywords Big data analytics · Novel approach · Business process · Big data · Conceptual model

1 Introduction

Huge information investigation is an innovation which helps transform concealed bits of knowledge in large information into business esteem by utilizing progressed examination strategies so as to help better business choices, as the most blazing new practice in business analytics today. As indicated by certain studies, about 80% of CEOs or official groups see that enormous information investigation activities can possibly drive business worth, for example, making new income streams, improving

M. J. S. Shabbir (✉) · C. M. Mankar
Department of Computer Science and Engineering, S.S.G.M.C.E, Shegaon, Maharashtra, India
e-mail: juned44@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_16

operational effectiveness, or cutting expense, and over 80% of the member associations do continuous ventures. Be that as it may, as indicated by another study, 55% of enormous information ventures don't get finished, and others miss the mark concerning their business goals.

Furthermore, the nature of large information examination must be as acceptable, or as awful, as the nature of the huge information it employs. The quality characteristics of enormous information, together with connections between information, ought to subsequently be characterized in a major information model. With a decent quality large information model, huge information examination can precisely distinguish significant business concerns, pattern openings, and valuable business experiences, which, thus, can prompt great business choices. However, no rules are accessible for how to build up a top-notch huge information model in a deliberate and objective manner.

At last, with the advances in information technology (IT), utilization of programming frameworks progressively has gotten predominant and basic more productively and successfully in completing the different undertakings and exercises in the quickly changing business space. In any case, adjusting a product framework well within its expected business process has been trying, because of the trouble in right off-the-bat understanding and exactly displaying a business procedure and also concocting a prerequisite detail of a framework for supporting the business procedure. Utilization of understood documentation helps business process model and notation (BPMN) for exact displaying of a business procedure, and UML use case for demonstrating programming prerequisites, yet the subsequent trouble still remains. There are a few issues: (1) since formal meanings of the models do not exist, various elucidations are conceivable which can prompt changes that go amiss from the first importance; (2) a business movement can be performed either by individuals or framework usefulness; (3) the granularity of use case is not really equivalent to that of a business action or undertaking; and (4) neither BPMN nor UML use cases consider non-functional requirements (NFRs).

2 Related Works

The examination [1] gives bits of knowledge into how standard large information investigative ability and insignificant ownership of huge information are bound to make conditions for business disappointment. The examination expands the current floods of research by revealing insight into choices and procedures in encouraging or hampering firms' capacity to saddle huge information to alleviate the reason for business disappointments. The examination prompted the categorization of various productive roads to look into information-driven ways to deal with business disappointment.

In this paper [2], the authors investigate how information and examination have been utilized so far in the advanced education part for upgraded learning or to help choices, and what openings and difficulties encompass BDA in this segment.

The problem [3] frames approach catches issues as observed by designers, however not by partners. This paper shows a system that expands the issue graph of the problem frames way to deal with in speaking to partner issues utilizing "delicate issue", a thought alluding to an unwanted circumstance that contrarily influences partner objectives and may have less obvious goals criteria. A delicate issue might be refined to increasingly explicit sub-issues and underlying drivers, which are then followed to relating arrangements in an outline called problem interdependency graph. The structure has been applied to the 1992 London emergency vehicle contextual analysis, which shows that the system serves to all the more exactly speak to the partner issues endured by the current framework ("as-seems to be") as one with the answers for being given by the new framework ("to-be") to such issues. The examination additionally shows that the system decides if the issues have been adequately tended to and, if not, why [4] business process model and notation.

Spark [5] has been set up as an alluring stage for large information investigation since it figures out how to conceal the vast majority of the complexities identified with parallelism, adaptation to non-critical failure, and bunch-setting from designers. Nonetheless, this comes to the detriment of having more than 150 configurable parameters, the effect of which cannot be comprehensively inspected because of the exponential measure of their mixes. The default esteems enable engineers to rapidly send their applications, yet leave the inquiry regarding whether execution can be improved open. In this work, we explore the effect of the most significant tunable Spark parameters with respect to rearranging, pressure, and serialization on the application execution through broad experimentation utilizing the Spark-empowered Marenostrum III (MN3) processing foundation of the Barcelona Supercomputing Center. The all-encompassing point is to manage designers on the most proficient method to continue to changes to the default esteems. We expand upon our past work, where we mapped our experience to an experimentation iterative improvement philosophy for tuning parameters in subjective applications dependent on proof from few test runs. The primary commitment of this work is that we propose an optional efficient system for parameter tuning, which can be effectively applied onto any processing framework and it appeared to yield tantamount if worse outcomes than the underlying one when applied to MN3; watched speedups in our approving experiment contemplates start from 20%. Furthermore, the new system can depend on runs utilizing tests rather than runs on the total datasets, which render it altogether progressively down to earth.

Business knowledge (BI) [6] offers colossal potential for business associations to pick up experiences into their everyday tasks, just as long-term openings and dangers. In any case, a large portion of the present BI instruments depends on models that are an excessive amount of information arranged from the perspective of business leaders. We propose a venture demonstrating the way to deal with connecting the business-level comprehension of the undertaking with its portrayals in databases and information distribution centers. The business insight model (BIM) offers ideas natural to business basic leadership—for example, objectives, procedures, forms, circumstances, impacts, and markers. Not at all like numerous endeavor models which are intended to be utilized to infer, oversee, or line up with IT framework usage, BIM means to help business clients compose and understand the tremendous measures of information about the venture and its outer condition. In this paper, we present center BIM ideas, concentrating particularly on thinking about circumstances, impacts, and markers. Such thinking supports the key investigation of business destinations considering current endeavor information, enabling investigators to investigate situations and discover elective methodologies. We depict how objective thinking systems from calculated displaying and necessities building have been applied to BIM. Strategies are likewise furnished to help to prevail upon pointers connected to business measurements, including situations where determinations of markers are fragmented. Assessment of the proposed displaying and thinking structure remembers a forgoing model execution, just as contextual investigations.

The genuine worth [7] of big data lies in its shrouded bits of knowledge, yet the present focal point of the big data people group is on the advances for mining experiences from monstrous information, as opposed to the information itself. The greatest test confronting ventures isn't the way to distinguish the correct information; however, rather, it is the way to utilize bits of knowledge acquired from big data to improve the business. To address this test, we propose GOMA, an objective arranged demonstrating way to deal with big data examination. Controlled by big data bits of knowledge, GOMA utilizes an objective arranged way to deal with catch business objectives, reason about business circumstances, and guide basic leadership forms. GOMA gives an efficient way to deal with coordinating two sorts.

Increasingly more data [8] about business forms are recorded by data frameworks as supposed "occasion logs". In spite of the ubiquity of such information, most associations analyze issues dependent on fiction instead of certainties. Procedure mining is a rising order dependent on process model-driven methodologies and information mining. It does not just enable associations to completely profit by the data put away in their frameworks, however, it can likewise be utilized to check the conformance of procedures, distinguish bottlenecks, and foresee execution issues. Will van der Aalst conveys the primary book on process mining. It plans to act naturally contained while covering the whole procedure mining range from process disclosure to operational help. In Part I, the writer gives the nuts and bolts of business process displaying and information mining important to comprehend the rest of the book. Part II centers on process revelation as the most significant procedure mining task. Part III moves past finding the control stream of procedures and features conformance checking, and authoritative and time points of view. Part IV manages the peruser in effectively applying procedure mining by including a prologue to the broadly utilized open-source instrument ProM. At long last, Part V makes a stride back, thinking about the material exhibited and the key open difficulties. In general, this book gives a far-reaching review of the cutting edge in process mining. It is expected for business process investigators, business specialists, process chiefs, graduate understudies, and BPM scientists.

Retailers frequently [9] concentrate on improving customer loyalty through customer-orientated sales strategies, and research by and large reflects that customer satisfaction is the precursor to customer loyalty. The profitability is enhanced due to increased loyalty, which in turn is impacted by increased satisfaction.

Analysis of consumer's [10] purchase pattern from each and every transaction in a retail store is used for developing a strategy for placement and promotion of products to improve customer satisfaction and sales revenue for the retailers.

The short-term [11] goals in the big data analytics market should aim at building technology foundation and developing a customer base for sustainable revenue generation. The medium-term goals should aim at supporting delivery schedules and strengthening customer bond with high-quality output.

Retailers have to [12] identify the moments when the customer is most receptive to their influence; it could be his home or when he is commuting for work or on the way to a mall or browsing through window displays at a store. Companies have to focus on relating to their customers in situations where their communication can be most relevant [13].

3 Proposed Work

To address the problems, a novel big data analytics framework using Spark as a novel approach for better business decisions is proposed, that is, supporting a comprehensive understanding of business and data, high priority, and fastness. This framework provides an applied model for enormous information investigation which comes to an obvious conclusion regarding significant ideas of the business side, for example, business objectives, issues, arrangements, business forms, and those of a major information side, for example, huge examination and huge questions to cross over any barrier between the two. It helps not just in far-reaching understandings of present and future business, yet additionally interchanges between partners by helping analyzers unequivocally model the ideas. Additionally, in the soul of objective direction, options in issues with the as-is business procedure and answers for the to-be business procedure can be estimated in the calculated model and approved by dissecting large information, and huge ones are chosen after exchange of examination through our proof-based assessment strategy. This procedure assists analyzers with concentrating on the best arrangements inside the given time and spending plan. To help our ideas, we executed novel approach which coordinates eclipse modeling framework (EMF) for the reasonable model and Apache Spark for a major information examination which empowers continuous preparing in a circulated and grouped registering stage, prompting quick business choices.

System Diagram (Fig. 1):

4 Result Analysis

A. *Comparison Graph* (Fig. 2):

B. *Comparison Table* (Table 1):

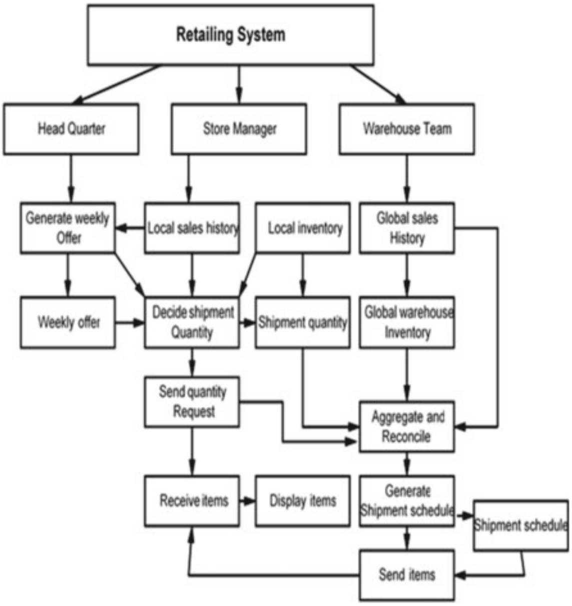


Fig. 1 System architecture

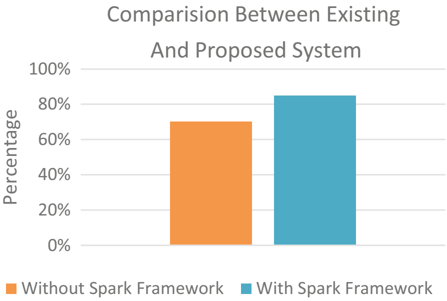


Fig. 2 Graph

Table 1 Comparative result

S. no.	Without spark framework (%)	With spark framework (%)
1	65	87

5 Conclusion

We have proposed an objective situated huge information examination system. All the more explicitly, the proposed system incorporates theoretical model which interfaces business and large information, a proof-based assessment strategy for choosing the best arrangements, a procedure for discovering business issues and arrangements—right off-the-bat by conjecturing them, and furthermore by approving them utilizing huge inquiries or huge investigation—and a supporting instrument which is actualized over Spark, constant huge information examination structure. Despite the fact that there are a few constraints, in any event through an exact investigation, the proposed system can help with huge information business examination in a worth included way, that is, exhaustive comprehension on business and investigation, high need, and quick choices.

References

1. Amankwah-Amoah J, Adomako S (2018) Big data analytics and business failures in data-Rich environments: an organizing framework. Science Direct
2. Al Hadwer A, Gillis D, Rezania D (2019) Big data analytics for higher education in the cloud era. In: 2019 the 4th IEEE international conference on big data analytics
3. Supakkul S, Chung L (2009) Extending problem frames to deal with stakeholder problems: an agent- and goal-oriented approach. In: Proceedings of ACM symposium on applied computing, pp 389–394
4. Business Process Model and Notation
5. Davidson A, Or A (2013) Optimizing shuffle performance in spark. Technical report, University of California, Berkeley-Department of Electrical Engineering and Computer Sciences
6. Horkoff J, Barone D, Jiang L, Yu E, Amyot D, Borgida A, Mylopoulos J (2014) Strategic business modelling: representation and reasoning. Softw Syst Model 1015–1041
7. Supakkul S, Zhao L, Chung L (2016) GOMA: supporting big data analytics with a goal-oriented approach. In: IEEE big data congress, pp 149–156
8. van der Aalst WMP (2011) Process mining: discovery, conformance and enhancement of business processes. Springer Science & Business Media, Heidelberg
9. Mittal V, Kamakura WA (2001) Satisfaction, repurchases intent, and repurchases behavior: investigating the moderating effect of customer characteristics. J Mark Res 38(1):131–142
10. Verma N, Malhotra D, Malhotra M, Singh J (2015) E-commerce website ranking using semantic web mining and neural computing. Proc Comput Sci (Science Direct, Elsevier) 45:42–51
11. Karthiban MK, Raj JS (2019) Big data analytics for developing secure internet of everything. J ISMAC 1(02):129–136
12. Palem G (2014) Formulating an executive strategy for big data analytics. Technol Innov Manag Rev 25–34
13. Gagnon JL, Julian Chu J (2005) Retail in 2010: a world of extremes. Strat Leadersh 33(5):13–23

High-Performance Digital Logic Circuit Realization Using Differential Cascode Voltage Switch Logic (DCVSL)

S. S. Kavitha and Narasimha Kaulgud

Abstract Most of dual-rail CMOS circuits are loosely based around differential cascade voltage logic switch. DCVSL provides dual-rail logic gates that have latching characteristics built into circuits itself. In CVSL logic output results are held until inputs induce a change, so that there is no loss of data, thereby saving energy and power. In today's digital application low power has become a key factor in high-speed computations. The proposed work gives an insight into the working of CVSL and the proposed method, showing a reduced number of gates, and thereby reducing area and power constraints. In this paper, exclusive detailed use of pass gate logic structure to put back the nMOS logic structure in conventional DCVSL circuit along with the implementation of adders are provided. The proposed circuit designs and results are compared and implemented using a cadence software tool and for quantum circuits QCAD tools are used. The study shows the optimization in case of power, area and speed achieved in comparison with conventional circuits.

1 Introduction

In terms of logic flexibility, power dissipation, circuit delay and layout area, the conventional CVSL is said to have more benefits over the existing static CMOS NAND/NOR design. For the dynamic implementation, DCVS shows superior performance over standard domino logic circuits. Floating node problem is the major issue in DCVS circuits. This process creates additional delay and current spikes similar to ratioed logic circuits, which results in false logic evaluation, and this makes dynamic circuits unreliable [1]. NMOS pass-gate network is used to eliminate this problem. Advantages of CVSL are:

S. S. Kavitha (✉) · N. Kaulgud
The National Institute of Engineering, Mysuru, India
e-mail: kavithaece@nie.ac.in

N. Kaulgud
e-mail: narasimhakaulgud@nie.ac.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_17

1. Low-load capacitance on inputs.
2. No static power consumption.
3. Automatic complementary functions.
4. A circuit is self-latching.

Disadvantages are they require complementary inputs and more transistors for a single function. However, they are much needed in some circuit blocks where complementary signals are generally needed, mostly in implementation of adders. Logic devises grip that is capable of handling complex Boolean logic expression within a circuit delay is accomplished in CVSL by cascading differential pairs of MOS devices into powerful combinational logic tree networks [2]. 2N-1 input variables are processed with N high-logic trees consisting of differential pairs of cascading NMOS logic devices. CVSL is more advantageous and provides the performance of up to four times compared to CMOS or NMOS primitive NAND or NOR logic families, without losing the expected low-power characteristics observed in CMOS circuitry. Theoretically and practically, CVSL is double denser similar to traditional NAND or NOR logic and is more flexible with available designs in Cadence tools. With cascaded improved performance NMOS devices with the absence of stacked PMOS devices, which are used separately as pull up devices in load and buffer circuits, these combinational NMOS logic structures can be built. The criticality of PMOS to the NMOS spacing and to optimize the PMOS devices is, therefore, be reduced, suppressing the device or process complexity load on CVSL designs.

Quantum circuit is a model of quantum computation. This is built by an interconnecting sequence of quantum gates. These circuits are acyclic and cannot copy qubits. A logic which erases a bit of information is known as irreversible logic. For erasing each bit the energy dissipation will be more. Gates like AND, OR, XOR and so on used in digital design are irreversible. Energy loss should be considered an important parameter in the design of digital circuits. The alternate method of reducing power dissipation is reversible logic. Reversible gates, like TOF-FOLI, FEYNMAN and so on, have an equal number of input logics and output logics. The most commonly used gates in reversible circuits are CCNOT, CNOT and NOT gates. Quantum computer normally differs from a classical computer. The computation in quantum technology uses quantum bits (qubits) which satisfy superposition, entanglement and tunneling process to perform reversible logic. In this paper, QCAD tool is used to realize such quantum gates, and half adder is designed based on it [2].

2 Cascode Voltage Switch Logic

The Cascode voltage logic swing uses two cross-connected branches where true network *f* and its complementary network fbar are present in the cross-connected output nodes. In order to construct a complementary network, all series connections in *f* are changed to parallel connection and all parallel connections to series connection and thus complementing all input signals. Therefore both true and complementary

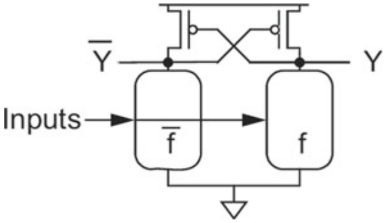


Fig. 1 CVSL logic

inputs are required in CVSL type of logic to obtain true and complementary outputs [3]. The NMOS network comprising two complementary NMOS switch networks is switched to a pair of cross-connected PMOS transistors. The block diagram of the basic Cascode voltage swing is as shown in Fig. 1.

Figure 1 consists of PMOS pull-up switches which are cross-connected to form the latching characteristics. The duality structure of each other is formed in the N pull-down tree structure. In this PMOS network "f" provides the logic same as in static CMOS logic gate whereas "fbar" makes use of inverted inputs, thus resulting in complementary output. The right side of the N pull-down network is provided with true input, while complimentary input is provided to left side of the N pull-down network. The basic working procedure is explained as follows:

1. One branch of the NMOS networks will be switched ON and the other branch will be switched OFF, for any given input pattern.
2. The output is pulled low by pull-down NMOS network that is ON. Opposite PMOS transistor will be turned high by this low output resulting in opposite high output.
3. This makes the opposite output to rise and the other pMOS transistor turns OFF, so there is no static power dissipation observed.

3 Existing Work

Conventional CMOS logic consists of *N* number of switches to realize the logic [4]. Defined conventional half-adder design is as shown in Fig. 2.

DCVS has lesser number of transistor counts compared to NAND/NOR implementations. DCVS provides better performance of about 4× as estimated to the CMOS or NMOS, NAND or NOR circuits and thus sustaining the less power dissipation circuitry [3]. The existing schematic diagram of the half-adder is as shown in Fig. 3.

The logic functions are evaluated by using NMOS logic tree. Signals indicate with "-" are complementary signals. Compared to conventional CMOS circuitry, these DCVS logic has less number of transistor counts and hence results in high-speed operation and low-input capacitance. The main advantage of these circuits is they consume less power due to less supply voltage level. This circuit implementation shows the separate circuitry for logic evaluation of adder sum and carry.

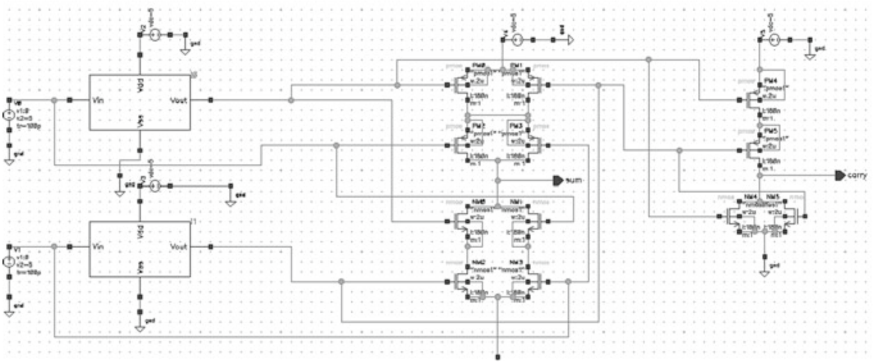


Fig. 2 Conventional CMOS logic

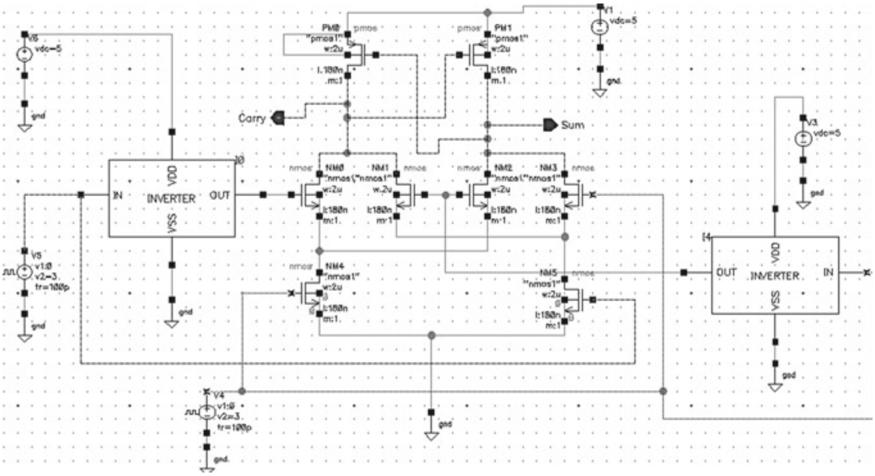


Fig. 3 Existing CVSL logic tree

4 Proposed CVSL/DCVS Design

Logic functions required for the computation of sum and carry of half-adder are evaluated just by using four NMOS transistors. Thereby one can justify that the gate count required to realize adder has been reduced and therefore increasing circuit performances [5]. Compared to conventional and DCVS circuitry, these new design provides less power dissipation and high speed of operation. The only difference observed from DCVS circuit is that in the logic variable connection. The next exquisite characteristic is that, without any inverter logic, DCVS circuits can produce complementary outputs. This makes the NAND and AND.

To be the same circuit but with the output, nodes are interchanged. Hence the DCVS structure is said to be the best-suited design for standard cell library development.

Along with CMOS implementation, QCA implementation is also proposed to crosscheck the performance evaluation with two technologies. Circuit implemented for adder circuit in QCA has different dimensions parameters as compared to VLSI tools. However, it is known that quantum circuits dissipate less energy and provide faster computation. Based on these circuit parameters results are evaluated and summarized below. Detail explanation on the working principle and results obtained are described below in subsections.

4.1 Circuit Working Principle

The leftmost part of NMOS logic evaluates the carry of half-adder whereas the rightmost part is used to evaluate the sum of half-adder. Function variables are provided to the source and gate terminals and logic are obtained similar to pass transistor working principle. Because of single tree structure evaluation, the problem relating to the floating node and charge leakage and charge sharing is completely eliminated. During both the inputs are "0" the rightmost NMOS M1 and M3 are ON, M2 and M4 are OFF and thus provides the output sum and carry as 0. Similarly, during $A = 0$ and $B = 1$, M1 and M3 are ON, and the switches being the source logic move toward the drain and thus resulting in sum = 1 and carry = 0. When $A = 1$ and $B = 0$, M2 and M3 are ON, producing sum = 1 and carry = 0. Finally, when both the inputs are "1", M2 and M4 are ON, producing the output as sum = 0, carry = 1. Therefore, all the logic states presented works similar to half-adder working principle 4 (Fig. 4).

4.2 Result and Analysis

According to a particular sequence, to compare the performance and power dissipation with the competing techniques such as CMOS [6], CVSL [7] and DCVS [8], the SUM circuits are built as the bench test. For differential cascade voltage swing logic circuit, the XOR circuit is used as SUM circuit and carry is AND circuit implemented using minimum NMOS logic switches.

The below result shows the working of adder circuit using the CVSL logic. The small glitches that were observed show the behavior of NMOS switch. Since NMOS is used as a pass transistor in evaluating logic, as it is known that NMOS suffers to pass complete logic 1, small switching characteristics are disturbed. From the result shown below, it is clear that the operation of the adder circuit happens to be at most evaluating all the logic levels, as shown in Fig. 5.

It is straightforward that DCVS has a lesser count of transistors and the full logic swing CMOS signals. For the low capacitive load which is less than 0.3 pF, DCVS

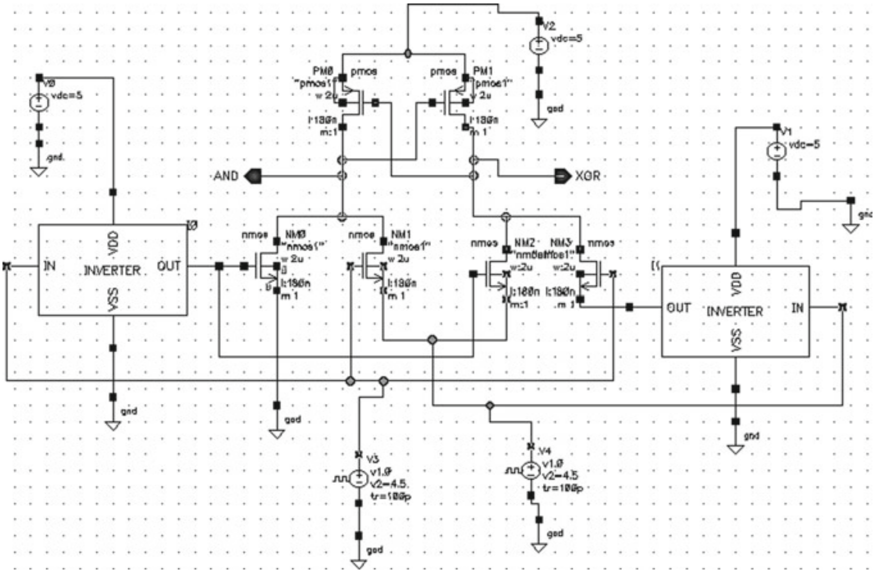


Fig. 4 Proposed CVSL logic

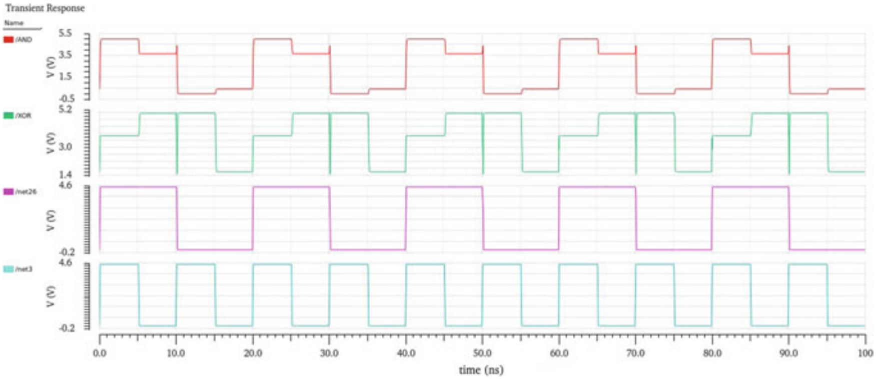


Fig. 5 CVSL results

has the best performance. The only serious fault star is that DCVS alone is di cult to carry complicated logic circuits because of the weak pull-up transistors [5]. This will deprave the superiority of DCVS quality at high capacitive loads. However, the control variable or functional variable can be used as input signals to pass gate. By this long chain of logic circuits can be terminated at the gate level, if a proper control variable is chosen.

4.3 Quantum Circuit and Results

Quantum computing is a vast area of research relating to nano-technological devices. Beyond VLSI, one can think of next generation as quantum's era. Quantum computing mainly focuses on reversible computing where there is one-to-one correspondence between the input variable and the output variable. Four quantum dots.

In a QCA cells are arranged in the form of a square pattern comprising electrons sites that can occupy adjacent places by tunneling through them (Fig. 6).

Polarization P = –1 represents logic "0" and P =+1 represents logic "1". Placing quantum dots in series to the sides of each other forms the wire. The interconnect wire consisting of quantum dot cells is formed to transmit polarization state. Majority gate and NOT gate are considered as the two most fundamental building blocks of QCA [9]. Majority of gate consists of three inputs and one output. It is a five-cell structure. Among three inputs, two inputs are based on logic and one input is a fixed polarization which detects the type of gate. If the fixed polarization is "–1" it works as AND gate and if it is "+1" it works as OR gate. Figure 7 shows the representation of AND and OR gates.

There are four stages of clocks for state transition. The first stage boosts the tunneling barriers to rise. In the second stage, the tunneling barriers are large enough to avoid electrons from tunneling. In the third stage, the high barriers start to lower. And in the last stage, the tunneling barriers permit the electrons to freely pass through them again. In short, it means that electrons will flow freely when the clock signal goes high and when the clock signal logic goes low, the cell is said to be latched.

Due to some of the limitations such as increasing switching speed and decreasing power consumption and increasing complexity observed in CMOS technology, QCA has become an alternative solution to all such limitations in the years to come [10].

Fig. 6 QCA cell

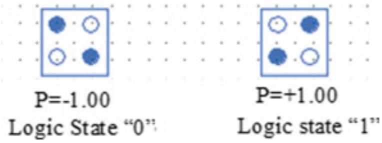
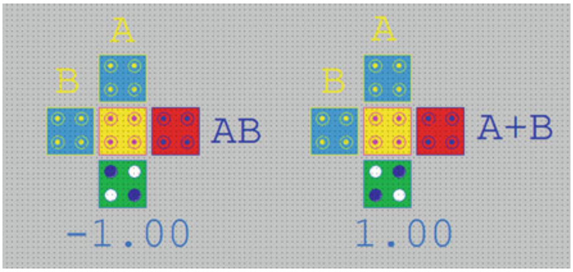


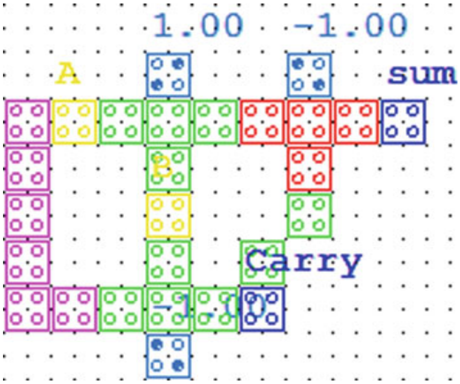
Fig. 7 QCA AND and OR gate



This paper presents the half-adder design using QCAD tool. The comparative analysis is made with DCVS CAD and DCVS QCAD tool simulation results (Fig. 8).

The results shown below give the clear picture of working of half-adder circuit as evaluated for SUM and CARRY. SUM is the evaluation of XOR gate and CARRY is the evaluation of AND gate. Output results are obtained after 0.25 delay in a clock cycle since the difference between each clock in QCA is 0.25 (Fig. 9).

Fig. 8 Quantum half-adder logic



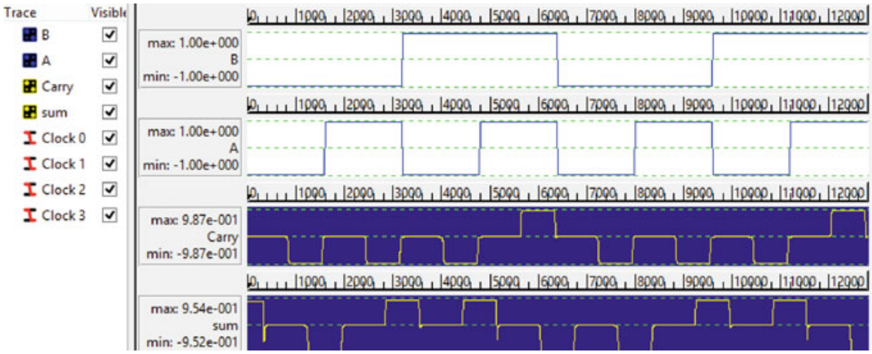


Fig. 9 Quantum half-adder results

4.4 Results and Comparison

The results obtained from CAD tool and QCAD tool are summarized below:

Parameters	V (V)	I (mA)	Power	Area	Delay (ns)	Transistor/cell count
Conventional CMOS	5.255	4.599	24.16 mW	32 μm	110	16 transistors
Existing DCVS design	5.224	3.225	17.84 mW	24 μm	67.22	12 transistors
Proposed DCVS design	5.000	3.168	16.80 mW	20 μm	55.34	10 transistors
Quantum half adder design	–	–	12.11 meV	0.03 μm^2	0.5	27 cells

From the above tabulation, it is clear that the proposed DCVS design has better performance than the conventional one. Summary of power, delay and area are reported. On glancing with power report the proposed design has 33.33% improvement over conventional CMOS, but a very small difference with the existing method. Similarly, area and delay have drastic changes in comparison. Due to the decrease in delay, the circuits are said to achieve high computation speed. However, there exists a trade between the circuit parameters. Lastly, the transistor count can be compared with the existing and proposed design.

5 Conclusion

In this paper, a new circuit design technique for the adder circuit is discussed. Due to the absence of NMOS logic tree floating node problem is eliminated, thus providing better performance. Proposed DCVS is actually the best of all at the light capacitive load range as compared to that of the existing design. For the more vigorous design, however, one can choose DCVS as the best design choice. A CAD tool simulation shows the feasibility of using DCVS to attain a superior performance design. DCVS also exhibits the best power-delay product, with addition to the dynamic circuit design. QCAD shows the minimum power utilization and less delay and area, hence used to build any logic circuits in the near future.

References

1. Hatano H (2013) SET immune spaceborne CVSL and C 2 VSL circuits. J Electr Control Eng 3:43–48
2. Sharma M, Mehra R (2016) Design analysis of full adder using cascade voltage switch logic. IOSR J VLSI Signal Process 6:18–23

3. Gupta K, Bagga S, Pandey N (2016) Efficient CVSL based full adder realizations. In: 2016 IEEE 1st international conference on power electronics, intelligent control and energy systems (ICPEICES). IEEE, pp 1–5
4. Kaur G, Kumar A, Singh J (2014) Design of high-speed full adder using improved differential split logic technique for 130 nm technology and its implementation in making ALU. Int J Comput Appl 96(18)
5. Lai F-S, Hwang W (1997) Design and implementation of differential cascode voltage switch with pass-gate (DCVSPG) logic for high-performance digital systems. EEE J Solid-State Circuits 32(4):563–573
6. Prashant Kumar et al. Design of low power and area efficient half adder using pass transistor and comparison of various performance parameters. In: 2017 International Conference on Computing, Communication and Automation (ICCCA). IEEE. 2017, pp. 1477–1482
7. Hatano H (2010) A single event effect analysis on static CVSL exclusive-OR circuits. IEICE Trans Electron 93(9):1471–1473
8. Bastami M, Mirzaee RF (2017) Integration of CTL, PTL, and DCVSL for designing a novel fast ternary half adder, pp 1477–1482
9. Cho H, Swartzlander EE (2007) Adder designs and analyses for quantum-dot cellular automata. IEEE Trans Nanotechnol 6(3):374–383
10. Kavitha SS, Kaulgud N (2017) Quantum dot cellular automata (QCA) design for the realization of basic logic gates. In: 2017 international conference on electrical, electronics, communication, computer, and optimization techniques (ICEECCOT). IEEE, pp 314–317

Analysis and Classification of Ripped Tobacco Leaves Using Machine Learning Techniques

M. T. Thirthe Gowda and J. Chandrika

Abstract Tobacco crop is one of the major crops in the world and plays a vital role in the international market. After cultivation of tobacco plants, classification of ripped tobacco leaves is one of the challenging tasks. Generally in India, the classification of ripped leaves is done by a manual process only. Machine learning-based classification model is introduced to nullify the human intervention in the classification of tobacco ripped leaves. The proposed model is designed according to classification of three important features like color, texture, and shape. For the experimental purpose the images of the leaves are captured using mobile sensor and the dataset was created. The dataset consists of 2040 tobacco leaf images; from the image dataset 1378 images are used for training and 622 images are used for testing. The proposed model is validated by many machine learning algorithms, where the classification model achieved an efficiency of 94.2% on validation accuracy and 86% in testing accuracy. The proposed model has significant high performance on the classification of ripped tobacco leaves.

Keywords Classification · Features · Tobacco leaf · Grading · FCV

1 Introduction

Tobacco crop is one of the commercial crops in India and it plays a significant role in the Indian economy and employment. Throughout the world, countries such as China, Brazil, USA, and India are the major producers of tobacco [1]. India is the third highest contributory in tobacco production. Around 0.47 million hectares of

M. T. Thirthe Gowda (✉)
Department of Computer Science & Engineering, Government Engineering College, Hassan, Karnataka, India
e-mail: thirthegowda.gech@gmail.com

J. Chandrika
Department of Computer Science & Engineering, Malnad College of Engineering, Hassan, Karnataka, India
e-mail: chandrikaramesh@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_18

land over 15 states in India grow tobacco, and 800 million kg of tobacco is produced every year and around 75,000 farmers are cultivating this crop. Many varieties of tobacco leaves, like flue-cured Virginia (FCV), beedi, burley, hookah, chewing and so on, are grown in India, but some of the varieties are exported and most of them are used for local consumption. The FCV and burley tobacco are major exportable tobacco types [1], where FCV-type crop is cultivated in southern states of India, that is, in Karnataka and Andhra Pradesh regions, out of which 60% of the tobacco crop is produced from Mysuru and Hassan district of Karnataka. In general, 100% of FCV tobacco crop products is purchased by well-known multinational cigarette manufacturing companies.

At the time of harvesting, it is difficult for labors to separate tobacco leaves for curing process and it is more time-consuming, thus requiring more numbers of labors. Manual classification is a very tedious job and accuracy of the manual classification is very less; it also affects the quality in tobacco leaf grading. To overcome the problem, an automated classification system for uncured tobacco leaves is proposed. The proposed system helps in improving classification accuracy, increases speed, and helps to get a good quality of tobacco leaves. Many more models have been proposed for harvesting and leaf classification.

2 Related Works

In recent years, tobacco leaf classification is emerging as the most significant concept from the computer vision paradigm [1], and a survey on various image-based classification techniques is proposed for grading various flue-affected tobacco leaves. Hence, this makes the research work in [1] to be more closely related to the proposed model.

In [2], Zhang et al. (1997) have proposed a novel method to classify the grade of the tobacco leaves on the basis of the features like their color, texture, and shape. With the help of various feature extraction mechanisms initiated by chroma and hue with standard deviation and mean values, the authors have extracted varied color feature for tobacco leaves and in (2011) also they proposed fuzzy comprehensive method for evaluation [3]. Yawootti et al. [4] have proposed a new approach for classification of Thai flue-cured tobacco leaves using color histogram technique and achieved 91% of accuracy on color features. Liqun Han [5] proposed a method by recognize a part of the growth of flue-cured tobacco leaves for classification using support vector machine. Image processing and fuzzy statistics have been used in the proposed method for classification.

To analyze the various colors of tobacco leaves [6], a transformation technique is proposed that transfers traditional RGB to the Munsell system. Furthermore, the standard deviation and the average value of hue obtained from FCV leaves cluster are used to obtain the extended color features to grade the tobacco leaves in a better way.

Followed by that, Guru et al. have proposed a new representation called min–max representation to filter the cured tobacco leaves. This newly proposed data representation [7] has been applied on the selected 887 images from where they are categorized into 12 different categories with an accuracy of 91%.

In [8], the proposed technique has achieved an accuracy of 86.9% with a novel machine vision-based tobacco leaves classification which helps mainly in the automated harvesting process. Here, a CIELAB model is preferred for analyzing the color, leaf texture, and their segmentation to unprecedently increase the classification accuracy.

Consequently, Guru et al. have proposed a classification for differentiating the flue-cure tobacco leaves by using the deep CNN model. CNN is used to train only three image classes which contain a total of 120 samples with a classification accuracy of about 85.1%.

Simonyan et al. proposed very deep CNN for large-scale image recognition. In this work, ConvNet model is developed to overcome the findings of ImageNet. It has 16–19 layers; localization and classification can be achieved; and reduction of top 1 error rate to 28.1% and top 5 error rate to 9.4% [9].

A generalized classification method using deep CNN has been proposed by Zhipeng et al. to classify the flue-cured tobacco leaves.

The CNN is used with 120 image samples for training only for three classes, and it achieved 85.10% of validation accuracy [10].

Victor et al. (2012 and 2018) proposed ImageNet classification with deep convolutional neural networks. In this method, around 1.2 million of static images are used for training; 150,000 images are used for testing, 50,000 images are used for validation with 1000 different classes; and 37.5% top 1 error rate and 17% top 5 error rate were achieved [11].

Szegedy et al. proposed going deeper with convolutions. In the proposed method the utilization of computational resource inside the network improved [12].

3 Proposed Method

The main objective of the proposed system is to develop a model for automatic tobacco leaf classification, and its generic architecture is shown in Fig. 1.

The proposed architecture has two modules, like image collection unit and the processing unit, where image collection unit has many modules to get ripped images of tobacco leaves, like datasets, preprocessing, segmentation, and feature extraction.

3.1 Acquire Sample Leaves

A total of 2040 FCV green (uncured) tobacco leaves were photographed using mobile camera with 4160 × 3120-pixel resolution and saved as JPEG file format. The snap of

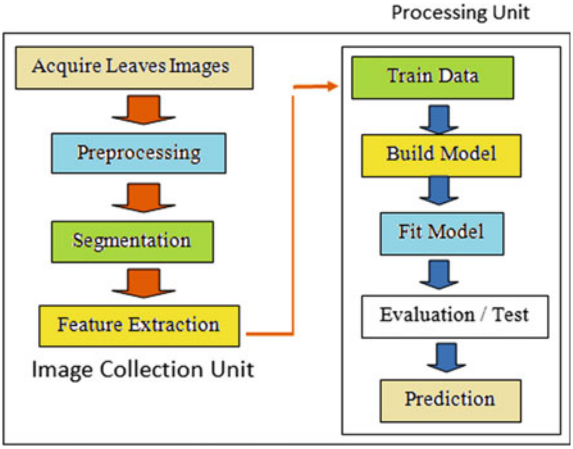


Fig. 1 Generic architecture of the proposed model

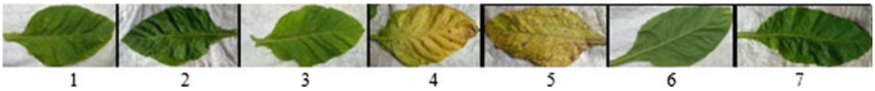


Fig. 2 Tobacco leaves samples, where sample No. 1 and 3 are ripe 1 leaves, No. 2 is ripe 2 leaf, No. 4 and 5 are overripe leaves, No. 6 and 7 are unripe leaves

these leaves was taken at the place of harvesting station, very close to barren, where we restore the tobacco leaves situated at Ramanathapura village, Hassan district, Karnataka state, India. Here we classify the uncured tobacco leaf images into four categories, like ripe 1, ripe 2, unripe, and overripe. Figure 2 shows a sample of tobacco leaf images belonging to four grades, where the grading of each leaf is defined by the grading expert.

3.2 Image Acquisition

The dataset consists of RGB images of size 4160 × 3120 pixels with high resolution, while grading it leads to an unexpected result. So in this regard, we apply a unified approach before attempting to detect leaves images. Hence we scaled down all images into 300 × 452 dimension, where the ratio of new image to the old image was first calculated before performing the actual resizing, so the aspect ratio of the original image will not be altered.

3.3 Image Segmentation

Image segmentation is the process of separating the object from the rest of the image [12]. In this section all images of tobacco leaves are detected, where the bounding box enclosing the ROI of first extracted or cropped leaf image.

3.4 Classification

In this study, TensorFlow was used to build a CNN-based tobacco leaves classifier. TensorFlow is an open-source deep learning framework developed by Google, allowing a user to quickly and efficiently implement various algorithms fundamental to neural networks. Given the wide range of functions already made available, as well as the community support, TensorFlow was chosen over other well-known frameworks at this time.

3.5 Performance Evaluation

The correct classification and misclassification in tobacco leaf detection and grading were summarized in a confusion matrix. The confusion matrix is an n × n matrix in which each column represents the number of instances in an actual class and each row represents the instances in a predicted class.

Accuracy is an evaluation metric, and it is derived from the confusion matrix and the accuracy rate is computed by Eq. (1). It contains various parameters like true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Finally, accuracy can be defined as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

3.6 Texture Feature

Texture feature is an important feature, and it plays a vital role in tobacco leaf classification. It contains various models like GLCM, LBP, LBPV, GLTP, and so on. In the proposed system, GLCM model is used to extract the texture feature in leaf.

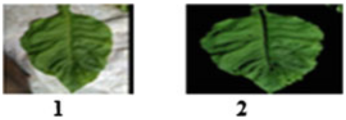


Fig. 3 Image 1 is the original image and Image 2 is the resultant image after segmentation [13]

4 Experimental Results

The implementation of the proposed model is performed for analysis, and the classification of tobacco leaf is performed using Windows 10, 64-bit operating system, 2.80 GHz Intel(R) Core(TM) i7, CPU, 16 GB RAM.

4.1 Segmentation

Segmentation of images was done to uniformly produce leaf images facing upward since the proposed algorithm can be able to detect multiple leaves under different angles. The results of the processing steps of image segmentation are displayed in Fig. 3.

4.2 Classification

Experiments on classification were done using three datasets, namely segmented (1020) images, non-segmented (1020) images, and the combination of segmented and non-segmented (2040) images. They were trained separately to determine which dataset can generate the highest classification rate in grading tobacco leaves. The performance of the classifier was obtained after 500 iterations.

4.3 Dataset

Dataset for the network is an image dataset. Image datasets are green uncured tobacco leaves, and these are acquired in agricultural tobacco processing area near Ramanathapura tobacco board, located in Hassan, Karnataka, India. All images are captured using a mobile camera of 14 MP and images are featured by three channels (red, green, and blue channel).

Images are acquired with different illumination. The original size of the acquired images is 4080 × 3680 pixels and 2040 images are collected, where all image samples belong to four categories such as

- overripe

- ripe 1
- ripe 2
- unripe.

The datasets are divided into training dataset of 1378 image samples, a dataset of 622 images as testing samples, and remaining 40 images are used for prediction. The training dataset consists of four categories.

- 268 images belong to category 1 (unripe)
- 255 images belong to category 2 (overripe)
- 398 images belong to category 3 (ripe 1)
- 457 images belong to category 4 (ripe 2).

Before learning of these images preprocessing methods are used. Testing dataset consists of four categories.

- 130 images belong to category 1 (unripe)
- 117 images belong to category 2 (overripe)
- 182 images belong to category 3 (ripe 1)
- 193 images belong to category 4 (ripe 2).

The testing dataset is used to evaluate the performance of the model. Before applying it to the model, preprocessing technique is used. We conducted the experiments using deep neural networks on green uncured tobacco leaves. Three groups of tobacco leaves image samples have been used in our experiments.

The first group of images belonging to four categories has been trained by neural networks. As per the results given by the model evaluation, the classification validation accuracy result is 94.2%. The second group of image samples is testing samples, and these samples are not trained by the neural networks. The testing accuracy result is 86%. The confusion matrix obtained on the results is given in Table 1.

The above confusion matrix dataset may have 398 samples of ripe 1 class, 18 samples are incorrectly classified, that is, 12 samples are misclassified as ripe 2 and 6 samples are misclassified as unripe. Among 457 samples of ripe 2 class, 36 samples are incorrectly classified, that is, 12 samples are misclassified as ripe 1, 17 samples are misclassified as overripe, 7 samples are misclassified as unripe.

In the third category overripe class, out of 255 samples 17 samples are misclassified as ripe 2 class. The fourth category is unripe class: 18 samples are misclassified

Table 1 Confusion matrix of ripped and un-ripped of tobacco leaves

Status	Ripe 1	Ripe 2	Over_Ripe	Un_Ripe
Ripe 1	380	12	0	6
Ripe 2	12	428	17	2
Over_Ripe	0	17	238	0
Un_Ripe	18	7	0	243
Accuracy (%)	94.20			

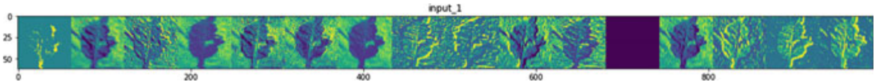


Fig. 4 Input file [7]

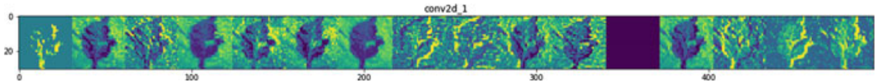


Fig. 5 Convolution output

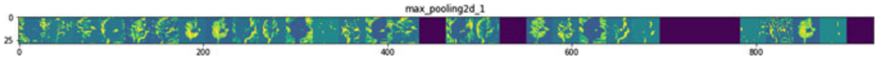


Fig. 6 Maximum pooling file

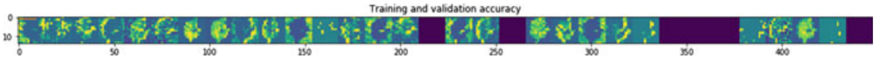


Fig. 7 Training and validation accuracy

as ripe 1 and 7 samples are classified as ripe 2 class. Figures 4, 5, 6, and 7 show the output of different convolution layer of this model.

Figures 4, 5, 6, and 7 show the output of different layers in a CNN model along with the training accuracy and validation. For preprocessing in the CNN model 256 × 256 size images were fed, where Fig. 4 is the input layer to the CNN model and the outcome of all testing samples with different layers are specified in Figs. 5, 6, and 7 (Fig. 8).

Figures 9 and 10 illustrate the training and testing accuracy based on ROC.

Figure 11 illustrates the training and testing loss using ROC.

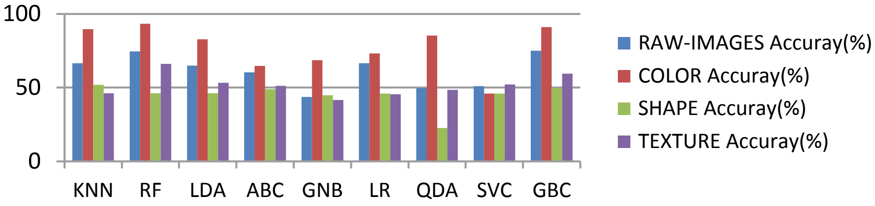


Fig. 8 Accuracy in different machine learning algorithms

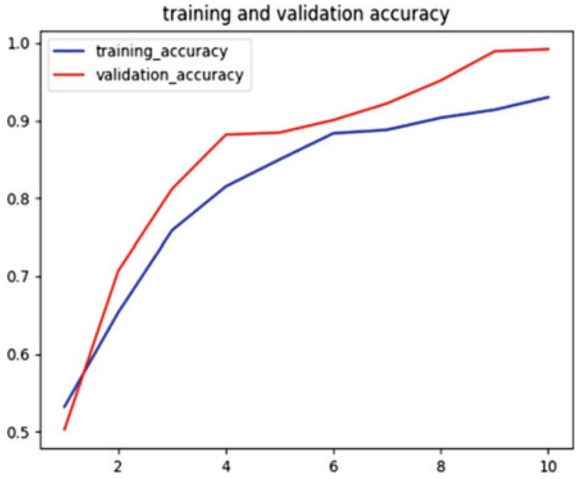


Fig. 9 ROC of accuracy validation

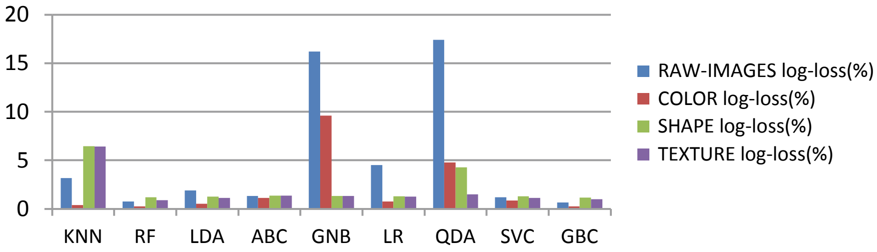


Fig. 10 Loss of different machine learning algorithms

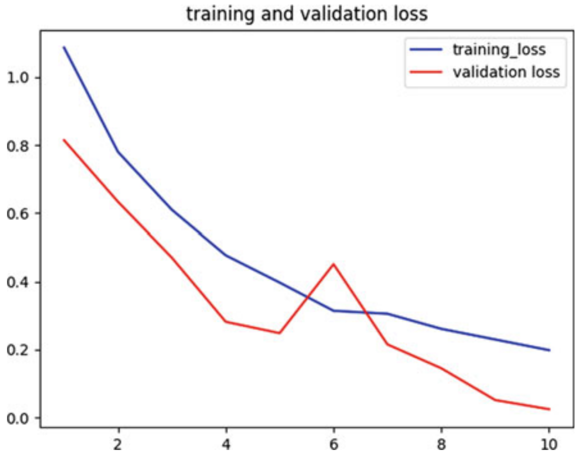


Fig. 11 ROC of loss of CNN

5 Conclusion

There are many methods which have been implemented for tobacco leaf classification using machine vision. But still, there is no automated system for separating the green uncured tobacco leaves. The main objective of the proposed model is to classify the green uncured tobacco leaves for the curing process. This helps the farmer to apply adequate temperature and load the tobacco sticks to the barren to get better quality of tobacco. The proposed model has a better performance and achieved 94.2% of classification and 86% of testing accuracy. A convolution neural network is used to classify the tobacco leaf. Around 2000 samples were used and the network is trained with 1378 samples and achieved 94.2% classification accuracy. Three features are used to classify the leaf. The color feature is the best feature to achieve the best accuracy compared to texture and shape feature. In future, we will enhance the proposed work to detect analysis of the diseases in a leaf as well as in the plant.

References

1. ICAR-CTRI (2019) Rajahmundry. Tobacco in Indian Economy. https://ctri.icar.gov.in/index.php
2. Zhang J, Sokhansanj S, Wu S, Fang R, Yung W, Winter P (1997) A trainable grading system for tobacco leaves. Comput Electron Agric 16(30):231–244
3. Zhang F, Zhang X (2011) Classification and quality evaluation of tobacco leaves based on Image processing and fuzzy comprehensive evaluation. In: Sensors-MDPI
4. Yawootti A, Kaewtrakulpong P (2005) A machine vision system for thai flue-cured tobacco classification
5. Han L (2008) Recognition of the part of the growth of Flew-cured tobacco leaves based on SVM. In: Proceedings of 7th world congress on intelligent control and automation, pp 25–27
6. Zhang J, Sokhansanj S, Wu S, Fang R, Yung W, Winter P (1998) A transformation technique from RGB signals to the Munsell system for color analysis of tobacco leaves. Comput Electron Agric 96:155–166
7. Guru DS, Mallikarjuna PB, Manjunath S (2011) Min-max representation of features for grading cured tobacco leaves. Stat Appl 9(1&2) (New Series):15–29
8. DS, Mallikarjuna PB, Manjunath S (2012) Machine vision-based classification of tobacco leaves for automatic harvesting. Intell Autom Soft Comput 18(5):581–590. ISSN: 1079-8587
9. Simonyan K, Vedaldi A, Zisserman A (2014) Deep inside convolutional networks: visualising image classification models and saliency maps
10. Liu J, Shen J, Shen Z, Lie R (2012) Grading tobacco leaves based on image processing and generalized regression neural network. In: IEEE conference
11. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: ICLR
12. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: CVPR(IEEE)
13. Gurumoorthy S et al Dasari SK, Chintada KR, Patruni M (2018) Flew-cured tobacco leaves classification: a generalized approach using deep convolutional neural networks. In: Cognitive science and artificial intelligence, Springer briefs in forensic and medical bioinformatics

Next-Generation WSN for Environmental Monitoring Employing Big Data Analytics, Machine Learning and Artificial Intelligence

Rumana Abdul Jalil Shaikh, Harikumar Naidu, and Piyush A. Kokate

Abstract A worldwide network of wireless sensors is used to monitor dynamic environmental changes with respect to time. Therefore, the data provided by these sensor networks are crucial for collecting specific information; hence data analytics is essential in such networks. For effective utilization of gathered data, big data analytics can be one of the prominent solutions since the data plays an important part in machine learning allowing the WSN to adapt the dynamic changes in environment to save cost and efforts of redesigning the present WSN. In this paper we present the advances of WSN to further develop the next-generation wireless sensor network by employing software-defined network (SDN), big data analytics, machine learning and artificial intelligence tool along with its benefits and challenges. We also discuss the software-defined wireless sensor network (SDWSN) and the possibility of application of artificial intelligence in it to meet the challenges of SDWSN and its advantages. And finally, we have discussed different problems associated with WSN network specifically for environmental monitoring and their respective solutions using different machine learning paradigms and how efficiently the adoption of big data analytics in ML and AI plays an important role to serve the improved performance requirements.

Keywords Wireless sensor network (WSN) · Software-defined WSN · Environmental monitoring · Big data analytics · Machine learning (ML) · Artificial intelligence (AI)

R. A. J. Shaikh (✉) · H. Naidu
G. H. Raisoni College of Engineering, Nagpur, India
e-mail: shaikh_rumana.ghrcemtechped@raisoni.net

H. Naidu
e-mail: harikumar.naidu@raisoni.net

P. A. Kokate
CSIR-NEERI, Nagpur, India
e-mail: pa_kokate@neeri.res.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_20

1 Introduction

In wireless sensor network, several independent, small, inexpensive and energy-saving sensor nodes are manipulated, which collect the data and send it to the controller used in the system. The controller then transmits the data via the wireless nodes used in the system to the base station or sink also known as end device [1]. These wireless nodes can be two or more, depending on the system requirement to form the wireless network. Wireless sensor networks (WSNs) have the potential to create powerful applications each with its own features and requirements [1–9]. Therefore, wireless sensor network is widely used in a variety of applications, such as defense, medical, environmental, agricultural and industrial. The primary goal of WSN is to wirelessly transmit or share data to achieve this. The available standards are ZigBee, WiFi, Bluetooth, Z-Wave, and so on. For industrial application wireless HART and ISA100 standards are popularly used. One of the different strengths of WSN are the sensor nodes in the network, which can be designed with different sensors such as air quality or weather, optical, thermal, pressure and acoustic sensors depending on the target environment [1]. To design the WSN for specific application, the wireless sensor network is divided into event-driven WSN, periodic- or time-driven WSN and demand-driven WSN, and each has its own corresponding communication mode.

I. **Event-driven WSN**: In event-driven mode, the sensors report acquisition data to the end device as soon as a specific event (e.g. fire, flood, radioactivity, and gas leakage, etc.) has been detected.

II. **Time-driven WSN**: In the period-driven mode (or time-controlled mode), the sensor nodes record information from the environment at predetermined samples of time and periodically send the data to the end device. For example, in the WSN for air quality monitoring the user defines the particular period of time to obtain the data, such as early hours in the morning and late hours at night. In such type of system the sensor will gather the data for predefined time span and transmit it to the receiver.

III. **Demand-driven WSN**: In the demand-driven wireless sensor network, the sensor senses the data only when the user commands it to transmit it to the receiver end and hence also known as query-driven mode.

Designing methodical structures that are worthy of number of application schemes is difficult to achieve, particularly for environmental monitoring (in agriculture, for air quality and critical areas where human accessibility is difficult to achieve, etc.) keeping account of various issues related to WSN, such as data reliability, node clustering, localization, energy-aware routing, fault detection, security, event scheduling and data aggregation [1, 2, 10–16]. Energy consumption is one of the key issues present in WSN and it is mainly due to the radio communication due to which wireless transmission and reception take place [2]. Routing is one of the major issues associated with the WSN. The three vital driving forces in order to achieve efficient routing techniques are quality of service (QoS), deregulated telecommunication industries and tremendous growth in network size and its respective usage [2, 15]. Another issue with WSN is its security. As far as the security of WSN is concerned, the attacks can be classified as goal-oriented. This type of attack can be active or passive, and the second type of attack is performance-oriented. It can be outside and inside attack. In the third type of attack which is layer-oriented, the physical layer of the network stack has been targeted [17].

Along with the challenges mentioned above and the increasing need of WSN-based applications, these challenges need to be resolved in the cost-effective way. There have been many researches listed and the software-defined wireless sensor network (SDWSN) is one of the prominent solutions [2, 18, 19].

SDN allows network administers to control and manipulate devices in the network [2]. In the network implementation SDN allows the separation of physical layer of forwarding plane from the control plane [2, 20]. SDN serves automation, virtualization through which localization of network resource is possible and provides the platform to manipulate number of devices in the network with a single command. Deployment of SDN in WSN allows the network to change its behavior accordingly at any instant of time. SDN also serves dynamic scaling which allows the network to change its size and quantity. It also improves performance by optimizing devices in the network by improving its load capacity and bandwidth management, along with service integration [2].

As the WSN deals with the dynamic environmental variation with respect to time, therefore sensor nodes must be able to adapt to the environmental changes and also must operate adequately. At the critical areas which are unpredictable and where human accessibility is difficult to achieve (volcano eruption, air quality monitoring at certain height, weather monitoring at glaciers, waste water monitoring etc.), to gather information in such location the sensor network must be designed with the ability of self-calibration, thus machine learning is the most popular solution to save the cost and efforts of redesigning the system network as it allows the system to calibrate and update with the variable environment requirement [1, 2].

Primarily, machine learning (ML) was introduced in the late 1950s as a tool for artificial intelligence (AI) [1, 21] and originated from early studies in cybernetics and robotics. With the passing time, the evolution of machine learning has been witnessed and its approach has been tilled more toward the algorithm which is computational paradigm. Owing to its robust nature, the ML is extensively used in a variety of applications such as genomics, voice recognition, spam detection, fraud detection (imposter recognition) and in advertisement network. The paradigm applied in ML comes from number of diverse fields which cover mathematics, computer science, statistics and neuroscience [1].

In the field of information and communication technology, huge amount of data is popularly termed as big data. The WSN consists of a large number of sensor nodes distributed in wide area even though the data sensed and transmitted by the single sensor node seems to be small but the data altogether obtained by the sensor network are huge [22]. As the time goes by, the data gathered by these WSNs experience an exponential growth. Therefore wireless sensor networks (WSNs) are among many possible large data sources, as it generates a large amount of data by various sensor nodes in networks. In large-scale WSN to process data, big data analytics is one of the effective solutions.

To make AI a reality, big data analytics plays an important role. To learn and calibrate itself, the system needs previously-driven data and this need of data from where system can process, learn, calibrate and on the basis of which system makes decision accordingly is provided by big data analytics.

The rest of the paper is organized in the following fashion. Sect. 2 describes the software-defined WSN and how SDWSN overcomes the basic issues associated with WSN. In Sect. 3 we discuss the utilization of machine learning as AI tool to target the major issues of large-scale wireless sensor network (LS-WSN). In Sect. 4 we elaborate further role of big data analytics in WSN. And finally, in Sect. 5 we conclude the article by explaining the advance technologies mentioned above and how they improve the performance in the WSN.

2 Software-Defined WSN

The fusion of software-defined network (SDN) with standard wireless sensor network (WSN) results in new paradigm known as software-defined wireless sensor network (SDWSN). Incorporation of SDN in WSN brings simplicity in network handling process, its management and configuration. It also provides flexibility to bring into effect instant changes (Fig. 1).

The SDN made implementation of AI in WSN an easy task, through which various challenges associated with WSN can be resolved efficiently. The application of SDN in large-scale WSN not only provided solution to various issues associated with WSN but also can be used for better management of WSN [23].

Many researches have been performed to develop a flexible platform for utilization of AI in large-scale WSN to improve its performance by targeting issues, including routing, traffic management security and energy-efficient monitoring. SDN alone targets the common issues (architecture, network topology and security, energy saving along with load balancing, data transmission and routing, node scheduling as well) associated with WSN to achieve effective deployment and improve performance as SDN uses algorithm which separates forwarding phase from control which enables to simplify management and configure the network extension. The SDN also contributes in resolving data transmission-related problems in the network [24].

2.1 Architecture

SDN brings flexibility in the network by managing the nodes in the network efficiently. Incorporation of SDN made illumination of old nodes and inclusion of new nodes in the network an easy task [25, 26].

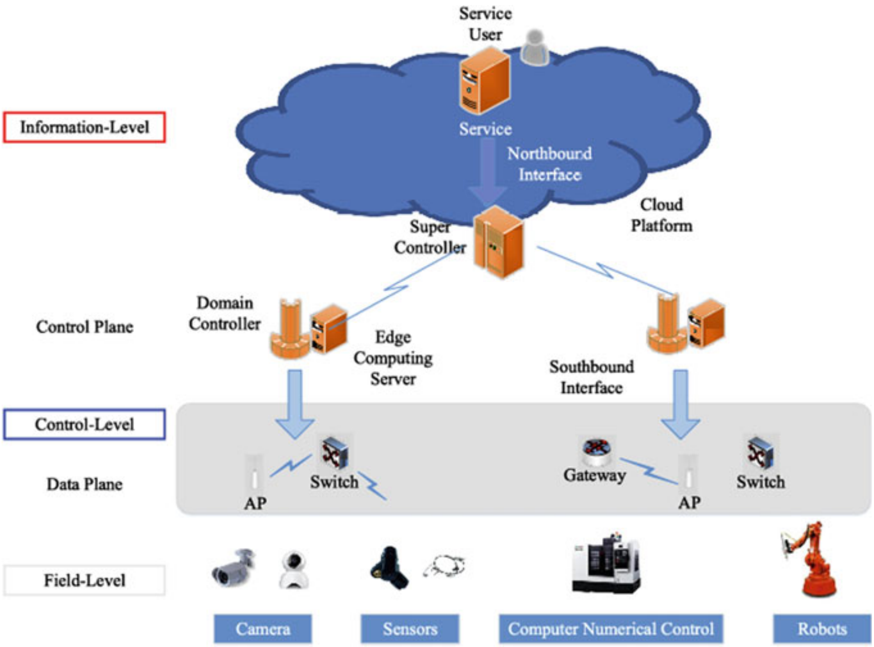


Fig. 1 Architecture of SDN in WSN (*Source* https://doi.org/10.1109/access.2019.2890854)

2.2 *Network Topology*

In the standard WSN network topology is dynamic (not fixed), due to demise of old nodes and addition of new nodes and results in variation in network topology. Incorporation of SDN in WSN makes sure the normal operation of the network communication regardless of change in the network topology [25–27].

2.3 *Routing Protocol*

Routing protocol ensures the faithful transmission of data packets among the nodes in the network by selecting and controlling the most preferable data path from source to destination nodes through various hops. In the SDWSN, the SDN obtained the data flow based on the suitable route selection [28–32].

2.4 Energy Consumption and Scheduling

In the WSN the energy consumption to achieve faithful communication between source and destination other than monitoring the network is an important aspect to take in consideration while designing the network. In the SDWSN as the shortest suitable route is provided by the SDN efficiently, it reduces the energy consumption. An optimized scheduling is important as it increases the life time of the nodes, which eventually contributes to prolong the network lifecycle and minimizes the network operating cost [33–40].

2.5 Transmission and Network Load

In the network the rate of node communication decreases with increase in network load which directly affects the processing time of packets. The SDWSN uses the centralized load balancing and proper nodes scheduling pattern which reduces the use of these nodes and the overall network communication [41–44].

2.6 Network Security

The standard WSN is vulnerable when it comes to security, as SDWSN utilizes centralized control and coupled network authentication along with encryption which effectively improves the network security [45–47].

3 Utilization of Machine Learning in WSN

In WSN, the machine learning is considered as a tool that generates algorithms and patterns which are utilized to provide prediction models [1]. In particular, for environmental monitoring applications these predictive models can be proved essential as it can provide notifications of future occurring events by processing previously available data. In WSN machine learning and AI can perform inevitable type of tasks such as decision making and control, data storage and compression, detection of repeated patterns sequence which help in finding new hidden patterns in complex data, and suggest optimized ways of sensor deployment in the network to achieve maximum data coverage. Among the many benefits of ML in WSN, one of them is ML techniques that perform analysis to find correlation between the nodes in the network, thus providing practical approaches to extract targeted information effectively [48]. On the basis of intended framework the machine learning algorithms

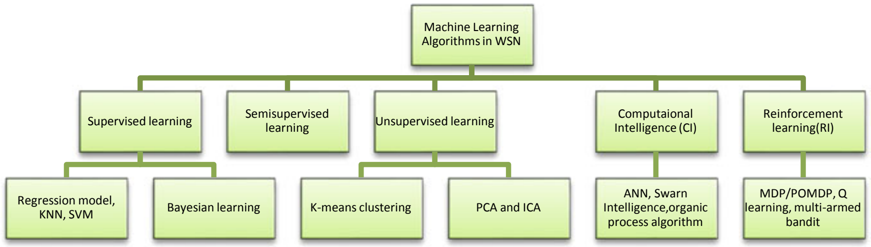

Machine Learning Algorithms in WSN
Supervised learning
Semisupervised learning
Unsupervised learning
Computaional Intelligence (CI)
Reinforcement learning(RI)
Regression model, KNN, SVM
Bayesian learning
K-means clustering
PCA and ICA
ANN, Swarn Intelligence,organic process algorithm
MDP/POMDP, Q learning, multi-armed bandit

Fig. 2 Machine learning algorithms in WSN

(approaches) are categorized as supervised, semi-supervised, unsupervised, reinforcement learning (RL) and computational intelligence (CI) approaches [1, 48, 49]. These algorithms generate cost-effective approach compared to numerical models [48]. Figure 2 presents the family of machine learning paradigms.

In machine learning the techniques and algorithms can also be segregated on the basis of similarities in terms of operation [50].

3.1 Supervised Learning Method

As the name indicates in supervised learning method the system model is designed with predefined input and output and termed as labeled set. This model describes the learned relationship between system parameters along with its input and output [1].

Figure 3 presents various supervised machine learning techniques along with its characteristics, such as Bayesian learning with key features of posteriori distribution calculation, Gaussians mixture (GM) model, expectation maximization (EM) and hidden Markov models (HMMs). Supervised ML algorithm is further classified as regression models (with key features of estimating the variables relationship, linear and logistics regression), K-nearest neighbor (KNN) and support vector machines (SVM) [50].

3.1.1 Bayesian Learning

In networks the probability of posterior distribution of targeted variables along with its status on input signals and all of the training instances is optimized by using Bayesian learning algorithm [50]. One of the many advantages of Bayesian learning algorithm is it requires less number of training samples [51]. The Gaussians mixture (GM) model, expectation maximization (EM) and hidden Markov models (HMMs) are general examples of generative models using Bayesian learning algorithm [52, p. 445]. In Gaussian mixture model the data points from different clusters are grouped and Gaussian distributed, whereas in EM model the two steps of operation are followed, where E step represents the function of lower bound likelihood and in M

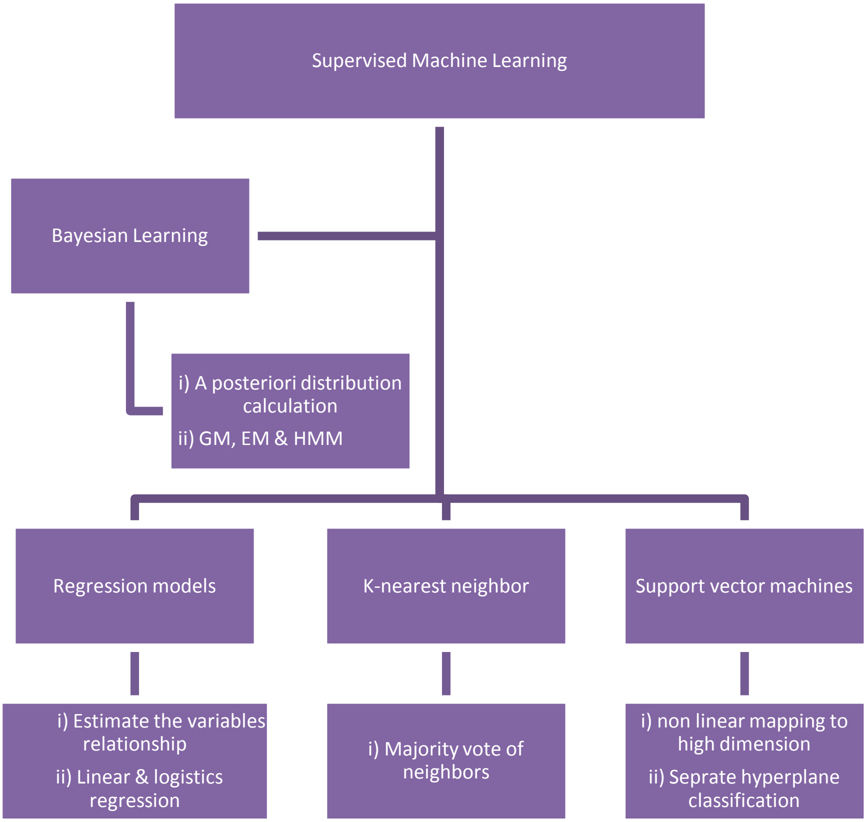

Supervised Machine Learning
Bayesian Learning
i) A posteriori distribution calculation
ii) GM, EM & HMM
Regression models
K-nearest neighbor
Support vector machines
i) Estimate the variables relationship
ii) Linear & logistics regression
i) Majority vote of neighbors
i) non linear mapping to high dimension
ii) Seprate hyperplane classification

Fig. 3 Supervised machine learning algorithms

step the selected function is being maximized [50]. Another type of popular model in Bayesian learning is HMM which is used to design the probability distribution of the observed sequences, where the unknown variables through which specific mixture of the components is being controlled in every observation and related through Markov process [50]. In WSN Bayesian inference can be applied in order to assess consistency of event by utilizing incomplete datasets.

3.1.2 K-Nearest Neighbor (K-NN)

K-NN algorithm falls under the supervised machine learning techniques as data sampled is classified on the basis of output values, that is, labels of the nearest samples of data; for example, if sensor node misses readings, then that can be predicted by taking into account average measurements of adjacent sensors within defined boundaries. Especially in query-driven WSN, K-NN algorithm is extensively used as it does not require high computational power as computed function is relative to local points [1], where K represents small positive integer value [53, 54].

3.1.3 Regression Model

Regression model is used to estimate the relationships among the variables in the system. The benefit of regression analysis is the predicting capability through which the value of continuous valued targets can be efficiently known by defining input variables D-dimensional signal vector "x". These estimated targets are the function of self-dependent variables. In supervised machine learning the regression models can be generated linearly and logistically. In linear regression the function is linear while the function in logistic regression is logistic considering similar sigmoid curve [50].

3.1.4 Supportive Vector Machines (SVMs)

The SVM performs classification of data points along with the objects utilizing the labeled training samples [55]. In contrast with the KNN the SVM model depends on nonlinear process of mapping which includes conversion of original training data to the higher dimension where it can be easily identified due to the separation. By utilization of these models in the network the radio parameters can be identified related to specific users [50] (Fig. 4).

Supervised learning method is the most preferable machine learning algorithm as it targets several WSN operation-related issues, such as localization and targeting objects [56–58], event identification and query processing [53, 54, 59, 60], media access management [61–63], security and invasion detection [64–67] and QoS along with information integrity and accountability [68–70].

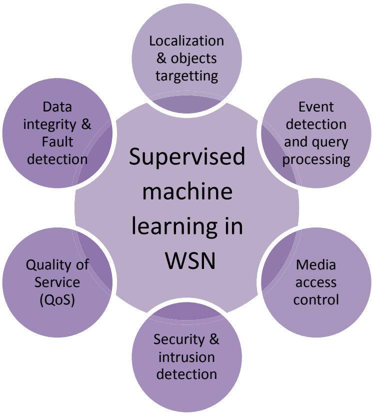
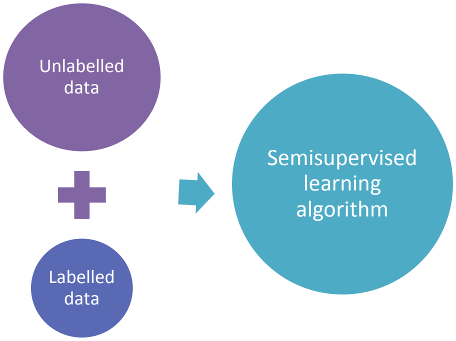

Supervised machine learning in WSN
Localization & objects targetting
Event detection and query processing
Media access control
Security & intrusion detection
Quality of Service (QoS)
Data integrity & Fault detection

Fig. 4 Application of supervised machine learning algorithms in WSN

Fig. 5 Semi-supervised machine learning algorithms



3.2 *Semi-Supervised Machine Learning Method*

The semi-supervised machine learning method is a technical combination of supervised and unsupervised learning comparatively with higher performance in which large amount of unlabeled data are combined with labeled data to generate the algorithm (Fig. 5).

3.3 *Unsupervised Machine Learning Method*

On the contrary to supervised learning method, in unsupervised learning the input and output are not predefined, and thus termed as unlabeled set. The aim of an unsupervised learning method is to categorize the set of samples into different groups by identifying similarities among them [50].

Figure 6 presents various unsupervised machine learning techniques along with their characteristics. Mainly these algorithms are K-means clustering, principal component analysis (PCA) and independent component analysis (ICA) through which hidden independent factors can be efficiently unveiled.

3.3.1 K-Means Clustering

The goal of K-means clustering is to group the observations (n) into clusters (k), in the manner where these observations belong to the nearest cluster in the network [50]. K-means clustering targets the sensor node clustering problems in the network [1].

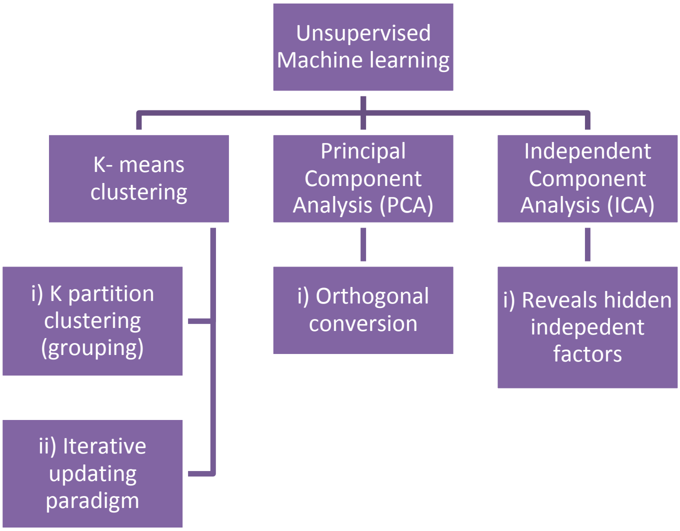

Unsupervised Machine learning
K- means clustering
Principal Component Analysis (PCA)
Independent Component Analysis (ICA)
i) K partition clustering (grouping)
ii) Iterative updating paradigm
i) Orthogonal conversion
i) Reveals hidden indepedent factors

Fig. 6 Unsupervised machine learning algorithm

3.3.2 Principal Component Analysis

PCA involves the conversion of potentially correlated variables to uncorrelated variables in the ratio where number of principle components should not exceed the number of original variables [50].

3.3.3 Independent Component Analysis

On the contrary, ICP involves statistics-based analysis technique which is used to unveil hidden factors that contain group of random variables in the network [50] (Fig. 7).

Unsupervised machine learning techniques target different WSN security-related issues by identifying rarity, liability and invasion in the network. It also resolved traffic problems in the network by providing better clustering techniques [50].

4 Big Data Analytics in WSN

LS-WSN generates large amount of data, particularly in environmental monitoring application as data is being fetched continuously sometimes to predict future events

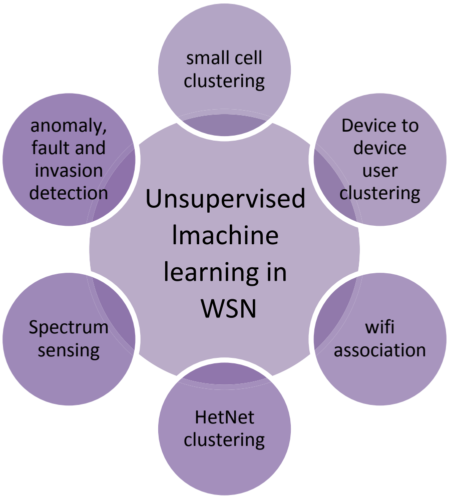

small cell clustering
anomaly, fault and invasion detection
Device to device user clustering
Unsupervised lmachine learning in WSN
Spectrum sensing
wifi association
HetNet clustering

Fig. 7 Application of unsupervised machine learning algorithms in WSN

with time. The amount of data increases further for the next-generation wireless sensor network and this data is essential in order to train and calibrate itself accordingly without human interference.

5 Conclusion

Design and implementation of wireless sensor network specifically for environmental monitoring at critical areas where interference and accessibility of human beings are difficult to achieve. Hence the need of self-calibration in WSN as per the change in environment is essential.

In this paper we discussed utilization of software-defined network, artificial intelligence and machine learning techniques to enable the WSN to calibrate itself according to the variation in the environment. In large-scale WSN the deployment of sensors to collect and monitor different aspects in environment and gathering data efficiently is one of the vital concerns to be taken into account while designing the network along with the inherent issues of localization and targeted objects. Event identification and query processing, media access management, security, invasion detection and QoS with integrity of information and its accountability can be improved and resolved by various AI techniques and machine learning algorithm.

SDWSN and AI together build a platform to improve performance of LS-WSN by improving reliability and security in networks and also improve response time to detect and solve occurring issues by analyzing, processing and updating the network as per the requirements.

Machine learning in LS-WSN performs certain types of tasks such as decision making and control, data storage and compression, identification of repeated patterns and analysis of pseudo-patterns sequence which help in finding new hidden patterns in complex data.

Big data analytics is one of the prominent solutions in LS-WSN to solve the processing and storage issues. As in WSN continuous data is received and this data is essential to predict the future events and get alarming notification for disaster.

References

1. Alsheikh MA, Lin S, Niyato D, Hwee-Pink T (2014) Machine learning in wireless sensor networks: algorithms, strategies, and applications. IEEE Commun Surv Tutor 16:4
2. Matlou OG, Abu-Mahfouz AM (2017) Utilizing artificial intelligence in software defined wireless sensor network. In: IECON-43rd annual conference of the IEEE industrial electronics society, Beijing China
3. Abu-Mahfouz AM, Olwa T, Kurien A, Munda JL, Djouani K (2015) Towards developing a distributed autonomous energy management system (DAEMS). In: Proceedings of the IEEE AFRICON 2015 conference on green innovation for african renaissancce, pp 1–6
4. Dongbaare P, Chowdhury SP, Olwal TO, Abu-Mahfouz AM (2016) Smart energy management system based on an automated distributed load limiting mechanism and multi-power switching technique. In: Proceedings of the 51st International universities power engineering conference
5. Mudumbe MJ, Abu-Mahfouz AM (2015) Smart water meter system for user-centric consumption measurement. In: Proceedings of the IEEE international conference on Industrial Informatics, pp 9993–9998
6. Abu-Mahfouz AM, Haman Y, Page PR, Djouani K (2016) Real time dynamic hydraulic model for potable water loss reduction. Proc Eng 154(7):99–106
7. Cheng B, Cui L, Jia W, Zhao W, Gerhard PH (2016) Multiple region of interest coverage in camera sensor networks for tele-intensive care units. IEEE Trans Ind Inform 12(6):2331–2341
8. Silva B, Fisher RM, Kumar A, Hancke GP (2015) Experimental link quality characterization of wireless sensor networks for underground monitorig. IEEE Trans Ind Inform 11(5):1099–1110
9. Phala KSE, Kumar A, Hancke GP (2016) Air quality monitoring system based on ISO/IEC/IEEE 21451 standards. IEEE Sens J 16(12):5037–5045
10. Abu-Mahfouz AM, Hancke GP (2013) Evaluating ALWadHA for providing secure localization for wireless sensor network. In: IEEE AFRICON conference, pp 501–505
11. Ntuli N, Abu-Mahfouz AM (2016) A simple security architecture for smart water management system. Proc Comput Sci 83(4):1164–1169
12. Louw J, Niezen G, Ramotsoela TD, Abu-Mahfouz AM (2016) A key distribution scheme using elliptic curve cryptography in wirelesssensor networks. In: Proceeding of the 14th IEEE international conference on industrial Informatics, pp 1166–1170
13. Abu-Mahfouz AM, Hancke GP (2017) ALWadHA localization algorithm: yet more energy efficient. IEEE Access 5(5):6661–6667
14. Abu-Mahfouz AM, Hancke GP (2017) Localised information fusion techniques for location discovery in wireless sensor networks. Int J Sens Netw
15. Silva B, Hancke GP (2016) IR-UWB-Based non-line-of-sight identification in harsh environments: principles and challenges. IEEE Trans Ind Inform 12(3):1188–1195
16. Chiwewe TM, Mbuya CF, Hancke GP (2015) Using cognitive radio for interference resistant industrial wireless sensor networks: an overview. IEEE Trans Ind Inform 11(6):1466–1481
17. Kaur H, Sahore S (2016) A survey on wireless sensor network (wsn) security using AI methods. Int J Latest Trends Eng Technol 7(4):234–239
18. Kobo HI, Abu-Mahfouz AM, Hancke GP (2017) A survey on software defined wireless networks: challenges and design requirements. IEEE Access 5(1):1872–1899

19. Modieginyane KM, Letswamotse BB, Malekian R, Abu-Mahfouz AM (2017) Software defined wireless sensor network application opportunities for efficient network management: a survey. Comput Electr, Eng
20. Xiang W, Wang N, Zhou Y (2016) An energy efficient routing algorithm for software defined wireless sensor networks. IEEE Sens J 16(20):7393–7400
21. Ayodele TO (2010) Introduction to machine learning, in new advances in machine learning. InTech, Rijeka
22. Ramesh Babu KR, Suja GJ, Samuel P, Jos S (2015) Performance analysis of Big data gathering in wireless sensor network using an EM based clustering scheme. In: IEEE Fifth international conference on advances in computing and communications
23. Ndiaya M, Hancke GP, Abu-Mahfouz (2017) Software defined networking for improved wireless sensor networking for improved wireless sensor network management: a survey. Sensors 17(5):1031, 1–32
24. Duan Y, Luo Y, Li W, Pace P, Fortino G (2018) Software defined wireless sensor networks: a review. In: Proceeding of the 2018 IEEE 22nd international conference on computer supported cooperative work in design
25. Luo T, Tan H, Quek TQS (2012) Sensor open flow: enabling software defined wireless sensor networks. IEEE Commun Lett 16(11):1896–1899
26. De Gante A, Aslan M, Matrawy A (2014) Smart wireless sensor network management based on software defined networking. In: IEEE 27th Biennial symposium on Communications, pp 71–75
27. Han Z, Ren W (2014) A novel wireless sensor network structure based on SDN. Int J Distrib Sens Netw 10(3):1–7
28. Mohapatra R, Mishra S, Mohapatra T (2012) Coverage problem in wireless sensor networks. Comp Cytogenet 2(1):67–72
29. Arumuganm G, Ponnuchamy T (2015) Ea-leach: development of energy efficient leach protocol for data gathering in wsn. EURASIP J Wirel Commun Netw 2015(1):1–9
30. Figueiredo CMS, dos Santos AL, Loureiro AAF, Nogueira JM (2005) Policy-based adaptive routing in autonomous wsns. In: IEEE ambient network international conference on distributed systems: operations and management. Springer, Berlin, pp 206–219
31. Shanmugapriya S, Shivakumar M (2015) Context based route model for policy based routing in wsn using sdn approach. In: iSRASE
32. Wang C, Sohraby K, Daneshmand M, Hu Y (2006) A survey of transport protocol for wireless sensor networks. IEEE Netw 20(3):34–40
33. Tian D, Georganas N (2003) A node scheduling scheme for energy conservation in large wireless sensor networks. Wireless Commun Mob Comput 3(2):271–290
34. Xing G, Wang X, Zhang Y, Lu C, Pless R, Gill C (2005) Integrated coverage and connectivity configuration for energy conservation in sensor networks. ACM Trans Sens Netw 1(1):36–72
35. Hua C, Yum TP (2007) Asynchronous random sleeping for sensor networks. ACM Trans Sens Netw 3(3):1–25
36. Kumar S, Lai TH, Balogh J (2004) On k-coverage in a mostly sleeping sensor network. In: Proceeding of the tenth annual international conference on mobile computing and networking. ACM, pp 144–158
37. Nath S, Gibbons PB (2007) Communication via fireflies: geographic routing on duty-cycled sensors. In: IEEE 6th international conference on information processing in sensor networks, pp 440–449
38. Wang Y, Chen H, Wu X, Shu L (2016) An energy-efficient sdn based sleep scheduling algorithm for wsn. J Netw Comput Appl 59:39–45
39. Yuan Z, Wang L, shu L, Hara T and Qin Z. (2011) A balanced energy consumption sleep scheduling algorithm in wireless sensor networks, IEEE in wireless communications and mobile computing conference (IWCMC), pp. 831–835
40. Wang Y, Chen H, Wu X, Shu L (2015) Improving wsns sleep scheduling mechanism with sdn-like architecture. In: International conference on information processing in sensor networks. ACM, pp 338–339

41. Levendovszky J, Tornia K, Treplan G, Olah A (2011) Novel load balancing algorithms ensuring uniform packet loss probabilities for wsn. In: IEEE in vehicular technology conference (VTC spring), pp 1–5
42. Zhang Y, Sun G, Li W (2011) Dehca: load balance clustering algorithm for energy heterogeneous wsn based on distance. Appl Mech Mater 44–47:3294–3298
43. Wang M, Li S-N, Li Z-G (2011) Multiple routing with load balancing based on ant colony algorithm in wsn. Comput Eng 37(14):1–4
44. Anatoliy S, Hu Z, Vasyl Y (2015) Increasing the data transmission robustness in wsn using the modified error correction codes on residue number system. Elektronika ir electrotechnika 21(1):76–81
45. Hu Z, Wang M, Yan X, Yin Y, Luo Z (2015) A comprehensive security architecture for sdn. In: IEEE in intelligence in next generation networks (ICIN), pp 30–37
46. Smeliansky R (2014) Sdn for network security. In: IEEE in science and technology conference (Modern networking technologies) (MoNeTec), pp 1–5
47. Yoon C, Park T, Lee S, Kang H, Shin S, Zhang Z (2015) Enabling security function with sdn: a feasibility study. Comput Netw 85:19–35
48. Prajapati J, Jain SC. (2018) Machine learning techniques and challenges in wireless sensor networks. In: Proceeding of the 2nd International Conference on inventive communication and computational technologies. IEEE
49. Abu-Mostafe YS, Magdon-Ismail M, Lin H-T (2012) Learning from data. AMLBook
50. Jiang C, Zhang H, Ren Y, Han Z, Chen KC, Hanzo L (2017) Machine learning paradigms for next generation wireless networks. In: IEEE International conference on communications (ICC)
51. Box GE, Tiao GC (2011) Bayesian inference in statistical analysis, vol 40. Wiley, Hoboken
52. Alpaydm E (2014) Introduction to machine learning, 3rd edn. The MIT Press, Cambridge
53. Winter J, Xu Y and Lee W-C. (2005) Energy efficient processing of k nearest neighbour queries in location-aware sensor networks. In: Proceeding 2nd international conference mobile ubiquitous systems: networking and services, pp 281–292
54. Jayaraman PP, Zaslavsky A, Delsing J (2010) Intelligent processing of k-nearest neighbors queries using mobile data collectors in location aware 3D wireless sensor network. Trend in applied intelligent systems. Springer, Berlin, pp 260–270
55. Steinwart I, Christmann A (2008) Support vector machines. Springer, New York
56. Morelande M, Moran B, Brazil M (2008) Bayesian node localisation in wireless sensor networks, In: Proceedings of IEEE international conference acoustics. Speech signal process, pp 2545–2548
57. Lu C-H, Fu L-C (2009) Robust location-aware activity recognition using wireless sensor network in an attentive home. IEEE Trans Autom Sci Eng 6(4):598–609
58. Shareef A, Zhu Y, Musavi M (2008) Localization using neural networks in wireless sensor networks. In: Proceedings of 1st international conference mobile wireless middleware. Operating systems, and applications, pp 1–7
59. Yu L, Wang N, Meng X (2005) Real-time forest fire detection with wireless sensor networks. In: Proceedings. 2005 ınternational conference on wireless communications, networking and mobile computing, vol 2, pp 1214–1217
60. Bahrepour M, Meratnia N, Poel M, Taghikhaki, Havinga PJ. (2010) Distributed event detection in wireless sensor network for disaster management. In: Proceedings of 2nd 2010 international conference on intelligent networking and collaborative, pp 507–512
61. Kim M, Park M-G (2009) Bayesian statistical modelling of system energy saving effectiveness for MAC protocols of wireless sensor network. Software engineering, artificial intelligence, networking and parallel/distributed Computing, vol 209. Studies in computational Intelligence. Springer, Berlin, pp 233–245
62. Shen Y-J, Wang M-S (2008) Broadcast scheduling in wireless sensor networking using fuzzy Hopfield neural network. Exp Syst Appl 34(2):900–907
63. Kulkarni RV, Venayagamoorthy GK (2009) Neural network based secure media access control protocol for wireless sensor network. In: Proceedings of IJCNN, pp 3437–3444

64. Janakiram D, Adi malikarjuna Reddy V, Phani Kumar A (2006) Outlier detection in wireless sensor networks using Bayesian belief networks. In: Proceedings 1st ınternational conference on communication systems software & middleware, pp 1–6
65. Branch JW, Giannella C, Szymanski B, Wolff R, Kargupta H (2013) In-network outlier detection in wireless sensor networks, knowl. Inf Syst 34(1):23–54
66. Kaplantzis S, Shilton A, Mani N, Sekercioglu Y (2007) Detecting selective forwarding attacks in wireless sensor networks using support vector machines. In: Proceedings 3rd ınternational conference on ıntelligent sensors, sensor networks and ınformation, pp 335–340
67. Rajasegarar S, Leckie C, Palaniswami M, Bezdek J (2007) Quarter sphere based distributed anomaly detection in wireless sensor networks. In: Proceedings IEEE ınternational conference on communications, pp 3864–3869
68. Snow A, Rastogi P, Weckman G, Snow A, Rastogi P and Weckman G. (2005) Assessing dependability of wireless networks using neural networks. In: Proceedings IEEE military communications conference, vol 5, pp 2809–2815
69. Moustapha A, Selmis R (2008) Wireless sensor network modelling using modified recurrent neural networks, application to fault detection. IEEE Trans Instrum Meas 57(5):981–988
70. Wang Y, Martonosi M, Peh L-S (2007) Predicting link quality using supervised learning in wireless sensor networks. ACM SIGMOBILE Mob Comput Commun Rev 11(3):71–83

Generating Automobile Images Dynamically from Text Description

N. Sindhu and H. R. Mamatha

Abstract Synthesis of a realistic image from matching visual descriptions provided in the textual format is a challenge that has attracted attention in the recent research community in the field of artificial intelligence. Generation of the image from given text input is a problem, where given a text input, an image which matches text description must be generated. However, a relatively new class of convolutional neural networks referred to as generative adversarial networks (GANs) has provided compelling results in understanding textual features and generating high-resolution images. In this work, the main aim is to generate an automobile image from the given text input using generative adversarial networks and manipulate automobile colour using text-adaptive discriminator. This work involves creating a detailed text description of each image of a car to train the GAN model to produce images.

Keywords Generative adversarial networks · Generator · Discriminator · GANS · Text to image synthesis

1 Introduction

Generation of images from matching visual descriptions provided in the textual format is a challenge that has attracted attention in the recent research community in the field of artificial intelligence. Powerful and generic neural network architectures have been developed to learn discriminative text feature representations. However, relatively a new class of convolutional neural networks referred to as generative adversarial networks (GANs) has provided compelling results in understanding textual features and generating high-resolution images. This new methodology can find its application in multimedia software for the synthesis of a wide range of images

N. Sindhu (✉) · H. R. Mamatha
PES University, Bengaluru, India
e-mail: sindhunarasimha07@gmail.com

H. R. Mamatha
e-mail: mamathahr@pes.edu

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_21

like room interiors, faces. The architecture of such a neural network comprises two networks generator and discriminator. A generator produces an image from the input noise vector and discriminator detects the fake image generated by the generator. The discriminator is capable of classifying the image generated as real or fake. Given the features of data, discriminator tries to classify the input data by predicting the category or label to which data belongs. Discriminator maps feature to labels and generator does the opposite, given certain data features instead of predicting a label generator attempts to predict features given labels. The solution to text to image synthesis involves two stages. In the first stage, the generator neural network must learn the visual features described in the textual feature. In the second stage, the features are used to generate the image. Effective generative adversarial networks model and training methods can be used to generate an image from the given text input. One of the major applications of generative adversarial networks is the manipulation of images which is a challenge faced in the area of artificial intelligence. Text-adaptive generative adversarial networks (TAGANs) are used to semantically modify the visual attribute and generate the image according to the given text input. The scope of this work involves generating automobile images and manipulating the colour and not on improving the quality of images.

1.1 Organization of Paper

The background needed to understand the problem is included in the introduction section. A few examples of previous works on text to image synthesis and attributes description are included in the related work section. The methods section describes the modules used in this work to generate images from a given text description. Implementation section describes the details of the dataset, GAN model to generate the image and manipulate the colour of the image. The results achieved are discussed in the results and discussion section.

2 Related Works

Realistic image synthesis from text is useful and interesting, but artificial intelligence systems are far from this objective. Effective neural network architectures have been developed in recent years to learn text representations. Novel deep architecture and generative adversarial networks are used in translating visual attributes from text, that is, characters to pixels generating realistic images from the given text input as described by Reed [1]. This work demonstrates the capability of generative adversarial networks to generate images from text input. The datasets used in this approach are CUB-birds dataset and the Oxford-102 flowers dataset with five text descriptions of visual attributes per image.

The approach presented by Farhadi et al. [2] is to change the objective of image recognition from labelling to a detailed visual description. The method of providing a detailed visual description of the objects aids the users to describe the obvious features along with other unusual features so that the model can learn how new objects can be recognized with few visual examples.

Generation of images from text is a new interesting happening in the domain of computer vision as we can design applications that can enable computers to understand images from textual description. GAN-INT-CLS as described by Huang et al. [3] attempts to generate images from given text input describing the visual attributes. This approach is similar to CGAN that uses a condition vector and a noise vector but embeds textual descriptions in place of class labels or attributes.

The multimodal recurrent neural network for generating text descriptions for an image is described by Mao et al. [4]. According to the probability distribution, the caption to images is generated. This proposed model consists of RNN for text description and CNN for images. These two models interact with each other to form a multimodal recurrent neural network. The model proposed in this work addresses the challenge of generating text description for images and retrieval of images and sentences.

Reed et al. [5] propose to overcome the limitations of visual attributes which describe the visual features. In this approach, the natural language model is trained from scratch using words without pre-training. Deep structured joint embedding approach is proposed to embed images and text descriptions describing visual attributes.

The experiments were conducted on CUB and Oxford-102 datasets. The method suggested here is to describe the visual attributes by avoiding the background of an image, species name and to avoid figures of speech.

The translation of the text to image using GAN as proposed by Viswanathan et al. [6] is used to generate flower images. The dataset used is Oxford-102 flower dataset with captions per each image. For text encoding, RNN-CNN is used. The experimental results show that accurate images were generated according to given text input.

To solve the problem of manipulation of images according to the given text input, Seonghyeon et al. [7] propose a method to use natural language to manipulate images. To describe the new visual attributes of an image, the approach here is to semantically change or modify the visual attributes of an object. Text-adaptive generative adversarial networks are used to semantically modify the visual attributes. The experiment was conducted using CUB dataset and Oxford-102 dataset with birds and flower images, respectively. The results from this experiment show that accurate images were generated as per the given text input.

2.1 Motivation

One of the major problems in the field of artificial intelligence is the synthesis and manipulation of an image from the given text input. In this current work, the aim is to generate car images and manipulate the colour of car images. Generative adversarial networks are implemented to generate the images from the given text input. For training the model, images and the corresponding text descriptions are used. To manipulate the colour of car images text-adaptive discriminator is implemented which semantically modifies the attributes to generate the images according to given text input.

3 Methods

In this work of translating text to image, the GAN model is trained with car images. It mainly consists of modules such as generator, discriminator and encoder for text encoding.

3.1 GAN

Goodfellow et al. [8] proposed a novel idea of GAN—generative adversarial network—which comprises two neural networks: generator and discriminator. The generator is a neural network which generates an image based on a given text description. Discriminator classifies the generated image as real image or fake image. The generator creates new images and passes it to discriminator as illustrated in Fig. 1.

Steps involved in GAN:

- The generator takes a random number as input and generates an image.
- The generated image is fed to the discriminator which will be trained with real images and text description.
- Both real and fake images are the inputs to discriminator which predicts the image produced by the generator as real or fake.
- The discriminator is trained with real images, hence can provide continuous feedback on the ground truth of images.
- The generator is in a feedback loop with discriminator.

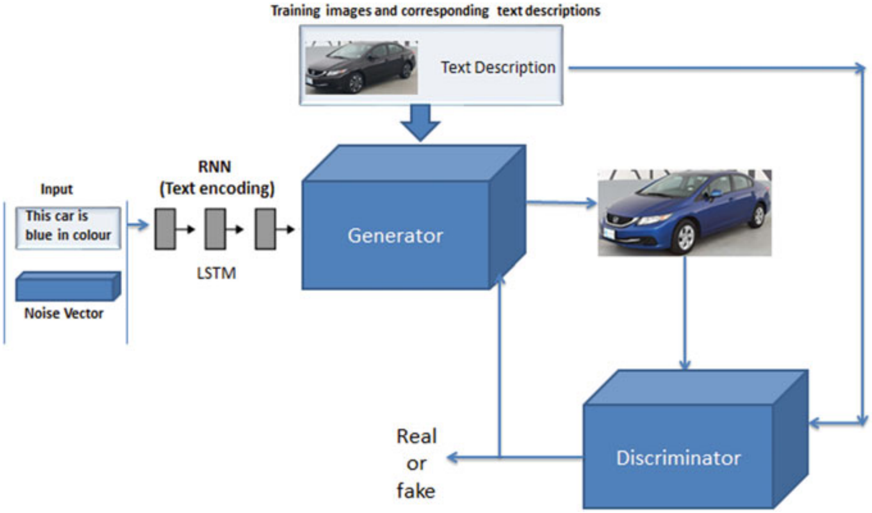


Fig. 1 GAN architecture

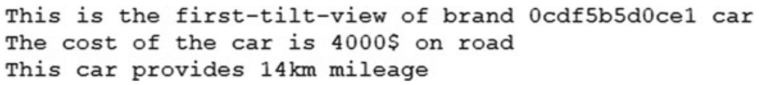
3.2 Encoding

RNN encoding is used to encode the captions of each image in a manner that can be understood by the model. For encoding the captions of each image in Char-CNN-RNN, RNN is a neural network in which the current step takes the output of the previous step as input. This type of neural network is helpful in certain scenarios where the previous step should be remembered. For example, to predict the next word in a sentence, the previous words are required. This problem is solved with the help of a hidden layer in RNN. The hidden state is an important feature of RNN which remembers the information.

3.3 Text-Adaptive Generator Adversarial Network

Text-adaptive generative adversarial network is used to generate the images that are semantically modified while the irrelevant text content will be preserved. Generator encodes the input image and generates a semantically modified image according to the given text input. To preserve the text irrelevant content reconstruction loss is used.

To classify the attributes independently, a word-level local discriminator is created by text-adaptive discriminator as described in [6]. With the help of this text-adaptive discriminator, only the region of the image will be modified according to the given text input. Each sentence-level discriminator is split into word-level discriminator, and with the help of this feedback is given to generator by discriminator to generate



```
This is the first-tilt-view of brand 0cdf5b5d0ce1 car
The cost of the car is 4000$ on road
This car provides 14km mileage
```

Fig. 2 Generated three-line text description

or modify image based on given text input. The text-adaptive discriminator is trained to identify attributes individually rather than sentence level. The image is classified as real or fake based on word-level matching.

4 Implementation

The accuracy and effectiveness of CNN-based image processing system depend on the input dataset used for training and validation of the system. The dataset should comprise a wide range of different images in order to represent the different classes. For this current work, we have chosen Kaggle Carvana dataset. This dataset contains a large number of car images (as .jpg files). Each car has exactly 16 images, each one taken at different angles. Each car has a unique id and images are named according to id_01.jpg, id_02.jpg … id_16.jpg. In addition to the images, some basic metadata about the car make model, year and trim are also provided. The next step is to provide a fine-grained description of visual attributes in text format for each of the image. The idea is to describe salient visual aspects manually in common language which provides a flexible and compact way of encoding features.

4.1 Dataset 1

In the first attempt, 6,573 car images are used for training testing. The captions are generated for each image. Most parts of the caption are generic with unique keyword per image like the cost of the car, model of car and so on. The generated text description file sample is illustrated in Fig. 2.

4.2 Dataset 2

Out of 16 images from each model, images with side and front view where the features of the car are clearly visible such as full car body, logo, headlight and tyre are selected. Two brands of car, GMC and Honda are chosen from Carvana dataset which consists of 1,952 images. The images are cropped to remove the background and these images along with the text descriptions are used to train both generator and discriminator. For the implementation of GAN to generate car images the model is trained with GMC car images and Honda Accord car image as illustrated in Figs. 3 and 4.

The captions for these images are generated using a python script but the colour of each car is manually included. This is different from the previous approach where one of the visual features, that is, the colour of the car is included. The text description contains only the visual attributes. According to the paper [4] figures of speech, naming the species and description of background should be avoided. The five-line text description generated is illustrated in Fig. 5.

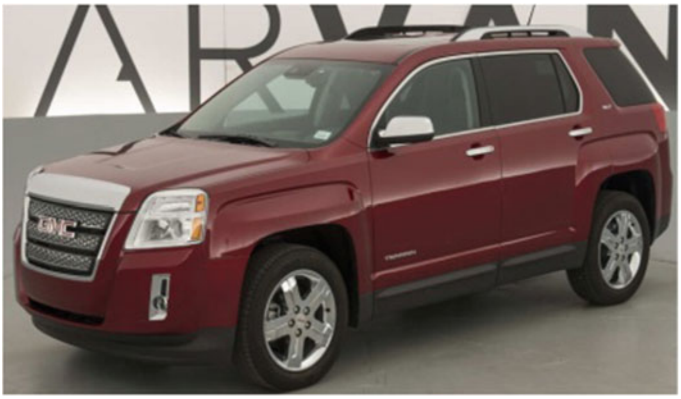


Fig. 3 Car sample of GMC [4]

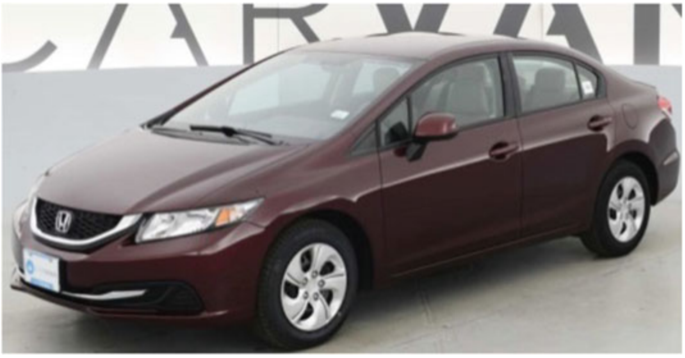
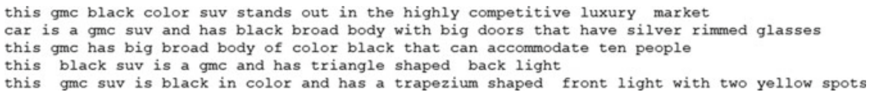


Fig. 4 Car sample of Honda [5]



```
this gmc black color suv stands out in the highly competitive luxury  market
car is a gmc suv and has black broad body with big doors that have silver rimmed glasses
this gmc has big broad body of color black that can accommodate ten people
this  black suv is a gmc and has triangle shaped  back light
this  gmc suv is black in color and has a trapezium shaped  front light with two yellow spots
```

Fig. 5 Detailed five lines of text descriptions

4.3 Dataset 3

For the interpolation of the colour of car images, the model is trained with 5,802 images with corresponding text descriptions. The dataset contains images of different brands of the car such as Acura, Audi, Buick, Chevrolet, Chrysler, Ford, GMC, Dodge, Hyundai, Jeep, Lexus, Honda, Mini, Toyota, Mitsubishi and Nissan. For training the text-adaptive generative adversarial network each image is described in ten lines. In this approach, a detailed text description is given, which includes more visual attributes such as the colour of the car, type of car, logo, grille type, rim, spokes and mirror colour. The sample of the Audi image and text description for an Audi car is illustrated in Figs. 6 and 7 (Table 1).

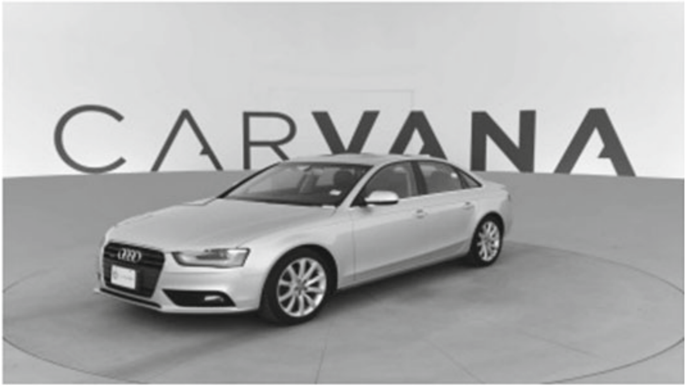
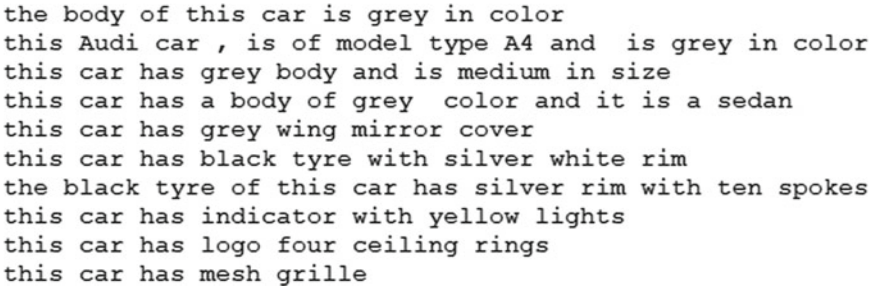


Fig. 6 Car dataset sample of Audi [1]



```
the body of this car is grey in color
this Audi car , is of model type A4 and  is grey in color
this car has grey body and is medium in size
this car has a body of grey  color and it is a sedan
this car has grey wing mirror cover
this car has black tyre with silver white rim
the black tyre of this car has silver rim with ten spokes
this car has indicator with yellow lights
this car has logo four ceiling rings
this car has mesh grille
```

Fig. 7 Text description of Audi car

Table 1 Dataset summary

Dataset	Number of images	Number of lines in the text description
Dataset 1	6,573	3 lines
Dataset 2	1,952	5 lines
Dataset 3	5,802	10 lines

5 Results and Discussion

In this experiment, the GAN model is able to generate the images according to the given text descriptions successfully. The model has been trained with car images using Carvana dataset and corresponding text description for each image. The major work in this experiment was to create a text description for each image. The images and metadata text files are processed to prepare the data for training generative adversarial network. The pre-processing steps are as follows:

1. The images are loaded and resized to 64 × 64 format.
2. All the text files are loaded, processed and vocabulary file is created.
3. The images and captions are divided into training and testing datasets and are saved in pickle format.

The conditional generative adversarial network can be trained to process text and images together and the discriminator is trained to classify pairs as fake or real. The discriminator observes real images with matching text and synthetic images with arbitrary text inputs. The model is trained for 600 epochs with a batch size of 115 for each epoch.

In the first attempt, the model was trained with Dataset 1 (data collection Sect. 4.1) where three lines of caption per image are given. The generated output image is shown in Fig. 8. The image generated is of low resolution as the visual attributes are not included in the text description of the image. For testing, one-line text input is given to generate the image.

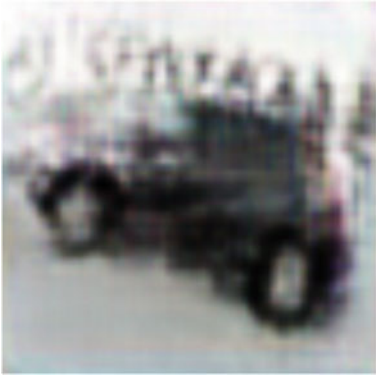
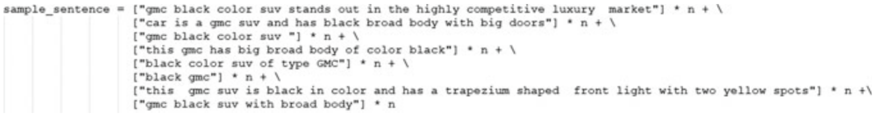


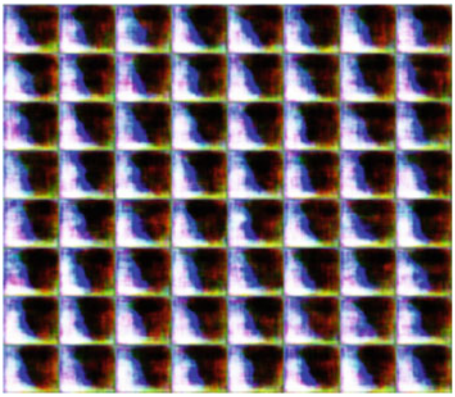
Fig. 8 Sample image generated [1]



```
sample_sentence = ["gmc black color suv stands out in the highly competitive luxury  market"] * n + \
                  ["car is a gmc suv and has black broad body with big doors"] * n + \
                  ["gmc black color suv "] * n + \
                  ["this gmc has big broad body of color black"] * n + \
                  ["black color suv of type GMC"] * n + \
                  ["black gmc"] * n + \
                  ["this  gmc suv is black in color and has a trapezium shaped  front light with two yellow spots"] * n +\
                  ["gmc black suv with broad body"] * n
```

Fig. 9 Text description

Fig. 10 Epoch 0



Input—This is the first-tilt-view of brand 0cdf5b5d0ce1 car.

In the second attempt as described by Reed et al. [8], the number of captions was increased and modified to include a rich visual description of the image in Dataset 2 (data collection Sect. 4.2). During training, the following input was provided for sample sentences that are used for the generation of images at the end of every epoch. The sample sentence list takes eight input sentences as illustrated in Fig. 9. For each of the input sentence, the top eight probable images are extracted from the model. Hence an 8 × 8 image matrix is saved after each epoch. The figure illustrated in 10 is of 0th epoch which is the first set of images created by a generator using initial random noise with five lines of caption per image during training.

The output of final epoch as per the given eight lines input is illustrated in Fig. 11.

The following results show the various test run by providing one-line text input to generate the images.

Input—This red colour Honda Accord sedan has a medium-size body that can accommodate four people.

The generated output image is illustrated in Fig. 12.

Input—This blue sedan has a triangle-shaped backlight.

The generated output image is illustrated in Fig. 13.

Input—This red SUV is a GMC and has a triangle-shaped backlight.

The generated output image is illustrated in Fig. 14.

The above image generated is not matching the given text input, that is, the interpolation of the colour of car images failed. To interpolate the colours, TAGAN is

Fig. 11 Epoch 600

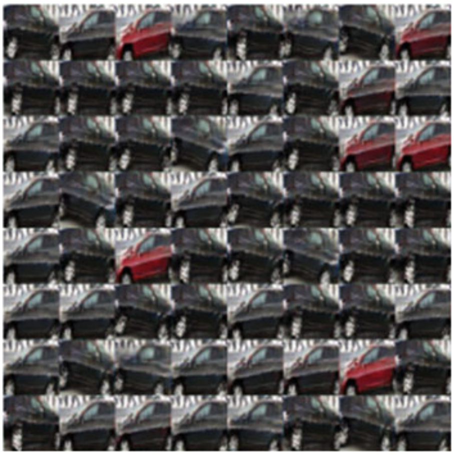


Fig. 12 Red colour Honda Accord [1]

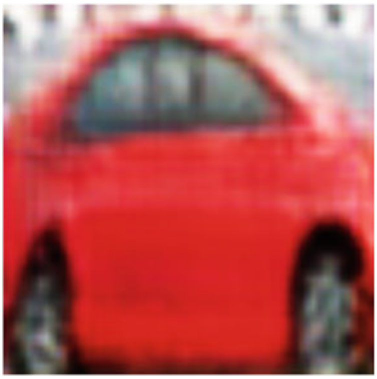


Fig. 13 Blue colour Honda Accord [2]

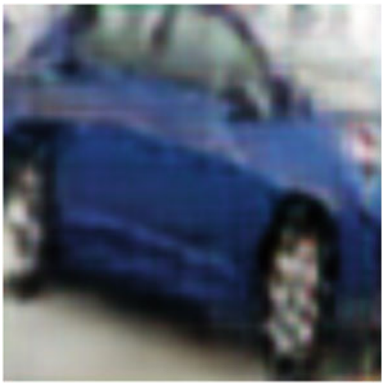


Fig. 14 Sample Honda Accord [3]

implemented and the results are shown in the below section. In the third attempt, the Dataset 3 (data collection Sect. 4.3) is used for training and testing the text-adaptive discriminator model to interpolate the colour of car images.

The selected image as shown in Fig. 15 is intended to be edited.

The generated images for the given original image and text are shown below:

Input—car with a white body.

The generated output image is illustrated in Fig. 16.

Input—car with a blue body.

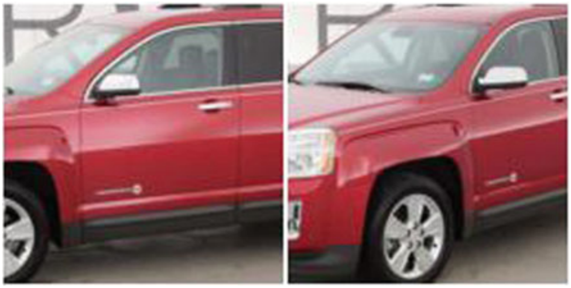


Fig.15 Original image [1]

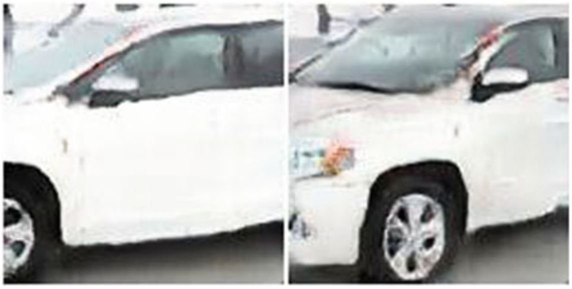


Fig. 16 White colour car [2]

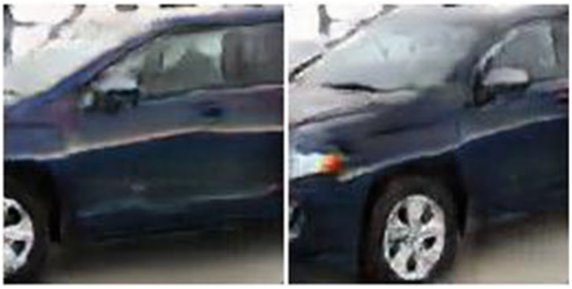
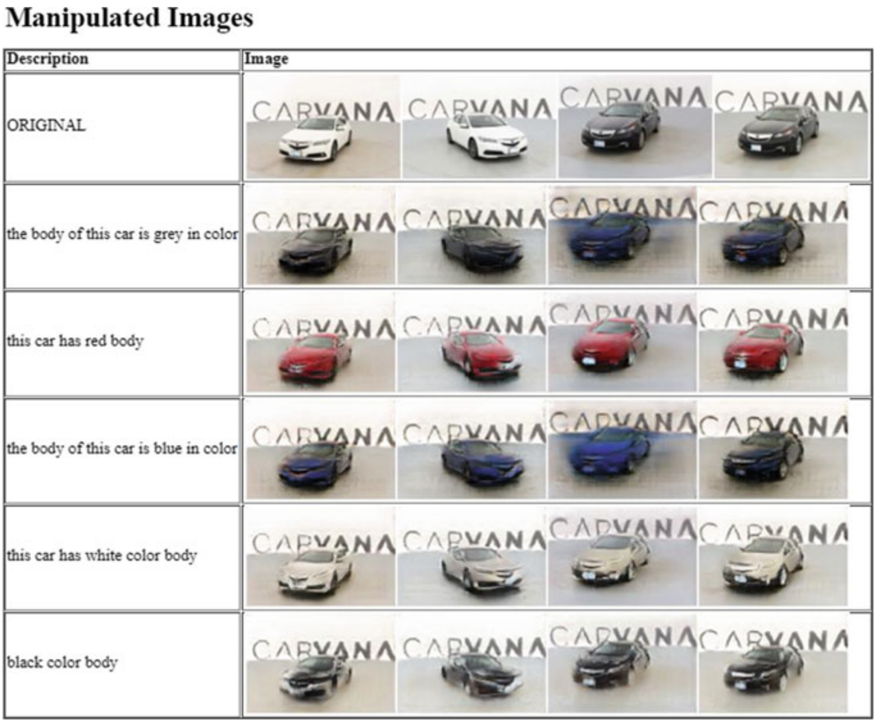


Fig. 17 Blue colour car [3].

The generated output image is illustrated in Fig. 17.

The results generated by using 5000 car images of different brands for training the TAGAN model are illustrated in Fig. 18. These results are generated after 100 epochs.



Manipulated Images

Description	Image
ORIGINAL	
the body of this car is grey in color	
this car has red body	
the body of this car is blue in color	
this car has white color body	
black color body	

Fig. 18 Manipulated images

6 Conclusion

As witnessed in the design and implementation of two approaches for text-directed generation/editing of the images, it can be concluded that the experiment is successful in achieving the targeted application feature.

In the beginning, different approaches for achieving the required results were evaluated, and the generative adversarial network is selected due to its self-validation of the functionality via discriminator.

The GAN model is able to generate images based on the given description successfully. The verification of generated images could be done by looking at the progress of images generated through the successive epochs. The interpolation of the colour of car images is also achieved using text-adaptive generative adversarial network which is able to generate the car images of the given colour.

7 Future Work

The images generated from our GAN-CLS experiment are of low resolution. The advanced alternative approaches for improving the quality of the images have been analysed. An advanced approach called Stack GAN which has two phases of GAN to improve the resolution of images and make the images realistic could be used. The current work can be extended to implement this approach to improve the quality of images generated by GAN-CLS. In the current TAGAN approach the basic interpolation of colours is achieved, and this could be extended to improve fine-grained interpolation of colours and other visual features. The limitation of this system is that it accepts the input only in the form of text descriptions; this can be extended in future to build the system that would accept the input from the user in the form of speech, that is, the user can give input by speaking to the system.

Compliance with Ethical Standards.

All author states that there is no conflict of interest.

Humans/animals are not involved in this work.

We used our own data.

References

1. Reed S, Akata Z, Yan X, Logeswaran L, Schiele B, Lee H (2016) Generative adversarial text to image synthesis. arXiv:1605.05396
2. Farhadi A, Endres I, Hoiem D, Forsyth D (2009) Describing objects by their attributes. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE, pp 1778–1785
3. Huang H, Yu PS, Wang C (2018) An introduction to image synthesis with generative adversarial nets. arXiv:1803.04469
4. Mao J, Xu W, Yang Y, Wang J, Huang Z, Yuille A (2014) Deep captioning with multimodal recurrent neural networks (m-rnn). arXiv:1412.6632

5. Reed S, Akata Z, Lee H, Schiele B (2016) Learning deep representations of fine-grained visual descriptions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 49–58
6. Viswanathan A, Mehta B, Bhavatarini MP, Mamatha HR (2018) Text to image translation using generative adversarial network. In: 2018 international conference on advances in computing, communications and informatics (ICACCI). IEEE, pp 1648–1654
7. Nam S, Kim Y, Kim SJ (2018) Text-adaptive generative adversarial networks: manipulating images with natural language. In: Advances in neural information processing systems, pp 42–51
8. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Advances in neural information processing systems, pp 2672–2680

Body Mass Index Implications Using Data Analysis in the Soccer Sports

Akash Dasmondal and P. K. Nizar Banu

Abstract Soccer is considered among the most popular sports in the world among the last few years. At the same time, it has become a prime target in developing countries like India and other Asian countries. As science and technology grow, we can see that sports also grow with science, and hence technology being used to determine the results sometime or sometimes it is used to grow the overall effect. This paper presents the attributes and the qualities which are necessary to develop in a player in order to play for the big-time leagues called Premier League, La Liga, Serie A, German Leagues and so on. Simple correlation and dependence techniques have been used in this paper in order to get proper relationship among the attributes. This paper also examines how the body mass index plays an effect on the presentation of soccer players with respect to their speed, increasing speed, work rate, aptitude moves and stamina. The point is likewise to discover the connection of the above credits concerning body mass index. As in universal exchange, football clubs can profit more in the event that they have practical experience in what they have or can make a similar bit of room to maneuver. In a universe of rare assets, clubs need to recognize what makes them effective and contribute in like manner.

Keywords Data analysis · Soccer · BMI · Co-relation · Soccer · Sports

1 Introduction

The affection for the professional game across the world is profoundly established inside the psyches of the individuals of India. With Manchester United, Chelsea F.C., Arsenal F.C., and Liverpool F.C. as the main establishments in the rundown of most popular football crews in India, British impact is still in place [1]. With

A. Dasmondal · P. K. Nizar Banu (✉)
Department of Computer Science CHRIST (Deemed to be University), Bangalore, Karnataka, India
e-mail: nizar.banu@christuniversity.in

A. Dasmondal
e-mail: akash961996das@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_22

the coming of advancement and globalization in the new monetary approach of 2015, the administration of India has concocted various plans to improve each part of the economy including sports. As, on one hand, India was called a monetary superpower and quickest developing economy post advancement in 1991; then again, it confronted extreme Criticism for its disintegrating sports execution until ongoing Commonwealth Games and Olympics in 2016. For example, India positioned fourth in the 1956 Olympics in Football and 165th in 2012 [2].

Somewhat, proficient football has turned out to be simply another part of the corporate world, and, thus, it has lost a lot of its neighborhood appeal and a considerable lot of its inalienably geological qualities. In any case, the area still appears to have a significant influence in present-day football. Undoubtedly, the geographic premise of this game is generally obvious at an assortment of geographic scales: worldwide, national and neighborhood, as it is exhibited by the global challenge among nations, the vocational examples of expert football crews, and fans' connection to groups in explicit areas, and so on [3].

Club achievement in games is a relative term. While a few groups are continually attempting to win the English Premier League, Bundesliga, La Liga and others are content with simply remaining in the challenge. Before each season begins, sports specialists investigate potential overachievers and underachievers, considering cash spent, squad experience, and other quantitative and subjective recognitions. As in worldwide exchange, football clubs can profit more on the off chance that they work in what they have or can make a similar bit of leeway. In a universe of rare assets, clubs need to distinguish what makes them effective and contribute appropriately; in playing better protectively or potentially upsettingly, in choosing players cautiously, in utilizing keen administration draws near, and so on.

We should remember that having a similar preferred position does not prompt achievement, notwithstanding a bounty of assets; additionally, not every person needs to have some expertise one way, similarly as rich nations don't represent considerable authority in a couple of items. Be that as it may, amateur clubs, as amateur nations, need to utilize the assets they have [4]. Sports are no more for spoilsports and Indian guardians are supporting their kids' yearnings to seek after football as a vocation more than ever. Things and market powers around the game are changing and there is a ton of positive buzz in India around soccer. The Indian football story has recently started with the cooperation of industry, the appearance of worldwide accepted procedures and professionalization of game. There are detours; however, there is a bound together assurance to beat these difficulties. This publication titled Indian Football Scenario displays an outline of market powers and open opinions around the football in the nation with devoted sections to activities by national and global pinnacle bodies, other than vital speculations by corporate from India and abroad, including driving head alliance clubs [2]. This production is a guide for speculators and global specialists to turn into the partners in developing Indian football division and exploit in this long haul improvement process as there are a large number of chances for specialists, instructive gatherings and enterprises to have a significant and enduring effect on the eventual fate of Indian football [2].

Currently, the ascending of overweight and weight is a noteworthy issue of general wellbeing which brings about various sicknesses and wellbeing aggravations. Different examiners have affirmed that around 6% of the absolute general wellbeing across worldwide is identified with physical dormancy, overweight and corpulence. Presently we can see it is becoming a major issue to understand sportsmen's life which effects themselves and others in sports by reducing their performace causing their auxiliary way of life leading to unproductive performance [11]. They are particularly going into the computerized world as opposed to the ground. So, it is a typical issue with the present understudy's life. Anthropometry is a technique for the appraisal of their physical make-up, for example, body mass index, player attributes such as speed, sprint, and stamina and ball control. As per the dictionary, it is the estimation of the size and extents of the human body. Body mass index is generally the most utilized strategy to check the dietary status in grown-ups. It is characterized as an individual's load in kilograms isolated by the square of the individual's stature in meters (kg/m^2). As indicated by the World Health Organization for 20 years of age, grown-ups typical body mass index is 18.5–26, underneath them, are underweight and over them are overweight and hefty. The body fat.% expansion is in direct connection to body mass index esteems, particularly among youngsters and teenagers; in any case, there are sure examinations that have discovered a few deviations in body piece esteems in that regard [5].

In order to justify the above, things like body mass index analysis and correlation techniques and the developments of the sports culture in developing countries like India and other countries at grassroots levels develop into a better level in sports [12].

2 Background

Achievement in football relies upon many interrelated angles—funds, ability choice, the executives draws near, preparing, and so forth, and keeping in mind that assets are restricted, groups need to settle on ideal choices about how to allow their assets to expand their odds of progress: Should they practice and increase a similar bit of leeway through their cautious or hostile style [4].

Additionally, the examination demonstrates that money-related assets impact the likelihood of achievement for "good" clubs, while they are absolutely inconsequential for base positioning clubs [4]. Thinking about the discoveries of the present examination, it was inferred that the predominance of overweight in futsal players ought to be a significant worry for professionals working in this group activity. Moreover, it was seen that expanded body mass index was identified with diminished scores in key game-related physical wellness parameters, for example, dashing, bouncing, and anaerobic power, particularly in youthful futsal players. In this manner, enhancing the body mass index ought to be considered as a preparation and sustenance objective so as to improve sports execution [6]. The speculations that remote elements have made in Indian football so far are empowering and genuinely necessary. Putting resources into the eventual fate of the game in India with a reasonable vision and a guide is the thing that will drive Indian football forward, ideally helping the nation satisfy its colossal football potential [8]. From structure quality youth advancement foundations, preparing offices, and arenas, to cooperating with Indian football alliances and individual I-League clubs for specialized and promoting coordinated efforts, there are various manners by which expert clubs and other remote elements can help become the delightful game on the planet's second-most crowded nation [7].

The Indian Super League was established in 2013 out of a push to make football a top game in India and to build the degree of Indian football around the world. The alliance at present highlights eight groups from around India, with four of the main five urban communities by the populace, spoken to in the class [7]. Not at all like other football associations from around the globe, the Indian Super League is the one of only a handful couple of classes to not utilize the advancement and transfer framework; however, rather it is an establishment framework along the lines of the Twenty20 cricket alliance, the Indian Premier League, Major League Soccer of the United States and Canada, and the A-group Australia. Since the debut season, two groups have been delegated to the bosses of the Indian Super League.

Similarly, as nations can have certain favorable circumstances in global exchange, football clubs can accomplish a near preferred position including certain parts of the game, for example, subtleties of guarded and hostile play, and components, for example, ownership, set-piece scoring, and counter-assault play [9]. Sportswriters discuss accomplishing upper hand in games; some take a gander at various parts of the game as a model for making near as opposed to an upper hand. Implicit learning is a method for making an upper hand in the National Basketball Association. They found that there is a positive connection between shared group involvement and group execution, showing that playing together for a while can make a preferred position for clubs. One of the primary investigations around there was that those nations may represent considerable authority in games the way they have some expertise in worldwide exchange, by making a near bit of leeway in a particular game. Also, these creators utilized the idea of an uncovered similar bit of latitude to look at the Olympic Games, examining both specializations in games and the idea and found that high-pay nations practice less [4].

Football is a major business; we just need to take a gander at the wages of the top players to see the sum of cash in the game. Football is likewise a fascinating point for financial specialists since it excites feelings and interests that regularly don't fit into perfect monetary models. Football is one of the most mainstream sports with an extraordinary onlooker base all through the world [10]. In the United States, be that as it may, in spite of the expanding prevalence of football as a game to play among youth, the observer base is still a long way behind that of different games, for example, American football, and baseball. Subsequently, Major League Soccer (MLS) is yet considered as a second-level game alliance. In any case, MLS has been attempting to extend its observer base through different promoting means, for example, supporting numerous grassroots competitions across the country to spike enthusiasm for the group among football members. In any case, a financial condition that the USA is confronting may bewilder proficient football participation. The development in the prevalence of English football has been founded on the capacity of the Premier League, and the Clubs need to understand the estimation of the football rivalry. Every one of the primary income streams, counting ticket deals, stock, sponsorship, promoting and extra employments of the arena, just as the clearance of communicating rights, gain from a solid group rivalry that arrives at a wide crowd [3].

3 Dataset Description

The purpose of this paper is to make use of player qualities and player movement. Considering the contrasting destinations of every one of these two views such as BMI and the other attributes necessary for the players to perform at the best level, there is a necessity for two, as portrayed underneath.

1. A CSV record involves player properties, which incorporates player stature and weight. Going ahead, the report will allude to this document as Player Attribute Dataset. This dataset is constrained to one year of player information and will help make a model perform a prescient arrangement.
   - This dataset contains various features (characteristics) and 18,147 records (players).
   - This dataset is sourced from Kaggle Repository and also with respect to the Franchise EA Sports Game called FIFA.

4 Methodology

4.1 Data Preprocessing

It is a significant advance in the information mining process. The expression "trash in, trash out" is especially pertinent to information mining and AI ventures. Information gathering techniques are frequently inexactly controlled, coming about in out-of-run esteems; unimaginable information mixes missing qualities, and so on. Breaking down information that has not been painstakingly screened for such issues can deliver misdirecting results. In this manner, the portrayal and nature of information are most important before running an examination. Frequently, information preprocessing is the most significant period of an AI venture, particularly in computational science.

Information arrangement and separating steps can take a lot of preparation time. Information preprocessing incorporates cleaning, instance choice, standardization, change, include extraction, and choice, and so forth. The result of information preprocessing is the last preparing set. Information pre-preparing may influence how the results of the last information handling can be translated. This perspective ought to be deliberately viewed as when the translation of the outcomes is a key point, such as in the multivariate handling of precious information. In the dataset, we have heights

Table 1 Classification of BMI

BMI	Classifications
<18.5	Underweight
18.5–25	Normal
25–30	Overweight
>30	Obese

Table 2 BMI distribution of 18,147 players

BMI	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	34
No. of players	2	7	67	304	1324	2789	5021	4942	2505	908	212	47	12	5	1	1

and weight in inches and feet and weights in kg so to convert it into cm and kg to find the BMI.

5 Results and Discussion

5.1 BMI Classifications

See Table 1.

Table 2 also tells how many players have been distributed in the respective BMI from the lowest to the highest value.

Table 1 classifies the BMI distribution according to world standard which is an important attribute for our discussion.

5.2 BMI and Player Positions of 18,147 Players

Table 3 shows the position distribution in each position of football players. It shows the lowest BMI in each position and also the highest BMI. The mean values show that all the players are normal and fit.

6 Bar Chart of BMI

Fig. 1 shows the overall BMI distribution of the 18,000 players. It is also showing that how the BMI is distributed among the players and that majority of the players are fit and the most common BMI is 22 and 23, which shows that they are normal.

Table 3 BMI distribution according to the position

Position	Playing sub position	No. of counts	Mean of BMI	Min	Max
Goalkeeper	Goalkeeper (GK)	2025	22.63	16.00	29.00
Defense position	Wing-backs (LB, RB, LWB, RWB)	2778	22.35	17.00	28.00
	Central defenders (CB, RCB, LCB)	3088	22.48	16.00	29.00
Midfield	Defensive midfielder (CDM, RDM, LDM)	1439	22.40	18.00	27.00
	Midfielder (CM, RCM, LCM, RAM, LAM, CAM)	3180	22.19	18.00	28.00
Forward	Forward (ST, CF, RF, LF, LS, RS)	2667	22.56	17.00	34.00

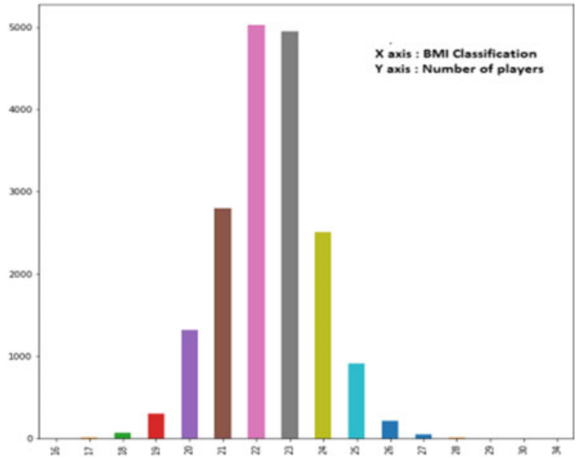


Fig. 1 Distribution of BMI using bar chart

There are other players also which are having BMI other than the mentioned one; they are also overweight as well as obese and underweight also.

Goalkeeper: Fig. 2 shows that majority of the BMI is normal, and none of the players are obese; only handful of players are underweight and overweight.

Center Back: Fig. 3 shows that almost all the players are fit and normal with their BMI and the fact is only little number of players are having the BMI in the overweight category.

Wing-Backs: In Fig. 4 we can see that majority of players have BMI as 22–23 but few numbers can be seen that with more than that also though they are wingbacks, they are fit as players.

Defensive Midfield: From the Fig. 5 we can see that in defensive midfield we often see that in this we can see that BMI is divided in all the category of classification.

Fig. 2 Goalkeeper

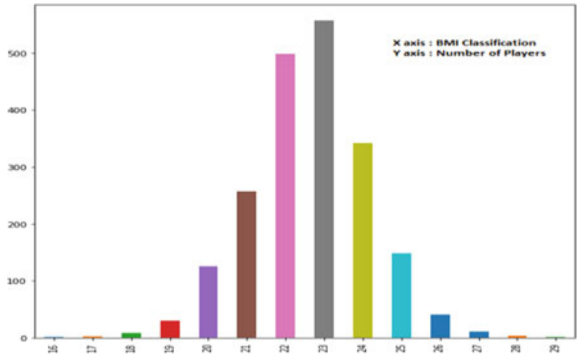


Fig. 3 Center back

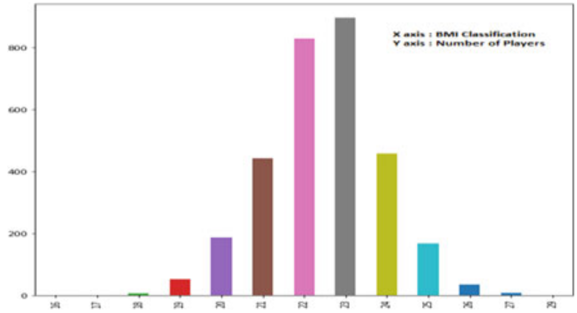


Fig. 4 Wing-back

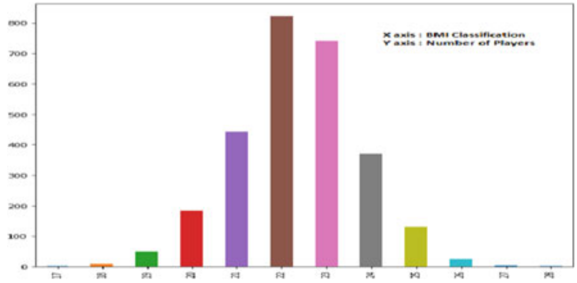
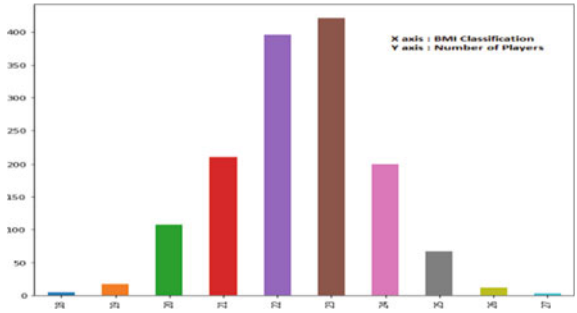


Fig. 5 Defensive midfield

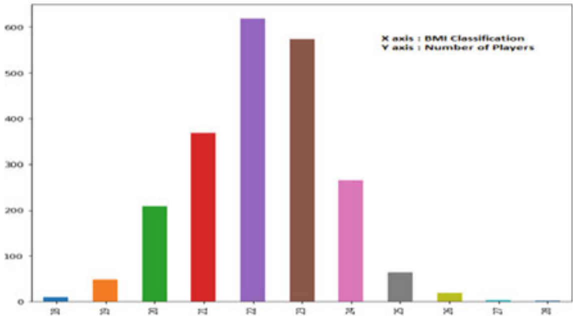


The highest BMI is 23 and also the lowest being observed is 18 and 27 with least count.

Central Midfield: From Fig. 6 in the case of central midfield it has been observed that the BMI is distributed in every aspect of classification and also the fact that none of the players are obese and the highest count with the BMI in midfield is 22 and lowest is 28.

Attacking Midfield: Fig. 7 shows the BMI distribution of attacking midfielder. We can see that there is a neck-to-neck competition in the BMI number (22–23) and most of the players have BMI in this category and few are in the underweight and overweight category.

Fig. 6 Central midfield



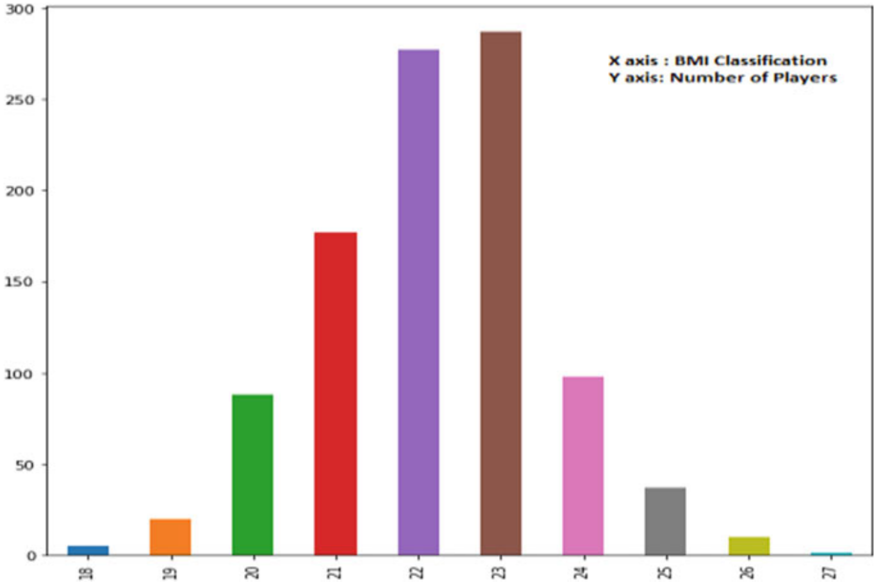


Fig. 7 Attacking midfield

Fig. 8 Winger

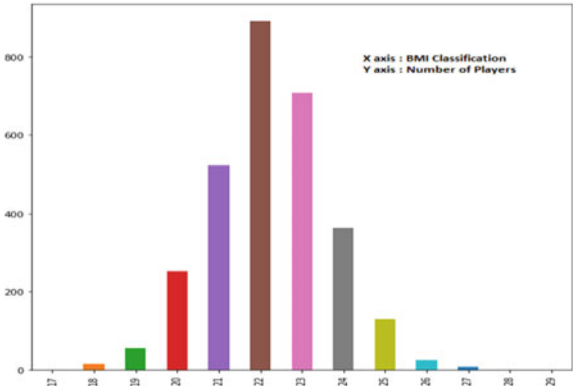
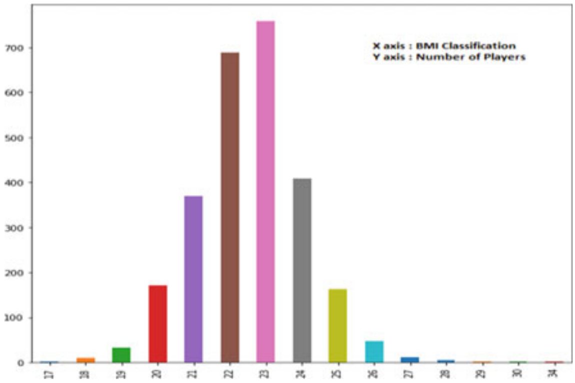


Fig. 9 Strikers



Wingers: From Fig. 8 we can conclude that in football wingers are known as game changers and it has been seen that 22 is the highest BMI in the distribution, but many of the players are having BMI with overweight as per the classification.

Strikers: From the Fig. 9 we can see that the strikers are classified into the lowest to the highest BMI of 17–34 but the majority are in 23 or 22 and others were below than that, which shows there are few strikers who are obese but they proved to be good strikers.

7 Inferences of the Above Graphs

From the above graphs, we can observe how the distribution of the BMI is shown position-wise in a football, and maximum and minimum count of the BMI is also given over there. This shows that how BMI is the crucial factor in this kind analysis and the most common BMI is 22–23 but there are players who have more BMI as compared to these two.

These results can be used by the academies or the grassroots-level schools where these sports are considered as the development of sports and are targeted as the important aspect which will lead to developing players with respect to BMI and other qualities which are necessary for the development of the sports and also to the future of sports.

8 Attributes Necessary for a Player to Play in a Particular Position

Forward Position: From Table 4 the forward position is the top attribute which is required by a striker to play at best level. This positioning is the most important for them and finishing.

Table 4 Forward position

S. No.	Player attributes	Correlation values
1	Positioning	0.909207
2	Finishing	0.902016
3	Ball control	0.901148
4	Shot power	0.874362
5	Reactions	0.869161
6	Composures	0.831669
7	Volleys	0.830210
8	Short passing	0.820211
9	Dribbling	0.804175
10	Long shots	0.794537

Table 5 Attacking wingers

S. No.	Player attributes	Correlation values
1	Ball control	0.917393
2	Dribbling	0.901025
3	Short passing	0.876960
4	Reactions	0.867029
5	Positioning	0.847970
6	Composure	0.826851
7	Vision	0.821689
8	Long shots	0.777837
9	Crossing	0.772218
10	Curve	0.726345

Table 6 Attacking midfielder

S. No.	Player attributes	Correlation values
1	Ball control	0.920842
2	Dribbling	0.872275
3	Vision	0.865493
4	Positioning	0.860432
5	Short passing	0.854687
6	Reactions	0.853099
7	Long shots	0.815866
8	Finishing	0.804176
9	Composure	0.792721
10	Crossing	0.791225

Attacking Wingers: Table 5 says that the wingers who are termed as magician need these attributes to develop themselves. In order to become a good player to play at wing position ball control and dribbling should be at the top level and rest depends on the top attributes.

Attacking Midfielder: Table 6demonstrates that the players who want to play at the attacking midfielder just behind the striker need to develop these attributes which are important for them in order to develop themselves, like ball control vision and positioning which are very much important in this position.

Central Midfielder: In Table 7 the attributes of central midfielder are presented. The most important attribute is ball control. If the player has that then other attributes can be developed easily with that. Long passing and short passing are the essential ones in this position.

Defensive Midfielder: Table 8 shows that the defensive midfield is an important position and the key attribute is short pass because they have to play with defense

Table 7 Central midfielder

S. No.	Player attributes	Correlation values
1	Ball control	0.919851
2	Short passing	0.900168
3	Long passing	0.873657
4	Reaction	0.870867
5	Vision	0.868832
6	Dribbling	0.839227
7	Composure	0.823537
8	Long shots	0.779927
9	Crossing	0.745074
10	Positioning	0.743965

Table 8 Defensive midfielder

S. No.	Player attributes	Correlation values
1	Short passing	0.854465
2	Reactions	0.852749
3	Long passing	0.834873
4	Ball control	0.833840
5	Composure	0.810332
6	Interceptions	0.805374
7	Standing tackle	0.769761
8	Vision	0.712175
9	Marking	0.693849
10	Dribbling	0.663375

as well as center midfield and wingers. One has to master the short pass attribute in order to develop the others.

Wing-Backs: Table 9 proves that in wing-backs short passing crossing is necessary because they play on either side of flanks of the pitch, so they support the wingers and also the defense. Tackling is also important as they are defenders.

Center Backs: From Table 10 we can say that central back are the heart of defense line, so in order to protect that line they need to be good in judging the game and also tackling is mandatory for them; marking and all other attributes are add-on for them.

Goal Keeper: Table 11 is about playing in goal keeper position. It requires that only a few attributes are important for being a keeper; in that reflexes are important as they have to dive as required and also reaction is important, but all these are important ones to pursue the role of being a keeper.

Table 9 Wing-back

S. No.	Player attributes	Correlation values
1	Short passing	0.885130
2	Crossing	0.878275
3	Reactions	0.876291
4	Ball control	0.876159
5	Marking	0.839429
6	Composure	0.814242
7	Standing tackle	0.808585
8	Sliding tackle	0.800808
9	Interceptions	0.797898
10	Dribbling	0.788264

Table 10 Center back

S. No.	Player attributes	Correlation values
1	Standing tackle	0.909005
2	Interceptions	0.901862
3	Marking	0.872058
4	Sliding tackle	0.871751
5	Reactions	0.861155
6	Heading accuracy	0.832236
7	Composure	0.799573
8	Aggression	0.729125
9	Short passing	0.724332
10	Ball control	0.703714

Table 11 Goalkeeper

S. No.	Player attributes	Correlation values
1	GK reflexes	0.941000
2	GK diving	0.937871
3	GK positioning	0.935355
4	GK handling	0.918714
5	Reactions	0.857489
6	GK kicking	0.774512
7	Composure	0.547686

9 Inference of the Above Tables

The above table shows that the attributes which are the topmost and high priority which a player has to focus in order to develop themselves into a better player in-game and these attributes are defined with respect to BMI which has been discussed earlier; and also the attributes play a necessary role and development becomes more effective if they focus on these attributes. The development will be at the optimum and effective level if many aspects can be improved, such as performance opportunities quality and at last achievements.

10 Conclusions

The purpose of this analysis is how the body mass index is distributed among world-class players and how the attributes are important in specifying the roles of a player who played the position. The body mass index distribution is classified position-wise to show which BMI is most suitable to a player in the world of football. It is also observed that some players also have BMI which is overweight as well as underweight, but they are fit as players, but majority of the professional players are fit and have normal BMI.

The graphical patterns of the BMI show a lot of information that is required. It portrays the importance of the BMI value and additional attributes are found to develop along with BMI to saturate the effectiveness of BMI. Then we have the dependence value as correlation values which can be used in order to see which attributes are necessary for a player to master the position. Not all attributes are important, but few are the most important attributes, but if these few of them are developed then rest can be developed using these attributes.

Even though these attributes can be deployed at the grassroots level of football, in order to develop the players from the root level so that in future they become the good players and technically strong players, the proper nurture of the players at the early age will be an additional asset to the future players, which will be more advanced and opportunistic These results can be used in areas or nations where the sports is under development or in good future-like countries such as India where these approaches can be taken in the form of trial and error method to test in order to get better results of these sports. Thus, data analysis can be used in sports in order to find the hidden areas, and also these areas can be explored using analysis techniques to get proper detailed results to the development of the sports to contribute to the future of Indian sports.

References

1. International football business institute, India a Future giant of the game
2. Mantri A (2013) India football: the Rising billion
3. Singh R, Singh Y, Mantri A, Grover V, Bhatnagar A, Sharma V, Economıcs of football business around the world
4. Georgıevskı B, Labadze L, Aboelsoud ME (2019) Comparative advantage as a success factor in football clubs: evidence from the English Premier League (EPL)
5. Dhar A (2019) Study of body composition among university level male football players of northeastern region: a comparative study
6. Mulholland J, Jensen ST (2014) The draft and career success of tight ends in the National Football League
7. Sharma AS (2017) A critical review on the Indian super league (ISL). IJPESH 4(2):257–261
8. Bissacco CA (2019) Analysis of the efficiency of the forwards in the Big 5
9. Ramaswamy A, Rajan S (2018) Soccer fan behavior in India
10. Pontaga I, Žīdens J (2011) Estımatıon Of Body Mass Index In Team Sports Athletes
11. Wragg M, World football report
12. Nikolaidis PT, Chtourou H, Torres-Luque G, Rosemann T, Knechtle B (2019) The relationship of age and BMI with physical fitness in futsal players

Biogeography-Based Optimization Technique for Optimal Design of IIR Low-Pass Filter and Its FPGA Implementation

K. Susmitha, V. Karthik, S. K. Saha, and R. Kar

Abstract A bio-inspired meta-heuristic biogeography-based optimization (BBO) algorithm, which imitates the migration and mutation processes of different species according to the habitat features, is used in this paper in order to get the optimal coefficients of an infinite impulse response (IIR) low-pass filter (LPF) of order 8. BBO mainly depends on the immigration rate (IR) and emigration rate (ER), through which the searching efficiency is enhanced. The simulation results have shown a better performance in terms of stopband attenuation, transition width, passband ripples (PBR), and stopband ripples (SBR). The optimized coefficients are utilized for the implementation of the IIR filter in the Verilog hardware description language (HDL) with the field-programmable gate array (FPGA).

Keywords BBO · Frequency response · Convergence · LPF · Magnitude response · Error fitness function · IIR filter · FPGA · Verilog HDL

1 Introduction

As the advancements are occurring in science and technology, the demand for signal processing is increasing enormously, which leads to the huge requirement of the filters. Signal processing involves in the assaying of the signals along with the reduction of noise in the raw signals. Generally, the principle of filtering is to exempt the

K. Susmitha (✉) · V. Karthik · S. K. Saha
Department of Electronics and Communication Engineering, NIT Raipur, Raipur, India
e-mail: susmithakomanapalli@gmail.com

V. Karthik
e-mail: karthik.valupadasu@gmail.com

S. K. Saha
e-mail: namus.ahas@gmail.com

R. Kar
Department of Electronics and Communication Engineering, NIT Durgapur, Durgapur, India
e-mail: rajib.kar@ece.nitdgp.ac.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_23

unnecessary components of the signal and allowing the desired part of the signal to pass. Depending on the physical structure and components used, filters are sorted as analog filters and digital filters. The electrical components like capacitor, inductor, and op-amp are utilized in the construction of the analog filters. The main drawback of the utilization of these components is they tend to high tolerance sensitivity, less accurate, liable to thermal drift, and large physical size. Obversely, digital filter accomplishes computational operations on a sampled, discrete-time signal to attain the desired features by utilizing the specifically designed digital signal processor (DSP) [1].

Digital filters are categorized into two types, namely: finite impulse response (FIR) and infinite impulse response (IIR) filters. FIR filter is non-recursive in nature due to the output is determined only from the present and past input values; on the other hand, IIR filter is recursive in nature as its output not only is determined from the former inputs but also from the former outputs; because of this, a huge memory is required in comparison with FIR filter [2]. As the size of memory is small in the FIR filter, its realization is simpler compared to the IIR filter. Obversely, IIR filter meets the certain specified features such as sharp transition width, high stopband attenuation, small PBR, and SBR. These specifications are ensured to be met for the IIR filter with the least order of the filter compared with FIR filter; as a result of this, appropriately designed IIR filter gives the frequency response (FR) near to the ideal value. Digital filters are utilized in various biomedical applications, namely in the medical equipment such as electrocardiogram by using differentiator-based preprocessor [3].

In order to attain the desired characteristics from the filter, optimization techniques are employed. Optimized design of the filters has better performances and are accepted in most of the fields of science and technology. For designing the digital filters, several constraints are to be followed to satisfy the desired design criteria:

- Filter order has to be minimum for the prescribed application;
- Enhancement in the computational speed of the filter;
- Stable filter.

There are several optimization techniques present in order to optimize the given problem; the heuristic search techniques are reported in the literature: Artificial bee colony algorithm [4], seeker optimization algorithm [5], genetic algorithm [6], harmony search [7], particle swarm optimization [8], gravitational search algorithm [9], and differential evolution [10]. In this paper, the filter coefficients are optimized by the BBO [11–14] which has a faster convergence to the optimal solution.

FPGA implementation of IIR filter is done by using the Verilog description, ISE design suite 14.6 is the platform utilized for this [15]. The optimized and truncated coefficients are given to the filter and the input of the filter is taken as a unit impulse signal.

This paper consists of five sections: Sect. 2 presents the problem definition of the LPF, Sect. 3 describes the optimization technique employed, Sect. 4 is of two parts; the first part deals with the optimizing the coefficients of IIR LPF and the second part deals with the FPGA implementation of the IIR LPF. Finally, Sect. 5 is the conclusion of the proposed work.

2 LPF Problem Definition

In this paper, an IIR LPF is designed with BBO technique. The difference equation IIR filter [2] is presented in (1):

$$y(k) + \sum_{i=1}^{m} a_i y(k-i) = \sum_{i=0}^{n} b_i x(k-i) \tag{1}$$

where $x(k)$ and $y(k)$ are the input and output, respectively; $n(\leq m)$ is the order of the filter; a_i is the denominator coefficient; and b_i is the numerator coefficient.

Transfer function (TF) of the IIR filter is stated by the following equation:

$$H(z) = \frac{\sum_{i=0}^{n} b_i z^{-i}}{1 + \sum_{i=1}^{m} a_i z^{-i}} \tag{2}$$

The FR of the IIR filter is given (2) assuming $z = \mathrm{e}^{j\omega}$:

$$H(e^{j\omega}) = \frac{\sum_{i=0}^{n} b_i e^{-ij\omega}}{1 + \sum_{i=1}^{m} a_i e^{-ij\omega}} \tag{3}$$

where $\omega \in (0, \pi)$ in radians.

In this optimization problem, the objective function, $J(\omega)$ is defined in (4). The error fitness value of each search agent is calculated on the basis of magnitude response deviation from its practical response which includes the PBR (δ_p), SBR (δ_s), and transition width of the designed filter, $|H_d(\omega)|$.

$$J(\omega) = \sum_{\omega} abs\left[abs(|H_d(\omega)| - 1) - \delta_p\right] + \sum_{\omega} abs[(|H_d(\omega)|) - \delta_s] \tag{4}$$

The first part and second part of (4), gives the error contribution in $J(\omega)$ for pass band and stopband regions, respectively.

3 Biogeography-Based Optimization Algorithm

BBO technique which mimics the distribution of species in nature is utilized in this paper to solve the optimization problem. The candidate solution (CS) which is an integral part of BBO has to be improvised with the EFF. The combination of the current population and newly obtained CS gives the optimal solution to a problem. Each habitat has a particular Habitat Suitability Index (HSI) on which the performance of a CS is dependent on high HSI refers to the better solution. The set of decision variables for an objective function is referred to here as suitability index variables (SIVs). SIVs are randomly initialized by using migration and mutations of the same SIVs. μ gives the apportioned probability of a solution features with remaining solutions and λ gives the acceptance probability of a solution from other solutions, are calculated by the following equations.

$$p(\infty) = \frac{\mu_j}{\sum_{i=1}^{np} \mu_i};\ p(\infty)\text{is the emigration probability;} \tag{5}$$

$$\text{immigration probability } \lambda_j = 1 - \mu_{j.} \tag{6}$$

where i gives the array and j denotes the element in the ith array. Mutation (M) is a probabilistic operator that randomly modifies habitat SIVs based on the habitats probability of existence is initialized in the range of [0, 1].

The flow chart of the BBO algorithm for getting the optimal set of coefficients is presented in Fig. 1.

4 Simulation Results and Analysis

4.1 Optimization of IIR LPF Coefficients

In this paper, the IIR filter of order 8 is optimally designed by using BBO in the MATLAB platform. BBO algorithm is run in MATLAB 2017a version, 2.30 GHz on Intel CORE i3 processor with 8-GB RAM.

Here, the study of IIR LPF of order 8 with an equal number of feedback coefficients and feedforward coefficients is simulated. These coefficients are optimized by using the BBO technique. The FR of the LPF is plotted by taking those optimal set of coefficients. The design parameters of the LPF are shown in Table 1. For the calculation of the fitness value of any candidate solution, 512 sample points are considered.

The optimal set of coefficients obtained by the BBO technique is tabulated in Table 2 and these coefficients are utilized for the implementation of the digital IIR LPF.

With the help of optimized coefficients, magnitude response plot in dB and normalized magnitude response plot are shown in Figs. 2 and 3, respectively.

From these plots, different qualitative parameters are calculated and presented in Table 3.

Figure 4 represents the convergence plot, i.e., error fitness value versus iteration cycles plot of the IIR LPF of order 8.

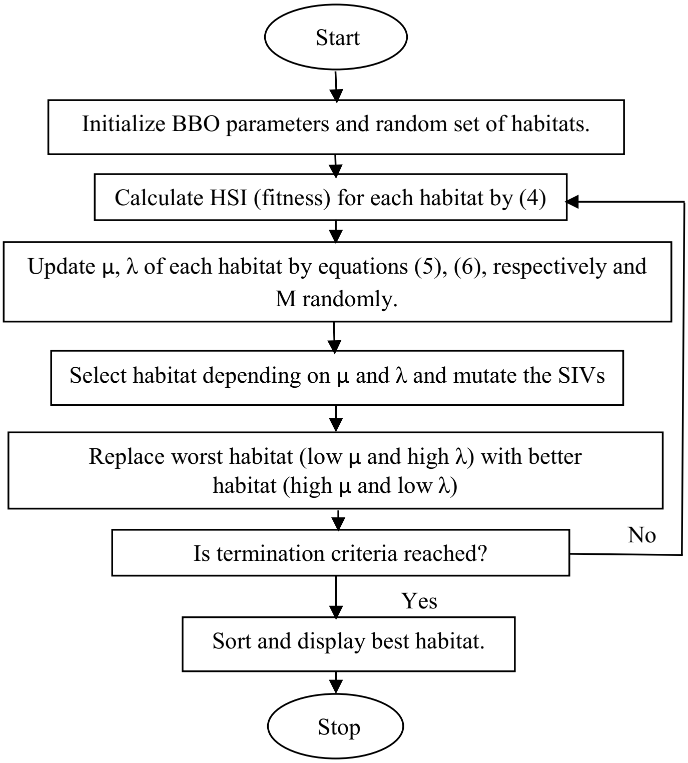

Start
Initialize BBO parameters and random set of habitats.
Calculate HSI (fitness) for each habitat by (4)
Update μ, λ of each habitat by equations (5), (6), respectively and M randomly.
Select habitat depending on μ and λ and mutate the SIVs
Replace worst habitat (low μ and high λ) with better habitat (high μ and low λ)
Is termination criteria reached?
No
Yes
Sort and display best habitat.
Stop

Fig. 1 Flow chart representing the BBO

Table 1 Design parameters

Design parameters	Values
Number of habitats	25
Maximum iteration cycle	500
M	0.1
σ	0.1
δ_s	0.001
δ_p	0.01
ω_p	0.35
ω_s	0.45

4.2 FPGA Implementation of IIR LPF

The IIR filter is realized by the Verilog coding and FPGA implementation is done. The hardware components used for the FPGA implementation of the IIR filter of order 8 are tabulated below in Table 4.

Table 2 Optimally obtained coefficients by BBO

Feedforward coefficients (b_k)	Feedback coefficients (a_k)
$b_0 = 0.0391$ $b_1 = 0.1254$	$a_0 = 1.1190$ $a_1 = -1.0223$
$b_2 = 0.2673$ $b_3 = 0.3799$	$a_2 = 2$ $a_3 = -1.2563$
$b_4 = 0.4195$ $b_5 = 0.3494$	$a_4 = 1.8066$ $a_5 = -1.3935$
$b_6 = 0.2190$ $b_7 = 0.0931$	$a_6 = 0.8584$ $a_7 = -0.1911$
$b_8 = 0.0212$	$a_8 = -0.0269$

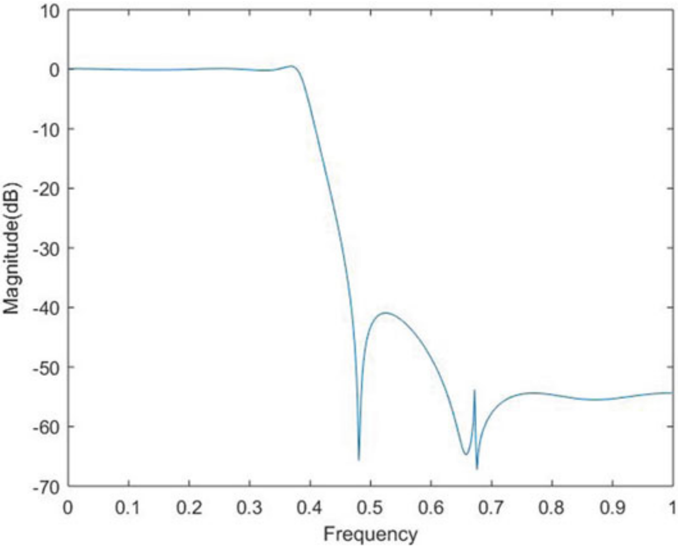


Fig. 2 Magnitude response (dB) plot of IIR LPF of order 8 designed by BBO algorithm

The optimal coefficients obtained are converted into the binary form. Figure 5 represents the simulation response of the Verilog HDL obtained by the coefficients through the optimization technique, this response is the general digital response of the filter considered. The response obtained is mathematically equated to the difference equation of the IIR filter. This simulation is done in the Verilog description and implemented in Spartan 3E FPGA. The ASIC implementation of the digital IIR filter can be done through the Verilog HDL realization.

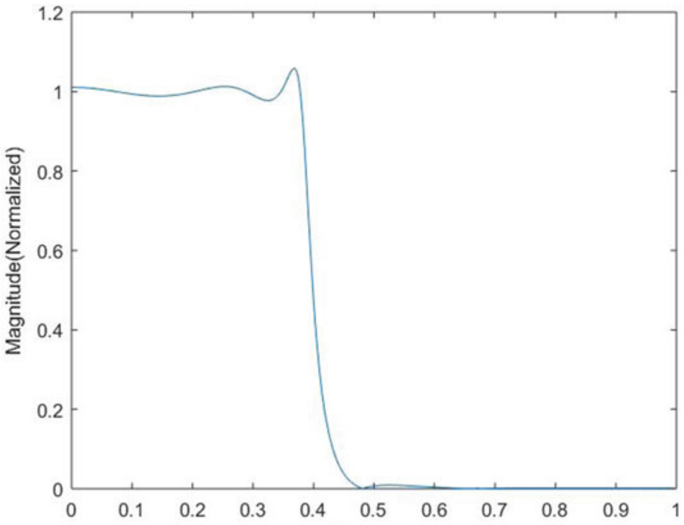


Fig. 3 Normalized magnitude response plot of IIR LPF of order 8 designed by BBO

Table 3 Analyzed data of designed IIR LPF of order 8 by BBO

Parameters of the filter obtained	Obtained values
Maximum stopband attenuation(dB)	−40.94
Maximum passband ripple	0.058
Maximum stopband ripple	0.008972
Average of stopband ripples	0.004703
Transition width	0.111

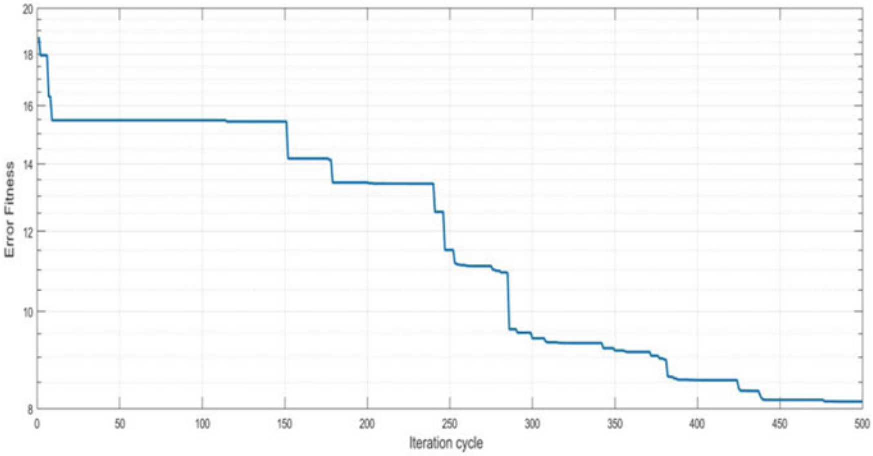
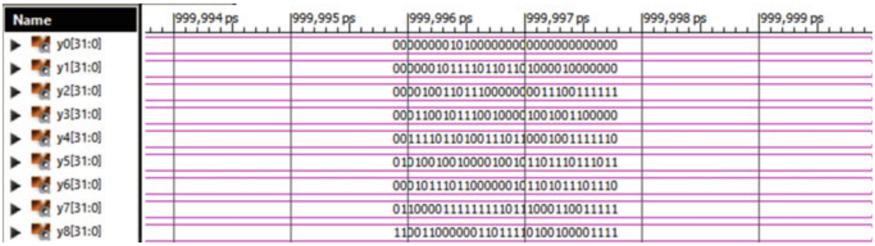


Fig. 4 Converging profile of IIR LPF of order 8 for BBO technique

Table 4 Number of Hardware components utilized for IIR filter of order 8

Components utilized	Count
Number of slice flip flops	505
Number of input–output blocks	578
Number of global clocks	1



```
y0[31:0]  00000000101000000000000000000000
y1[31:0]  00000010111101101101000010000000
y2[31:0]  00001001101110000000011100111111
y3[31:0]  00011001011100100001001001100000
y4[31:0]  00111101101001110110001001111110
y5[31:0]  01010010010000100101101110111011
y6[31:0]  00010111011000000101101011101110
y7[31:0]  01100001111111110111000110011111
y8[31:0]  11001100000011011110100100001111
```

Fig. 5 Simulation response of the optimized coefficients for LPF IIR

5 Conclusion

In this paper, IIR LPF is designed by using the BBO technique which follows the mutation and migration processes of different species depending on HSI. BBO uses the random search method to find an optimal solution and improvises the CS after each iteration and converges to an optimal solution. The coefficients obtained by the BBO algorithm are used to design the IIR LPF. This algorithm attains the optimal solution in a reasonable number of iterations. FPGA implementation of the IIR LPF is realized in the Verilog HDL and the optimal coefficients are utilized for the filter design as well as for reporting the binary response.

References

1. Oppenheim V, Schafer RW, Buck JR (1999) Discrete-time signal processing. Prentice-Hall, NJ
2. Hussain ZM, Sadik AZ, O'Shea P (2011) Digital signal processing-an introduction with Matlab applications. Springer
3. Nayak C, Saha SK, Kar R, Mandal D (2019) An optimally designed digital differential based preprocessor for R-peak detection in electrocardiogram signal. Biomed Signal Process Control 49:440–464
4. Karaboga N (2009) A new design method based on artificial bee colony algorithm for digital IIR filters. J Frankl Inst 346(4):328–348
5. Saha SK, Kar R, Mandal D, Ghoshal SP (2014) Digital stable IIR bandpass filter design using seeker optimization technique. Adv Mater Res 905:406–410
6. Karaboga N, Cetinkaya B (2004) Design of minimum phase digital IIR filters by using genetic algorithm. In: IEEE 6th nordic signal processing symposium, Finland, pp 29–32
7. Saha SK, Kar R, Mandal D, Ghoshal SP (2014) Harmony search algorithm for infinite impulse response system identification. Comput Electr Eng 40(4):1265–1285

8. Saha SK, Sarkar S, Kar R, Mandal D, Ghoshal SP (2012) Digital stable IIR low pass filter optimization using particle swarm optimization with improved inertia weight. In: International joint conference on computer science and software engineering (JCSSE), Bangkok, pp 147–152
9. Saha SK, Kar R, Mandal D, Ghoshal SP (2014) Gravitation search algorithm: application to the optimal IIR filter design. J King Saud Univ-Eng Sci 26(1):69–81
10. Karaboga N (2005) Digital IIR filter design using differential evolution algorithm. EURASIP J Appl Signal Process 2005:1269–1276
11. Simon D (2008) Biogeography-based optimization. IEEE Trans Evol Comput 12(6):702–713
12. Rahmati SH, Zandieh M (2012) A new biogeography-based optimization (BBO) algorithm for the flexible job shop scheduling problem. Int J Adv Manuf Technol 58(9):1115–1129
13. Lohokare MR, Pattnaik SS, Panigrahi BK, Das S (2013) Accelerated biogeography-base optimization with neighbourhood search for optimization. Appl Soft Comput 13(5):2318–2342
14. Simon D (2011) A dynamic system model of biogeography-based optimization. Appl Soft Comput 11(8):5652–5661
15. Sergiyenko A, Serhienko A (2018) Digital filter design using VHDL. In: International conference: high performance computing, Kiev, pp. 123–126

Invasive Weed Optimization-Based Optimally Designed High-Pass IIR Filter and Its FPGA Implementation

V. Karthik, K. Susmitha, S. K. Saha, and R. Kar

Abstract A meta-heuristic algorithm named as Invasive Weed Optimization (IWO) approach is considered in this paper for the design of infinite impulse response (IIR) high-pass filter (HPF) of order 8. This particular optimization technique is inspired by nature and it mainly depends on the colonizing characteristic feature of weeds. Unlike other optimization techniques, IWO converges fastly to the optimal solution and results in accurate solution parameters such as stopband attenuation, transition width, passband ripples (PBR), and stopband ripples (SBR). The optimally obtained coefficients are employed in the design of IIR HPF of order 8 which is realized in the Verilog hardware description language (HDL) and thus dumped on the field-programmable gate array (FPGA).

Keywords IWO · IIR filter · HPF · FPGA · Convergence · Magnitude response · Verilog HDL

V. Karthik (✉) · K. Susmitha · S. K. Saha
Department of Electronics and Communication Engineering, NIT Raipur, Raipur, India
e-mail: karthik.valupadasu@gmail.com

K. Susmitha
e-mail: susmithakomanapalli@gmail.com

S. K. Saha
e-mail: namus.ahas@gmail.com

R. Kar
Department of Electronics and Communication Engineering, NIT Durgapur, Durgapur, India
e-mail: rajib.kar@ece.nitdgp.ac.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_24

1 Introduction

Nowadays in the fields of science and engineering, there is a requirement for signals to be analyzed, synthesized, encoded, decoded, enhanced, and transported. The branch of science that takes care of these is termed as signal processing. Due to the advancement in the field of circuit integration technology and rapid increment in the computational power of systems, signal processing applications are enhanced and employed in many areas of science and technology [1]. The filtering process involves in the manipulation by changing, reorganizing, or converting the spectrum of the signal. The basic principle involved in the filter is it operates on the frequency domain termination and passage of frequencies, i.e., it allows to pass the specific band of frequency and blocks the others. This particular frequency at which the blocking of the frequencies occurs is termed as the cutoff frequency. This process of discriminating the frequencies has the prime importance because there may occur the mixing of unwanted signals, usually noise with the information-carrying signal. Noise either may be naturally created or by human beings, in terms of frequency characteristics, signals are mainly affected by thermal noise, flicker noise, avalanche noise, etc.

Filters can be designed by the signal which is continuous in time given as input with the analog components such as resistor, inductor, capacitor, operational amplifiers; these types are termed as analog filters. Whereas digital filters are employed to perform computational operations on a discrete-time signal to acquire the aimed characteristics. Analog filters have been replaced by digital filters due to extended applications and enhanced efficiency. While compared with the analog filters, the physical size of digital filters is minimized, highly accurate, reliable, and exempt to changes occur in environment.

Infinite impulse response (IIR) and finite impulse response (FIR) are the two types of filters. In terms of design particulars such as cutoff frequencies, stopband attenuation, PBR, SBR, and filter order, IIR filter is more reliable compared to FIR filter. As the order of the filter design decreases considerably in IIR filters, delay elements and multipliers required will be very less and also the computational time required will be less during hardware implementation and software realization, respectively. IIR filter is recursive in nature, its yield relies on the past input and output. But FIR filter yield relies explicitly on the past input values; hence, it is non-recursive in nature [2]. Filters are the main components in various biomedical applications such as in electrocardiogram (ECG) [3].

Gradient-based (GB) iterative search algorithms are usually employed in minimizing the error between the desired response of the filter to the predicted response of the filter. But the GB iterative search algorithms cannot outperform convergence to the global minimum when the error is multimodel. Hence, evolutionary algorithms are utilized for optimizing problems. Various heuristic search techniques are reported in the literature, particle swarm optimization [4–7], genetic algorithm [8], gravitational search algorithm [9, 10], seeker optimization algorithm [11], and BAT algorithm [12], etc. In this paper, an optimizing algorithm named IWO which is inspired from the growth of the weeds in nature was developed by Mehrabian et al. [13], various researches are undergone in this particular optimization technique [14–16], is employed in order to optimize the filter coefficients.

FPGA implementation of the filter is done by realizing the IIR filter in the Verilog HDL [17]. The coefficients obtained from the optimization technique are considered and the unit impulse signal is considered as input. The simulated output response of the filter in binary form is reported. ISE Design Suite 14.6 is utilized for the implementation and Vertex 5 is the FPGA used.

This paper consists of five sections, Sect. 2 briefly illustrates the problem definition for the HPF design. Section 3 discusses the IWO algorithm for optimizing the filter coefficients. Section 4 presents the results and simulation part in which the first part consists of optimizing the filter coefficients and the second part is FPGA implementation of the designed IIR HPF. Finally, Sect. 5 concludes the proposed work.

2 HPF Design Problem Definition

In this paper, IIR HPF is designed by utilizing IWO. The input–output dependency of the IIR filter [2] can be given in (1):

$$y(p) + \sum_{k=1}^{n} a_k y(p-k) = \sum_{k=0}^{m} b_k x(p-k) \tag{1}$$

where $x(p)$ is the filter input; $y(p)$ is the filter output; $n(\geq m)$ represents the order of filter; a_k is the denominator coefficient, and b_k is the numerator coefficient.

$$H(z) = \frac{\sum_{k=0}^{m} b_k z^{-k}}{1 + \sum_{k=1}^{n} a_k z^{-k}} \tag{2}$$

$H(z)$ is the transfer function (TF) of the IIR filter. The frequency response of the IIR filter with the assumption $z = e^{j\Omega}$ is given in (3):

$$H(e^{j\Omega}) = \frac{\sum_{k=0}^{m} b_k e^{-jk\Omega}}{1 + \sum_{k=1}^{n} a_k e^{-jk\Omega}} \tag{3}$$

where $\Omega \in (0, \pi)$ in radians.

In this paper, the objective function, $J(\omega)$ is defined in (4) is employed. The EF value of each search agent is calculated on the basis of magnitude response deviation from its practical response which includes the PBR (δ_p), SBR (δ_s), and transition width of the designed filter, $|H_d(\omega)|$.

$$J(\omega) = \sum_{\omega} abs[abs(|H_d(\omega)|) - \delta_s] + \sum_{\omega} abs\big[abs(|H_d(\omega)| - 1) - \delta_p\big] \tag{4}$$

The EF function which is mentioned in (4) is the generalized EF function that is to be optimized by IWO algorithm. The first part and the second part of (4) give the error contribution in $J(\omega)$ for stopband and passband regions, respectively.

3 IWO Algorithm

In this paper, IWO technique, an efficient numerical stochastic optimization algorithm inspired by the colonizing characteristic feature of weeds [14–16], is employed. Generally, weeds are plants whose growth leads to a serious threat to the desirable plants. The main inspiring features for utilizing the IWO as the optimizing problem are fast growth, fast reproduction, and distribution, robustness, and adaption to the change in the environment. Initialization—random generation of the population, reproduction—based on fitness values of the plants they are allowed to reproduce seeds, spatial distribution—generated seeds are distributed over search space, with zero mean and the standard deviation (σ) decreases linearly from initial SD (σ_{initial}) to final SD (σ_{final}) such that the seeds are abode to the parent plant, SD of iterations are calculated by the equation given in (5).

$$\sigma_{\text{iteration}} = \frac{(\text{iteration}_{\max} - \text{iteration})^n}{(\text{iteration}_{\max})^n} * (\sigma_{\text{initial}} - \sigma_{\text{final}}) + \sigma_{\text{final}} \tag{5}$$

where $\sigma_{\text{iteration}}$ is the SD at the present step; n is non-linear modulation index and finally, exclusion of the seeds—based on fitness values the plants are terminated [14]. The flow chart of IWO to get the optimal set of coefficients is presented in Fig. 1.

4 Simulation Results and Analysis

4.1 Optimization of IIR HPF Coefficients

The IWO-based optimally designed filter coefficients are obtained after 500 iterations with the sampling rate of 512. This algorithm is executed on MATLAB 2017a version platform on Intel CORE i3 processor, 2.30 GHz with 8-GB RAM. The design parameters are given in Table 1.

In Table 2, optimized coefficients approximated to 5 decimal points of IIR filter of order 8 are presented, by taking these optimal coefficients the FR of HPF are plotted.

With these optimized coefficients, magnitude response plot in dB and normalized magnitude response plot are shown in Figs. 2, 3, respectively.

Through these plots, different qualitative parameters are calculated and presented in Table 3.

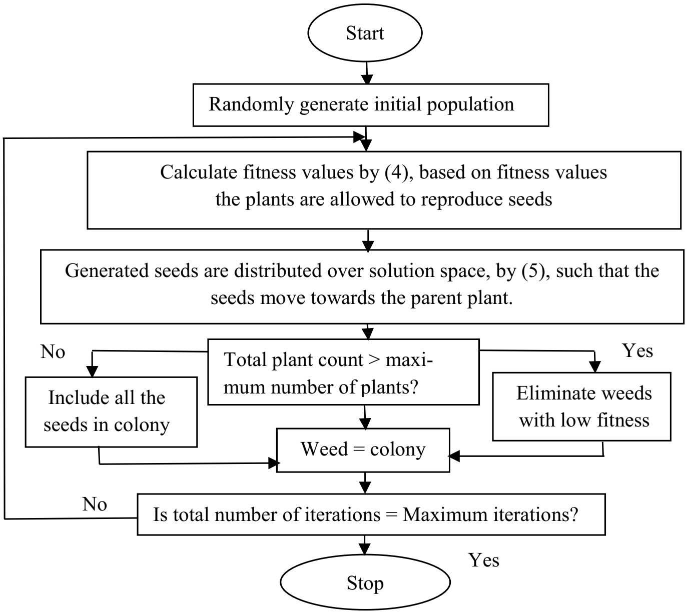

Start
Randomly generate initial population
Calculate fitness values by (4), based on fitness values the plants are allowed to reproduce seeds
Generated seeds are distributed over solution space, by (5), such that the seeds move towards the parent plant.
No
Total plant count > maximum number of plants?
Yes
Include all the seeds in colony
Eliminate weeds with low fitness
Weed = colony
No
Is total number of iterations = Maximum iterations?
Yes
Stop

Fig. 1 IWO flow chart

Table 1 Design parameters

Design parameters	Values
Population size	25
Maximum iteration cycle	500
σ_{initial}	0.5
σ_{final}	0.001
n(exponent)	2
δ_s	0.001
δ_p	0.01
ω_p	0.35
ω_s	0.45

Table 2 Optimally obtained coefficients by IWO

Numerator coefficients (b_k)	Denominator coefficients (a_k)
$b_0 = 0.19633$ $b_1 = 0.08928$ $b_2 = –0.82695$ $b_3 = 0.17072$ $b_4 = 1.74854$ $b_5 = –2$ $b_6 = 0.25928$ $b_7 = 0.73328$ $b_8 = –0.33229$	$a_0 = 0.37694$ $a_1 = 0.60153$ $a_2 = 0.91623$ $a_3 = 1.87253$ $a_4 = 1.51975$ $a_5 = –1.45639$ $a_6 = 1.70410$ $a_7 = 0.17844$ $a_8 = –1.25148$

Fig. 2 Magnitude response (dB) plot of IIR HPF of order 8 designed by IWO algorithm

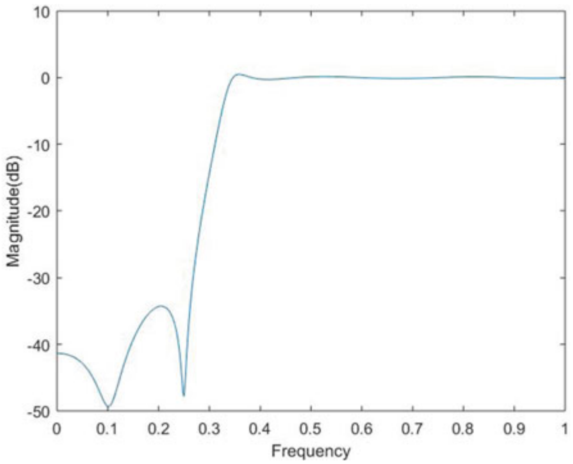


Fig. 3 Normalized magnitude response plot of IIR HPF of order 8 designed by IWO

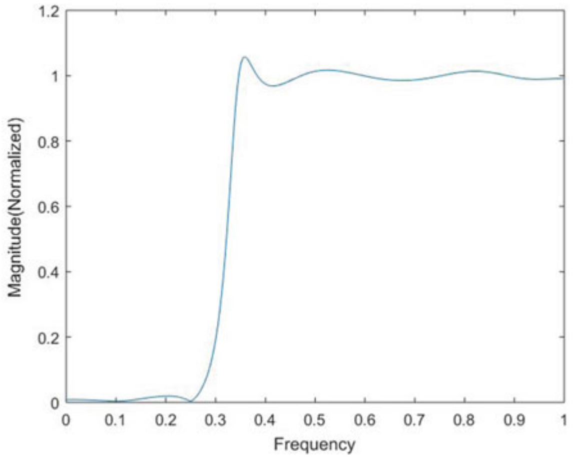


Table 3 Analyzed data of IIR HPF of order 8

Parameters of the filter obtained	Obtained values
Maximum stopband attenuation(dB)	−34.27
Maximum passband ripple	0.057
Maximum stopband ripple	0.01938
Average of stopband ripples	0.011384
Transition width	0.0975

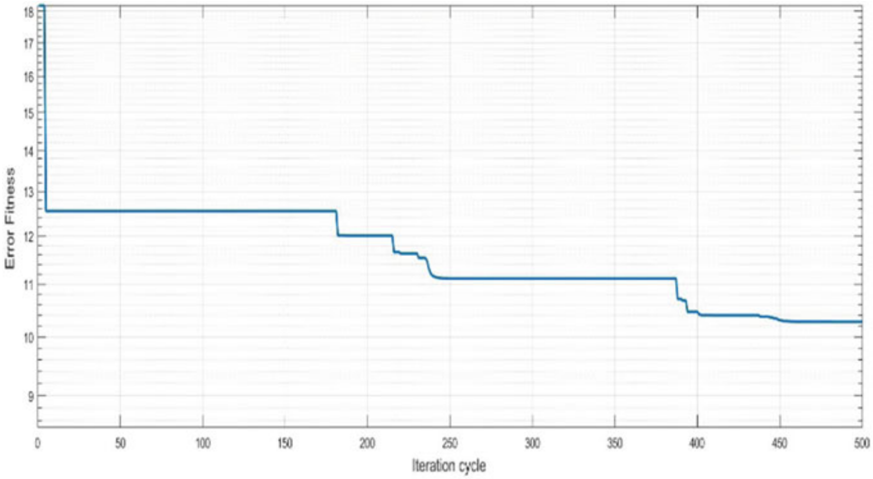


Fig. 4 Converging profile of IIR HPF of order 8 for IWO technique

Figure 4 represents the convergence curve of order 8 of IIR HPF plotted between the EF versus iteration cycles.

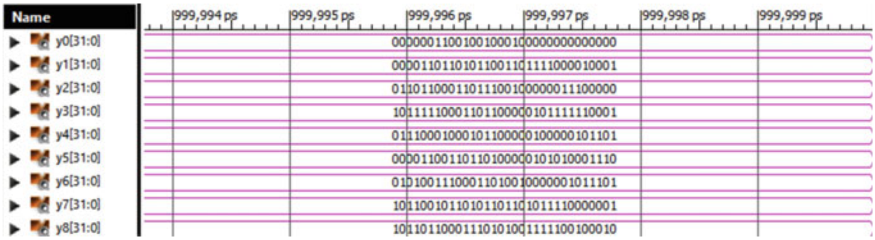
4.2 *FPGA Implementation of IIR Filter*

The hardware components utilized for the design of order 8 IIR filter when implemented in Vertex 5 FPGA are presented in Table 4.

The optimal set of coefficients of Table 2 is truncated and the binary response is plotted in Fig. 5 for the unit impulse input signal. The binary response of the IIR HPF is equaled to the general HPF output calculated mathematically. Hence, digital filter is realized in the Verilog HDL and FPGA through this ASIC design of the digital filter can be implemented.

Table 4 Hardware utilization of IIR filter of order 8

Components utilized	Count
Number of slice flip flops	505
Number of input output blocks	578
Number of global clocks	1



```
y0[31:0]  00000011001001000100000000000000
y1[31:0]  00001101101011001101111000010001
y2[31:0]  01101100011011100100000011100000
y3[31:0]  10111110001101100000101111110001
y4[31:0]  01110001000101100000100000101101
y5[31:0]  00001100110110100000101010001110
y6[31:0]  01010011100011010010000001011101
y7[31:0]  10110010110101101101011110000001
y8[31:0]  10110110001110101001111100100010
```

Fig. 5 Binary form representation of the output obtained of IIR HPF through IWO

5 Conclusion

In this paper, IIR HPF is designed using a weed colonization-based meta-heuristic algorithm known as IWO. The robustness and randomness properties of weed colonization are used for better exploration and exploitation to find a better solution in the given search space. The obtained result is effective in terms of various desired parameters namely, stopband attenuation, PBR, SBR, and better control over transition width. IWO converges to the better optimal solution in the acceptable execution time and attains the least values of EF function with a reasonable number of iterations. The output response IIR HPF which is realized in the Verilog HDL, the optimal coefficients are utilized for the IIR HPF design, and also the binary response is reported.

References

1. Oppenheim AV, Schafer RW, Buck JR (1999) Discrete-time signal processing. Prentice-Hall, NJ
2. Hussain ZM, Sadik AZ, O'Shea P (2011) Digital signal processing-an introduction with Matlab applications. Springer
3. Nayak C, Saha SK, Kar R, Mandal D (2019) An optimally designed digital differential based preprocessor for R-peak detection in electrocardiogram signal. Biomed Signal Process Control 49:440–464
4. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: IEEE ınternational conference on neural network, vol 4, pp 1942–1948
5. Saha SK, Sarkar S, Kar R, Mandal D, Ghoshal SP (2012) Digital stable IIR low pass filter optimization using particle swarm optimization with improved inertia weight. In: International joint conference on computer science and software engineering (JCSSE), Bangkok, pp 147–152
6. Saha SK, Kar R, Mandal D, Ghoshal SP (2012) Optimal IIR filter design using novel particle swarm optimization technique. Int J Circuits Syst Signal Process 6(2):151–162
7. Saha SK, Yangchen S, Mandal D, Kar R, Ghoshal SP (2012) Digital stable IIR high pass filter optimization using PSO-CFIWA. In: IEEE symposium on humanities, science and engineering research, Kuala Lumpur, pp 389–394
8. Karaboga N, Cetinkaya B (2004) Design of minimum phase digital IIR filters by using genetic algorithm. In: IEEE 6th Nordic signal processing symposium, Finland, pp 29–32
9. Saha SK, Kar R, Mandal D, Ghoshal SP (2014) Gravitational search algorithm: application to the optimal IIR filter design. J King Saud Univ-Eng Sci 26(1):69–81

10. Rashedi E, Nezamabadi-pour H, Saryazdi S (2011) Filter modelling using gravitational search algorithm. Eng Appl Artif Intell 24(1):117–122
11. Dai C, Chen W, Zhu Y (2010) Seeker optimization algorithm for digital IIR filter design. IEEE Trans Ind Electron 57(5):1710–1718
12. Saha SK, Kar R, Mandal D, Ghoshal SP, Mukherjee V (2013) A new design method using opposition–based BAT algorithm for IIR system identification problem. Int J Bio-Inspired Comput 5(2):99–132
13. Mehrabian AR, Lucas C (2006) A novel numerical optimization algorithm inspired from weed colonization. Ecol Inform 1(4):355–366
14. Zhang X, Xu J, Cui G, Wang Y, Niu Y (2008) Research on invasive weed optimization based on the cultural framework. In: International conference on bio-inspired computing: theories and applications, Adelaide, pp 129–134
15. Ramu Naidu Y, Ojha AK (2018) A space transformation invasive weed optimization for solving fixed-point problems. Appl Intell 48(4):942–952
16. Ouyang A, Yang Z (2016) An efficient hybrid algorithm based on harmony search and invasive weed optimization. In: International conference on natural computation, fuzzy systems and knowledge discovery, Changsha, pp 167–172
17. Sergiyenko A, Serhienko A (2018) Digital filter design using VHDL. In: International conference high performance computing, Kiev, pp 123–126

Identification of Online Auction Bidding Robots Using Machine Learning

Pooja Maan and R. Eswari

Abstract The aim of this project is to identify the bidding robots using machine learning, which bids in an online auction. Bidding robot is basically an application which helps to place a bid or click automatically on a website. So, this project will help the site owners to prevent unfair auction by easily flag the robots and remove them from their sites. The major steps are feature extraction, feature selection, model implementation, and classification. Feature engineering is done which includes feature extraction, dropping unnecessary features, and selecting necessary features. Various machine learning classification models are applied with new features to classify human and robot online auction bids and the best performance achieved is ROC score 0.954 using Random Forest.

Keywords Bids · Robots · Machine learning · Classification · Feature engineering

1 Introduction

Online auctions are the auctions which held on the Internet. The Internet has pushed the scope and range of these auctions to a point beyond what the original suppliers had expected. This is mainly due to the breakdown and removal of the physical limitations of traditional auctions such as geographical location, appearance, time, and a specific target audience. The largest online auction site is eBay, the first to facilitate transactions between individuals. WebStore, OnlineAuction, and Overstock are other common examples of online auction sites.

But these online auctions and bidding sites are becoming increasingly populated by auction snipers or online bots which are controlled by software. Auction bot [1] manipulation prevents actual human bidders from winning auctions and also enables

P. Maan (✉) · R. Eswari
National Institute of Technology, Tiruchirappalli, Tamil Nadu, India
e-mail: poojamaan2211@gmail.com

R. Eswari
e-mail: eswari@nitt.edu

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_25

sellers to use bots to bid the item to hike prices, making it almost impossible for real users to buy listed items at a reasonable price. This leads to unsatisfied bidders. For good customer base and better customer experience, classification is required in real time to eliminate the false bidders.

1.1 Problem Statement

The goal of this project is to classify whether the bidding done in an online auction is through a human or robot. This problem is a previous Kaggle competition [2]. In this competition bidding and bidder dataset is already given. For evaluation of the result Receiver Operating Characteristic curve (ROC-AUC) [3, 4] is used as this was the required metric in the competition.

Data exploration is required in this to get better insight how the robot's behavior is different from the human while bidding in an auction and also to gain the domain knowledge. Later, as per the requirement, various new features are formed on the basis of given features to train the model effectively and get better results. Feature selection techniques like confusion matrix and Random Forest "feature importance" are used. Finally, various machine learning classification algorithms are applied for comparison of results and get the best model.

2 Related Work

The previous work done used the same dataset and models for training and testing, in order to find the online bidding auctions robots which make it unfair to win an auction for a human. The only thing varies in previous papers is the feature engineering in order to improve the accuracy, the better the accuracy better will be the trained model.

In [5], as feature extraction 27 dense features with dimension varying from 1 to 6 and 3 sparse features with dimension 432,198, and 10 are formed. They trained the model with all predefined features and also with selected features to get the best performance. The best performance of this paper is 0.94 AUC CV score using Random Forest. With selected features and hypertuning of parameters AUC testing score is 0.92. Tree-based model worked very well in this case. They have not given importance to the time when bidding was done, as time-series features will help to track the time steps of a user. For each model, feature selection is done separately to reduce the dimensions but that is time consuming and not affecting the final result much.

In [6], as per feature engineering some new features were made. And they reduced the data points (row) before model training. It also used AUC ROC score as an evaluation metrics. Exploratory analysis is done to find out how robots are different from humans during an auction. Four models are used for training, which includes Logistic Regression, AdaBoost, SVM, and Random Forest. From which AdaBoost gave the best result with AUC score 0.874. Tenfold cross-validation ($K = 4$) is used in which nine are used for training and one for testing. But the reduction of data points leads to wrong classification during testing because there will be less data to train a model and for better results training data must be big in machine learning. Some computations constraints are also present due to which other classifiers are not implemented, which can work well on this type of data.

3 Proposed Work

In the proposed work, feature engineering is done which is a key part of this project, feature engineering includes feature extraction, dropping of unnecessary features, and features selection. Six different machine learning classification algorithms are applied.

3.1 Feature Engineering

Feature engineering is divided into three parts. By data exploration (dataset explanation is in Sect. 4.1), came to the conclusion that humans and robots behavior is different while bidding in an auction, e.g., robots will bid faster and more number of times as compared to genuine bidder and also a human cannot use a large number of URL, IP, country, and devices at the same time. So, the goal is to come up with best features by features extraction and feature selection techniques.

3.1.1 Feature Extraction

In feature extraction, for each bidder id new features are made. Some of the features are dense which includes number of auction, country, URL, IP, device, auction played, won, etc. While some are sparse features like merchandise and time series. Merchandise is a feature with ten categories in which auction occurs, e.g., jewelry, computers, etc. In time-series features, bin is created for estimation of time steps and 14 features are made which includes six Fast Fourier Transform, six Wavelet RMS, median, and max feature.

Table 1 contains all the new features formed on the basis of given features along with reasoning.

3.1.2 Drop Features

Two features are dropped, which are “Payment Account” and “Address”. The reason behind this is that these are given in alphanumeric format (db9b78e9629861ac

Table 1 New features extracted

Feature	Reasoning
nbActions (1)	Number of auction participated by a user (counts in the bids dataframes)
Time response (3)	Timing between the last bid and the action of the user on an auction (mean, min, and max time response)
Bid value (3)	Number of bids before the action of the user on an auction (mean, min, and max bid value)
Auction played and won (2)	Estimation of the winner by the last action and finding the number of auction played by users
Nb of IP address, country, URL(1)	Count of triple
Merchandise (10)	One hot encoding of ten categories
nbCountry (1)	Number of countries used by user
nbIp (1)	Number of IP used by user
URL features (2)	Number of URL used by user and ratio of actions from the main URL "vasstdc27m7nks3"
nbDevice (1)	Number of devices used by user
Time series (14)	Creating bin for estimation of time step, the creation of a time series by users (Fast Fourier Transform, Wavelet RMS, Median, Max)

699f105), i.e., cannot be considered as numeric label or categorical label to give in a machine learning model. As machines can only understand numbers, we can give numerical features to the machine learning model and convert categorical to numerical but converting alphanumeric in numeric or categorical form is not feasible and will not affect the training of model in a better way.

Total number of features finally formed after extraction and dropping are 39.

3.1.3 Feature Selection

Feature selection is a technique in which we select only the feature with more importance and discard the least one. The requirement of feature selection is more when we have lots of features and model taking lots of time in training or when the least important features are affecting the performance of more important features. Techniques used for feature selection are confusion matrix and Random Forest feature importance model.

Figure 1 is a plot of confusion matrix among all the features. Here, the scale is between 0 and 1, where 1 represents highly correlated (dark color) and 0 (light color) represents no correlation between two features. When two features are highly correlated then we can discard one to reduce the biasing in model and to reduce the

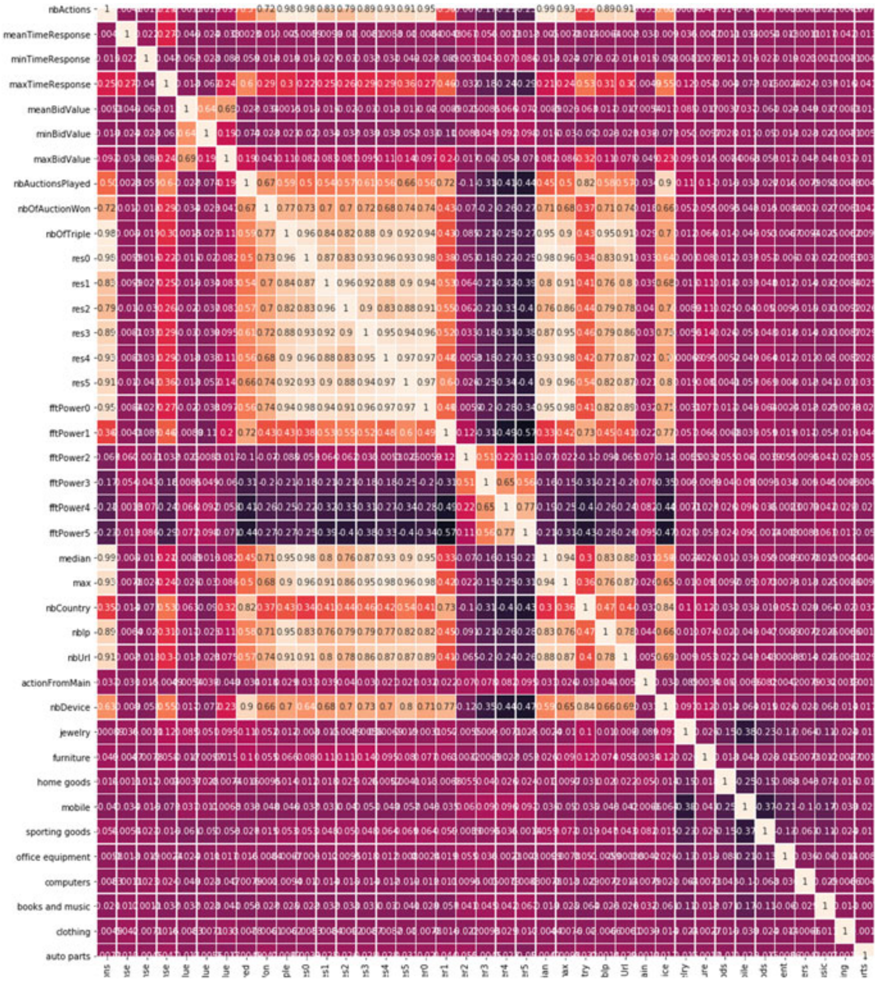


Fig. 1 Confusion matrix between all the features extracted

number of features. Discarding one in two correlated features will help in improving the accuracy. From Fig. 1, all the time series features are correlated and nbAction is correlated with most of the features.

Figure 2 represents the importance of each feature with respect to target variable using Random Forest Classifier. Tree-based classifiers have inbuilt class “feature importances” which takes all the features, train the model, and then give importance of each feature against the target variable.

Using Figs. 1 and 2, feature selection is done and dropped few features to reduce the total number of features and improve the AUC score. Dropped features in feature

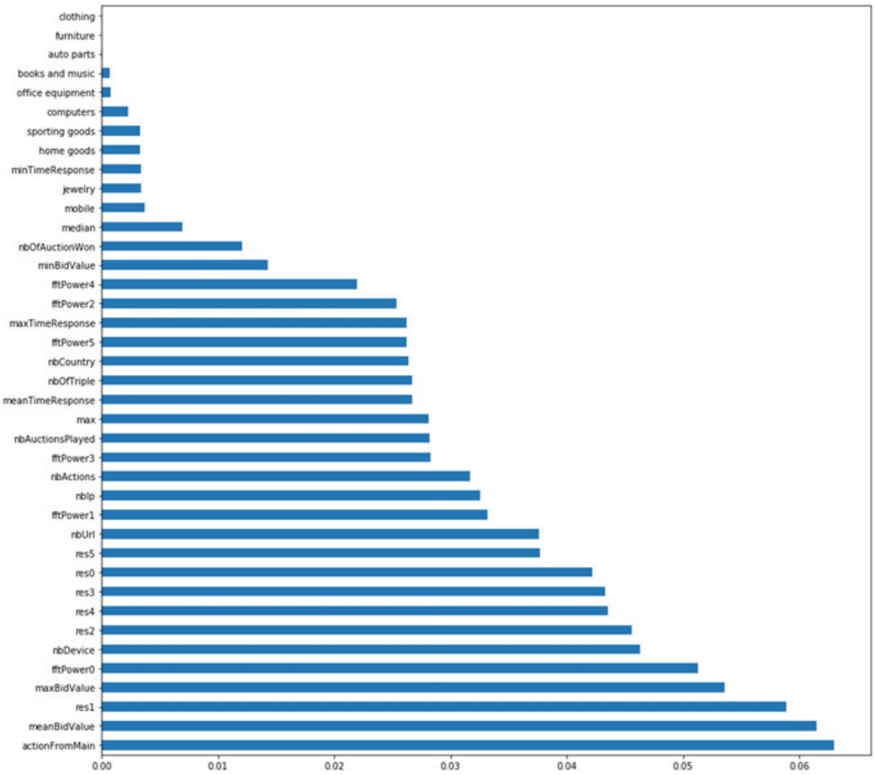


Fig. 2 Feature importance using Random Forest Classifier

selection are “furniture”, “office equipment”, “books and music”, “clothing”, and “auto parts”.

3.2 Models Used

In this paper, six different types of machine learning classifiers [7–9] are used to compare the performance with each other and also for comparison with [5].

Baseline classifier used is as follows:

- Logistic Regression (LR) [10]: It is naturally binary classification algorithm, can be used for regression, where we take an assumption that data is linearly separable. In this, we try to find the best plane, such that, it classifies most of the points correctly.
  Logistic regression equation is given in Eqs. (1) and (2).

$$ln\left(\frac{P}{(1-P)}\right) = \beta_0 + \beta_1 \cdot x \tag{1}$$

$$P = \frac{\exp^{\beta_0} + \beta_1 \cdot x}{1 + \exp^{\beta_0 + \beta_1 \cdot x}} \tag{2}$$

- Support Vector Machine (SVM) [11]: It is one of the most popular classification algorithms and also it can be used for regression. The key idea is that we need to find the plane that separates one class from another class as widely as possible. In other words, we try to find a hyperplane that maximizes the margin. Margin is distance between positive hyperplane and negative hyperplane.
  Support vector machine equation is represented in Eqs. (3) and (4).

$$h(x_i) = \text{sign}\left(\sum_{j=1}^{s} \alpha_j y_j K(x_j, x_i) + b\right) \tag{3}$$

$$K(v, v^{`}) = \exp\left(\frac{||v - v^{`}||^2}{2y^2}\right) \tag{4}$$

- Decision Tree (DT) [12]: Decision tree is a set of excess parallel hyperplanes that divide the whole reason into hyperplane cube or cuboid. Decision tree is a nonlinear classifier like neural networks. It is generally used for nonlinearly separable data, even in regression decision tree is nonlinear. Unlike linear regression, there is no equation to express the relationship between independent and dependent variable.

Also used classifiers which are extension of decision tree as follows:

- Random Forest (RF) [13]: It is the most popular bootstrapped aggregation intuition ensemble model (bagging model) where we do bootstrap sampling. In this, we randomly take “n” number of points from data points from dataset and built a model for each base learner. The core idea of Random Forest is instead of doing only row sampling (bagging) we do row sampling as well as column (feature) sampling. After training each model, majority vote is done in classification for aggregation.
  RF = DT (Base learner) + Bagging + Column Sampling
- AdaBoost (Ada) [14]: It is a boosting ensemble model where we try to reduce the bias by keeping variance low. It is one of the most popular algorithms for face recognition. Here, the key idea is using the error from previous model and passing that error to the next model to predict how much error each data point would have in previous stage.
- Gradient Boosting (GB) [15]: Gradient Boosting algorithm is same as AdaBoost, the only difference is that its base learner is gradient boost decision tree.

Decision tree-based models are also used above in feature selection for finding the important features.

Table 2 Runtime environment in Google Colab

Environment	RAM (GB)	Disk space (GB)
None	25	48
GPU	25	358
TPU	35	48

4 Experimental Setup

4.1 Platform

For execution of the models, Google Colab[?] is used. Google Colab is a free cloud-based service. It provides us with an environment to run code of python(.py) or ipynb notebook cells. One can access this from any system at any time.

Table 2 shows the types of runtime environment given in Google Colab.

For this project, GPU is used and it gave good computation to run the program in less time.

4.2 Dataset

Dataset is taken from Kaggle [2], it is given as a part of competition. The given dataset is in two parts.

5 Results and Analysis

All the above classifying algorithms are applied to all the features extracted from the given data, i.e., 39 and also applied Random Forest on the selected features which are selected in feature selection Sect. 3.1 above.

As evaluation metric ROC curve is used, ROC stands for Receiver Operating Characteristic [3, 4]. It is a curve in which True Positive Rate (TPR) is plotted against False Positive Rate (FPR) and the score or area lies between 0 and 1. Figure 3 represents how a curve is plotted between true positive rate and false positive rate. The area under the curve, also called AUC, is what we will try to find in this paper.

In Fig. 4, TPR and FPR are given with confusion matrix. TPR is also known as sensitivity and FPR is (1-specificity).

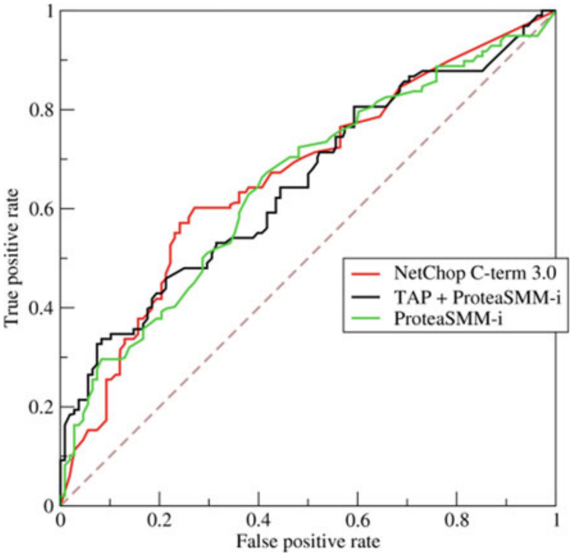
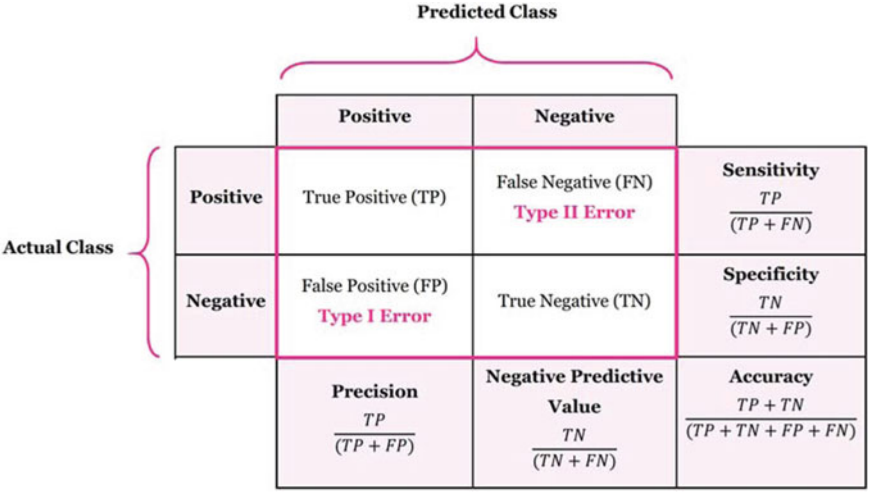


Fig. 3 Receiver Operating Characteristic curve



		Predicted Class		
		Positive	**Negative**	
Actual Class	**Positive**	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{(TN + FP)}$
		Precision $\frac{TP}{(TP + FP)}$	**Negative Predictive Value** $\frac{TN}{(TN + FN)}$	**Accuracy** $\frac{TP + TN}{(TP + TN + FP + FN)}$

Fig. 4 TPR (sensitivity) and FPR (1-specificity)

Table 3 ROC score using the proposed features on different Models

Model applied	Training score	Cross-validation score ($K = 4$)
SVM	1.0	0.9488
Logistic regression	0.9468	0.9468
Decision tree	1.0	0.9384
Random forest	0.992	**0.9543**
Adaboost	1.0	0.944
Gradient boosting	1.0	0.942
Random forest **(feature selection)**	0.98609	0.95379

5.1 Evaluation Score

In this paper, only training and cross-validation score are present where number of folds is four ($K = 4$). All the models are trained on all features extracted, i.e., 39 and also on selected features to get these scores. Table 3 shows the training score and cross-validation score ($K = 4$) of each model.

From Table 3, we can observe that Random Forest gave the best result out of all with AUC score 0.9543. Same like [5], here also tree-based model worked well. Last row of Table 3 contains the training and cross-validation score of Random Forest with only selected features as the Random Forest gave the best result on all extracted features (39), so only Random Forest is applied on selected features. Here, the model which worked best on all the features is Random Forest, so applied that model on selected features and it gave score 0.9537 which is less than the score with all features extracted.

5.2 Result Comparison

Out of all the features extracted, 19 features are same as base paper [5] and rest 20 features are different from them. Also, the features selected are different in both cases to reduce the number of less important features.

In [5], same models are applied and the same evaluation metric is used. Feature extraction and selection is different in both the cases. Table 4 shows the comparison of cross-validation score with [5].

From Table 4, we can compare that taking different features improved the accuracy pretty much. The main difference in features is because of the time-series features which created bin for estimation of time step, then created time series by users. Fast Fourier Transform and Wavelet RMS are used. For each machine learning classifier accuracy is increased, especially in decision tree, it is increased by 0.28%.

Table 4 Comparison of previous work and proposed work

Model applied	Previous work result (CV score)	Proposed work result (CV score)
SVM	0.8874	0.9488
Logistic regression	0.8082	0.9468
Decision tree	0.6711	0.9384
Random forest	0.9404	**0.9543**
Adaboost	0.9373	0.944
Gradient boosting	0.9180	0.942
Random forest **(feature selection)**	0.928	0.95379

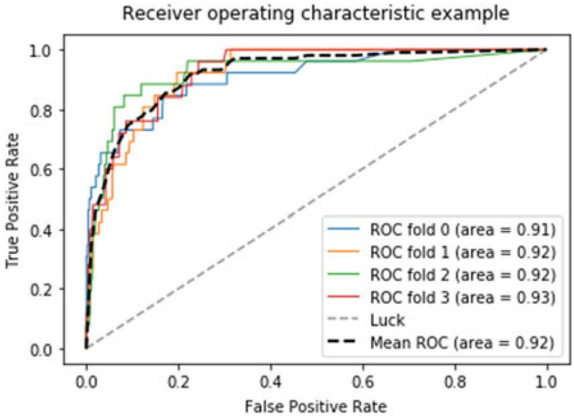


Fig. 5 ROC curve of final result using Random Forest

5.3 Final Result

Random Forest gave the best score out of all models with all the features. For this result, total 39 features are used and these parameters are used (n-estimators = 600, max-depth = 15, min-samples-leaf = 2).

After dropping five features in feature selection, accuracy did not increase much. Even with all features got better results.

Figure 5 shows the ROC curve of Random Forest for each of the fourfold from which the final result came.

6 Conclusion and Future Scope

Different machine learning models give different results in real-time problem, but one can get better results by doing feature engineering effectively. We have seen that by feature extraction and feature selection, tree-based models give better accuracy. Time-series features helped to improve the accuracy of each model which was not done in [5].

In future, parameters hypertuning of models can be done to improve the results more and deep learning can be applied to check if that can improve the accuracy of this type of dataset.

References

1. Ito T, Fukuta N, Shintani T, Sycara KP (2000) BiddingBot: a multiagent support system for cooperative bidding in multiple auctions. In *Proceedings fourth international conference on multiagent systems* (pp. 399–400).
2. Dataset. KAGGLE. https://www.kaggle.com/c/facebook-recruiting-iv-human-or-bot
3. Davis J, Goadrich M (2006) The relationship between Precision-Recall and ROC curves. In *ICML '06*
4. Hanley JA, Mcneil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143(1):29–36
5. Gu X, Shi S (2017) Human or robot.
6. Packer C, William H (2015) Bid-war: human or robot? 2 June 2015
7. Alpaydin E (2004) Introduction to machine learning. Adaptive computation and machine learning. MIT Press, Cambridge
8. Kotsiantis SB (2007) Supervised machine learning: a review of classification techniques. Informatica (Slovenia) 31:249–268
9. Gron A (2017) Hands-on machine learning with scikit-learn and tensorflow: concepts, tools, and techniques to build intelligent systems, 1st edn. O'Reilly Media
10. Haifley T (2002) Linear logistic regression: an introduction. In IEEE international integrated reliability workshop final report, Lake Tahoe, CA, USA, pp 184–187
11. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B (1998) Support vector machines. IEEE Intel Syst Appl 13(4):18–28
12. Ding H, Wang X (2002) Research on algorithm of decision tree induction. In Proceedings of international conference on machine learning and cybernetics, Beijing, China, vol 2, pp 1062–1065
13. Paul A, Mukherjee DP, Das P, Gangopadhyay A, Chintha AR, Kundu S (2018) Improved random forest for classification. IEEE Trans Image Process 27(8):4012–4024
14. Shaowen L, Yong C (2014) A kind of improved AdaBoost algorithm. In 7th international conference on intelligent computation technology and automation, Changsha, pp 16-18
15. Wen Z, He B, Kotagiri R, Lu S, Shi J (2018) Efficient gradient boosted decision tree training on GPUs. In 2018 IEEE international parallel and distributed processing symposium (IPDPS), Vancouver, BC, pp 234

Machine Learning-Based Green and Energy Efficient Traffic Grooming Architecture for Next Generation Cellular Networks

Deepa Naik, Pothumudi Sireesha, and Tanmay De

Abstract In the year 2015, the United Nation has adopted 17 Sustainable Development Goals (SDG) to ending poverty, saving the planet and bringing prosperity for all by the year 2030. Universal broadband connectivity is considered as one significant contributing factor to achieving these goals. There is a close correlation between the national Gross Domestic Product (GDP) and broadband availability. Broadband access has great potential in opening up work opportunities and boosting income for poverty-stricken people in the remote and underdeveloped countries. It is estimated that there are still about 1.2 billion people who are still not connected to the Internet. Broadband requirements from this segment along with rising broadband demand from urban consumerism have put pressure on the available frequency spectrum. The optical fiber communication has an abundance bandwidth. The Internet Service Provider (ISP) cannot provide the optical network in remote areas, due to cost constraints, climate, weather, and high investment costs. Hence its wireless counterpart WiMAX has short set up time and low deployment cost. Hence universal broadband connectivity can be achieved by Hybrid Optical WiMAX networks. Further, the abovementioned remote areas suffer from low infrastructure and unreliable power supply. In this paper, we have used alternative sources of energy to mitigate the problem of unreliable electricity supply, particularly in the areas. The proposed machine learning-based renewable energy prediction depends on the geographical location of the network node. The predicted renewable energy can be used as a source for serving traffic demands. The traffic aggregation methods were used to minimize network resource consumption. The unpredictability in harnessing renewable energy is mitigated by using backup nonrenewable energy. The simulation results show that the proposed algorithm reduces nonrenewable energy consumption.

Supported by organization National Institute of Technology, Durgapur.

D. Naik (✉) · P. Sireesha · T. De
Department of Computer Science and Engineering, National Institute of Technology, Durgapur, India
e-mail: naiksavantdeepa@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_26

Keywords Base Station (BS) · OLT (Optical Line Terminal) · ONU (Optical Network Unit) · WiMAX (World Interoperability for Microwave Access Network) · WDM (Wavelength Division Multiplexing) · PON (Passive Optical Network) · Traffic grooming

1 Introduction

Passive Optical Network (PON) offers large bandwidth, long-distance reach, lower power consumption, and it is very reliable. At the same time, it has high deployment cost [1]. The Wireless Mesh Network (WMN) is flexible, less costly but offers low bandwidth capacity [2, 3]. The hybrid network combines reliable high capacity advantage of PON with flexible ubiquitous nature of WiMAX. The high energy consumption is one of the major issues in hybrid networks. The study related to optical and wireless network is implemented in an isolated way. This motivated us to work in this area. The question of energy generation from renewable sources is currently considered as one of the important topics that are growing rapidly as a result of its many advantages. Renewable energy has started to compete with nonrenewable energy sources for many reasons [4]. Renewable energy generation capacity is significantly affected by fluctuations in the atmosphere. Solar energy can be generated in the morning hours in the presence of sunlight during day time and the energy generated from the wind depends on the wind speed and geographic location. These fluctuations also depend on the seasons. The fluctuation of renewable energy resources may affect the power grid generation units. Therefore, to overcome this issue, the energy produced by natural resources should be utilized properly to limit the nonrenewable energy usage. The proposed traffic grooming algorithms will help to minimize energy consumption. The nonrenewable energy has been provided in each network node due to handling the fluctuation in the renewable energy resources in the network. To deal with the best sources of renewable energy depending on the geographical location machine learning algorithms can provide accurate results. Machine learning algorithms can be learned from previous data and help find a prediction of current and future data [5, 6]. Various machine learning approaches are used to design and implement different phases of renewable energy systems based on the specifications of the issue and their characteristics [7]. Developing the optimal position for renewable power plants play a significant role in reducing nonrenewable energy resources.

The main sources of power consumption are the network devices (routers and base stations). The link category absorbs 80% of total electricity used in wired networks [8]. Energy-efficient network and network components save electricity and have lower operational and maintenance costs [9]. Different methodologies have been proposed for reducing the power consumption. Many of them are based on Sleep and Active modes of optical network units to save energy.

To minimize brown energy consumption renewable energy sources are used. Water, wind, sunlight, tides, and geothermal heat are sources of renewable energy.

These are available naturally, continuously, and in plenty. They do not cause any pollution, hence are called green energy sources. For reliable, low cost, environment-friendly operations renewable energy is to be harvested in a decentralized way [10]. The author in [11] presents options, technologies, to deploy renewable energy in a decentralized manner. Wind and solar energies are unreliable because of the unpredictable weather. Similar approaches to harvest the energy are found in [12]. The hybrid system can handle the unpredictable nature of traditional conventional energy resources. However, the hybrid system increases complexity associated with the system hence optimization techniques are required. The further hybrid system requires a huge amount of power to operate.

The ONUs in PONs and Base Station in wireless networks consume maximum energy. As such in a hybrid network, ONU integrated with BS will consume more power. The traffic grooming approach used in [13] optical network to reduce nonrenewable energy. In traffic grooming, the low traffic demands of individual users are aggregated into high-speed light path to maximize the usage of network resources [14]. The traffic grooming ensures efficient usage of channel capacity offered by PON. Further, it also lowers equipment costs. Existing energy-efficient algorithms deal with optical and WiMAX networks separately. Such energy-efficient techniques are very rarely proposed for hybrid networks. This motivated us to apply the nonrenewable energy minimization technique for hybrid networks.

In this article the problem is defined in Sect. 2. Algorithm for energy available routes is described in Sect. 3. Analysis of result is done in Sect. 4 and conclusions in Sect. 5.

2 Problem Definition

In a hybrid network, the hybrid requests are bidirectional. The source and destination nodes may be optical or wireless, i.e., traffic requests may originate in optical/wireless node and their destination may be optical/wireless. The constraints referred to are stated by the authors in [15].

The hybrid network denoted by a bidirectional graph $G(V, E)$, where V is the set of nodes and E is the set of links. V may be optical or wireless nodes. E is the set of optical or wireless links. The optical nodes are connected by optical links and the Optical Line Terminal (OLT) is connected to the Optical Network Unit (ONU) by an optical cable. The ONU is integrated with WiMAX base station to provide a wireless connection to the relay station. The relay nodes are used to enhance the network coverage area.

- Given: A traffic matrix between node pairs (s, d), the availability of renewable energy per node (which is predicted from machine learning method for solar or wind energy).

- The OLT uses the WDM-PON to assign wavelengths to each ONUs in the network. The dedicated channels are used between OLT and ONU to serve the wavelengths depending on the loads in the ONU.
- The objective is to reduce the total nonrenewable energy consumption of hybrid network on that day.

3 Proposed Model

In this section, we have presented a machine learning-based renewable energy prediction technique for identifying the sources of renewable energy. The wind and solar energy depend on the climate as well as on the geographical location of the site. In our work, the network nodes have the provision for both solar panels as well as wind turbines. Depending on the time, the machine learning algorithms will predict which one can be used as a source of energy during the day. The output from the predicted source of energy will be used as the source to route the traffic. We have observed that if both are used as the source of energy huge energy is going as drained if it is not utilized within the two days. The architecture for wind and solar energy is depicted in Fig. 1. We have used the decentralized techniques for renewable energy generation due to challenging issues related to centralized energy generation [10]. The steps involved in Algorithm.

3.1 The steps involved in Algorithm discussed in Section 3.1

- Step 1: We apply the machine learning tool to identify the nodes which are powered by renewable energy. The nonrenewable energy also provided in each network node as backup energy.
- Step 2: We divide the traffic matrix into two matrices. The first matrix contains the number of requests which consume the full capacity of the wavelength channel. These traffic demands cannot be groomed with others. These need an entire wavelength channel and transponders at source and destination nodes.
- The second matrix consists of the traffic requests that consume the traffic the partial capacity of the wavelength channels. These requests can be groomed with other requests to maximize the network usage and minimize energy consumption. The traffic requests are groomed over the existing active transponders.
- Step 3: Select the traffic demands from the first matrix and find the renewable energy available routes between the s-d pairs. If the nodes are not powered by renewable energy then use the nonrenewable energy to serve the traffic demands.
- If the traffic demands are between the optical nodes, i.e., source and destination nodes are optically connected, the wavelength continuity constraints and transponders' availability is checked to establish the connection requests.

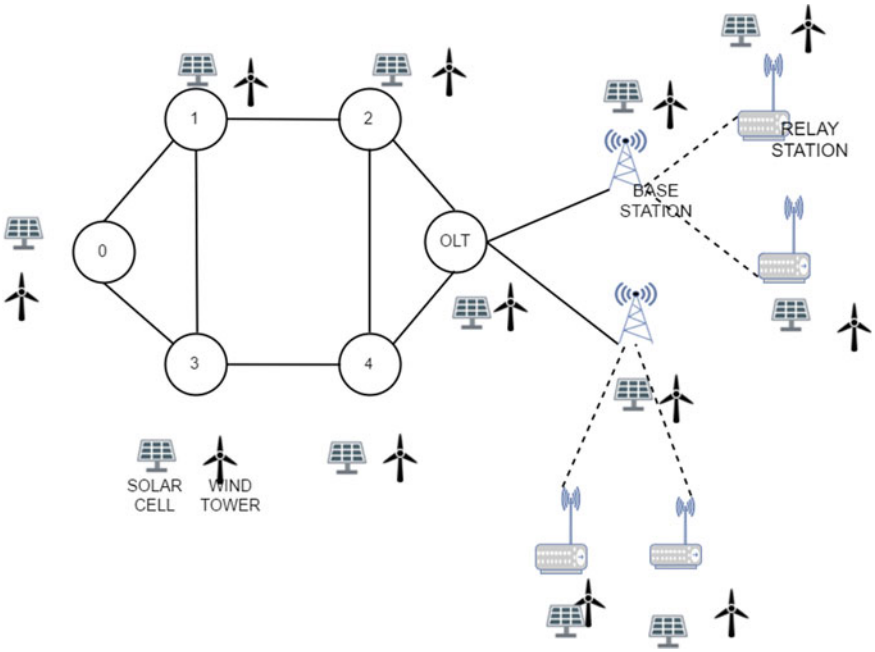


Fig. 1 Eight-node hybrid network with renewable energy

- If the source and destination nodes are wireless relay nodes, then the relay nodes should be in the communication range of the base station. Here sufficient frequency slots should be available to serve the demands.
- If the requests are hybrid, i.e., source and destination nodes are optical or wireless, both the optical and wireless constraints should be satisfied to serve the requests.
- After serving the first matrix, sort the contents of the second matrix depending on the traffic demands. Then select the top requests from the list. Find the renewable energy route to serve the requests. If there exist similar common requests, groom them with existing requests without violating the channel capacities.
- The requests which are unable to serve are served by nonrenewable energy. We have limited the number of transponders and wavelength channels available in the network to meet the operational and capital expenditure in a limit.
- Finally total energy consumption for servicing traffic is calculated for the given network.

3.2 Harvesting Renewable Energy from Sun and Wind for Hybrid Network

Renewable energy production depends on sky cover [16] (i.e., cloud cover) and wind speed [17]. Each node in the network is capable of producing enough renewable sources of energy either from wind or from the sun. As such each node is equipped with solar panels and wind turbines. We assume that the area of silicon solar panels and windward configured at each node is 100 m^2. The weather reports for the average sky cover and the average wind speed are available at the (www.worldweatheronline.com) for each network node.

4 Result Analysis

The proposed algorithm is implemented using C++ coding. Machine learning algorithms are implemented using Python2. To evaluate the performance of proposed algorithm we have used the hybrid network with one OLT and two ONU's integrated with WiMAX base station (as depicted in Fig. 1) Traffic demand between s-d pair follows uniform distribution with random variable over a given range of OC-[1–48]. The energy consumption of the transponders for wireless and optical networks are explained in [18]. With the help of machine learning algorithms, complex relationships or patterns are evolved from empirical data and accurate decisions are taken [19]. For studying future characteristics of a system Support Vector Machines (SVM) and Artificial Neural Networks (ANN) are found to be effective [20]. As such we have used ANN and SVM algorithms for accurately predicting weather forecast reports. Here the Multi-Layer Perceptron model is trained with the following parameters (hidden layers = 4, the number of hidden neurons = 4). This produced an accuracy of 0.93% whereas SVM produced 0.86 accuracy in training data.

Figure 2 depicts the renewable energy available per node in the network. The month of May is the highest source of solar energy. The machine learning algorithm predicts the source's energy available depending on the geographical location of the network node.

The relation between the traffic demanded versus traffic served in the network nodes are shown in Fig. 3. The shortest path traverse the shortest route to serve the requests. The normalized approach uses the energy aware route considering the hop count and the energy available along the path. This algorithm performs better in serving the traffic demands. The network resources get exhausted in the shortest route algorithm as the traffic demands increases.

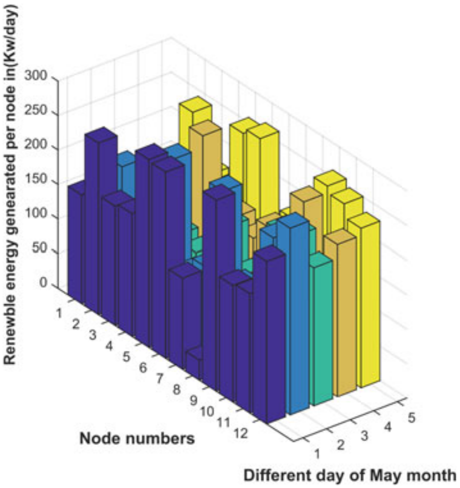


Fig. 2 The nodes powered by renewable energy in the month of May 1–May 5, 2018

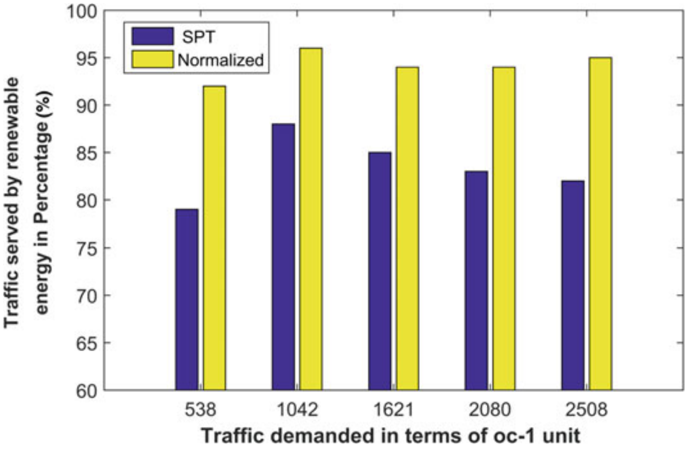


Fig. 3 Traffic demanded versus traffic serve in eight-node hybrid network

Figure 4 depicts the renewable energy resource used to serve the traffic demands in the network versus the traffic demands. The normalized energy aware route performed better in heavy load scenario. This will utilize the entire network resource to serve the traffic demands.

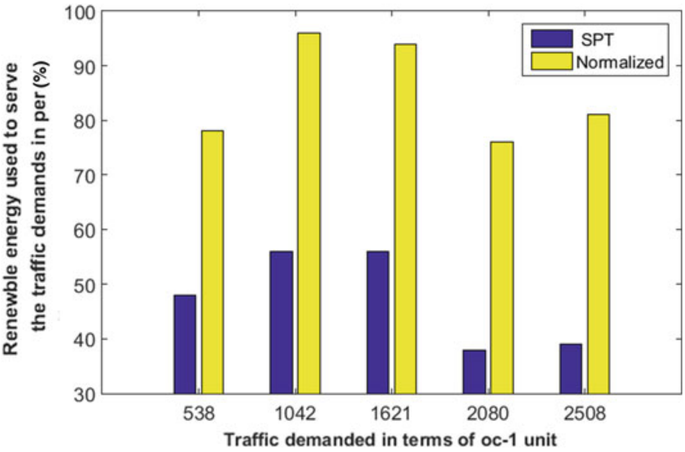


Fig. 4 Renewable energy used to serve the requests versus traffic demands

5 Conclusion

In this work, for hybrid networks renewable energy is predicted using machine learning techniques. A prediction model can predict the best source of renewable energy at a particular geographical location. Results show that the predicted model minimizes the nonrenewable energy used for serving the traffic. In the future, this work will be extended to predict network failure detection in the hybrid network.

References

1. Zhou H, Mao S, Agrawal P (2015) Optical power allocation for adaptive transmissions in wavelength-division multiplexing free space optical networks. Digit Commun Netw 1(3):171–180
2. Peng M, Li Y, Jiang J, Li J, Wang C (2014) Heterogeneous cloud radio access networks: a new perspective for enhancing spectral and energy efficiencies. arXiv:1410.3028
3. Luo C, Guo S, Guo S, Yang LT, Min G, Xie X (2014) Green communication in energy renewable wireless mesh networks: routing, rate control, and power allocation. IEEE Trans Parallel Distrib Syst 25(12):3211–3220
4. Panwar N, Kaushik S, Kothari S (2011) Role of renewable energy sources in environmental protection: a review. Renew Sustain Energy Rev 15(3):1513–1524
5. Voyant C, Notton G, Kalogirou S, Nivet ML, Paoli C, Motte F, Fouilloy A (2017) Machine learning methods for solar radiation forecasting: a review. Renew Energy 105:569–582
6. Pelekanou A, Anastasopoulos M, Tzanakaki A, Simeonidou D (2018) Provisioning of 5g services employing machine learning techniques. In: 2018 international conference on optical network design and modeling (ONDM). IEEE, pp 200–205
7. Zeng J, Qiao W (2013) Short-term solar power prediction using a support vector machine. Renew Energy 52:118–127

8. Nikoukar A, Hwang IS, Liem AT, Wang CJ (2015) Qos-aware energy-efficient mechanism for sleeping mode onus in enhanced EPON. Photonic Netw Commun 30(1):59–70
9. Vacca JR (2007) Optical networking best practices handbook. Wiley, Hoboken
10. Hiremath R, Kumar B, Balachandra P, Ravindranath N (2011) Implications of decentralised energy planning for rural india. J Sustain Energy Environ 2:31–40
11. Hiremath R, Shikha S, Ravindranath N (2007) Decentralized energy planning; modeling and applicationa review. Renew Sustain Energy Rev 11(5):729–752
12. Deshmukh M, Deshmukh S (2008) Modeling of hybrid renewable energy systems. Renew Sustain Energy Rev 12(1):235–249
13. Mukherjee B, Ou CS, Zhu H, Zhu K, Singhal N, Yao S (2004) Traffic grooming in mesh optical networks. In: Optical fiber communication conference, Optical Society of America, ThG1
14. Huang S, Dutta R (2007) Dynamic traffic grooming: the changing role of traffic grooming. IEEE Commun Surv Tutor 9(1):32–50
15. Chowdhury P, Tornatore M, Sarkar S, Mukherjee B (2009) Towards green broadband access networks. In: GLOBECOM 2009-2009 IEEE global telecommunications conference. IEEE, pp 1–6
16. Sharma N, Gummeson J, Irwin D, Shenoy P (2010) Cloudy computing: leveraging weather forecasts in energy harvesting sensor systems. In: 2010 7th annual IEEE communications society conference on sensor, mesh and ad hoc communications and networks (SECON). IEEE, pp 1–9
17. Feng C, Cui M, Hodge BM, Zhang J (2017) A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. Appl Energy 190:1245–1257
18. Baliga J, Ayre R, Hinton K, Tucker RS (2011) Energy consumption in wired and wireless access networks. IEEE Commun Mag 49(6):70–77
19. Maimó LF, Gómez ÁLP, Clemente FJG, Pérez MG, Pérez GM (2018) A self-adaptive deep learning-based system for anomaly detection in 5g networks. IEEE Access 6:7700–7712
20. Bankole AA, Ajila SA (2013) Cloud client prediction models for cloud resource provisioning in a multitier web application environment. In 2013 IEEE Seventh International Symposium on Service-Oriented System Engineering. IEEE, pp 156–161

Robust Image Encryption in Transform Domain Using Duo Chaotic Maps—A Secure Communication

S. Aashiq Banu, M. S. Sucharita, Y. Leela Soundarya, Lankipalli Nithya, R. Dhivya, and Amirtharajan Rengarajan

Abstract The increase of cyberattacks in the field of information security leads to high attention in the digitalized environment. There is a demand for well-organised methods to transmit data securely over the Internet. In the proposed algorithm, a greyscale image is secured by an encryption technique by two stages, i.e. confusion and diffusion. Digital images can be expressed in terms of numeric values called pixels. The images are converted by the integer wavelet transform-domain technique, where the images are separated by approximation and detailed sub-bands. The proposed encryption algorithm is implemented for the approximated sub-bands which contain significant information of the digital image. 1D logistic and 2D tent maps generate chaotic random keys. By various statistical and differential analyses, the robustness of the proposed algorithm is examined by achieving a maximum entropy of 7.997, near-zero correlation and larger keyspace of 10^{80}. With the results, it has confirmed that the enhanced chaos-based Integer Wavelet Transform (IWT) encryption has succeeded to resist any cyberattacks.

Keywords Image encryption · Chaotic map · IWT · Confusion · Diffusion

1 Introduction

In the past few years, the world has hugely transformed towards digitalisation. A tremendous number of data like text documents, multimedia are broadcasted in digital patterns over a public network. While the technologies enhanced, cyber crimes have also been raised due to its valuable data. So there must be an extreme assurance to guard the digital information from hackers. It is an efficient method to transmit the digital images securely and privacy is encryption technique so that merely authorised members can access it.

S. Aashiq Banu · M. S. Sucharita · Y. L. Soundarya · L. Nithya · R. Dhivya · A. Rengarajan (✉)
Department of Electronics and Communication Engineering, School of Electrical & Electronics Engineering, SASTRA Deemed University, Thanjavur 613401, Tamil Nadu, India
e-mail: amir@ece.sastra.edu

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_27

In the literature, two standard techniques are adopted, i.e. the spatialdomain and transform-domain methods. The encryption in the spatial domain has the numerical properties of the image even later confusion. The hackers may utilise this characteristic to retrieve the original digital image. Encryption in the transform-domain method produces enhanced security and recovery the low-level components of the transferred image also when noise or any attacks affect.

There are various transform methods accessible, i.e. Discrete Cosine Transform, Integer Wavelet transforms (IWT) [1], Discrete Wavelet Transform [2] and Fourier transform [3]. As digital images data are huge, the traditional algorithm is not suitable for encryption. To secure the digital images from cyberattacks several algorithms have been projected based on chaos theory, optical transform and DNA coding techniques. To improve encryption algorithms, many researchers have been urged to execute a combination of chaos and cryptography techniques. The chaotic map has unique qualities like periodicity, high sensitivity to initial states, ergodicity and randomness.

For a secured communication [4] proposed a reciprocal time-domain disruption in DWT with synthesis image generated by 14-bit LFSR and diffusion with Lorenz attractor and achieved an entropy of 7.9965. Analyses like chosen plain text key sensitivity are performed and claimed as a robust algorithm. A hyper-chaos key is used for image encryption by combining the spatial and transform domain for more security [5, 6]. Performing encryption in the spatial domain is not secure which is easily attractable and in a transform domain, it is robust, in [7] DNA-based image encryption was performed in the frequency domain with hyper-chaos key and achieved entropy of 7.9923.

By the advantages of chaos and the transform domain, a unique encryption method has been proposed by using a chaotic logistic map and tent map. The plain image is transformed into the IWT domain in which the blocks are split into Low-Low, Low-High, High-Low and High-High sub-bands. The low-frequency sub-bands are divided into 16 blocks separately and confusion and diffusion are performed by logistic and tent map, consequently for L-H and H-L sub-bands. The HH band is less sensitive so it is not carried out in this algorithm [18].

The manuscript is designed as follows: the basics of the IWT domain and chaotic maps in Sect. 2. The proposed method in Sect. 3. Further Sect. 4 results and analyses with the conclusion of this paper in Sect. 5.

2 Preliminaries

2.1 Integer Wavelet Transform

The fundamental level wavelet transform is the Haar wavelet to get the frequency domain of the image. By the lifting scheme, IWT is accomplished which provides integer coefficients. Low-Low, Low-High, High-Low and High-High are the four components decomposed from this domain. In the High–High band, it is of less sensitive components where the encryption technique is not carried out. For the L-L, L-H, H-L sub-bands the encryption algorithm is implemented.

2.2 Chaotic Maps

In the proposed algorithm duo chaotic maps are implemented, i.e. logistic map and tent map which is one-dimensional chaotic maps. The mathematical representation of the logistic map is given in Eq. 1 with the control parameter $p \in (3.56, 4]$ and x $(1) \in (0, 1)$.

$$x_n = px_n(1 - x_n) \tag{1}$$

The chaotic tent map mathematical representation is given in Eq. 2 with the control parameter (1, 2]

$$x_n(i + 1) = \beta xn \quad for \quad x_n > 1/2 \tag{2}$$

$$y_n(i + 1) = \beta(1 - x(i)) \quad for \quad 1/2 < x_n \tag{3}$$

3 Proposed Methodology

Transmission of data must be protected from hackers, which becomes a significant factor. To protect information, encryption technology must be adopted. In this paper, an encryption technique is performed in the transform domain with duo chaotic map for permutation and substitution. The low frequencies components are split into 16 sub-blocks in which even and odd are performed with chaotic maps to achieve the final encryption image Steps (1–8) are as follows and overall architecture is shown in Fig. 1.

Step 1: By IWT, convert the image *I1* from the spatial domain to the transform domain to obtain four frequency bands as Low-Low (LL), Low-High (LH), High-Low (HL) and High-High (HH).

Step 2: A 128 grey level LL image of size (128 × 128) is taken and decomposed into 16 number of blocks with block size (32 × 32).

$$\text{img\_blocks1} = (\text{LL},\ \text{size}(\text{LL}, 1)/4 \times \text{ones}(1, 4),\ \text{size}(\text{LL}, 2)/4 \times \text{ones}(1, 4)$$

Step 3: Pixel permutation is applied to each block based on odd and even block

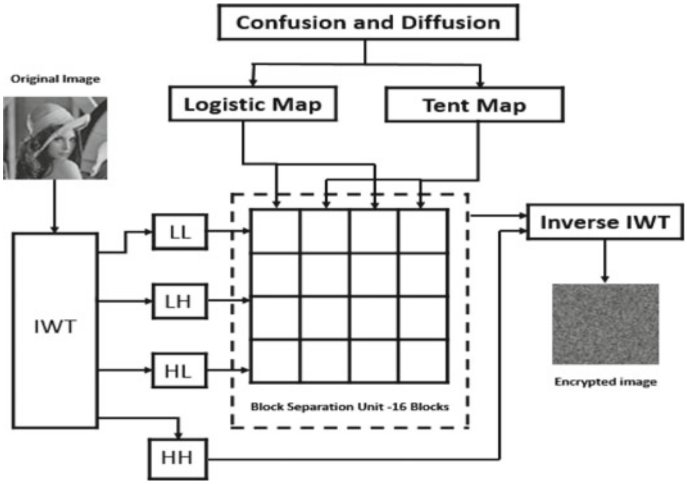


Fig. 1 The architecture of the proposed process

Step 4: For an odd block, pixel permutation is performed by the chaotic logistic map and then substitution is implemented with the initial values *Xn* = (0, 1) for generating pseudo-random sequences. The sequences are sorted in ascending order

$$[\mathrm{I1},\ \mathrm{J1}] = \ \mathrm{sort}\ (\mathrm{S1})$$

It is the indexing task *I1* is a new index after sorting data *S1* and sorted sequence after *S1* is *J1*. And diffusion is implemented by XOR ing with a logistic map.

$$F(i, j) \ = \ \mathrm{floor}\left[\mathrm{mod}\left(x(k) \times 10^{17}, \mathrm{block\_size}\right)\right]$$

$$\mathrm{Diff}(i, j) = C1(i, j)\mathrm{XOR}F(i, j)$$

Step 5: For an even block, pixel permutation is performed by chaotic tent map which is sorted in ascending order, finally confused image Con1 is achieved. And diffusion is implemented by XOR ing with tent map.

Step 6: By combining all the blocks on the subsection will result in the confused-diffused image (*D1*).

Step 7: Repeat the same Steps (2–7) for LH and HL sub-band. Therefore the confused-diffused images *D2* and *D3* is achieved

Step 8: By Inverse Integer Wavelet Transform (IIWT) of images *D1*, *D2*, *D3* and the frequency band HH, respectively, the ultimate encrypted image E is obtained.

4 Results and Discussion

Experimental analysis is carried on several greyscale images of size 256 × 256 which are executed using the personal computer of Windows 7 operating systems, 3 GHz CPU, 4 GB and Mat Lab 2014b is practiced as compiling software. Figure 2a–h shows the original image and its corresponding encrypted images

4.1 Statistical Analysis

4.1.1 Entropy

To prove the encryption technique the main analysis is entropy, which is used to calculate the random value transpired in the image. The entropy should produce

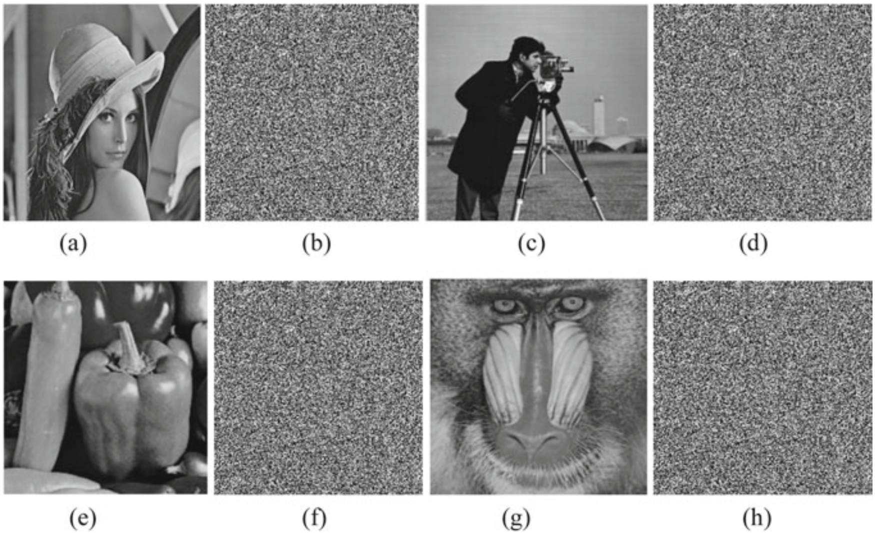


Fig. 2 **a**, **c**, **e**, **g** Original image and its corresponding encryption image (**b**, **d**, **f**, **h**)

Table 1 Images entropy

Entropy	Original	Encrypted
Pepper	7.6013	7.9968
Lena	7.4818	7.9965
Cameraman	7.1238	7.9964
Baboon	7.2340	7.9969

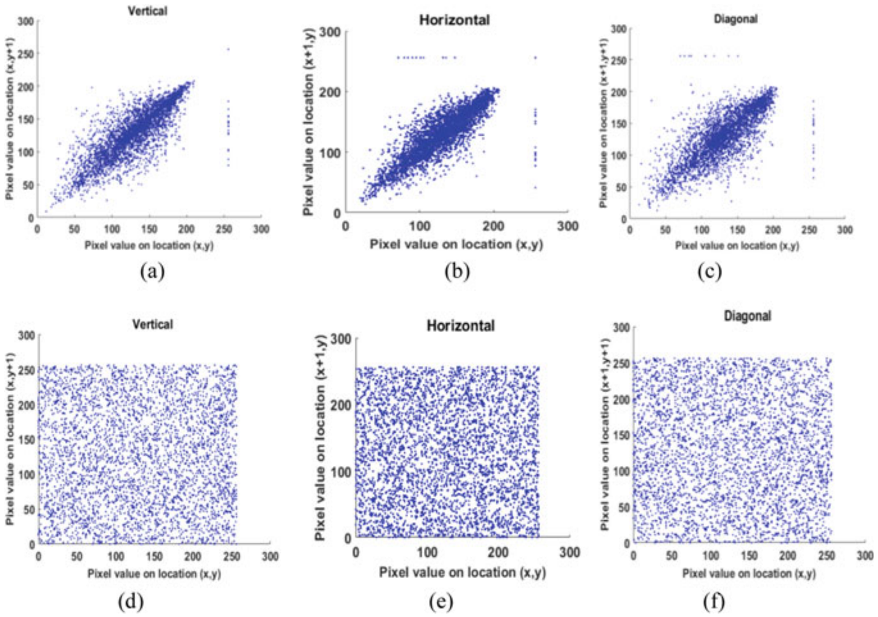


Fig. 3 **a**–**c** Original image baboon correlation analyses **d**–**f** encrypted correlation analyses

close to 8 for a more reliable technique [8–10]. From Table 1 the proposed method has attained the utmost entropy of 7.9968.

$$En = -\sum_{i=1}^{N1} P(x_j)\ \log_2\ P\left(x_j\right) \tag{4}$$

where the entropy is *En*, the greyscale level of *N1* image is x_j and *P* is the probability of each greyscale level.

4.1.2 Correlation Analysis

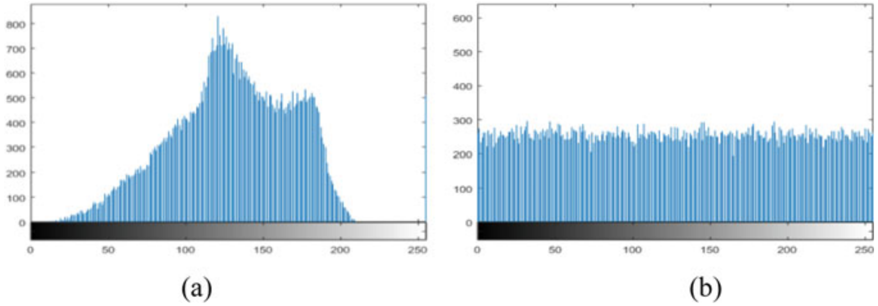
The main parameter of cryptanalysis is a correlation coefficient of the adjacent pixels. For the encrypted image, the correlation coefficient should be near to zero for horizontal, vertical and diagonal [11]. By Fig. 3 and Table 2 the correlation coefficient analysis has insisted that the algorithm can oppose the statistical attacks.

4.1.3 Histogram

A desirable encryption technique must have a uniform histogram for the encrypted image from the illustration and Fig. 4 confirms that it provides uniform patterns.

Table 2 Correlation coefficient analyses of different images

Correlation analyses	Original			Encrypted		
Images	H	V	D	H	V	D
Pepper	0.9460	0.9538	0.9141	0.0539	0.0134	0.0047
Lena	0.9106	0.9507	0.8849	0.0209	0.0156	0.0986
Cameraman	0.9303	0.9590	0.9048	0.0496	0.0741	0.0859
Baboon	0.8319	0.9483	0.7880	−0.0059	0.1690	0.0069



(a)

(b)

Fig. 4 **a** Histogram of original image. **b** Histogram of an encrypted image

4.2 *Differential Analyses*

The differential attack is to determine the relationship within the original and encrypted images [13, 14]. To determine the differential attacks number of pixel change rate NPCR and UACI must be intended by using the Eqs. (5) and (6). $C1$ and $C2$ are a different encrypted image. The cipher image of the original image is $C1$ and the cipher image of one-pixel change in the original image is $C2$. Table 3 shows the values of NPCR and UACI.

$$\text{NPCR} = \frac{\sum_{i,j} D(i,j)}{R \times C} \times 100\% \tag{5}$$

Table 3 NPCR and UACI value

Images	NPCR	UACI
Cameraman	99.622	33.41
Lena	99.594	33.37
Pepper	99.606	33.41
Baboon	99.633	33.45

Table 4 MSE and PSNR-encrypted image

Encrypted images	MSE	PSNR
Lena	7505.3	9.6891
Barbara	8357.2	8.8942
Pepper	8245.1	8.9734
Cameraman	9439.2	8.3563

$$\text{UACI} = \frac{1}{R \times C}\left[\sum_{i,j} \frac{|C1(i,j) - C2(i,j)|}{255}\right] \times 100\% \tag{6}$$

4.3 Error Metric Analysis

To find the difference between the original and encrypted image the Peak Signal to Noise Ratio (PSNR) must be below 10 dB. Table 4 shows the PSNR and MSE results of the images and has obtained adequate value.

4.4 Key Strength Testing

The keyspace needs to be higher than in 2^{128} to oppose the brute force attack [15, 16]. The suggested technique utilised duo chaotic maps, each having initial values and a control parameter. The keyspace of this method is 10^{80} which is sufficient to oppose any attack that provides a high-security level. An encryption algorithm must be sensitive to the unknown key. If a slight change occurs in secret key then it resultant produces a different encrypted image.

4.5 Chosen Plain Text Attack Analysis

Any techniques which are performing an XOR for diffusion method should examine the chosen plain text analysis to verify its potential of the proposed method [17]. It is estimated for two of the original plain images by XOR ing it and corresponding encryption images to confirm the chosen plain text analysis as shown in Fig. 5. By the outcomes of the images which prove that it can withstand such attack and the Eq. (7) is as follows.

$$\text{XOR}(\text{Plain\_1}, \text{Plain\_2}) \neq \text{XOR}(\text{Cipher\_1}, \text{Cipher\_2}) \tag{7}$$

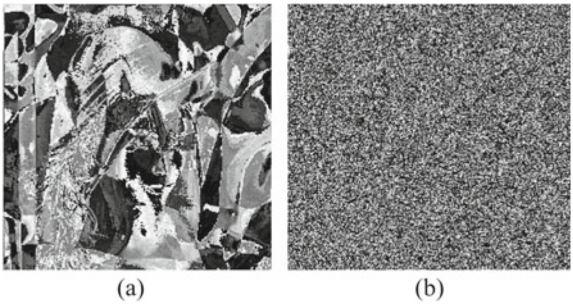


Fig. 5 **a** Original Peppers and Lena with XOR operation, **b** encrypted Peppers and Lena with XOR operation

Table 5 Performance comparison

Metrics	Keyspace	Horizontal correlation	Vertical correlation	Diagonal correlation	Entropy
Aashiq et al. [4]	2^{354}	0.0313	0.0733	0.0846	7.996
Guan et al. [7]	10^{58}	0.001158	0.000198	−0.000226	7.992
Dagadu et al. [11]	2^{128}	0.0003	0.0009	−0.0001	7.999
Ramasamy et al. [12]	2^{256}	−0.0237	−0.0178	−0.0284	7.995
Yavuz [14]	2^{572}	0.001987	0.004498	−0.008735	7.999
Proposed	10^{80}	0.0313	0.0733	0.0846	7.997

4.6 Performance Comparison

The performance of the proposed algorithm is correlated with the state-of-the-art methods with regard to entropy, correlation analysis and the keyspace which are exhibited in Table 5.

5 Conclusion

A new method is executed on the integer wavelet transform domain by duo chaotic maps for digital image encryption. The confusion and diffusion are performed in the low-low, low-high, high-low frequencies sub-bands by splitting them into 16 blocks. Even and odd columns are permuted and substituted by duo chaotic maps and final encryption is achieved. Through the experimental analyses, it is determined that it can resist brute force attacks. Further, it is examined by several statistical and differential attacks. The outcomes of the proposed method are secure in performances when compared with an existing encryption scheme. Forthcoming work will be executed on digital medical image encryption on the transform domain.

Acknowledgments Authors thank the Department of Science & Technology, New Delhi for the FIST funding (**SR/FST/ET-II/2018/221**). Also, the authors wish to thank the Intrusion Detection Lab at School of Electrical & Electronics Engineering, SASTRA Deemed University for providing infrastructural support to carry out this research work.

References

1. Tao RTR, Meng X-YMX-Y, Wang YWY (2010) Image encryption with multiorders of fractional Fourier transforms. IEEE Trans Inf Forensics Secur 5(4):734–738
2. Patra JC, Phua JE, Bornand C (2010) A novel DCT domain CRTbased watermarking scheme for image authentication surviving JPEG compression. Digit Signal Process. https://doi.org/10.1016/j.dsp.2010.03.010
3. Makbol NM, Khoo BE (2014) A new robust and secure digital image watermarking scheme based on the integer wavelet transform and singular value decomposition. Digit Signal Process. https://doi.org/10.1016/j.dsp.2014.06.012
4. Aashiq Banu S, Sridevi A, Sivaram R, Rengarajan A, Paramasivam VM (2019) Reciprocal time domain disruption of image for secured communication. In: IEEE international conference on computer communication and informatics (ICCCI), pp 1–4. https://doi.org/10.1109/iccci.2019.8822167
5. Yuan W, Yang X, Guo W, Hu W (2017) A double-domain image encryption using hyper chaos. In: International conference on transparent optical networks, pp 1–4
6. Singh DK, Tomar K (2018) A robust color image encryption algorithm in dual-domain using chaotic map. In: IEEE ınternational conference on ınventive communication and computational technologies, pp 931–935
7. Guan M, Yang X, Hu W (2019) Chaotic image encryption algorithm using a frequency - domain DNA encoding. IET Image Process 13(9):1535–1539. https://doi.org/10.1049/iet-ipr.2019.0051
8. Khan M, Shah T (2015) An efficient chaotic image encryption scheme. Neural Comput Appl 26(5):1137–1148
9. Kumar S, Panna B, Jha RK (2019) Medical image encryption using fractional discrete cosine transform with chaotic function. Med Biol Eng Comput 57(11):2517–2533
10. Zhang Y (2019) Security analysis of chaos triggered image encryption scheme. Multimed Tools Appl 78(22):31303–31318
11. Dagadu JC, Li JP, Addo PC (2019) An image cryptosystem based on pseudorandomly enhanced chaotic DNA and random permutation. Multimed Tools Appl 78(17):24979–25000
12. Ramasamy P, Ranganathan V, Kadry S, Damaševičius R, Blažauskas T (2019) An image encryption scheme based on block scrambling, modified zigzag transformation and key generation using enhanced logistic-tent map. Entropy 21(7), art. no. 656
13. Alawida M, Samsudin A, Teh JS, Alkhawaldeh RS (2019) A new hybrid digital chaotic system with applications in image encryption. Sig Process 160:45–58
14. Yavuz E (2019) A novel chaotic image encryption algorithm based on content-sensitive dynamic function switching scheme. Opt Laser Technol 114:224–239
15. Manoharan S (2019) A smart image processing algorithm for text recognition, information extraction and vocalization for the visually challenged. J Innov Image Process (JIIP) 1(01):31–38
16. Parvees MYM, Samath JA, Bose BP (2019) Chaotic sequences are cryptographically secure now-an improved chaotic approach. J Comput Theor Nanosci 16(4):1438–1443

17. Li B, Liao X, Jiang Y (2019) A novel image encryption scheme based on improved random number generator and its implementation Nonlinear Dyn 95(3):1781–1805
18. Dhall S, Pal SK, Sharma K (2018) Cryptanalysis of image encryption scheme based on a new 1D chaotic system. Signal Process 146:22–32

Analysis of Attention Deficit Hyperactivity Disorder Using Various Classifiers

Hensy K. George and P. K. Nizar Banu

Abstract Attention Deficit Hyperactivity Disorder (ADHD) is a neurobehavioral childhood impairment that wipes away the beauty of the individual from a very young age. Data mining classification techniques which are becoming a very important field in every sector play a vital role in the analysis and identification of these disorders. The objective of this paper is to analyze and evaluate ADHD by applying different classifiers like Naïve Bayes, Bayes Net, Sequential Minimal Optimization, J48 decision tree, Random Forest, and Logistic Model Tree. The dataset employed in this paper is the first publicly obtainable dataset ADHD-200 and the instances of the dataset are classified into low, moderate, and high ADHD. The analysis of the performance metrics and therefore the results show that the Random Forest classifier offers the highest accuracy on ADHD dataset compared to alternative classifiers. With the current need to provide proper evaluation and management of this hyperactive disorder, this research would create awareness about the influence of ADHD and can help ensure the proper and timely treatment of the affected ones.

Keywords ADHD · Bayes net · Classification · Data mining · J48 decision tree · Logistic model tree · Naïve bayes · Random forest · Sequential minimal optimization

1 Introduction

ADHD is predominantly a neurodevelopmental disorder of childhood with symptoms like hyperactivity, lack of attention or concentration, and an inclination toward impulsive behaviors [1]. It makes a person to be restless and constantly active, unable to manage their impulses. They may also have problems like sleep and anxiety issues

H. K. George · P. K. Nizar Banu (✉)
Department of Computer Science, CHRIST (Deemed to be University), Bangalore, India
e-mail: nizar.banu@christuniversity.in

H. K. George
e-mail: hensygeorge07@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_28

typically noticeable before the age of six. It is not simply a childhood disorder but can continue through adolescence and childhood.

The main cause of ADHD is due to the low transmission of neurotransmitters, dopamine, and noradrenaline [2], within the brain regions—prefrontal cortical area and basal ganglia. Dopamine is responsible for the reward centers of the brain and it regulates mood. And so, the individual drive to seek out for the rewarding feeling, when the dopamine level goes down. Dysfunction of the prefrontal cortex region ends up in a shortened attention span, lack of alertness, and decreased efficiency of working, unable to begin and continue activities, and to differentiate and avoid unnecessary or disrupting events. Due to all these ADHD-affected individuals have diminished focus on everything. ADHD can also be due to the premature birth of the baby or due to the genes inherited from their parents or due to any brain damage [2].

ADHD can be categorized as those predominantly hyperactive, highly inattentive or sometimes hyperactive, and inattentive type [3]. ADHD predominantly hyperactive-impulsive kind show both impulsive and hyperactive behavior; however, they will not show enough symptoms of inattention to be added into the combined type. Those predominantly inattentive types have inattention but not hyperactive or impulsive behavior. This category was earlier referred to as Attention Deficit Disorder (ADD). Among the three categories, ADHD combined type is the most common type of disorder with signs like inattention, hyperactivity, and impulsivity. Children with inattentiveness get easily distracted, often forgetful, switching from one activity to another. They get easily bored and cannot focus even on one task properly. Hyperactive and impulsiveness shows perverse physical movement, cannot sit still or wait their turn. They interrupt conversations and have little or no sense of fear. They get into inappropriate situations, impatience, lack of emotional restraint and act without regard to consequences, and speak out of turn [3].

Anxiety disorder leads to a kind of uneasy behavior most of the time. It may increase heart beat, sweating, and dizziness. Oppositional deficit disorder may show undesirable and troublesome behavior especially toward people at an authoritative level. Conduct behavior results in harmful behavior like robbery, aggressiveness, causing damage, hurting people, or animals. Depression symptoms show sleep disorders; difficult to get sleep at night. Epilepsy affects the brain causing repeated fits or seizures. Dyslexia leads to learning difficulties.

Although there is no cure for ADHD, early diagnosis of the same can help the individual to effectively manage this disorder. The affected individuals can easily recover through psychotherapy, counseling, behavioral training, and medications. The timely and correct treatments, healthy food, enough sleep and exercise, and above all supportive parents who know how to respond to ADHD can make the individual focus, work, and learn, reducing their hyperactive and impulsive nature [4].

Many diagnoses and evaluations were done in the previous years among different age groups to find the existence of ADHD in children and adults. Primary school children of age 6–11 from four different schools in Coimbatore district was tested for ADHD and was found that ADHD is present more in male than in females and is highest among the age group 9–10 [5]. The children age 5–12 years old was tested and found that the prevalence of ADHD is 1.3% with a male and female ratio of 1.6:1 which constitutes a major health problem [6]. ADHD may affect a good part of a child's life. It affects not only the child but all the members of his/her family as well. The adverse effects of ADHD change with different aspects of the disorder being more common at different stages from kindergarten years to primary school and adolescence [7]. A large population-based sample of middle-aged adults was investigated for ADHD symptoms and was found that there is a compelling need for a better understanding of advanced development of age-appropriate approaches, to the diagnosis of ADHD in mid to late adults [8]. The negative effect of ADHD symptoms on the daily life of adults, few methods, and medications as the first-line treatment for adult ADHD is also suggested [9]. The aim of the work in this paper is to identify the presence and influence of ADHD using classifiers such as Naïve Bayes, Bayes Net, Sequential Minimal Optimization, Random Forest, J48 decision tree, and Logistic Model Tree.

2 Literature Review

Most of the individuals especially children show some kind of abnormality in their behaviors from time to time. Often it is difficult to differentiate their behaviors that express learning disabilities, from normal behavior. To prove the strong evidence of ADHD, a number of studies using various techniques and algorithms were carried on a different dataset based on age, risk factors, and other components related to ADHD.

Data mining techniques were applied to find the learning disabilities of school-going children. Signs and symptoms identified from the real dataset were used for the general assessment of learning disabilities [10]. Artificial intelligence techniques like Support Vector machine Neural networks and decision trees were implemented to find out the learning disabilities in school-going children. An ADHD dataset of children of age 5–6 age was classified into ADHD types using Naïve Bayes and J48 classifiers. The main intention was to make alertness regarding the effect of ADHD among the school-going children of age 5–6 years. The evaluation of the data using confusion matrix, ROC curve, and various accuracy measures showed J48 classifiers with 100% accuracy on ADHD than Naïve Bayes [11]. The analysis of ADHD was carried on using MLP and SVM classifiers among the children of 5–9 years of age. The result was evaluated with kappa statistics, mean absolute error, and ROC area and found that MLP is comparatively good for ADHD data classification than SVM [12]. Naïve Bayes and Decision tree classifiers were used to classify children and adolescence as ADHD and Non-ADHD through the supermarket game. Using only two classes these algorithms achieve a satisfactory accuracy with sensitivity and specificity of around 0.7 [13]. Diagnosis of ADHD dataset of 105 samples in the age group of 6–9 and gender was classified using statistical measures like mean and standard deviation. Two parameters, namely, age and gender were taken to check its significance on ADHD. The dependence of these variables was carried out using the Chi-square test. Analysis proved that gender is one of the important aspects [14] for the cause of ADHD. Understanding the risk factors that influence ADHD proves that, there is strong evidence of an inherited contribution to ADHD, although non inherited factors like environmental risks and change events are also important [15]. Association Rule Mining technique was used to explore the comorbidity of ADHD for Korean insurance data (NHID) [16]. It was found that the most prevalent comorbid psychiatric disorder of ADHD youths was mood/affective disorders and the highest association was found between disorders of scholastic skills and ADHD. Tic disorder which is the sudden, rapid, and non-rhythmic movement of the individual was an important role in the association between ADHD and other comorbid diseases [16]. The influence of ADHD in children age 2–15 was checked. The data samples were classified using the C5.0 algorithm and the result was analyzed using a confusion matrix. It was proved that any child can be influenced by ADHD due to brain abnormalities and birth risk factors [17]. Research on ADHD due to brain damage was done with 105 samples and found that a child will unquestionably have ADHD indications if it has brain damage or anomaly [18]. Imaging genetics is the advanced test of ADHD that links the genetic background of ADHD and the brain structure of the individual. It primarily recognizes genes that cause the brain disparities, provides a more clear understanding of how the genes outline the brain variation [19, 20].

This paper focuses on the analysis of ADHD disorder applying various classifiers like Naïve Bayes, Bayes Net, Sequential Minimal Optimization, Random Forest, J48 decision tree, and Logistic Model Tree classifiers on ADHD200 dataset. The results are analyzed and discussed based on different performance metrics and error measures like accuracy, time, kappa statistics, mean absolute error, root mean squared error, relative absolute error, and root relative squared error.

3 Experimental Analysis

3.1 Dataset

Data collection plays a vital role in every field. The dataset used here is taken from the ADHD-200 dataset. The ADHD-200 Global Competition made an attempt to assign ADHD-affected individuals into different categories. Teams were organized to provide the best method for diagnosing individuals with ADHD from their resting-state fMRI scans. As a result, a large dataset ADHD-200, with data from participants scanned at eight different sites was made available by ADHD-200 Consortium. It is the first publicly available dataset. The Peking University (PekingU), Kennedy Kriegar Institute (KKI), Bradley Hospital/Brown University (BrownU), NeuroIMAGE Sample (NeuroIMAGE), Oregon Health and Science University (OHSU), University of Pittsburg (UPitt), New York University Child Study Centre (NYU), and Washington University in St. Louis (WashU) are the different sites [21].

Table 1 ADHD dataset attributes

S. No	Attributes	Description
1	Age	Age of the patient
2	Gender	Male/female (1/0)
3	Handedness	Left/right/ambidextrous (0/1/2)
4	ADHD index	ADHD symptom rating of the patient (numeric)
5	Inattentiveness	Patients score of inattentiveness (numeric)
6	Hyper/Impulsiveness	Patients score of hyper/impulsiveness (numeric)
7	VerbalIQ	The score derived from selected subtests from the Wechsler Intelligence Scales (numeric)
8	PerformanceIQ	The score derived from selected subtests from the Wechsler Intelligence Scales (numeric)
9	ADHD	noADHD/ADHDmod/ADHDhigh (ADHD measure$\leq$30/ADHD>30/ADHD>50)

3.2 Data Preparation and Attribute Selection

When the data collected is of good quality the decisions made will be proper and accurate. But if the data contains missing values, impossible data combination, out of range values it can produce misleading results leading to a huge problem.

As discussed in [21], the ADHD-200 dataset has 24 attributes and 973 records. While experimenting we found few attributes like quality control and intelligent quotient have partially missing and null values and it does not participate in providing good classification accuracy. So, these attributes were ignored and the ADHD dataset is reduced to Gender, Age, Handedness (left, right, Ambidextrous), ADHD-Index, Inattentiveness, Hyper/Impulsiveness, Verbal IQ, Performance IQ, and ADHD (Output classes-no ADHD, ADHD mod, ADHD High) attributes.

The dataset with its attributes and classification is listed in Table 1. ADHD dataset with 406 records was given as input for data preprocessing. During the data cleaning and preprocessing stage, duplicate records, missing data, impure data, and unpredictable data are removed. The quality of the data can be measured in terms of accuracy, completeness, consistency, timeliness, believability, and interpretability. Figure 1 shows the output of preprocessed data.

3.3 Classification Methods

When the data is ready, the next step is the classification of data using various classifiers. This means to assign a collection of data into various categories for more accurate predictions which makes the analysis effective. The main goal of the classification is to predict the target classes. Here the dataset is divided as testing and training dataset and the classifier is trained using the training dataset. The correctness

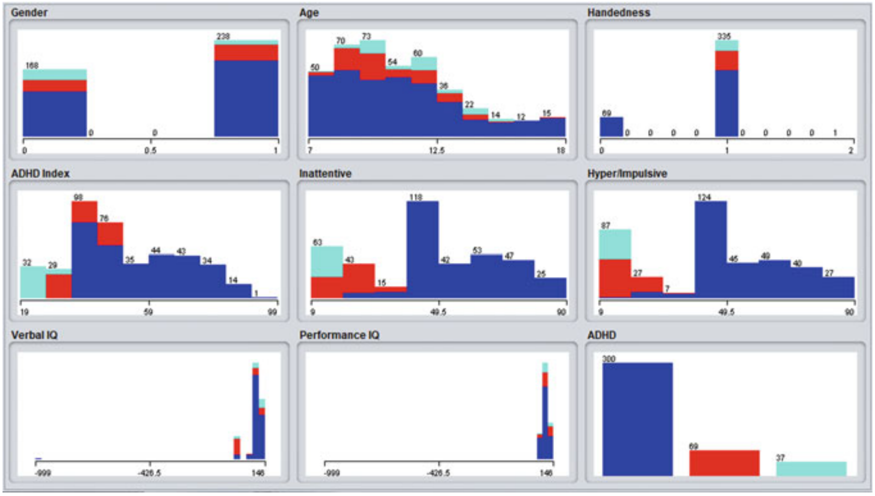


Fig. 1 Preprocessed output

of the classifier could be tested using the test dataset. The data mining classifiers, namely, Naïve Bayes, Bayes Net, Sequential Minimal Optimization, Random Forest, J48 decision tree, and Logistic Model Tree are applied to the ADHD dataset and analyzed.

3.3.1 Naive Bayes Algorithm

The Naïve Bayes algorithm is a collection of classification algorithms based on Bayes probability theorem. It has independent assumptions between the predictors and calculates the probabilities of every factor to bring a better outcome with the highest probability. This algorithm is highly matched when the measure of the input is high.

The Bayes rule states that

$$P(A|B) = P(B|A)P(A)/P(B) \tag{1}$$

where $P(A|B)$ is probability of occurrence of B when A is given, $P(A)$ is the probability of A, $P(B/A)$ is probability of occurrence of A when B is given, $P(B)$ is the probability of B. Bayes theorem allows to find the latter probability P($A|B$) when individual prior probabilities $P(A)$, $P(B)$, and $P(B|A)$ is known.

The result of the Bayesian classifier is good when the attributes of the training dataset are categorical and already classified. When the new pattern without the class label is given to the Naïve classifiers it obtains the new pattern and gives the output class label for which the probability calculated is the maximum according to the probabilistic calculations [11].

3.3.2 Bayes Net

A Bayes net classifier signifies a graphical model with a set of nodes as variables and edges as their conditional dependencies. They are represented through a directed acyclic graph [22]. The variables that are conditionally independent of each other are the nodes that are not connected. Each node is linked with a function that takes nodes parent variables values as input and gives the possibility of the variable represented by the node as output. Bayesian networks are very accurate in guessing the possibility by getting into an already existing event and having a look at numerous possible reasons and examining if any one of several possible known sources was the contributing feature. For example, when the symptoms are clear the algorithm can easily identify the chances of the presence of various diseases thus representing the relationships between diseases and symptoms.

3.3.3 Sequential Minimal Optimization

Sequential Minimal Optimization (SMO) is an easy and effective productive algorithm that can easily find a good solution for the Support Vector machine problem. It is possible through Quadratic Programming libraries, removing additionally added matrix storage and numerical QP optimization steps. It needs no added matrix storage at all. The SMO breaks the complete QP problem into sequences of sub-problems and resolves the smallest optimization problem at every step [23]. It selects two Lagrange multipliers to mutually optimize, find, and update the SVM to produce the new finest values analytically. Numerical QP optimization is evaded fully. This is the great advantage of SMO. There is no need for an entire QP library routine here, but the inner loop of the algorithm can be expressed in a very short quantity of C code. Thus, the complete QP problem is solved quickly.

3.3.4 Random Forest Algorithm

Random Forest is a classification technique where individual multiple trees are merged together to form the Random forest. It uses bagging methods to create an entire forest of random uncorrelated decision trees to reach the best productive results. Every tree here produces some class results and the model's prediction is the class with the most votes [24]. There is no random forest model without the decision trees. In the Random Forest theorem, it can select only from a random subset of features that gives more variations among the trees in the model and lower correlation across trees and more diversification. Compared with the other classification algorithms Random Forest algorithm has the following advantages like when this algorithm is used in any classification problem the overfitting problem will never occur and classification and regression task can be solved through the Random Forest algorithm. It can be used for identifying the most important features out of the existing features from the dataset.

Table 2 Features and applications of the classifiers

Classifiers	Features	Applications
Naive Bayes algorithm	Classification process	Intrusion detection [27]
Bayes Net algorithm	Classification process	Tool for expert systems [28]
Sequential minimal optimization	Quadratic programming	Training support vector machine [23]
Random forest algorithm	Classification and regression	Breast cancer diagnosis [29]
J48 decision tree	Classification process	Prediction of diabetes [30]
Logistic model tree	Predictive analysis	Fault diagnosis [31]

3.3.5 J48 Decision Tree

The J48 decision tree is a technique which brings out all the possible solution to a decision, based on some condition. In other words, it is to explore the method the attributes-vector performs for a number of records. Here on the basis of the training dataset, the newly generated instances are evaluated and their category is found. Larger programs are generally split into more than one class and the data is classified based on the model. The main objective is the simplification of a decision tree until it increases steadiness in flexibility and accuracy [25].

3.3.6 Logistic Model Tree

A ‘logistic model tree’ is made by growing a typical classification tree and building logistic regression models for all nodes. Some of the subtrees are pruned using a pruning measure, and all the logistic models are joint along a path into one model in some fashion [26].

The features and applications of various classifiers applied and analyzed for the ADHD dataset are shown in Table 2.

4 Results and Discussion

The results and comparison of the classifiers are studied and discussed based on measures like accuracy, time, kappa statistics, root mean squared error, mean absolute error, root relative squared error, and relative absolute error.

Table 3 Performance results

Classifiers	Accuracy (%)
Random forest	99.75
LMT	99.51
J48	99.01
Bayes net	97.31
Naïve Bayes	96.80
SMO	93.35

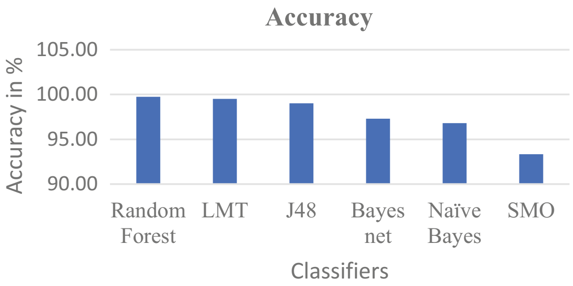


Fig. 2 Accuracy of the classifiers

4.1 Accuracy

Accuracy is the closeness of a measure to the actual value of what is being measured. It is the capability of the model to correctly predict the class labels of new or previously unseen data. Table 3 shows the performance results of all the classifiers.

From Table 3 we understand that the Random Forest classifier gives more accuracy in the analysis of ADHD dataset taken in this paper. The graphical representation in Fig. 2 depicts the same. The Random Forest classifier does an effective classification among other classifiers.

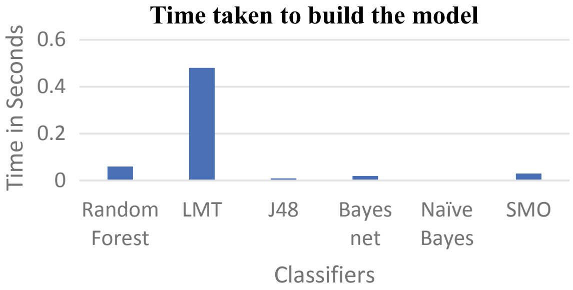
4.2 Time

Time refers to the amount of time spent to build the model. Table 4 gives the time taken by each classifier to produce the results. Figure 3 shows that the Random Forest classifier takes a reasonable time of 0.06 s to build the model and bring out the accurate results.

Table 4 Time taken to build the model

S. No.	Classifiers	Time (s)
1.	Random forest	0.06
2.	LMT	0.48
3.	J48	0.01
4.	Bayes Net	0.02
5.	Naïve Bayes	0.00
6.	SMO	0.03

Fig. 3 Time taken to build the model



4.3 Kappa Statistics

Kappa statistic is a very good measure to identify how closely the instances are classified by the machine learning classifier. It matches the data labeled as ground truth, controlling for the accuracy of a random classifier as measured by the predicted accuracy. The graphical representation of the kappa statistics measure in Fig. 4 shows that the best classifier on the ADHD dataset is the Random Forest classifiers.

Fig. 4 Kappa Statistics metrics

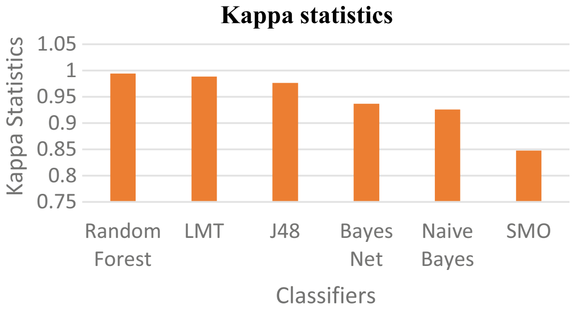
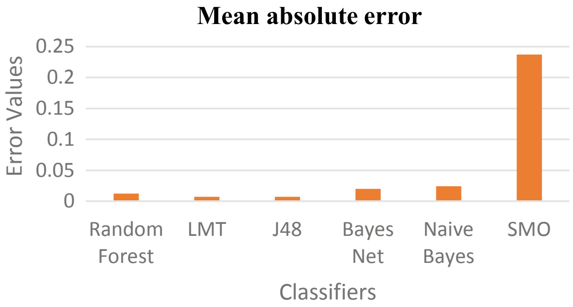


Fig. 5 MAE metrics



4.4 Mean Absolute Error

Mean Absolute Error measures the average of the absolute value of the difference between the predicted and the actual values. It is a little smaller than the Root Mean Square Error. Figure 5 shows the analysis of all the classifiers based on Mean Absolute Error.

4.5 Root Mean Squared Error

Root Mean Squared Error is used to calculate the difference between the predicted and actual value. It is used to measure accuracy. It is the standard deviation of the predicted errors. The RMSE will continually be greater or equal to the Mean Absolute Error. Figure 6 shows the analysis of all the classifiers based on Root Mean Squared Error.

Fig. 6 RMSE metrics

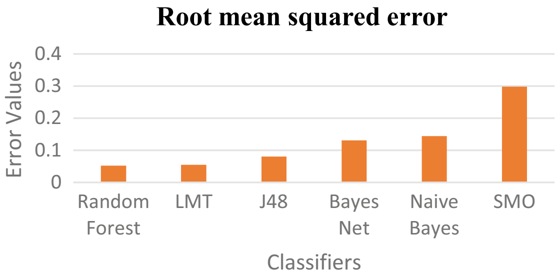
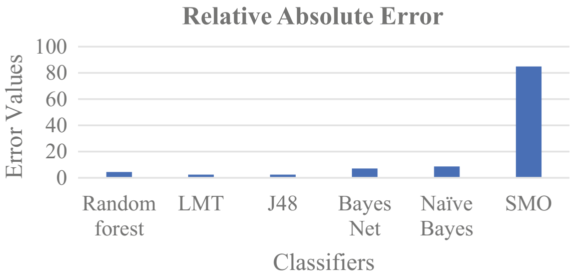


Fig. 7 RAE metrics



4.6 Relative Absolute Error

The root absolute error is the difference between the actual value and the individual measured value. It is the ratio of dividing the absolute error by the magnitude of the exact value. Figure 7 shows the analysis of all the classifiers based on Relative Absolute Error.

4.7 Root Relative Squared Error

The root relative squared error is relative to what the error would have seen if the average of the actual values had been used. Figure 8 shows the analysis of all the classifiers based on Root Relative Squared Error. Representation of the kappa statistics and the error measures of all classifiers in Table 5 shows that the Random Forest classifier has the minimum error based on RMSE and RRSE among other classifiers. The MAE, RAE, and RRSE metrics of SMO classifier is very high because it does not match exactly with the prior probabilities of the classes observed in the dataset.

Fig. 8 RRSE metrics

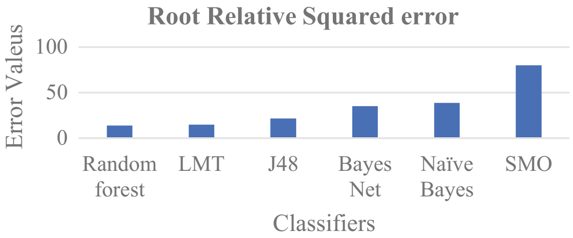


Table 5 Kappa statistics and error measures

Parameters	Random forest	LMT	J48	Bayes Net	Naive Bayes	SMO
Kappa statistic	0.994	0.988	0.976	0.937	0.926	0.848
Mean absolute error	0.013	0.007	0.007	0.02	0.024	0.24
Root mean squared error	0.052	0.055	0.082	0.131	0.145	0.298
Relative absolute error	4.487	2.580	2.539	7.147	8.725	84.872
Root relative squared error	14.012	14.812	21.655	35.113	38.819	79.963

5 Conclusion

In this paper, the classification of Attention Deficit Hyperactivity Disorder using different data mining classifiers is analyzed and discussed. The dataset used here is prepared from the ADHD-200 dataset. The knowledge obtained through this work using data mining methods and various classifiers can surely improve the diagnosis of ADHD disorder. Analysis and performance metrics prove the Random Forest classifier to give the best accurate results. As ADHD is one of the dangerous diseases spreading across India with 11.32% ADHD-affected primary school children, and ADHD data collection is a challenging task, researches can focus on developing systems that generate data efficiently and algorithm for further diagnosis and analysis.

References

1. https://www.news-medical.net/health/ADHD-History.aspx
2. https://www.news-medical.net/health/How-does-ADHD-Affect-the-Brain.aspx
3. https://www.cdc.gov/ncbddd/adhd/index.html
4. https://www.downtoearth.org.in/news/health/study-finds-genetic-variants-that-increase-ADHD-risk-62281
5. Venkata JA, Panicker A (2013) Prevalence of attention deficit hyperactivity disorder in primary school children. Indian J Psychiatr 55(4): 338–342. https://doi.org/10.4103/0019-5545.120544
6. Ramya HS, Goutham AS, Pandit LV (2017) Prevalence of attention deficit hyperactivity disorder in school going children aged between 5–12 years in Bengaluru. Curr Pediatr Res 21(2):321–326. ISSN 0971-9032
7. Harpin VA (2005) The effect of ADHD on the life of an individual, their family, and community from preschool to adult life. Arch Dis Child 90(Suppl I):i2–i7. https://doi.org/10.1136/adc.2004.059006
8. Das D, Cherbuin N, Butterworth P, Anstey KJ, Easteal S (2012) A population-based study of attention deficit/hyperactivity disorder symptoms and associated impairment in middle-aged adults 7(2):e31500
9. Rosler MR, Casas M, Konofal E, Buitelaar J (2010) Attention deficit hyperactivity disorder in adults. World J Biol Psychiatr 11:684–698
10. Mary MT, Hanumathappa M (2014) Diagnosis of learning disabilities in school going children using data mining techniques: a survey. IJISET-Int J Innov Sci Eng Technol 1(8)

11. Radhamani E, Krishnaveni K (2016) Diagnosis and evaluation of ADHD using Naïve Bayes and J48 classifiers. In: International conference on computing for sustainable global development (INDIACom). IEEE. 978-9-3805-4421-2/16/$31.00
12. Radhamani E, Krishnaveni K (2016) Diagnosis and evaluation of ADHD using MLP and SVM classifiers. Indian J Sci Technol 9(19):93853. ISSN 0974-5645
13. Santos FEG, Bastos APZ, Andrade LCV, Revoredo K, Mattosy P (2011) Assessment of ADHD through a computer game: an experiment with a sample of students, p 104. IEEE. 978-0-7695-4419-9/11$26.00. https://doi.org/10.1109/vs-games.2011.21
14. Radhamani E, Krishnaveni K (2018) Diagnosis of ADHD using statistical measures. Int J Eng Res Comput Sci Eng (IJERCSE) 5(3). ISSN: 2394-2320
15. Thapar A, Cooper M, Jefferies R, Stergiakouli1 E (2012) What causes attention deficit hyperactivity disorder? Arch Dis Child 97:260–265. https://doi.org/10.1136/archdischild-260 2011-300482
16. Leejin KIM, Myoung S (2018) Comorbidity study of attention-deficit hyperactivity disorder (ADHD) in children: applying association rule mining (ARM) to Korean national health insurance data. Iran J Public Health 47(4):481–488
17. Radhamani E, Krishnaveni K (2017) Prognosis of ADHD using R programming and MATLAB Tools. Int J Emerg Technol Adv Eng 7. ISSN 2250-2459, ISO 9001:2008
18. Radhamani E, Krishnaveni K (2018) Analysis of brain data attributes to detect the prevalence of ADHD using R programming. Indian J Sci Technol 8(XI). ISSN 2249-7455
19. Wu Z, Yang L, Wang Y (2014) Applying imaging genetics to ADHD: the promises and the challenges. Mol Neurobiol 50:449–462. https://doi.org/10.1007/s12035-014-8683-z
20. Skogli EW, Teicher MH, Andersen PN, Hovik KT, Øie M (2013) ADHD in girls and boys – gender differences in co-existing symptoms and executive function measures. BMC Psychiatr 13:298
21. Brown MR, Sidhu GS, Greiner R, Asgarian N, Bastani M, Silverstone PH, Greenshaw AJ, Dursun SM (2012) ADHD-200 global competition: diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements. Front Syst Neurosci 28. https://doi.org/10.3389/fnsys.2012.00069
22. Devarakonda N, Pamidi S, Kumari V, Govardhan A (2012) Intrusion detection system using Bayesian network and hidden Markov model. Proc Technol 514(4):4,506
23. Platt JC (1998) Sequential minimal optimization: a fast algorithm for training support vector machines. Microsoft Research jplatt@microsoft.com. Technical report MSR-TR-98-14, © 1998 John Platt
24. Biau G (2012) Analysis of a random forests model. J Mach Learn Res 13:1063–1095
25. Arora R, Suman I (2012) Comparative analysis of classification algorithms on different datasets using WEKA. Int J Comput Appl 54(13):0975–8887
26. Landwehr N, Hall M, Frank E Logistic model trees. University of Freiburg, University of Waikato
27. Mukherjee S, Sharma N (2012) Intrusion detection using Naive Bayes classifier with feature reduction, 2212-0173. Elsevier Ltd. https://doi.org/10.1016/j.protcy.2012.05.017
28. Wiegerinck W, Kappen B, Burgers W (2010) Bayesian networks for expert systems: theory and practical applications. Springer, Berlin
29. Nguyen1 C, Wang Y, Nguyen HN (2013) Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic. J Biomed Sci Eng 6:551–560
30. Kaur G, Chhabra A (2014) Improved J48 classification algorithm for the prediction of diabetes. Int J Comput Appl 98(22):0975–8887
31. Nachiappan MR, Sugumaran V, Elangovan M (2016) Performance of logistic model tree classifier using statistical features for fault diagnosis of single point cutting tool. Indian J Sci Technol 9(47). https://doi.org/10.17485/ijst/2016/v9i47/107940

A Technique to Detect Wormhole Attack in Wireless Sensor Network Using Artificial Neural Network

Moirangthem Marjit Singh, Nishigandha Dutta, Thounaojam Rupachandra Singh, and Utpal Nandi

Abstract Wormhole attack is a harmful attack that disrupts the normal functioning of the network by manipulating routing protocols and exhausting network resources. The paper presents a technique that detects the presence of wormhole attack in wireless sensor network (WSN) using artificial neural network (ANN). The proposed technique uses connectivity information between any two sensor nodes as the detection feature. The proposed technique has been implemented considering the deployment of sensor nodes in the wireless sensor network area under uniform, Poisson, Gaussian, exponential, gamma & beta probability distributions. The proposed technique does not require any additional hardware resources and gives a comparatively high percentage of detection accuracy.

Keywords Wireless sensor network · Probability distribution · Wormhole attack · ANN

1 Introduction

Wireless sensor network (WSN) is defined as a set of sensor nodes deployed in a geographical area without any proper planning to sense, transmit, and process information. Originally, WSN was developed for use in military operations, but later

M. M. Singh (✉) · N. Dutta
Department of Computer Science & Engineering, North Eastern Regional Institute of Science & Technology, Nirjuli 791109, Arunachal Pradesh, India
e-mail: marjitm@gmail.com

N. Dutta
e-mail: nishidtt9@gmail.com

T. R. Singh
Department of Computer Science, Manipur University, Canchipur, Manipur, India
e-mail: rupachandrath@manipuruniv.ac.in

U. Nandi
Department of Computer Science, Vidyasagar University, Midnapore, West Bengal, India
e-mail: nandi.3utpal@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_29

on, it found a wide range of applications in industry, transportation, machine health, etc. The use of a wireless sensor network in these applications involves handling of sensitive information due to which the sensor network needs to be protected against possible security threats and attacks. Factors such as limitation of resources and poor deployment of sensor nodes in unattended environments make WSN susceptible to several kinds of dreadful attacks, such as black hole attacks, wormhole attacks, Sybil attacks, sinkhole attacks, selective forwarding attacks, etc. [1, 2].

One such dreadful attack is the wormhole attack, a severe attack in ad hoc networks that is particularly challenging to defend against, in which an attacker captures packets at one end of the network and replays them at the other end using illegitimate or malicious nodes. The wormhole attack can take place in spite of the authentication and encryption measures being deployed, therefore, detecting wormhole attack and defending against it is a necessary step for ensuring the overall security of the WSN [3–5].

A wormhole attack detection method has been proposed in this paper using artificial neural network (ANN) along with connectivity information as the detection feature. The paper is organized as follows: Sect. 2 describes the wormhole attack in brief. In Sect. 3, the concept of artificial neural network has been discussed in brief. Section 4 presents the related work. Section 5 describes the proposed technique while Sect. 6 provides the simulation results and analysis. Conclusion is given at Sect. 7 and the references at the end.

2 Wormhole Attack

In wormhole attack, the malicious node traps the data packets from a definite location in the network so as to tunnel them to itself that is located at a remote location [6]. For setting up a wormhole attack, an attacker needs at least two or more malicious nodes that establish a low latency and high bandwidth link between them. This wormhole link is located between two nodes that are placed with large mutual distance [7]. The wormhole link is fitted with radio transceivers which facilitates the attacker to capture packets at one end of the tunnel and replays them at the other end [8]. Although wormhole attack facilitates a high speed, long link that can be used to transfer packets between distant nodes without traversing several hops, making it a useful networking service, the attacker makes use of it for its own advantage. Using this wormhole link, the attacker attracts traffic in the network and then drops or modifies data packets selectively to disrupt the flow of data packets [9]. The attacker can record data traffic and use the gathered information to analyze and break security of the network. It can then turn off the wormhole link to create unnecessary routing events and also adversely influence network routing by creating a list of fake neighbors and affecting neighborhood discovery process [10].

The classification of wormhole attack is done by modeling it into three types so as to facilitate development of detection and prevention techniques in the network. This classification is done on the basis of three criteria: visibility of the attackers on the communication path; tendency of the wormhole nodes whether to show or hide their identities and their behaviors in terms of packet forwarding. The three wormhole attack models are Open wormhole, Half-open wormhole, and Close wormhole. In Open wormhole attack model, the source and destination nodes along with the two malicious nodes are visible while the legitimate nodes on the transmission path are kept hidden. After completing the route discovery process, the attacker's identity is then included in the packet header. The normal nodes think of the wormhole nodes as their neighbors even though they are aware of the wormhole nodes' presence during the packets transmission time [3, 7]. In Half-open wormhole attack model, out of the two malicious nodes present in the network, only the one near the source node is visible. The modification of data packets does not take place in this type of attack, instead the malicious nodes simply transmit the packets from one side of the link to the other side [3]. In the closed wormhole attack model, the identities of the malicious nodes along with the nodes in the path of transmission are kept hidden and the source and the destination nodes are made to believe as one hop away from each other. Here the contents of the data packet are not tampered with and the malicious nodes simply listen to the conversation between normal nodes [3, 11].

3 Artificial Neural Network

In recent years, artificial intelligence has been extensively used in intrusion detection mechanisms for protection of wireless sensor network against possible attacks due to its ability to improve detection accuracy. Artificial neural network (ANN) is however a simple network of computing cells capable of solving complex problems [12]. The architecture of an ANN can be categorized into three basic types, viz, Single-layer feed-forward networks, Multilayer feed-forward networks, and recurrent networks [13, 14]. The back propagation (BP) algorithm is a supervised learning (SL) algorithm that is used for training the multilayer neural network where the artificial neurons send the signals forward and propagate the error backwards [15, 16]. Training a neural network with BP will require the computation of the expected gradient of relevant error function with respect to weights and biases. Based on the learning rate, each iteration of BP updates the weights and the biases. The deployment of sensor nodes considered in the WSN area for the implementation purpose is based on six different types of probability distributions, namely, uniform, Poisson, normal, exponential, gamma, and beta probability distributions.

4 Related Work

Few methods have been proposed for detecting wormhole attack in wireless sensor networks as indicated by the current literature. A review of various techniques to detect wormhole attacks in WSN can be seen in [3]. Authors in [17] present a general method for wormhole attack detection named packet leashes and introduce two types of leashes, viz, temporal and geographical leashes. This method detects and thus defends against wormhole attacks by restricting the transmission distance of the data packets to a maximum allowable range; however, it works only for single wireless transmission. In temporal leash, the time at which a packet is sent is appended by the sending node, and the receiving node compares this value with the time at which it received the packet. A limit on the distance between the sender and the receiver is imposed in geographical leash, wherein the location of the sending node as well as the time at which the packet was sent is appended in the data packet. The drawback of geographical leash is that it always requires loosely synchronized clocks plus the nodes should be aware of their own locations.

Shaon and Ferens have proposed a wormhole attack detection method based on ANN using neighborhood hop count as the detection feature [18]. An approach of detecting wormhole attack in WSN based on connectivity information is found in [8]. The proposed algorithm is implemented in three various communication models (unit disk graph, quasi unit disk graph, and the model employed in TOSSIM simulator) using two different sensor node distributions (random distribution and grid distribution). The authors in [19] have presented another wormhole attack detection method based on the number of neighboring nodes and the round trip time. The proposed scheme is developed for wormhole attack detection in Ad hoc On-demand Distance Vector (AODV) protocol where the simulation is carried out using network simulator (ns-2). A wormhole attack detection mechanism called Multidimensional Scaling-Visualization of Wormhole (MDS-VOW) that makes use of multidimensional scaling and anomaly visualization can be seen in [20]. This method follows a centralized approach that leads to a single point of failure.

5 Proposed Work

The proposed algorithm uses connectivity information as the detection feature in place of neighborhood hop count to detect the presence of wormhole attackers in the WSN. It uses connectivity information as a measure of distance between two nodes where a limit on the communication range is imposed on the sensor nodes. The deployment of sensor nodes in an area of 1000 m × 1000 m is considered. A mobile sensor node designated as M_0 is randomly deployed in the WSN area that moves around the network area and collects the connectivity information between two random nodes in each location. The M_0 will continue to collect data samples even after enforcement of wormhole attack in the WSN. The maximum communication range of wormhole nodes is 150 m. The data samples collected by M_0 in both cases, i.e., WSN without wormhole attack as well as WSN with wormhole attack will be mixed up and stored in a dynamic array for future use. The attack detection percentage is obtained after training the neural network using the proposed algorithm. The proposed algorithm is given below.

```
Input: WSN area of 1000m X1000m, length of   wormhole link = 350m.
Output: Percentage detection accuracy.

1: Create a wireless sensor network
2: Deploy a mobile node, Mo that visits every location in the network area randomly and collects connectivity information between two random nodes in each location to form samples.
3: Enforce wormhole attack at random location in WSN area.
    3.1: Introduce 2 malicious nodes in the WSN area having a maximum communication range of 150m.
    3.2: Establish a link between them having a maximum length of 350m
4: Repeat Step 2 to collect data samples (wormhole samples) in the presence of wormhole attack.
5: Mix up the data samples collected in Step 2 and Step 4 in a data set called data in a dynamic array.
    5.1  : Denote the class of number of wormhole samples as 0
    5.2  : Denote the class of wormhole samples as 1
6: Store the number of wormhole and wormhole data samples
7: Sequentially select 70% data samples from the gathered data set, data obtained in Step 6 and store it in a data set, training data (dynamic array) for training purpose.
8: Sequentially select the remaining data samples (30%) from data and store it in a data set, testing data (dynamic array) for testing purpose.
9: Train the neural network with appropriate parameters.
10: Test the neural network
    10.1: If output ≥0.8, wormhole attack exists
    10.2: Else if training output < 0.8, wormhole attack does not exist.
11: Update the training data with testing data for further training of the neural network.
12: Update testing data with new wormhole samples collected by Mo
13: Repeat Step 9 and Step 10.
```

6 Result Analysis

The proposed algorithm has been simulated using Matlab with the simulation parameters given in Table 1. Simulation has been carried out for WSN configurations having 100, 200, 300, 400, and 500 nodes for the proposed algorithm as well as the one proposed by Shaon & Ferens.

The results of the simulations are given in Table 2. It can be observed from the results of simulation and the graphical plots that the proposed technique delivers equal or higher detection accuracy in almost all WSN configurations.

The graphical plot that provides comparison of detection accuracy for the proposed technique and existing technique (Shaon & Ferens) for WSN having 100, 200, 300, 400, and 500 nodes is given in Figs. 1, 2, 3, 4, and 5 respectively.

In WSN with 100 nodes, the highest detection accuracy of 97% is obtained using the proposed method when nodes are deployed in normal probability distribution as shown in Fig. 1. The detection accuracy is also the highest (96%) in normal node distribution for Shaon & Ferens' method. For WSN 100 nodes deployed in uniform, Poisson, exponential, and gamma distributions, the detection accuracy of the proposed method is equal to or higher than that of Shaon & Ferens' method except for beta distribution.

From Fig. 2, it can be seen that the highest detection accuracy of 95% is obtained using the proposed method when 200 nodes of WSN are deployed following normal probability distribution, while Shaon & Ferens' method gave detection accuracy of 91%. The detection rate using the proposed method is higher than that of Shaon & Ferens' method for WSN having 200 nodes.

It is observed from Fig. 3 that in WSN having 300 nodes the highest detection accuracy of 100% is obtained using proposed method under normal distribution of nodes. While the Shaon & Ferens' method gave an accuracy of 97% in the same WSN

Table 1 Simulation parameters

Parameters	Value
Detection feature	Connectivity information
Types of nodes distribution	Uniform, Poisson, normal, exponential, gamma, beta
Network field area	1000 m× 1000 m
Number of nodes	100, 200, 300, 400, 500
Max. communication range	50 m (normal nodes), 150 m (wormhole nodes)
Length of wormhole link	350 m
ANN architecture	[10, 10, 1]
Number of data points (Training)	140 × 100
Number of data points (Testing)	60 × 100

Table 2 Simulation results

Nodes	Detection accuracy (%)											
	Uniform		Poisson		Normal (Gaussian)		Exponential		Gamma		Beta	
	Proposed	Shaon & FerensJ	Proposed	Shaon & FerensJ	Proposed	Shaon & FerensJ	Proposed	Shaon & FerensJ	Proposed	Shaon & FerensJ	Proposed	Shaon & FerensJ
100	78	75	80	70	97	96	85	85	87	86	80	85
200	76	74	87	70	95	91	80	75	81	79	82	80
300	72	73	90	80	100	97	92	85	87	87	89	87
400	84	82	95	90	92	92	96	89	95	91	95	92
500	90	90	90	97	100	98	96	95	97	96	99	98

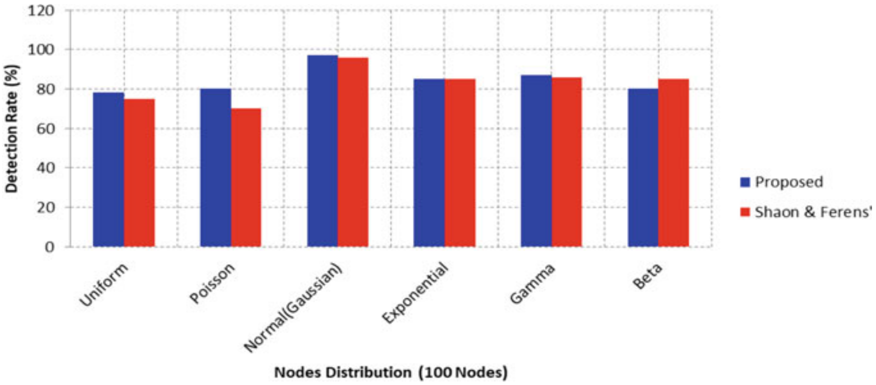


Fig. 1 Detection accuracy (%) for WSN having 100 nodes

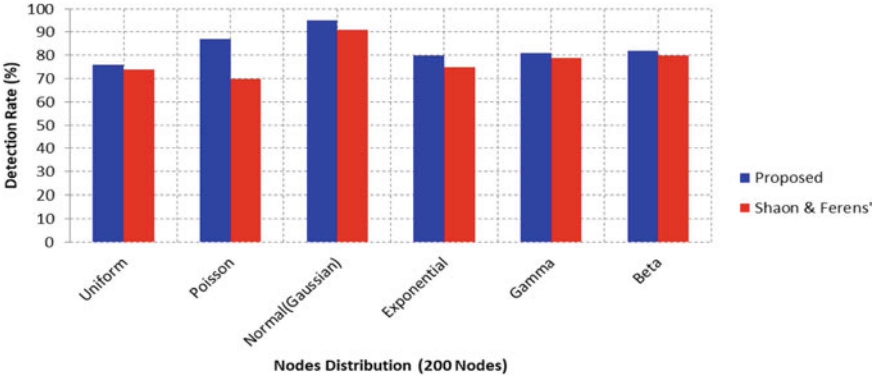


Fig. 2 Detection accuracy (%) for WSN having 200 nodes

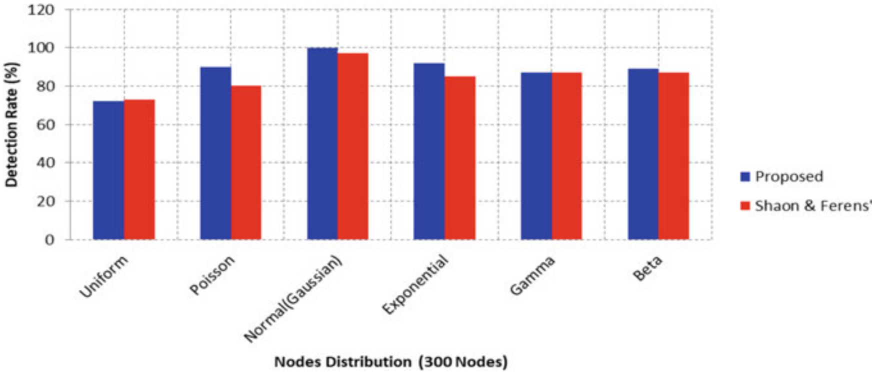


Fig. 3 Detection accuracy (%) for WSN having 300 nodes

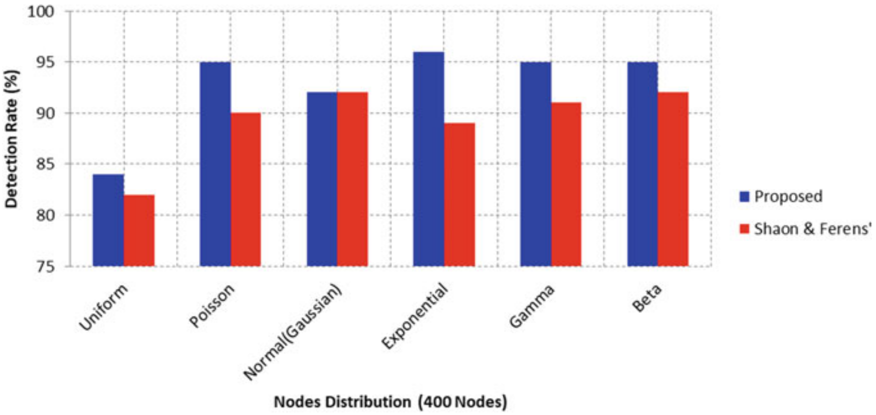


Fig. 4 Detection accuracy (%) for WSN having 400 nodes

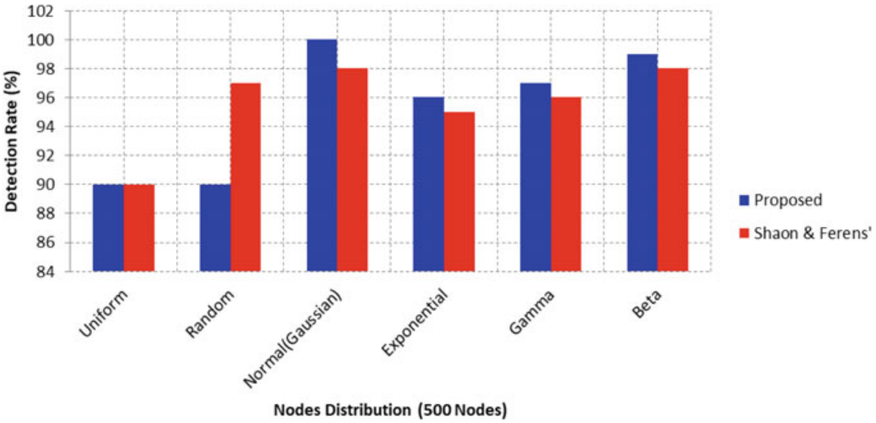


Fig. 5 Detection accuracy (%) for WSN having 500 nodes

configuration in normal distribution. In all types of nodes distributions, the detection accuracy rate of the proposed method is higher than that of Shaon & Ferens' method except in the case of uniform distribution of nodes.

Figure 4 shows the percentage detection accuracy for WSN with 400 nodes. In this case, the highest detection accuracy is obtained using the proposed method when the nodes are deployed in exponential distribution. In case of WSN with 500 nodes, the highest detection accuracy is obtained using the proposed method when the nodes are deployed in normal distribution as shown in Fig. 5.

7 Conclusion

This paper presents a technique to detect the presence of wormhole attacks in WSN using ANN where connectivity information is used as the detection feature. The deployment of sensor nodes has been done based on six different types of probability distributions, namely, Uniform, Poisson, Normal, Exponential, Gamma and Beta probability distributions. It can be seen from the simulation results that the proposed technique gives a higher percentage of detection accuracy for almost all WSN configurations having nodes 100–500 nodes. Further improvements in detection accuracy can be achieved by applying deep learning techniques and by training with large datasets.

References

1. Tomic I, McCann JA (2017) A survey of potential security issues in existing wireless sensor network protocols. IEEE Internet of Things J 4(6):1910–1923
2. Yang G, Dai L, Wei Z (2018) Challenges, threats, security issues and new trends of underwater wireless sensor networks. Sensors 18(11):3907
3. Dutta N, Singh MM (2019) Wormhole attack in wireless sensor networks: a critical review. In: Advanced computing and communication technologies. Springer, pp 147–161
4. Singh MM, Dutta N (2017) Security issues in wireless sensor networks. Int J Distrib Cloud Comput 05(02):7–16
5. Xie H, Yan Z, Yao Z, Atiquzzaman M (2018) Data collection for security measurement in wireless sensor networks: a survey. IEEE Internet of Things J 6(2):2205–2224
6. Singh R, Singh J, Singh R (2016) WRHT: a hybrid technique for detection of wormhole attack in wireless sensor networks. Mob Inf Syst 2016:13pp. Hindawi Publishing Corperation. Article ID 8354930
7. Maidamwar P, Chavhan N (2012) A survey on security issues to detect wormhole attack in wireless sensor network. Int J AdHoc Netw Syst (IJANS) 2(4):37–50
8. Maheshwari R, Gao J, Das SR (2007) Detecting wormhole attacks in wireless networks using connectivity information. In: Proceedings of 26th IEEE international conference on computer communications, Barcelona, Spain, May 2007
9. Pooja, Chauhan RK (2017) Review on security attacks and countermeasures in wireless sensor networks. Int J Adv Res Comput Sci 8(5):1275–1283
10. Bendjima M, Feham M (2016) Wormhole attack detection in wireless sensor networks. In: Proceedings of SAI computing conference, London, UK, 13–15 July 2016
11. Sharma N, Singh U (2014) Various approaches to detect wormhole attack in wireless sensor network. Int J Comput Sci Mob Comput 3(2):29–33
12. Russell SJ, Norvig P (2003) Artificial intelligence - a modern approach, 2nd edn. Prentice Hall series in artificial intelligence
13. Baig ZA, Khan, AI (n.d.) DDoS attack modeling and detection in wireless sensor networks. Mob Intell 595–626. https://doi.org/10.1002/9780470579398.ch26
14. Ray S (2019) Essentials of machine learning algorithms. https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/. Accessed 30 Oct 2019
15. Buntine WL, Weigend AS (1991) Bayesian back-propagation. Complex Syst 5(6):603–643
16. Hecht-Nielsen R (1992) Theory of the backpropagation neural network. In: Neural networks for perception. Elsevier, pp 65–93
17. Hu YC, Perrig A, Johnson DB (2006) Wormhole attacks in wireless networks. IEEE J Sel Areas Commun 24:370–380

18. Shaon MNA, Ferens K (2015) Wireless sensor network wormhole detection using an artificial neural network. In: Proceedings of the international conference on wireless networks (ICWN), the steering committee of the world congress in computer science, computer engineering and applied computing (World-Comp), pp 115–120
19. Tun Z, Maw AH (2008) Wormhole attack detection in wireless sensor networks. World Acad Sci Eng Technol 46:545–550
20. Wang W, Bhargava B (2004) Visualization of wormholes in sensor networks. In: Proceedings of 3rd ACM workshop on wireless security (WiSe'04), Philadelphia, USA, October 2004, pp 51–60

A Survey on Methodologies and Algorithms for Mutual Authentication in IoT Devices

Rashmi R. Sonth, Y. R. Pranamya, N. Harish Kumar, and G. Deepak

Abstract Authentication refers to the process of proving the identity of the IoT device or simply, it is an act of verifying the identity of the system. The scenario where two communicating parties authenticate each other at nearly the same time is called two-way authentication or mutual authentication. Mutual authentication provides trusted communication between remote IoT devices. This paper is aimed at performing a survey on various algorithms used for mutual authentication of IoT devices. The classification has been done based on the parameters used for authentication, its simplicity, optimality, and efficiency of the algorithm. An abstract analysis of the methodologies is illustrated.

Keywords Elliptic Curve Cryptography (ECC) · Internet of Things (IoT) · Mutual authentication · Confidentiality

1 Introduction

In today's world, where no network is immune to external attacks, a stable, secure, as well as an efficient system is essential for protecting the client data. As technology has improved and the network is becoming bigger and bigger, data security, confidentiality, and authentication have become one of the most important aspects of the organization to consider.

R. R. Sonth (✉) · Y. R. Pranamya · N. Harish Kumar · G. Deepak
Department of CSE, Dayananda Sagar College of Engineering, Bengaluru, India
e-mail: rashmisonth1999@gmail.com

Y. R. Pranamya
e-mail: pranamyabhat03@gmail.com

N. Harish Kumar
e-mail: nhk015@gmail.com

G. Deepak
e-mail: deepak.dsce@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_30

Network security is the process of securing communication and preventing intrusion into the corporate network. It focuses on how those devices interact with each other. Data security deals with protecting privacy measures that are applied to restrict unauthorized access for devices that are not secure. While IoT attacks in businesses are inevitable, we can become proactive in avoiding threats to the network and safeguarding valuable data from hackers. The security of IoT is a domain of technology that aims at protecting the IoT devices and network connecting them. Emerging technologies like Cryptography and Blockchain can be used to provide this high level of security.

Figure 1 shows the IoT architecture and how the objects are connected through the Internet. Figure 2 shows the interaction of objects and users in IoT Environment. While interacting with the objects authentication [1] of those objects and privacy has to be preserved. This is the main goal in the IoT environment. Section 2 discusses some authentication approaches discussed by various authors and their methodologies.

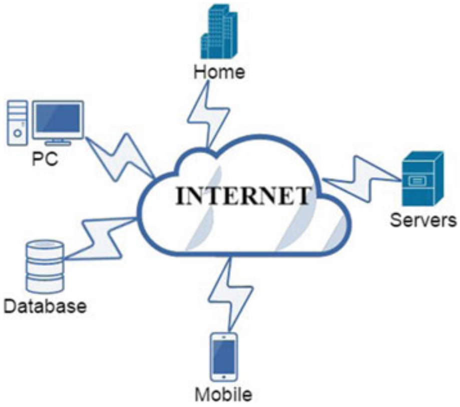


Fig. 1 IoT architecture

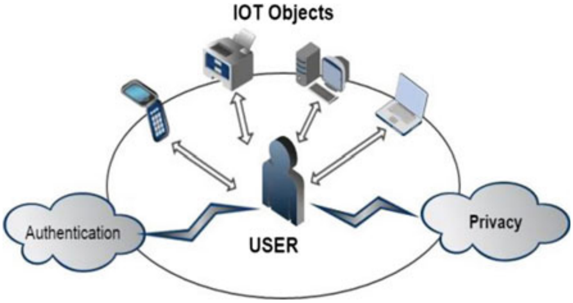


Fig. 2 User interaction with IoT objects

2 Literature Survey

Aman et al. [2] have presented a lightweight method for mutual authentication of IoT devices using Physical Unclonable Function (PUF). Here, it is assumed that there is a secure way for the server to obtain the initial Challenge-Response Pair (CRP) in offline mode. The server stores this CRP as a hashed value for the protection of user identity. Here, an IoT device sends the hash of its ID along with a nonce 1 (which is randomly generated) to the server. Now, the server searches its memory for that hashed value and if the match is not found, authentication fails. Else, the server will select the CRP stored against this hash value and generates another random nonce 2. The sender compares the hash function obtained by inputting parameters (sent by the server) with the hash function obtained from the server. This compared value gives either a successful or failed authentication.

Li and Cao [3] have provided One-Time Signature (OTS) using ECC. Firstly, the signer (sender) generates t random strings, picks a timestamp, applies the hash function, and sends it along with the message to find a valid signature. The timestamp and the message are concatenated to obtain a pointer for the hash function. Using this hash function, t hashed strings are obtained. Now, the signer is sought for a three-way collision within t hashed values. On encountering a collision, it is added to the signature set and the process is terminated. Else, on finding the signature, a temporary public key is generated using Elliptic Curve (EC). The signer then sends the message, the signature, the timestamp, the temporary public key, and the signer's ID to the verifier. The verifier checks the freshness of the message using a timestamp and then computes the values using the hash function indicated by the signer's ID from the signature. It then verifies them against the received value and the validity of the message. Otherwise, the verifier rejects the message and the process is terminated.

He et al. [4] have proposed a method that is based on broadcast encryption method in which a sender encrypts a message and broadcasts them over all channels, only the authenticated receivers will be able to decipher the message. The paper focuses on identity-based broadcast encryption that includes four tuple algorithm namely–Setup that includes setting up of master key ms and public parameters ps. Extract-inputting master key ms along with identity ID it produces a private key skid for identity ID. Encrypt, which gives a cipher text ct on a combination of ps, receiver set rs, and message M. Decrypt Message DM is received on ps, private key skid, and ciphertext ct. This brings in many practical applications like satellite TV subscriptions and social network service.

Mehmood et al. [5] proposed an authenticated certificates encryption scheme using public key topped by Diffie–Hellman (CDH) and bilinear Diffie–Hellman methods. The paper proves confidentiality by the INDCCA scheme where is a third party is involved in issuing the private key and the rest of the method is the same as the public key encryption that includes setup encryption query, decryption queries. The forgettability is achieved by considering an adversary key in setting up phase followed by an attack stage where adversary key makes encryption and decryption query of a particular device.

Bala et.al [6] worked on session initiation protocol and instantiated by applying ECC-based mutual authentication. The methodology comprises three phases, namely, System setup where the server selects two points on the elliptic curve one being the private key and other being public key also secret key is kept safe by applying one-way hashing technique. Registration phase user computes a key by hashing User ID and several other keys. Authentication phase where the sender and receiver authenticate the data between themselves. This paper resists replay attack, ensures anonymity and privacy, and masquerading attack.

Lyubashevsky and Masny [7] have worked on the safety of mobile applications using ECC-based mobile RFID. Paper mainly focuses on solving OTRUTS, i.e., one-time reading and unlimited service that means once a reader reads the data of a particular tag from one of the servers, and the reader can read it for innumerable times. The proposed protocol considers security patrolling and has four phases, i.e., the Initialization phase—here server is the Security Management Center (SMC) which chooses points on the elliptic curve as private and other as the public key. Authentication phase-Sentry post's tag (SPT) to patrolman's reader (PMR) authentication is the second phase where PMR extracts value and sends a request to the sentry post's tag at the client-side. From the server-side reverse process takes place and checks whether that Id is present in the database. One more authentication phase happens from the sentry tag to the reader. These two phases of authentication happen in a single session of communication. Data sending phase-server fetches the data, encrypts it, and transmits it to the reader. Once the reader obtains the data it computes certain values and recovers data and accepts only if the value matches else it aborts the session.

Fan et al. [8] propose a method that works on Certificate-Less Public-Key Cryptography (CL-PKC) that is immune to node imitation as well as replay attack in a network. The paper describes three modules, i.e., Network initialization module-the network administrator generates a master secret key from Key Generation Center (KGC) and global parameters that all other nodes can avail in the network. Node registration module-the end-users submit their identity to the key generation center and request a partial private key. Once the user is verified, the user runs the algorithm for the generation of a public and private key. Once the user enters a network, a secret key is shared for long term with each workstation nodes. The session-key establishment module establishes a shared session between a particular node and end-user. It ensures a safe channel between them with corresponding public keys.

Park et al. [9] proposed a lightweight RFID-based mutual authentication supported with cache in the reader which ensures that recently visited tags are authenticated by the reader itself. The correctness of this method is proved by GNY logic. The tag stores a random key by one-way hashing technique that is saved in the database. The reader generates the random number and sends an authentication query to the tag; the tag further generates a random number and sends two keys to the reader. Readers store the tag key set in the cache that already has done the previous set of transactions, reader calculates tag values and verifies it from the cache if not present then it is added to the cache. This technique greatly reduces the computation and storage space.

Tewari and Gupta [10] have come up with a mutual authentication protocol using four algorithms—System setup, Key generation, Initialization, and mutual authentication. After taking a security parameter as input, the setup phase produces system parameters which in turn, taken as input for key generation, produces a pair of private and public keys. Now, the two connected devices exchange identities and their corresponding public keys. To provide strong authentication, the message is divided into smaller pieces and the authentication algorithm is applied for each piece. The sender (A) encrypts a piece of the message with receiver (B)'s public key, who on receiving the ciphertext, decrypts it. If the decrypted message does not match with A's identity, authentication terminates. If successful, B encrypts it back and A performs the decryption and comparison. The process continues for all pieces of the message, providing authentication at each step.

Lars Kulseng, Zhen Yu et al. [11] have come up with a mutual authentication protocol using RFID tags between a pair of a reader and an RFID tag of the IoT devices. Here, the reader perpetually broadcasts the 'req' message. Upon receiving 'req', the tag replies with an index of the tag's ID without revealing the actual tag's ID. On receiving this index, the reader searches the database for the corresponding value of greeting and their XORed value is returned to the tag. This tag then confirms with the obtained value and the actual ID of it. Hence, the tag is verified without the reader knowing the ID of the tag. Now, the tag computes two greeting values from the sent greeting value and sends back two keys generated using them. The reader then generates back the greetings using the same algorithm and compares the values stored and obtained. This authenticates the tag for its identity. This is carried out for many rounds since the same pair of reader and tag has to continuously authenticate one another.

Mian Ahmad Jana, Fazlullah Khana et al. [12] have come up with a mutual authentication technique based on the payload for securing IoT connections. It uses the concept of payload to trade the messages between the server and the client. A provisional phase includes the exchange of a secret key by the clients to the server. The server maintains a table for all these secret keys corresponding to each client. Once verified from the table, the session initiation process begins where a request message is sent by each of the clients containing object ID to the central server. Upon successful verification of the ID, a secure session begins. During the server challenge phase, this server looks-up for the secret key corresponding to object ID and encrypts the payload (on successful verification) using the AES algorithm. In the message, the server includes the ID shared to it during the session initiation phase as well as the secret session key for the intended client to decrypt. The client decrypts using the secret key. A potential intruder can only eavesdrop ID but not the initial secret key without which the session ID is not accessible.

Shaoquan Jiang and Guang Gong [13] have extended on the existing Key Exchange (KE) protocol, where the two communicative parties hold long-term secrets and while interaction sessions a temporary key is created and encrypted and hence authentication happens. But this protocol is prone to exhaustive search attacks. To address this problem, a paper has come up with a Carrier Routing System model where the nodes in the network have a set of public parameters, but no party knows about the secret key. This methodology helps in authenticating the initiator in the network and the above method is proved by Decisional Diffie–Hellman (DDH).

Tewari, A, Gupta, B. B [14] have proposed an extremely lightweight protocol for mutual authentication which deals with bitwise operations and this methodology is applied on a medium between RFID tag and a reader. Here each device has two sets of 96-bit tag ID and key-value, i.e., {ID, K} that are exchanged by database server at the backend and the tag. Based on the older values {IDold, Kold} from the previous session the tag can be verified. The XORing operation is done in the generation of the 96-bit keys. Since the tag values keep getting updated, even if the attacker obtains IDs, which will be invalid for future protocol runs. This method has very low overhead and ensures untraceable.

Namje Park, Marie Kim [15] has worked on mutual authentication along with the distributive session key system for the devices where it calculates the session key in advance hence improving the performance. This methodology especially works for resource-limited devices. The device produces a random number and sends it along with its own ID to the receiver (R). R encrypts the received key and generates another number sends it to the next device along with its own ID and hence these values are XORed and this produces a session key. This key is used to encrypt the message. The method is immune to replay attacks, man-in-the-middle attacks.

Ahmad Khoureich Ka [16] have proposed a new authentication technique called RMAC which is very much applicable to highly constrained IoT devices based on the mirror-mac method. RMAC uses a 128-bit key and operates with a 64-bit bit response. Encryption is done using a one-time pad before sending it to the prover an n-bit string is returned to the verifier. Also, the XOR-Cascade encryption technique is used that consists of two rounds. Further, the key is renewed by a lightweight key establishment technique by mixing the function of gossamer protocol. This method proves to be secure against a middle-man attack.

References

1. Harish Kumar N, Deepak G (2019) Authentication and privacy for IoT objects: a survey. J Emerg Technol Innov Res 6(4):268–271
2. Aman MN, Chua KC, Sikdar B (2017) Mutual Authentication in IoT Systems Using Physical Unclonable Functions. IEEE Internet of Things Journal 4(5):1327–1340
3. Li, Q., Cao, G, "Multicast Authentication in Smart Grid with One-Time Signature. IEEE Trans. smart grid," pp. 686–696, 2011
4. He K, Weng J, Liu J-N, Liu JK, Liu W, Deng RH (2016) Anonymous identity-based broadcast encryption with chose ciphertext security. In: Proceedings of the 11th ACM on Asia conference on computer and communications security—ASIA CCS'16
5. Mehmood Z, Chen G, Li J, Li L, Alzahrani B (2017) A robust ECC based mutual authentication protocol with anonymity for session initiation protocol. PLoS ONE 12(10):e0186044
6. Bala DQ, Maity S, Jena SK (2017) Mutual authentication for IoT smart environment using certificate-less public key cryptography. In: 2017 third international conference on sensing, signal processing and security (ICSSS)

7. Lyubashevsky V, Masny D (2013) Man-in-the-middle secure authentication schemes from LPN and weak PRFs. In: Canetti R, Garay JA (eds) Advances in cryptology—CRYPTO 2013. Springer, Berlin, pp 308–325
8. Fan K, Jiang W, Li H, Yang Y (2018) Lightweight RFID protocol for medical privacy protection in IoT. IEEE Trans Ind Inf 14:1656–1665
9. Park N, Kim M, Bang HC (2015) Symmetric key-based authentication and the session key agreement scheme in IoT environment. Lecture notes in electrical engineering, pp 379–384
10. Tewari A, Gupta BB (2016) Cryptanalysis of a novel ultra-lightweight mutual authentication protocol for IoT devices using RFID tags. J Supercomput 73(3):1085–1102
11. Jan MA, Khan F, Alam M, Usman M (2017) A payload-based mutual authentication scheme for the Internet of Things. Futur Gener Comput Syst
12. Kulseng L, Yu Z, Wei Y, Guan Y (2010) Lightweight Mutual authentication and ownership transfer for RFID systems. In: 2010 proceedings IEEE INFOCOM
13. Li N, Liu D, Nepal S (2017) Lightweight mutual authentication for IoT and its applications. IEEE Trans Sustain Comput 2(4):359–370
14. Chen M, Chen S, Fang Y (2017) Lightweight anonymous authentication protocols for RFID systems. IEEE/ACM Trans Netw 25:1475–1488
15. Aghili SF, Mala H (2018) Security analysis of Fan et al. lightweight RFID authentication protocol for privacy protection IoT. Cryptology ePrint archive, report 2018/388
16. Chaudhry SA (2015) Comment on 'robust and efficient password-authenticated key agreement with user anonymity for session initiation protocol-based communications. IET Commun 9(7):1034

Emotion Scanning of the World's Best Colleges Using Real-Time Tweets

Sanjay Kumar, Yash Saini, Vishal Bachchas, and Yogesh Kumar

Abstract With the advent of technology and the popularity of social media, people have started sharing their points of view with the masses. Views of the people matter a lot to analyze the effect of dissemination of information in a large network such as Twitter. Emotion analysis of the tweets helps to determine the polarity and inclination of the vast majority toward a specific topic, issue, or entity. During elections, film promotions, brand endorsements/advertisements, and in many other areas, the applications of such research can be easily observed these days. The paper proposes performing the emotion scanning of people's opinions on the three top colleges and universities of the world according to various surveys and indices (Harvard, MIT, and Stanford) using real-time Twitter data. Also, a comparison has been drawn between the accuracies of a few machine learning techniques used, for instance, K-nearest neighbor (KNN), support vector machines (SVM), and Naïve Bayes.

Keywords Emotion analysis · SVM · KNN and Naïve Bayes

1 Introduction

Social media which is free of charge and user-friendly, needing only a working Internet connection has reached masses in recent times. The individuals use this forum widely to voice their concerns. One of the microblogging sites is Twitter, garnering with about 500 million tweets each day. Users of Twitter post (tweet) on

S. Kumar · Y. Saini · V. Bachchas (✉) · Y. Kumar
Department of Computer Science and Engineering, Delhi Technological University, Delhi, India
e-mail: vishalbachchas08@gmail.com

S. Kumar
e-mail: sanjay.kumar@dtu.ac.in

Y. Saini
e-mail: yash.0245113@gmail.com

Y. Kumar
e-mail: 210yogesh@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_31

different topics, services, day-to-day activities, places, personalities, etc. every day. Therefore, Twitter data is of utmost importance as it can be used in several cases, for instance, discussing the ongoing political issues, building up pressure on the government, and issues from various parts of the world. Viewpoints and reviews in the form of tweets from customers, potential users, and critics can easily influence the demand for a product. A lot of well-known colleges and universities all over the world try to provide the best education to their students. They wish to attract several more to pursue studies in their institutions. Thus, it becomes quintessential for the students in deciding from the options available to them when choosing a college that not only provides them the academic or professional skills they want but also equips them with the right learning resources depending on their ability. In order to perform the emotion analysis, the use of data/tweets of colleges/universities of the world is made. For the sake of convenience, top three colleges according to the recent QS ranking (Harvard, MIT, and Stanford) are taken and certain sentiments, for example, happiness, anger, sadness, fear, and disgust were analyzed to give a glimpse of the state of students in these premier colleges and universities. Finally, the Python language was used to formulate the proposed method.

2 Related Work

This section consolidates and summarizes a few scholarly works proposed in the field of data mining with the machine learning to analyze emotions on the tweets and to prepare prediction models for various applications. Bhatt et al. [1] analyzed three of the top colleges of India (IIT, NIT, and AIIMS). Their paper proposed the use of SVM using different kernels to predict the sentiments. The polarity of the sentiments was recorded and the results were obtained for each kernel. Sentiment analyses on "Digital India" using the inbuilt tool NLTK to preprocess the data is advocated by Mishra et al. [2]. And thus the polarity classification was carried out. The prediction of results was done by the Zhang and [3] wherein ELM (extreme learning machine) and SVM were used and contrast was drawn between the results obtained by them. The use of deep learning networks for effective analysis of sentiments was carried out by Ramadhani and [4] to record the accuracy and precision of the model. The use of SVM, Naïve Bayes, and ME (maximum entropy) was proposed by Hassan et al. [5]. The method proposed depression measurement using the voting model and feature selection technique. The SVM showed better results as compared to other mentioned algorithms. The successful sentiment analysis of movie reviews was done by Bandana [6], wherein NLTK (natural language processing tool kit) was employed for preprocessing and analyzing the data. And various metrics were evaluated for different classifiers (Naïve Bayes, SVM). Krishna et al. [7] did the opinion mining of a few top colleges in India using various classifiers (Naïve Bayes, KNN). The accuracy metric of the techniques was evaluated. In the data of one college, KNN proved to be accurate and in the other one, Bayes had higher accuracy. The use of hybrid of two ML techniques was illustrated by Gupta et al. [8], wherein both KNN and SVM were combined. The model resulted in much better results than any of the techniques done in isolation. The combined hybrid was applied to the data after the successful preprocessing of the data was carried out which included stemming, stop word removed, abbreviation expansion, tag word identification, spelling correction, and positive and negative word identification. The model registered a 9% more accuracy than both the models separately did. Thus, the shortcoming of less accuracy was successfully overcome by them. In the work proposed by Naiknaware et al. [9], MAE (mean absolute error) was recorded for different classification techniques done on different datasets. It was observed that less MAE resulted in much higher accuracy. Kumar and Abraham [10] proposed a different approach of assigning sentiments to the tweets making use of a group of adverbs and verbs placed before the adjectives. The datasets had the sentiment weights for five different sentiments and were used to calculate the score. The data was fed to the classifiers (different kinds of Naïve Bayes and SVM) and the accuracies were recorded. The formula proposed by them is used in this paper. A novel approach of forming positive, negative, and neutral dictionary after preprocessing the data was proposed by Alaoui et al. [11] which helped in finding the polarity. The data was fed to the classifier and the results were compared. Woldemariam [12] dealt with comparing lexicon-based and machine-learning-based methods and combining them. The method used a recursive neural tensor network (RNTN) giving 9.88% more accuracy than the lexicon-based method on the forum discussion dataset.

3 Proposed Work

Since the sentimental score of a sentence is dependent on the adjectives present in it, this paper proposes that when an adverb/verb is encountered before an adjective, the sentimental intensity of the adjective is enhanced and it facilitates the accurate assignment of the sentiment. The following steps were taken for the successful formulation of this methodology:

1. **Tweet extraction**: Raw/unclean tweets for the analysis are retrieved with the help of the "tweepy" package of Python, which provides functions for real-time twitter API. The API requires one to register a developer account with Twitter and fill in parameters such as consumer key, consumer secret, access token, and token secret. This API allows us to get all random tweets or filter data by using keywords. The user enters the college name with ("#") as a prefix. And the tweets are downloaded from Twitter and saved to a file. 3000 + tweets were considered for one college.
2. **Data Processing**: The step involves manually implemented steps such as removal of @, Http, #, RT, and various stop words like was, were, is.
3. **POS Tagging**: Now the cleaned data obtained is used for the colleges' part of speech tagging. As per the finding, only adverbs, verbs, and adjectives are considered in a tweet. This is done because the adverb/verb groups placed before

the adjective can enhance the intensity of the emotion of a tweet as mentioned above.

Example

(a) I can do swimming.
(b) I can do very good at swimming. Sentence (b) has a high intensity of emotion among these two as very good improves the emotional intensity of the swimming.

4. **Scoring**: A dataset of 1000 + adjectives with their happiness, anger, fear, sadness and disgust index is obtained through a survey conducted amongst the undergraduate students. The dataset of a few verbs and adverbs are also obtained. The algorithm used for the calculation of scores is:
   (1) If an adverb or verb is encountered after another adverb or verb then it is added. The process is repeated until an adjective (root form (stemmed) or original) is encountered.
   (2) On the encountering of the adjective:
      (a) If sum (verb, adverbs) < 0: then value of adjective = 5-value of adjective.
      (b) If sum (verb, adverbs) ≥ 0.5: then value of adjective = sum (verb, adverbs)* Value of adjective.
      (c) Else: value of adjective = 0.5 * value of adjective.

The score and sentiment of a tweet is the highest score among the five emotions and its corresponding sentiment.

5. **Classification**: The final step is the application of SVM, KNN, and Naïve Bayes machine learning classifiers to the data obtained after the sentiment analysis. The features are formed using the bag-of-words technique and the dependent feature being the sentiment. The labels are assigned as {'Happiness':1, 'Neutral':0, 'Anger':2, 'Fear'-3, 'Disgust'-4, 'Sadness'-1}. Further, the training and test data are divided in the ratio of 7:3.

4 Result

After the assignment of sentiments to each tweet, the above graph is prepared to give the number of tweets belonging to each college as per the five emotions. It is found that certain tweets could not be classified because of the lack of vocabulary, and hence they are labeled as neutral (Figs. 1 and 2, Table 1).

The accuracies of all the classifiers used are compared and SVM is found to be the best classifier among all the colleges. The paper thus proposes a method that gives a glimpse of the ongoing emotions on the campuses expressed through Twitter.

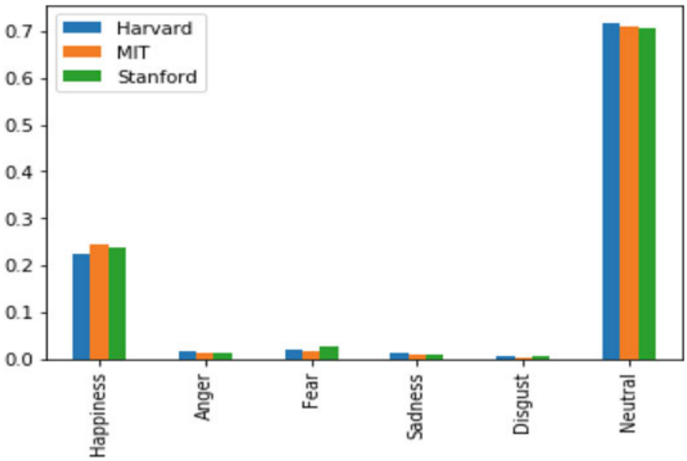


Fig. 1 The graph of colleges, sentiments, and their normalized numbers

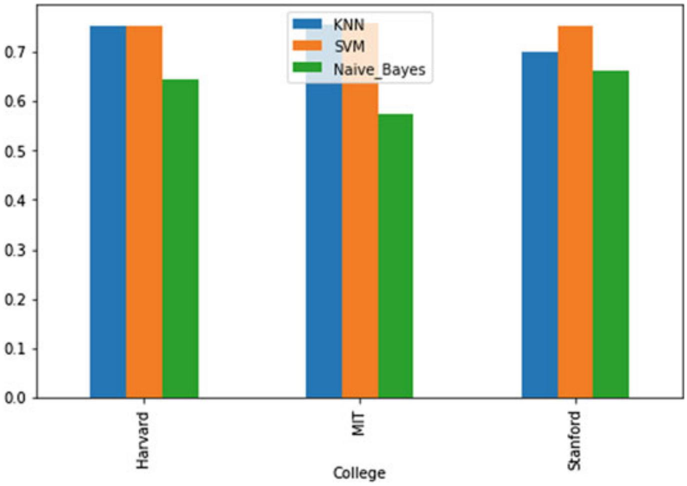


Fig. 2 The graph depicts the accuracies of machine learning classifiers

Table 1 Accuracy scores of various machine learning classifiers for the colleges' test data

College	KNN	SVM	Naïve Bayes
Harvard	0.7516	0.7526	0.6442
MIT	0.7566	0.7576	0.5729
Stanford	0.6993	0.7529	0.6604

5 Conclusion

Twitter data can be used to analyze and study various patterns and trends. Placement of adverb/verb before an adjective enhances its emotional intensity. In the above project, tweets of three different colleges are taken as the dataset, preprocessed, and assigned the sentiment; finally, accuracies of different classifiers are compared for different colleges. For the future scope of the project, more words in the adverbs, adjectives, and verbs need to be incorporated to reduce the neutral sentiment. Deriving emotions from emoticons in the tweets to determine the final sentiment could be done. Categorizing the tweets referring to either male or female can be incorporated. One could find out the dependency of consecutive tweets to influence the final scoring and improving on the proposed methodology.

Competing Interest The authors declare that they have no conflict of interest.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Bhatt G, Mamgain N, Mehta E, Mittal A (2016) Sentiment analysis of top colleges in india using twitter data. In: 2016 international conference on computational techniques in information and communication technologies (ICCTICT). IEEE
2. Mishra P, Rajnish R, Kumar P (2016) Sentiment analysis of twitter data: case study on digital India. In: 2016 international conference on information technology (InCITe)—The next generation IT summit on the theme—internet of things: connect your worlds. IEEE
3. Zhang X, Zheng X (2016) Comparison of text sentiment analysis based on machine learning. In: 2016 15th international symposium on parallel and distributed computing (ISPDC). IEEE
4. Ramadhani AM, Goo HS (2017) Twitter sentiment analysis using deep learning methods. In: 2017 7th international annual engineering seminar (InAES). IEEE
5. Hassan AU, Hussain J, Hussain M, Sadiq M, Lee S (2017) Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression. In: 2017 international conference on information and communication technology convergence (ICTC). IEEE
6. Bandana R (2018) Sentiment analysis of movie reviews using heterogeneous features. In: 2018 2nd international conference on electronics, materials engineering & nano-technology (IEMENTech). IEEE
7. Krishna CN, Sagar PV, Moparthi NR (2018) Sentiment analysis of top colleges. In: 2018 fourth international conference on advances in electrical, electronics, information, communication and bio-informatics (AEEICB). IEEE
8. Gupta A, Pruthi J, Sahu N (2017) Sentiment analysis of tweets using machine learning approach. 2017 IJCSMC 006(4):444–458
9. Naiknaware B, Kushwaha B, Kawathekar S (2017) Social media sentiment analysis using machine learning classifiers. 2017 IJCSMC 006(6):465–472
10. Kumar A, Abraham A (2017) Opinion mining to assist user acceptance testing for open-beta versions. 2017 J Inf Assur Secur 012:146–153. ISSN 1554-1010

11. Alaoui IE, Gahi Y, Messoussi R, Chaabi Y, Todoskoff A, Kobi A (2018) A novel adaptable approach for sentiment analysis on big social data. 2018 J Big Data 005:1–18 (Springer)
12. Woldemariam Y (2016) Sentiment analysis in a cross-media analysis framework. In: 2016 international conference on big data analysis (ICBDA). IEEE

Generating Feasible Path Between Path Testing and Data Flow Testing

C. P. Indumathi and A. Ajina

Abstract Software testing plays a major role in developing error-free software. The scope of testing the software is to identify the errors and errors present in the software. The data for testing are generated at the initial stage of software testing, which is a complex task during the testing process. There are several techniques that are available to generate test data. The paper puts forth the method to produce the cases that are used in testing from the control flow graph which is based on the path-oriented approach. This technique is to determine the feasible path upon all the possible paths. To solve this technique efficiently, a genetic algorithm is applied to identify the path that is optimal. The results from the pathwise approach are compared to the data flow testing approach. The comparative result shows that the produced data set for path testing produces a more feasible path than the data flow testing technique.

Keywords Software testing · Genetic algorithm · Data flow testing · Cyclomatic complexity · Path testing · Test data generation

1 Introduction

Testing software is the most important phase in the software industry to deliver the error-free software. Software testing is a very broad area that includes requirement specification, design, etc. The importance of testing is to verify the product before getting delivered. Nowadays, most of the industries are ready to develop the software, however, the challenging work is to deliver the software without any faults. The most challenging task is the testing part that only acquires 40–60% of the cost of

C. P. Indumathi (✉)
Department of Computer Science, University College of Engineering, (Bit Campus), Tiruchirapalli, Tamilnadu, India
e-mail: inducp@gmail.com

A. Ajina
Department of Computer Science, SMVIT, Bengaluru, Karanataka, India
e-mail: ajinajaya@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_32

overall software development. Software testing is not only finding the errors in the program, but it also verifies the behaviors and identify the defects and validate which specification is needed. There are various testing techniques available to test the software [1]. The testing process starts with input data. The generation of input data specifically to test the software is said to be the test data generation. There are many ways to produce data for testing. Input information is generated by the process of assessment data generation to test the adequacy of the software. It is one of the complex tasks in the field of testing. Generation of data to be assessed for a small program is quite easy, however, when dealing with large programs, it is not easy to produce possible data input [2]. The generation process is classified based on the testing process. They are random "test data generator, goal oriented test data generator, intelligent test data generator etc." [3, 4].

The approach behind the path oriented for generating feasible cases of test which covers all the possible paths in the program [5]. For large programs, infinite numbers of paths are available. All the paths are not needed to test as input. So the program is transformed as an intermediate-graph and formerly selects the path in which the bases have occurred. To get the optimized solution, genetic algorithm is applied. The path is converted into binary value as the initial population and then performs the "selection, crossover and mutation" for the inhabitants, in the last stage highest value of the fitness function is considered to be the feasible path. Then the results are compared with the data flow testing approach. The concept behind the data flow testing is how the variables are defined and shows how those variables assign the values. "For each variable, the definition use path and definition clear path are identified. Test data generated by path-oriented approach using genetic algorithm produce better results when compared to data flow testing strategy."

2 Background

Test input generation is significant in the testing of software. The error would be reduced when the input to test the software is correct. Thus the time, cost, and human workload will be reduced automatically. The authors Matthew Schmid and Frank Hill [6] describe the importance of test data and they introduce two techniques for generating test data. The data are generic and intelligent. The authors Rajappa and Panda [7] generate the test data by using the intermediate graph and optimal results are produced by applying a genetic algorithm.

The author Raluca [8] describes that the data for testing is generated according to the state from any application. He supports the automatic generation of test data. The authors Latiu and Cret [9] describe the importance of software testing and deal with the test methods that are used the internal structure of the program to generate the test data. The author Mohammad Reza and Hajar Homayouni [10] describe the challenges that are faced in test case generation in the form of framework.

The authors Pargas and Harrold [11] establish the Tgen tool to generate the test cases. Genetic algorithm is used to generate test data. This approach covers both the statement coverage and branch coverage. The author Sapna Varshney and Monica Mehrotra [12] describe the test data flow testing strategy for test data generation. So that the time and cost are to be reduced. To solve the method effectively, a genetic algorithm is used to find out the feasible path among all the paths.

The author Ahmed [13] discusses that how to manage the time and cost during multiple test data generation. They used multiple test data, which covers multiple target paths. Thus, the authors are to be taken as the reference for generating the test data based on the -oriented approach. Input data generation is used in the GA to produce the optimal solutions. In recent paper, the author Marashdih and Zaaba [14] investigates the current approaches to detect infeasible paths in the static analysis and also point out the challenges.

3 Proposed Approach

A. *Generation of Data For Testing Based on Path-Oriented Approach*

Test data generating using path-oriented method aims to find out the test cases from the path. There are many paths that are available from one program but the most suitable test cases are used to test the software and also many infeasible paths are generated for the program. So that the time and cost are to be reduced. To solve the method effectively, genetic algorithm is used to find out the feasible path among all the paths.

The program to generate test data is transformed as intermediate-graph followed by identifying the bases in the intermediate-graph. While applying this to the genetic algorithm, the predicates are converted into binary digits as chromosomes. Then the path should be determined based on the predicates in the graph.

i. *Calculation of fitness function:*

The fitness function for the genetic algorithm is

$$F(x) = (C' * C') + (W) \tag{1}$$

$F(x)$ Fitness
C' Cost for each path
W Weight calculated from the path in the program.

ii. *Selection*

Roulette wheel selection was utilized in the selection of the individuals among the populace. Formula to calculate individual probability and cumulative probability is given below, generation is based on the data flow testing strategy. It includes the genetic algorithm to generate the test

$$\mathrm{IP}(x) = F(x) / \sum \mathrm{mn} = (F(X)) \tag{2}$$

data, and then the results are to be determined with the coverage. The authors Gong and Yoa [15] discussed the problems that are faced during the generation of test data using the path-oriented method, where m is the size of the population.

$$\mathrm{CP}(x) = \sum{}^{L} n = 1\ \mathrm{IP}(X) \tag{3}$$

iii. *Crossover*

After the selection of chromosomes, the next process is the generation of new chromosomes by applying crossover function. It uses the binary representation to perform a crossover. There are many types of crossover methods. Some of them are uniform crossover, single-point crossover, arithmetic crossover, multi-point crossover, etc. In this, single-point crossover is used, with the probability value of 0.8.

iv. *Mutation*

After performing crossover, any one of the bit is replaced as 0 or 1. The conversation of bit takes place in the mutation process. Mutation is mainly used to maintain the diversity of the genetic algorithm. The mutation probability value is 0.2.

v. *Termination*

The genetic algorithm process will terminate until the fitness function has the highest value.

B. *Steps for the Proposed Approach*

(1) Write a program and convert into control flow graph.
(2) Convert the control flow graph into an intermediate graph.
(3) Identify and generate the test sequence from the predicate.
(4) For every path of graph,

- Compute the “fitness function”, Individual and cumulative probability.
- Perform selection, crossover, and mutation operation.

C. *Generation Of Data For Testing based on data flow approach.*

The global search technique is now a recent topic for researchers. The main aim of data flow testing is to generate the test data that covers the all-use criteria. Test data generation is one of the important tasks. The paper proposes the genetic algorithm for test data generation using data flow testing that produces the result for reducing time and complexity.

D. *Steps for this approach is given below.*
(1) The program to be tested is changed into a control flow graph. That indicates the flow of the program.
(2) Identify the variables in the program and then classify the variables into “c-use” and “p-use” variables.
(3) Then the most important is to determine the def-use path to identify the paths which are all satisfying all-use variable criteria.
(4) Generate the test data applying the GA.

$$\mathrm{F(x)} = \frac{\text{fresh dcu path} + \text{dpu path enclosed}}{\text{Aggregate num of dpu as } well \text{ as dcu path}''}$$

- Calculate the fitness function from the above equation and then perform three operations until all the requirements are to be satisfied.
- Perform selection, crossover, and mutation operation.

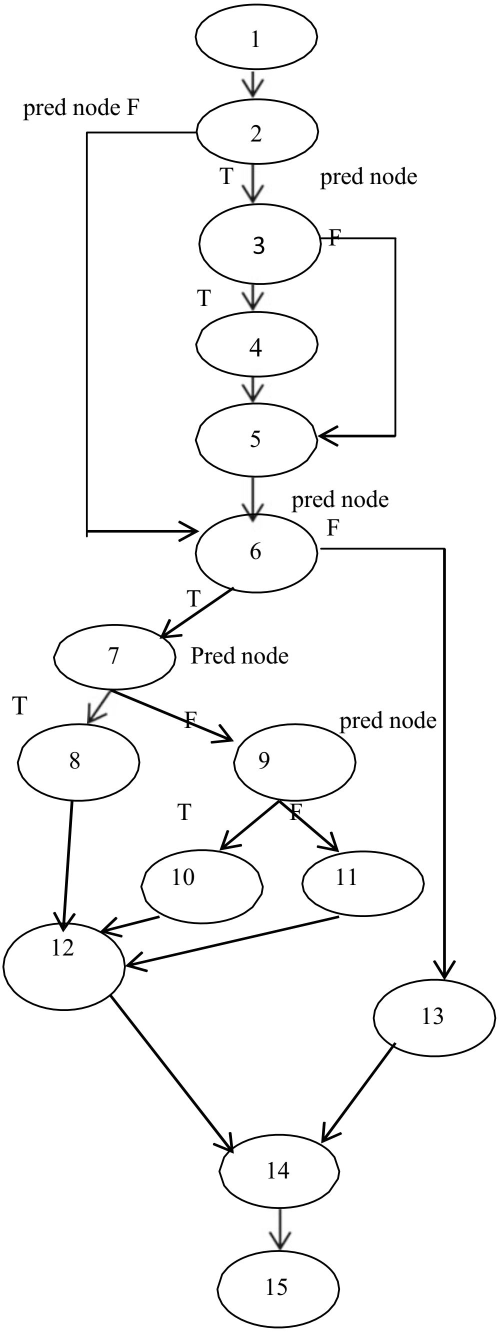
Path testing focuses on the paths to generate test cases in which the variables are to be used to cover all possible criteria.

4 Experimental Study

A. *Path-oriented test data generation*

We performed an experimental study for a detailed understanding of the concepts. Consider an example of a triangle program. The source code for the triangle program is given below. Control flow graph for the triangle program is shown in Fig. 1.

Fig. 1 Flow graph of triangle program



```
S1: int e,f,g,v;
S2: printf("Enter three sides");                                      (1)
S3: Scanf("%d%d%d",&a,&b,&c);
S4: v=0;
S5: if((e>=0)&&(e<=100)&&(f>=0)&&(f<=100)
    &&   (g>=0)&&(g<=100))                                                  (2)
S6: {
S7: if(((e+f)>g)&&((g+e)>f)&&((f+g)>e))
S8: {                                                                       (3)
S9: v=1;                                                                    (4)
S10: }
                                            (5)
S11: }
S12: if(v==1)                               (6)
S13: {
S14: if((e==f)&&(f==g))                     (7)
S15: printf("Equilateral")                  (8)
S16: else if((e==f)||(f==g)||(g==e))        (9)
S17: printf("isosceles")                    (10)
S18: else                                            (11)
S19: printf("scalene");
S20: }                                      (12)
S21: else
S22: {                                                   (13)
S23: printf("not a triangle")
S24: }                                      (14)
S25: }                                      (15)
```

Then identify the predicates in the graph. There are five predicates in the triangle program namely 2, 3, 6, 7, and 9. Random numbers are generated as the binary digit 0's and 1's as the population. It takes 8 bits to represent the chromosomes. Then find out the fitness function, individual probability, and cumulative probability for each path and then perform genetic operations.

Method is described as follows:

Table 1 Weight calculation

Nodes	M	S	$W = S - M$
15	14	15	1
14, 15	13	14	2
13, 14, 15	12	13	3
12, 14, 15	11	12	4
11, 12, 14, 15	10	11	5
10, 12, 14, 15	9	10	6
9, 8, 12, 14, 15	8	9	7
8, 12, 14, 15	7	8	8
7, 8, 12, 14, 15	6	7	9
6, 7, 8, 12, 14, 15	5	6	10
5, 6, 7, 13, 14, 15	4	5	11
4, 5, 6, 7, 13, 14, 15	3	4	12
3, 5, 6, 7, 13, 14, 15	2	3	13
2	1	2	14
1	0	1	15

Step1: Consider initial population as 01010101, 00000101, 01001001, 01010100, and 01010101.
Step2: Initial-chromosome 01010101, the path is 1-2-3-5-6-7-8-12-14-15. Charge of path is $E - N + 2 = 9 - 10 + 2 = 1$.
Step3: Fitness function calculated by stack weight is given in Table 1. In the table M represents the incoming nodes for a particular path. S represents the size. $W = 15 + 14 + 13 + 36 + 46 + 55 + 35 + 54 + 90 + 91 = 449$.
The "fitness function" of first chromosome is as follows.

$$F(x) = \left(C' * C'\right) + (W)$$
$$F(x) = (1 * 1) + 449 = 450$$

The initial population and the corresponding path for path testing is given in Table 2. And then the first iteration is given in Table 3.

Table 2 Predicate path

Initial population	Path
01010101	1-2-3-5-6-7-8-12-14-15
00000101	1-2-6-7-9-10-12-14-15
01001001	1-2-3-5-6-13-14
01010100	1-2-6-7-9-11-12-14-15

Table 3 Iteration 1

S.no	Individual pop	Fitness value
1	01010101	450
2	00000101	410
3	01001001	380
4	01010100	425

Step4: Then the IP value and the Cumulative value is calculated from the fitness function.
Step5: Then the crossover operation is performed to generate the new offspring.

2 ➔00000101	cross over point is 4 then the
3 ➔01001001	offspring's are 01000101, 00001001

Step6: Then the mutation is performed in the individual 2.

$$00001001 \rightarrow 00011001.$$

Step7: Thus the process will continue for four iterations. Then finally, the feasible path for the final iteration is the first population path. The feasible path is 1-2-3-5-6-7-8-12-14-15. Because it has the highest fitness value of 448.

B. *Data Flow Test Data Generation*

Step 1: The same triangle program is considered for test data generation based on the data flow approach.
Step 2: Determine the predicates in the graph and then identify the c-use and p-use variables in the program.
Step 3: Then find out the def-use path that covers all the paths in the programs. This can be displayed in Table 4.
Step 4: Then genetic algorithm is applied to generate test data. The fitness function is calculated by

$$F(x) = \frac{\text{fresh path for dcu} + \text{dpu path enclosed}}{\text{Aggregate num of dpu as well as dcu path}}$$

Table 4 Def-Cu and def-Pu path

Def-use path num	Def-use path (terminates with −1 for c-use paths)
1	1-2-3
2	1-2-6
3	1-3-4
4	1-3-5
5	1-7-8
6	1-7-9
7	1-9-10
8	1-9-11
9	1-6-7
10	1-6-13
11	4-6-7
12	4-6-13

The binary representation of the initial population and the associated path is given in Table 2. The fitness function for the data flow testing is given below.

$$\begin{array}{l} 01010101 \rightarrow 0.813 \\ 00000101 \rightarrow 0.75 \\ 01001001 \rightarrow 0.588 \\ 01010100 \rightarrow 0.833 \end{array}$$

Step5: Thus the process will continue for four iterations. Then finally, a feasible path for the final iteration is the last population. The path is 1-2-6-7-9-11-12-14-15.

The same path is not obtained for the path and testing of data flow. The testing of data flow moves through the false path. Else the path testing moves the path that covers all the possible nodes.

5 Conclusion

Generation of dataset based on the path-oriented method produces a more infeasible path. It leads to consuming more cost and time. We present path testing along with a genetic algorithm that produces the most optimal path among all other paths. The result from the path testing is compared with the data flow testing. The comparative result shows that the produced data set by path testing produces results in better performance for the data flow testing.

References

1. Trivedi SH (2012) Software testing techniques. Int J Adv Res Comput Sci Softw Eng 2(12)
2. Korel B (2000) Automated software test data generation. IEEE Trans Softw Eng 16
3. Segura S, Automated test data generation on the analyses of feature models: a metamorphic testing approach. In: European Commission and Spanish Government under CICYT project
4. HongChun Z (2013) Research on new techniques and envelopment trend of software testing. In: 2nd international conference on computer science and electronics engineering, Atlantis Press, Paris
5. Minj J, Belchanden L (2013) Path oriented test case generation for UML state diagram using genetic algorithm. Int J Comput Appl 82(7):0975–8887
6. Schmid M, Hill F, Data generation techniques for automated software robustness testing. In: Reliable software techniques corporation, defence advanced research project agency (DARPA)
7. Rajappa V, Biradar A (2008) Efficient software test case generation using genetic algorithm based graph theory. In: IEEE in proceedings of the 1st international conference on emerging trends in engineering and technology (ICETET 08), pp 298–303
8. Lefticaru R, Ipate F (2008) Automatic state-based test generation using genetic algorithms. In: IEEE 9th international symposium on symbolic and numeric algorithms for scientific computing

9. Latiu GI, Cret OA (2012) Automatic test data generation for software path testing criteria evolutionary algorithm. In: 2012 IEEE international conference on emerging intelligent data and web technology (EIDWT), pp 1–8
10. Keyvanpour MR, Homayouni H (2012) Automatic software test case generation: an analytical classification framework. Int J Softw Eng Appl 6(4)
11. Pargas RP, Peck RR (1999) Test data generation using genetic algorithms. J Softw Test Verif Reliab (Wiley)
12. Varshney S, Mehrotra M (2014) Automated software test data generation for data flow dependencies using genetic algorithm. Int J Adv Res Comput Sci Softw Eng 4(2)
13. Ahmed MA, Hermadi I (2007) GA based multi-paths test data generator. Elsevier Comput Oper Res 35:3107–3124
14. Marashdih AW, Zaaba ZF (2018) Infeasible paths in static analysis: problems and challenges. In: Proceedings of the AIP conference 2016, p 020079. https://doi.org/10.1063/1.5055481
15. Gong D, Yao X (2010) Automatic detection of infeasible paths in software testing. IET Softw 4(5):361–370

Soft Constraints Handling for Multi-objective Optimization

Md. Shahriar Mahbub, Fariha Tahsin Chowdhury, and Anika Salsabil

Abstract Most real-world search and optimization problems naturally involve multiple objectives and several constraints. In this work, an idea for a generalized new approach for handling both hard and soft constraints in Multi-Objective Optimization Problems (MOOP) is demonstrated. The main purpose is to fully satisfy all the hard constraints and satisfy soft constraints as much as possible. A modification to the binary tournament parent selection approach is proposed. The proposed approach is integrated with the two most widely used multi-objective evolutionary algorithms, i.e., NSGA-II and SPEA2. A test is conducted on four benchmark problems and satisfactory results are achieved.

Keywords Multi-objective optimization · Hard constraint handling · Soft constraint handling

1 Introduction

Optimization problems can be found in most of the practical fields. Moreover, most of the practical problems come with a number of constraints. There is a huge number of economic and financial problems that are directly related to encompassing both hard and soft constraints. As the feasibility of any solution is defined by sat-

Md. S. Mahbub · F. T. Chowdhury · A. Salsabil (✉)
Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh
e-mail: anika.aust.cse@gmail.com

Md. S. Mahbub
e-mail: shahriar.cse@aust.edu
URL: http://www.aust.edu

F. T. Chowdhury
e-mail: farihacseaust@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_33

isfying the hard constraints, the satisfaction of the problem can be defined by soft constraints. Mathematically, an optimization problem with hard and soft constraints can be formulated as

$$\begin{array}{lll}\textbf{Minimize/Maximize} & \mathrm{f}_m(X) & ;\ m = 1, 2, \ldots, M \\ \textbf{Subject to} & \textbf{:}\ g_j(X) \geq 0 & ;\ j = 1, 2, \ldots, J \\ & s_l(X) \geq 0 & ;\ l = 1, 2, \ldots, L \\ & x_i^{(L)} \leq x_i \leq x_i^{(U)} & ;\ i = 1, 2, \ldots, n \end{array}$$

Here, X is a solution vector of n decision variables: $x = (x_1, x_2, \ldots, x_n)^T$. The lower and upper bounds of the decision variable x_i are, respectively, $x_i^{(L)}$ and $x_i^{(U)}$. $g_j(X)$ presents J number of hard inequality constraints and $s_l(X)$ presents L number of soft inequality constraints. Moreover, there are M number of objective functions $f_m(X)$ that need to be minimized or maximized.

Most of the time the soft constraints are ignored. There exists hardly any literature where techniques for handing hard constraint are developed. In a few papers [1, 2], the soft constraints are considered as an additional objective or neglected. In this study, a generalized approach for handling soft constraints is developed. The main focus of the research is to handle soft constraints in parallel with hard constraints. For handling soft constraints, the binary tournament selection [3] is modified. Moreover, a metric called **Soft Constraint Violation** (SCV) is introduced. The solution which has greater SCV violates the soft constraints more than the other solutions. On this basis, the solutions which violate less soft constraints will get higher priority than the other ones when selecting a parent using binary tournament selection.

The proposed idea is tested in four benchmark problems concerning two algorithms **Non-Dominated Sorting Genetic Algorithm** (NSGA-II) [4] and **Strength Pareto Evolutionary Algorithm** (SPEA2) [5]. The results in the paper have evaluated the decrease of SCV over time. In most of the cases, the proposed approach performs well as desired. In the experiments, both for NSGA-II and SPEA2, the obtained results are mostly identical, as both of them use the same parent selection procedure. Some results show the undesired result which is a limitation. So, there lies a potential space for further research to overcome this limitation.

2 State of the Art

Please recall that very few papers propose an approach to handle both hard and soft constraints. Two papers are discussed in this aspect.

Jin et al. proposed a Constraint-Guided Method (CGM) with evolutionary algorithms for economic problems [1]. CGM is a technique that is used to handle both hard and soft constraints. It proposes a technique to convert the objective of a solution to a modified objective by considering hard and soft constraints by adding some penalty according to the status of a solution. The different penalty values are added on the priority basis (if a solution violates a hard constraint, more penalty value is added than the solution which violates only a soft constraint). Moreover, the ordering of soft constraint violations needs to be performed. Therefore, a solution violates a higher prioritized soft constraint that has more penalty value than a solution that violates a lower prioritized soft constraints. There are some problems regarding this approach. Firstly, the approach is only applicable to single-objective optimization. Secondly, the penalty values for hard and different soft constraints have to be defined. Defining penalty values require problem-specific knowledge which is not trivial at all.

Ray et al. [6] proposed an algorithm called Infeasibility-Driven Evolutionary Algorithm (IDEA) for handling hard constraints. IDEA proposes a significant difference in terms of ranking and selection of the solutions while comparing with traditional evolutionary algorithms. A new ranking technique is proposed which is based on the original objectives along with additional objective, which is called Constraint Violation Measure (CVM).

Singh et al. proposed a modified IDEA method for solving problems with a mix of hard and soft constraints called IDEA-M [2]. IDEA-M aims to deliver a set of solutions which

- satisfies hard constraints.
- achieves trade-offs concerning the soft constraints.
- has improved the rate of convergence.

To achieve the trade-off solutions, the CVM for soft constraints is considered as an additional objective. The approach may suffer from higher dimensional problems. Because the approach adds objective with the existing objectives, which may increase search space significantly.

3 Proposed Idea

In the context of handling hard constraints, many generalized techniques have been successfully used to get a feasible solution for constrained optimization problems. However, all these methods are not enough to handle both hard and soft constraints simultaneously as soft constraints are being neglected mostly. In this section, the main idea of our proposed techniques is described.

3.1 Existing Approach for Handling Hard Constraints

The most common existing technique is binary tournament selection for selecting parents [7]. While selecting parents, three different cases could arise as follows:

Case 1: Both solutions are infeasible

- The selected solution will be the solution that violates less hard constraints.

Case 2: One is feasible and the other one is infeasible

- The solution will be the feasible one.

Case 3: When both the solutions are feasible then

- **Sub-case 1**: One solution on the better front than the other

- Solution on the better front is selected.

- **Sub-case 2**: Both on the same front

- A solution is selected randomly.

Please note that the solution which violates hard constraints is called infeasible solution. Moreover, the shown approach is applicable for multi-objective optimization problems. For the case of multi-objective optimization, individuals are ranked according to domination [3]. Besides, the individuals are placed in fronts based on which individual dominates other individuals.

In the **Third Case**, there are two sub-cases. The first sub-case is shown in Fig. 1, if a solution is on the better front than the other solution, the solution which is located in the better front is selected as a parent. Without losing the generality, please consider two fronts, f_1 and f_2, in a multi-objective minimization problem. Front 1 and front 2 are two non-dominated fronts. There are two feasible solutions circled in green and red, situated on front 1 and front 2, respectively. As the green solution has been situated on the better Pareto front, it is selected as the Pareto Optimal solution. In the second sub-case, the solutions are situated on the same front and they have the same priority for getting selected as a next parent. Here, the next parent selection will be picked randomly. So, this random selection procedure of this case is modified to avoid uncertainty.

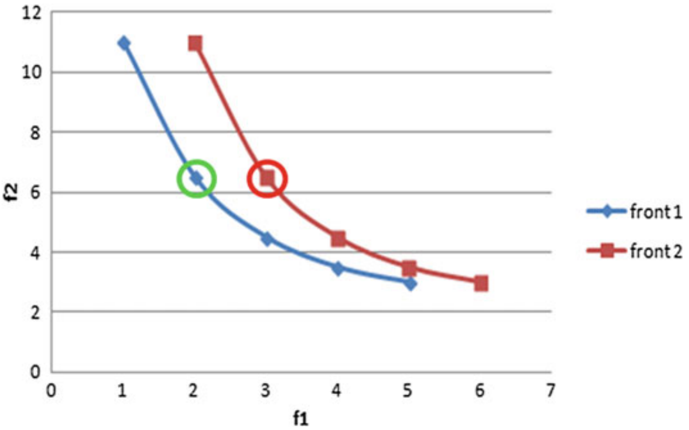


Fig. 1 Sub-case 1 of case 3

3.2 Modified Approach for Handling Both Hard and Soft Constraints

Please recall that hard constraints have to be satisfied. The solutions that do not violate hard constraints are valid. However, the valid solution that violates less soft constraints is a preferable solution than a valid solution that violates more soft constraints. Therefore, while selecting a parent that both are in the same front, a solution with less soft constraints violation is selected. Hence, the sub-case 2 of case 3 is modified which is shown below:

Case 3 (Modified): When both the solutions are feasible then

- **Sub-case 1**: One solution on the better front than the other

– Solution on the better front is selected.

- **Sub-case 2**: Both on the same front

– Less soft constraint violated solution is selected.

Let us consider only one front obtained which is shown in Fig. 2. There are two feasible solutions in the front and they are circled in blue and orange. Both solutions are situated on the same front. Considering, the Soft Constraint Violation (SCV) of the blue circled solution is 10 and the orange one is 15. As 10 < 15, blue one violates less soft constraints than the orange one. Hence, the blue solution is selected as a parent.

Finally, Algorithm 1 presents the details of the technique.

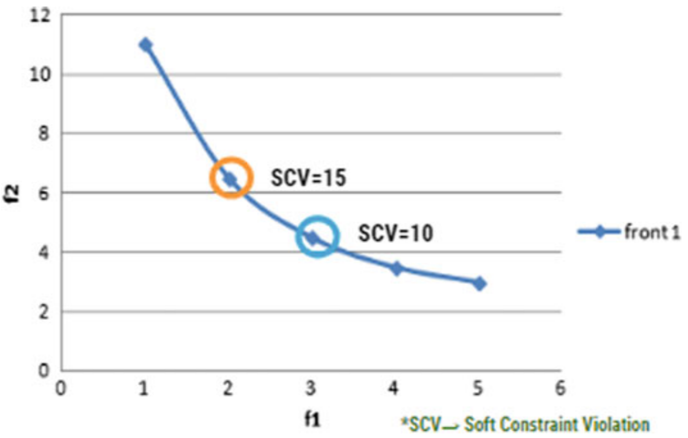


Fig. 2 Sub-case 2 of case 3

Algorithm 1 Soft Constrained Binary Tournament Selection

```
Sol1 {Solution-1}
Sol2 {Solution-2}
if Sol1 violates less Hard Constraints than Sol2 then
    return Sol1
else if Sol2 violates less Hard Constraints than Sol1 then
    return Sol2
else
    if Sol1 dominates Sol2 then
        return Sol1
    else if Sol2 dominates Sol1 then
        return Sol2
    else
        if Sol1 violates less Soft Constraints than Sol2 then
            return Sol1
        else if Sol2 violates less Soft Constraints than Sol1 then
            return Sol2
        else
            return a random solution
        end if
    end if
end if
```

3.3 Soft Constraint Violation (SCV)

There can be more than one soft constraint. In that case, a new metric called overall SCV is proposed. The metric adds all the violations. The mathematical formulation for calculating overall SCV is given below:

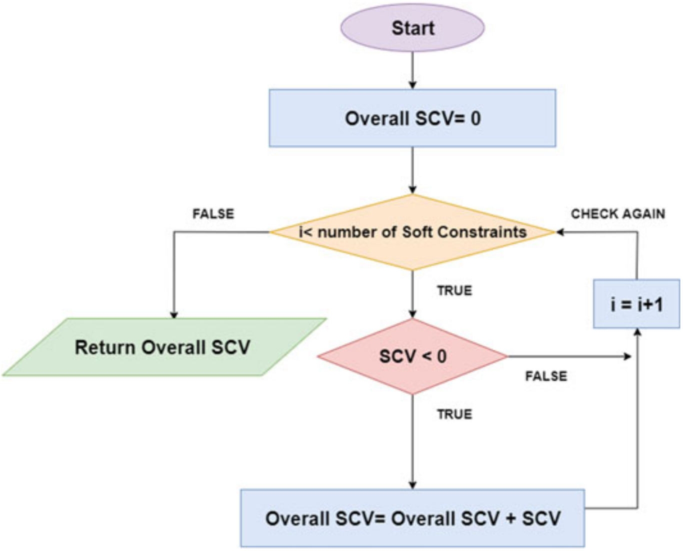


Fig. 3 Soft constraint violation (SCV) calculating flowchart

$$SCV = \sum_{i=1}^{L} s_i(x) \quad \text{if } s_i(x) < 0 \tag{1}$$

The flowchart for calculating the Soft Constraint Violation (SCV) is given below (Fig. 3):

For example, considering a problem which has four soft constraints. For one solution, the obtained values are given below:

$$s_1(x) = -0.1 \tag{2}$$

$$s_2(x) = 0 \tag{3}$$

$$s_3(x) = 1 \tag{4}$$

$$s_4(x) = -0.5 \tag{5}$$

To calculate the overall SCV, the negative values are added. Here, s_1 and s_4 are negative, which means the solution violates two soft constraints. And the overall SCV will be −0.6.

Table 1 Benchmark problems

Problem name	No. of objectives	No. of constraints
Tanaka	2	2
Srinivas	2	2
Viennet4	3	3
Osyczka2	2	6

4 Experiments and Results

Several experiments have been conducted to analyze the performance of our proposed approach. In the experiment, we select two most widely used MOEAs (i.e., NSGA-II [4] and SPEA2 [5]) and four benchmark problems (Tanaka [8], Srinivas [9], Viennet4 [10], Osyczka2 [11]). In the proposed approach, individual monitoring of each constraint will be analyzed, considering as hard and soft simultaneously. Finally, a discussion of the obtained results will be presented for better visualization and detailed analysis.

4.1 Benchmark Problems

Table 1 presents the details (name, number of objectives, number of constraints) of four benchmark problems.

In each experiment, we have considered each constraint as soft and the rest of them as hard constraints. We have continued this process in all the problems for both NSGA-II and SPEA2.

4.2 Evaluation Criteria

For each experiment, a graph is generated by plotting generation numbers (X-axis) and average SCV (Y-axis). Average SCV is calculated by averaging SCV of all the solutions of the generation. It is expected that the average SCV will decrease over time. In the first phase of the algorithm, the only priority is to find a feasible solution. Afterward, when the solutions are matured (mostly are in the same front), then binary tournament selection tries to decrease SCV. If the trend of average SCV reduces over time, i.e., it goes toward 0, then it can be said that our proposed technique has performed better.

Table 2 Parameter settings for NSGA-II and SPEA2

	NSGA-II	SPEA2
Initial population	100	100
Crossover	SBXCrossover [3]	SBXCrossover
Crossover probability	0.9	0.9
Mutation	PolynomialMutation [3]	PolynomialMutation
Mutation probability	1/n, n = the number of the decision variable	1/n
Generations	250	250
Archive size	–	100

4.3 Experimental Setup

Our proposed methodology is implemented in jMetal framework [12]. Table 2 shows the standard parameter settings for our experiments.

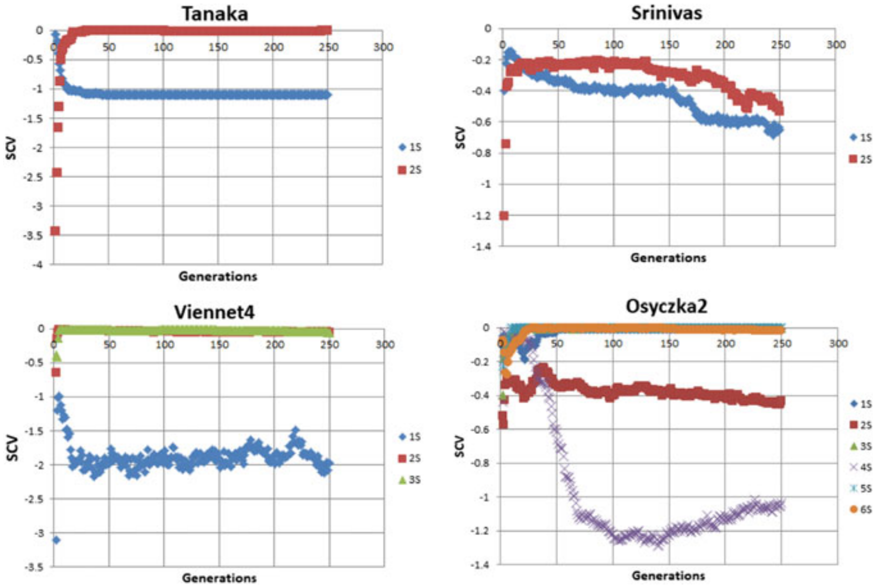


Fig. 4 Simulations for NSGA-II

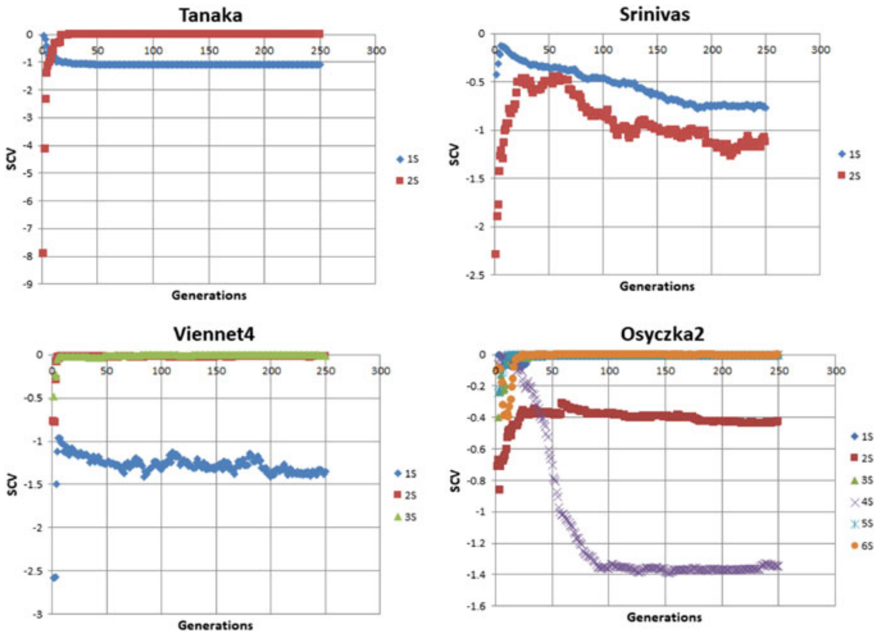


Fig. 5 Simulations for SPEA2

4.4 Results

In this section, the results of the two algorithms are presented.

4.4.1 Simulated Results of NSGA-II and SPEA2

Figures 4 and 5 show the results for NSGA-II and SPEA2, respectively. Each graph represents a problem and within the graph, all the soft constraints are presented by different series. Please recall that the experiments are conducted by considering one constraint as soft and rest of the constraints as hard. For example, consider that a benchmark problem has three constraints. Therefore, three runs are conducted, in each run, one of the three constraints is considered as soft and the other two are considered as hard constraints. Besides, the result of these three runs is plotted in a figure. The soft constraints are labeled as 1S, 2S, and so on.

4.4.2 Summary Results of NSGA-II and SPEA2

By observing four graphs, identical results are found. In **Tanaka**, for 1S desired trend is achieved, on the other hand, for 2S opposite is observed. For the second test problem, **Srinivas**, undesired patterns are observed. In the third test problem **Viennet4**, perfect curves are achieved for 1S and 2S. But in the case of 3S, some fluctuations in the trend are observed. Lastly in **Osyczka2**, for 1S, 3S, 5S, and 6S, desired trends are achieved and average results are observed for 2S. But in the case of 4S, poor results are observed.

4.5 Discussion

It can be observed that good results are not obtained for some soft constraints. Please recall that this approach is only applicable to parent selection techniques. However, the continuous improvement of the population is done in two phases (parent selection and survivor for next generation). In the later phase, Pareto ranking [3] and crowding distance [3] are applied to select the survivor for the next generation. The main preference for selecting a survivor is given on ranking. When most of the solutions are mature (having the same rank), crowding distances are considered for selecting survivors. Therefore, it could happen that more soft constraints violated solutions have better crowding distance; therefore, those solutions are selected for the next generation. Hence, soft constraints handling has secondary preference through parent selection. For this reason, we have not got the desired results for some soft constraints. More experiments are required to prove this conjecture.

5 Conclusion and Future Work

There are many hard constraint techniques available in the field of multi-objective optimization. However, handling soft constraint techniques are rarely found. Therefore, a soft constraint handling technique compatible with existing hard constraints handling technique is developed here. Binary tournament selection is modified to incorporate soft constraint handling techniques. Experiments with four benchmark problems on two algorithms (NSGA-II and SPEA2) are conducted. The test results show good results for most of the cases; however, some cases show poor results. The poor results could be explained by the selection mechanisms of survivors for the future generations.

There remains a vast area to work for evaluating and enhancing the proposed approach. One approach could be incorporating SCV into the survivor selection mechanism. For the selection of the next generation, ranking and crowding distance metrics are considered. However, SCV has been considered only in the case of parent selection. To achieve better results, integrating SCV alongside with ranking and crowding distance could improve the results for all the soft constraints.

References

1. Jin N, Tsang E, Li J (2009) A constraint-guided method with evolutionary algorithms for economic problems. Appl Soft Comput 9(3):924–935
2. Singh HK, Asafuddoula M, Ray T (2014) Solving problems with a mix of hard and soft constraints using modified infeasibility driven evolutionary algorithm (idea-m). In: 2014 IEEE congress on evolutionary computation (CEC). IEEE, 983–990
3. Kalyanmoy D (2001) Multi-objective optimization using evolutionary algorithms, vol 16. Wiley
4. Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multi-objective genetic algorithm: Nsga-ii. IEEE Trans Evol Comput 6(2):182–197
5. Zitzler E, Laumanns M, Thiele L (2001) Spea2: improving the strength pareto evolutionary algorithm. TIK-report, vol 103
6. Tapabrata R, Singh HK, Isaacs A, Smith W, Infeasibility driven evolutionary algorithm for constrained optimization. In: Mezura-Montes E (ed) Constraint-handling in evolutionary optimization. Studies in computational intelligence. Springer, Berlin, Heidelberg, pp 145–165
7. Deb K (2000) An efficient constraint handling method for genetic algorithms. Comput Methods Appl Mech Eng 186(2–4):311–338
8. Tanaka M, Watanabe H, Furukawa Y, Tanino T (1995) Ga-based decision support system for multicriteria optimization. In: IEEE international conference on systems man and cybernetics, vol 2. Institute of Electrical Engineers INC. IEEE, pp 1556–1561
9. Srinivas N, Deb K (1994) Muiltiobjective optimization using nondominated sorting in genetic algorithms. Evol Comput 2(3):221–248
10. Vlennet R, Fonteix C, Marc I (1996) Multicriteria optimization using a genetic algorithm for determining a pareto set. Int J Syst Sci 27(2):255–260
11. Osyczka A, Kundu S (1995) A new method to solve generalized multicriteria optimization problems using the simple genetic algorithm. Struct Optim 10(2):94–99
12. Durillo JJ, Nebro AJ, Alba E (2010) The jMetal framework for multi-objective optimization: design and architecture. In: CEC 2010, Barcelona, Spain, pp 4138–4325

Parking Management System Using Internet of Things

Aditya Sarin and Deveshi Thanawala

Abstract The raid population growth and increasing number of private vehicle ownership in countries like India have largely contributed to vehicle congestion. Inappropriate vehicle parking areas are considered as one of the root causes for slow moving traffic and has invoked the attention of the administration officials. The model proposed here serves a three-stage parking system using the Internet of Things. The first stage consists of an automated gate connected to an onscreen vehicle counter. The second stage consists of sensors that are installed at each parking slot, connected with a cloud platform by using a microprocessor board. These sensors send real-time data to the cloud by keeping track of the status of each parking slot and provide an online platform for the user to check the availability of parking slots. The final stage of the system provides an application interface that sends the user's parking details through e-mail and displays the same on the application.

Keywords Parking management system · Internet of Things · Cloud management · Parking security management

1 Introduction

Private vehicle ownership has engendered a major setback to the public transport community in India resulting in a vast proliferation of private automobiles over the last two decades. The number of registered vehicles in India has increased from 55 million in 2001 to 210 million in 2015 and is subject to escalate [1]. Being the second most populated country in the world these new private automobiles have also incurred a setback in its development.

A. Sarin (✉) · D. Thanawala
Mukesh Patel School of Technology Management and Engineering, Electronics and Telecommunication, Mumbai 400054, India
e-mail: aditya28sarin@gmail.com

D. Thanawala
e-mail: deveshi.m.thanawala@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_34

One of the primary effects due to this population explosion is space management that has given rise to an inevitable problem of vehicle parking. Vehicle management at the parking lots has become an important aspect of the best utilization of existing parking area capacity. Wastage of time and fuel at the parking area due to an increase in vehicle count has become a major issue for everyone [2].

Internet of Things (IoT) relates to a vast community of devices connected to the Internet which form a broad network where data from these billions of devices are shared [3]. The main purpose of developing an IoT-enabled parking management system is to put an end to the endless parking area problems such as finding availability of slots and the irregularity of entry and exit of the vehicles that cause congestion in such parking lots.

The proposed system works in three different stages. The first stage is designed to provide an automatic gate control mechanism that detects the presence of any vehicle and lets it in via the automatic gate. Once the vehicle has entered, an LCD screen keeps track of the number of vehicles that has entered. The second stage comprises the parking slots being enabled with sensors that detect if a vehicle is parked or not, sending this data to the cloud server that can then be monitored by the security officials in the control room. The security personnel can then track the unwanted congestion of vehicle movement inside the parking space by keeping a check on the number of vehicles entering and exiting the parking lot, which has to be same at all instants with a delay of some minutes acceptable for a car to move out of the parking lot. The third stage displays the cloud data on a mobile application and also sends an e-mail to the concerned vehicle owner about his parking details.

The structure of the paper is as follows: Sect. 2 explains the literature survey. Section 3 describes the system layout along with hardware description and the design flow which derives a stepwise procedural order in which the model works. Section 4 enumerates results and discussion. Section 5 includes the conclusion.

2 Related Work

Outdoor parking slots are available with GPS to keep a track on the position of the vehicle whereas for indoor parking slots Wi-Fi-enabled signatures were used to provide a similar mechanism [4].

Several models have been proposed to enable vehicle detection by interfacing sensors at the parking site [5]. Alternate solutions have been provided using RFID technology to automate car parking systems [6]. A large number of parking models have been proposed and implemented using Bluetooth, wireless networks, Zigbee, and RFID technologies. Among such works, the idea proposed in paper [7] explains the use of RFID for a smart parking system where the parking slot availability is managed by using RFID and periodically saved in the database.

Paper [8] provides an efficient interface for user interaction but uses infrared sensors for detection at a parking spot but such sensors have a low range of detection (about 20 cm). However, the paper gives an additional solution using advanced driver assistance which provides an efficient algorithm for charging the driver on the basis of the time for which the car is parked.

Paper [9] provides a solution for parking challenges being faced in Hyderabad by providing online booking of parking slots along with entry based on a barcode scanner. Paper [10] explains the usage of smart parking with the help of Zigbee communication and interfacing IR sensors for the detection of vehicles also using RFID technology. Paper [11] uses MQTT protocol which provides three levels of security levels on top of the TCP layer. It also integrated temperature sensors to enable the safety of the parking lot. Ultrasonic sensors are used for indoor parking and IR sensors are used to keep a check in outdoor area parking, sending this data to a web page.

Paper [12] provides a comparative study of different methods that are employed in improving exiting parking conditions, it provides methods such as using ultrasonic sensors, managing parking status using the online cloud, and RFID to detect the vehicles and sensor nodes. Paper [13] developed a queuing mechanism additionally which works in a first-in first-out manner, which is structured for peak hour and non-peak hours timing as well, providing statistical parameters to function in a multiple queue system. In Paper [14], the authors explain a model involving an Android application to ease the parking slot booking system and used a deep learning algorithm to detect the vehicle number plates.

3 Proposed Work

The proposed model illustrates a modular design of the parking management system integrating the hardware components with the software architecture. Figure 1 describes the system layout proposed that forms the foundation of the model illustrating the location and working of each device.

3.1 System Layout

3.1.1 Microprocessor

The microprocessor forms the core of the model that is primarily concerned with the sensor data being uploaded to the online cloud platform. It updates the sensor data and keeps the users updated from time to time. The microprocessor used is a Raspberry Pi B + model. Three ultrasonic sensors are connected that detect the presence of the vehicle in each parking spot.

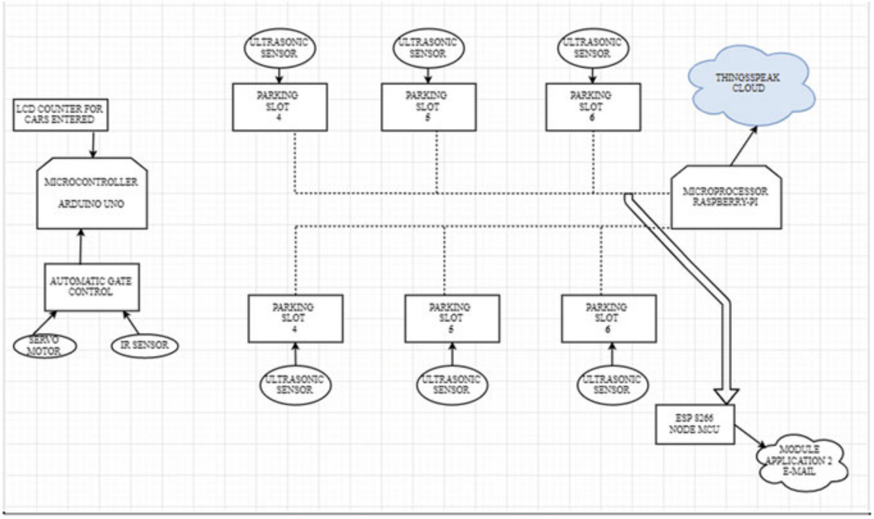


Fig. 1 System layout

3.1.2 Embedded Microcontroller

The microcontroller forms the nucleus of stage 1 of the proposed model that controls the working of the automatic gate control as well as the counter display at the entry of our parking lot. This board is connected to the IR sensor, servomotor, and LCD display. It facilitates the entry of the vehicle inside the parking lot and keeps track of the number of vehicles entering. The microcontroller board used in the proposed model is an Arduino UNO based on the microchip ATmega328P.

3.1.3 Node MCU ESP8266

Node MCU works on the ESP8266 Wi-Fi system-on-chip firmware. This module is connected to an LED at each parking slot indicating that the slot is filled. This data is then displayed on the BLYNK application.

3.1.4 Counter

The counter uses an LCD display which is present at the entry gate of the parking lot displaying the number of vehicles entering the area thereby notifying the driver if any parking lot is available or not thus preventing unnecessary traffic inside the parking lot.

3.1.5 IR Sensor

This sensor transmits an infrared wave from its transmitter terminal which hits the obstacle and returns at its receiver terminal, and hence it can determine the presence of an entity. IR sensor emits infrared radiation and helps in detecting the presence of an obstacle. However, an ultrasonic sensor emits ultrasonic sound waves that can accurately determine how far the obstacle is.

This sensor is not used as a parking sensor because IR sensors are highly sensitive to dust, smoke, and light. Moreover, IR sensors have a small detection range, and hence it can be used at the entry where vehicles move close to them but could fail to detect vehicles with their base at a greater height from the ground.

3.1.6 Servomotor

The servomotor forms the mechanism of movement for the entry gate, this servomotor is connected to the microcontroller and is programmed to provide rotation by 90° in clockwise as well as in counterclockwise direction.

3.1.7 Ultrasonic Sensor

This sensor detects the vehicle when the ultrasonic waves are blocked also calculating the distance at which the obstacle is present. It is placed with the sensor's transmitter and receiver facing upward, installed at the center of each parking spot. The orientation of the sensor is illustrated in Fig. 2. The sensor has a maximum range of 4 m and a minimum range of 2 cm with an angle that can be varied by 15°.

$$D = (V \times t)/2 \tag{1}$$

Formula 1 is used to calculate the distance between the sensor and the obstacle.

where

D required distance,
V speed of sound, and
t time sound wave takes after striking the object.

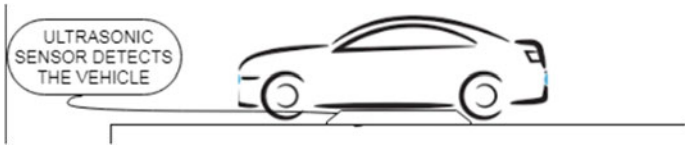


Fig. 2 Ultrasonic sensor orientation

The distance “*D*” is calculated using the above formula and this is how the sensor calculates the distance between an obstacle and itself.

3.1.8 ThingSpeak

ThingSpeak is an open-source cloud platform designed for the Internet of Things application developed by Mathworks. This online platform lets us store and collect such sensor data and interact with different forms of data stored. We can either keep our data public or private.

3.1.9 Blynk Application

Blynk is a user application specially designed for the Internet of Things. It is used to control hardware remotely, display sensor data, and also to store the data. In the proposed model, Blynk is used as an application through which the users can keep a track on parking availability. The application also provides notification service to the user via e-mail stating the location of the user’s parked vehicle.

3.2 Design Flow

Figure 2 above explains the flow of system architecture.

The entire flow can be described in four steps:

Step 1

The vehicle arrives at the parking lot entry, the driver can then check for the availability of parking spaces on the LCD counter, if any. If there are empty parking slots the vehicle enters the area through an automated gate that is a combination of an IR sensor and a servomotor connected to the Arduino Uno microcontroller. This microcontroller programs the threshold of the IR sensor. The sensor detects the entry of the vehicle that then sends signals to the microcontroller. After receiving the signal, the servo gate rotates and opens the entry for the incoming vehicle, and simultaneously the LCD counter updates the number of vehicles entered in the parking space (Fig. 3).

This forms Stage 1 of the proposed model.

Step 2

Here the vehicle that has already entered moves to the vacant parking spot. Each parking slot has an ultrasonic sensor engraved that detects if a car is parked or not. The sensor is programmed to a preset threshold. The transmitter sends the ultrasonic wave which is then received at the receiver after being reflected from the underside of the vehicle. If the distance of the obstacle is less than the given threshold, the sensor gets triggered and sends data accordingly.

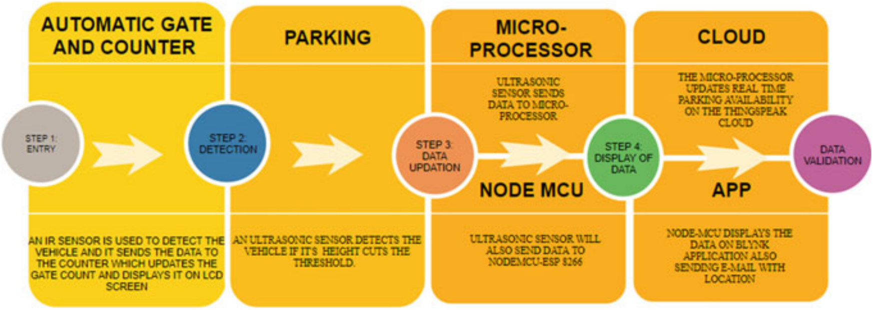

AUTOMATIC GATE AND COUNTER
PARKING
MICRO-PROCESSOR
CLOUD
STEP 1: ENTRY
STEP 2: DETECTION
STEP 3: DATA UPDATION
STEP 4: DISPLAY OF DATA
DATA VALIDATION
ULTRASONIC SENSOR SENDS DATA TO MICRO-PROCESSOR
THE MICRO-PROCESSOR UPDATES REAL TIME PARKING AVAILABILITY ON THE THINGSPEAK CLOUD
NODE MCU
APP
AN IR SENSOR IS USED TO DETECT THE VEHICLE AND IT SENDS THE DATA TO THE COUNTER WHICH UPDATES THE GATE COUNT AND DISPLAYS IT ON LCD SCREEN
AN ULTRASONIC SENSOR DETECTS THE VEHICLE IF IT'S HEIGHT CUTS THE THRESHOLD.
ULTRASONIC SENSOR WILL ALSO SEND DATA TO NODEMCU-ESP 8266
NODE-MCU DISPLAYS THE DATA ON BLYNK APPLICATION ALSO SENDING E-MAIL WITH LOCATION

Fig. 3 Design flow

Step 3 and Step 4

Step 3 and 4 are further divided into two parts.

Path 1

The data from the sensor is sent to the microprocessor which runs an algorithm to calculate the distance between the obstacle and the sensor with the help of the time taken for the wave to return. This data is then sent to an online cloud platform where it displays the number of parking slots and even the real-time parking availability so users can keep a check on the availability of parking slots conveniently. If the distance falls under the pre-determined threshold then the indicator on the cloud indicates that the particular slot is filled. The cloud platform used is the ThingSpeak cloud platform.

Path 2

The data from the sensor activates the on-site led which sends data to the Node-MCU ESP 8266 module. This module is programmed to connect with the Blynk application where the parking chart is displayed in a similar manner to the cloud platform, which provides users an alternate way to keep a check on the availability of the parking slots. The application is also programmed to provide an additional feature in which the user receives an e-mail notification as soon as he parks his car, specifying the users' vehicle location (block number and the floor number).

3.3 *Logic Flow*

Figure 4 explains the algorithm for vehicle detection using an ultrasonic sensor and sending data to the server.

The proposed model describes an algorithm for the operation of ultrasonic sensor. The sensor is first initialized by keeping the trigger pin low, this pin is an output pin

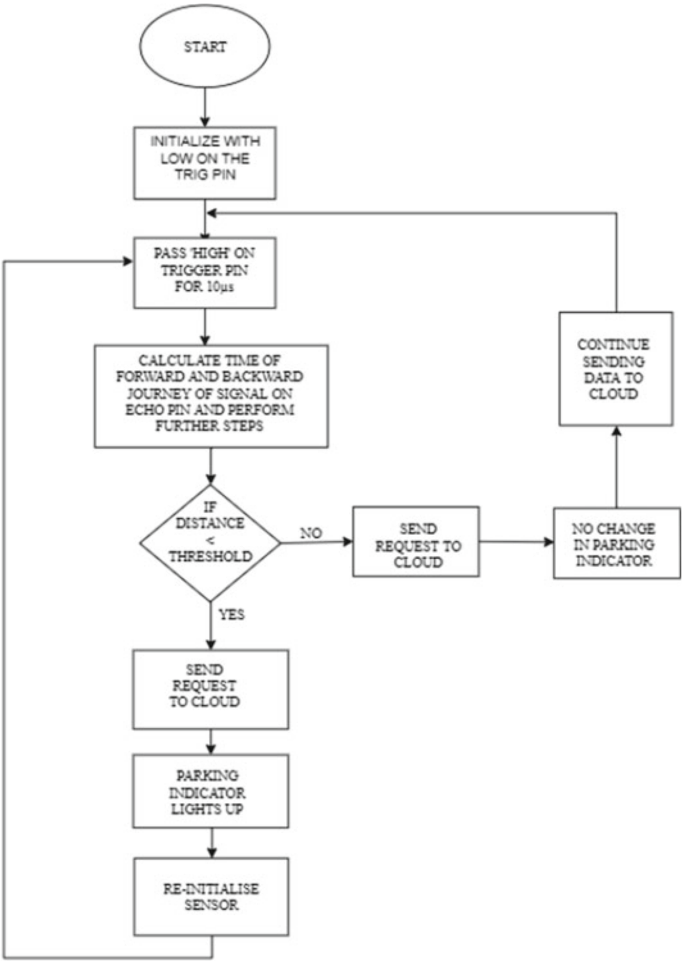

START
INITIALIZE WITH LOW ON THE TRIG PIN
PASS 'HIGH' ON TRIGGER PIN FOR 10µs
CALCULATE TIME OF FORWARD AND BACKWARD JOURNEY OF SIGNAL ON ECHO PIN AND PERFORM FURTHER STEPS
IF DISTANCE < THRESHOLD
NO
SEND REQUEST TO CLOUD
NO CHANGE IN PARKING INDICATOR
CONTINUE SENDING DATA TO CLOUD
YES
SEND REQUEST TO CLOUD
PARKING INDICATOR LIGHTS UP
RE-INITIALISE SENSOR

Fig. 4 Logic flow of ultrasonic sensor detection

and sends ultrasonic signals. The trigger pin is then made high for 10 µs for the signal to transmit. After 10 µs the trigger pin is made low again.

Now a threshold is set in the sensor, suppose "*x*", this threshold is set at 25 cm. If a car with a ground clearance of 10 cm is parked, the sensor detects an obstacle within its threshold range and indicates a deviation from its ideal state. The cloud platform detects a change in the sensor state and indicates the same on the screen. This demonstrates that the parking slot is occupied. The indicator turns off when the car leaves the parking slot, and hence the sensor has no obstacle in its threshold range.

Figure 5 explains the algorithm for vehicle entry using gate control and counter.

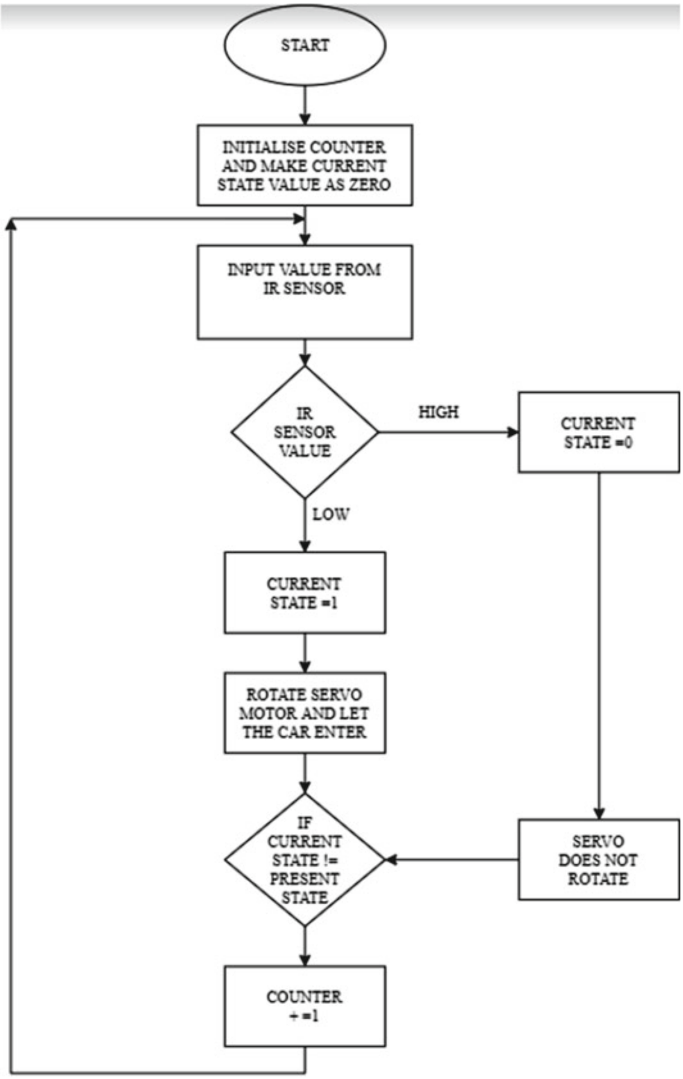

START
INITIALISE COUNTER AND MAKE CURRENT STATE VALUE AS ZERO
INPUT VALUE FROM IR SENSOR
IR SENSOR VALUE
HIGH
CURRENT STATE =0
LOW
CURRENT STATE =1
ROTATE SERVO MOTOR AND LET THE CAR ENTER
IF CURRENT STATE != PRESENT STATE
SERVO DOES NOT ROTATE
COUNTER + =1

Fig. 5 The logic flow of stage 1

Stage 1 of the proposed model consists of the counter, automatic gate control, and an IR sensor.

The IR sensor emits infrared radiation continuously. A counter variable is created that keeps track of the number of vehicles entering the parking lot. The counter has current as well as a previous state variable. These variables are initialized first and then the input from the IR sensor is taken. A preset threshold is applied to the IR. As soon as the microcontroller board receives a signal when the obstacle is detected, the current state increments by one and the microcontroller sends a signal to the servomotor specifying the degree of rotation it has to make.

Figure 6 explains the algorithm for vehicle detection and sends data to Blynk application.

As soon as the ultrasonic sensor detects a change it transmits a signal to the node MCU which acknowledges this change and displays the same on the application and an e-mail notification is sent to the user specifying the parking slot number. There is an on-site led which indicates if a slot is empty or not.

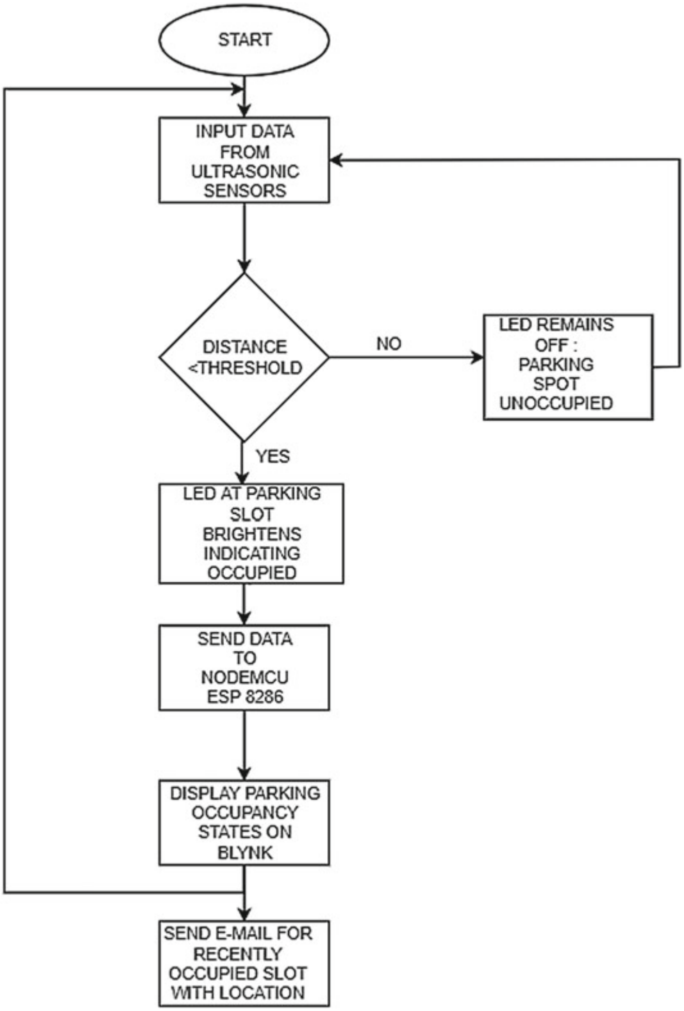

START
INPUT DATA FROM ULTRASONIC SENSORS
DISTANCE <THRESHOLD
NO
LED REMAINS OFF : PARKING SPOT UNOCCUPIED
YES
LED AT PARKING SLOT BRIGHTENS INDICATING OCCUPIED
SEND DATA TO NODEMCU ESP 8286
DISPLAY PARKING OCCUPANCY STATES ON BLYNK
SEND E-MAIL FOR RECENTLY OCCUPIED SLOT WITH LOCATION

Fig. 6 Logic flow for node MCU with app

4 Results Analysis

The proposed model is designed in such a way that minimum delay is maintained throughout the process to provide users an easier and faster solution to their daily parking problems.

The stage I of the model shows communication between the IR sensor, microcontroller, LCD display, and the automatic gate which occurs in approximately 1.2 s. The prototype for the same is shown in Fig. 7.

The model serves the following advantages:

- It provides early access to the parking lot for users to check parking availability through the cloud platform.
- This model minimizes congestion inside the parking lot as the driver can check for a number of empty parking slots, from the entry gate screen counter.
- Users need not remember the location of their vehicle as an e-mail notification is provided as soon as the user parks his vehicle.
- Security personnel can keep track of the parking history of any particular slot.
- The model also provides a system to keep track of the number of vehicles entering and the number of vehicles being parked at the same time. If the number of vehicles parked is not equal to the number of vehicles that have entered then the security member can infer that a vehicle has left its slot, therefore keeping a better track on vehicle congestion.

Figure 8 shows the parking data displayed on the online cloud platform (ThingSpeak). The data is available in a graphical form as well as in the form of an indicator. The graphical form also displays data history. The indicator signifies the availability of the parking slot. The data takes 15 s to be displayed on the cloud platform.

Figure 9 shows the application interface that displays similar data as on the cloud platform. This data, however, follows its path through the node MCU. The data here signifies two parking slots out of which three are occupied currently.

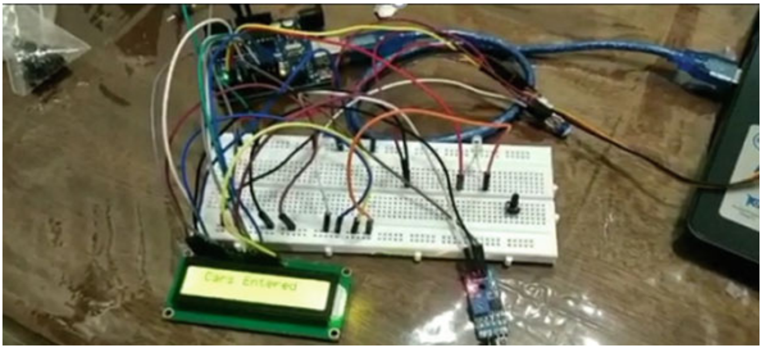


Fig. 7 Prototype for stage I

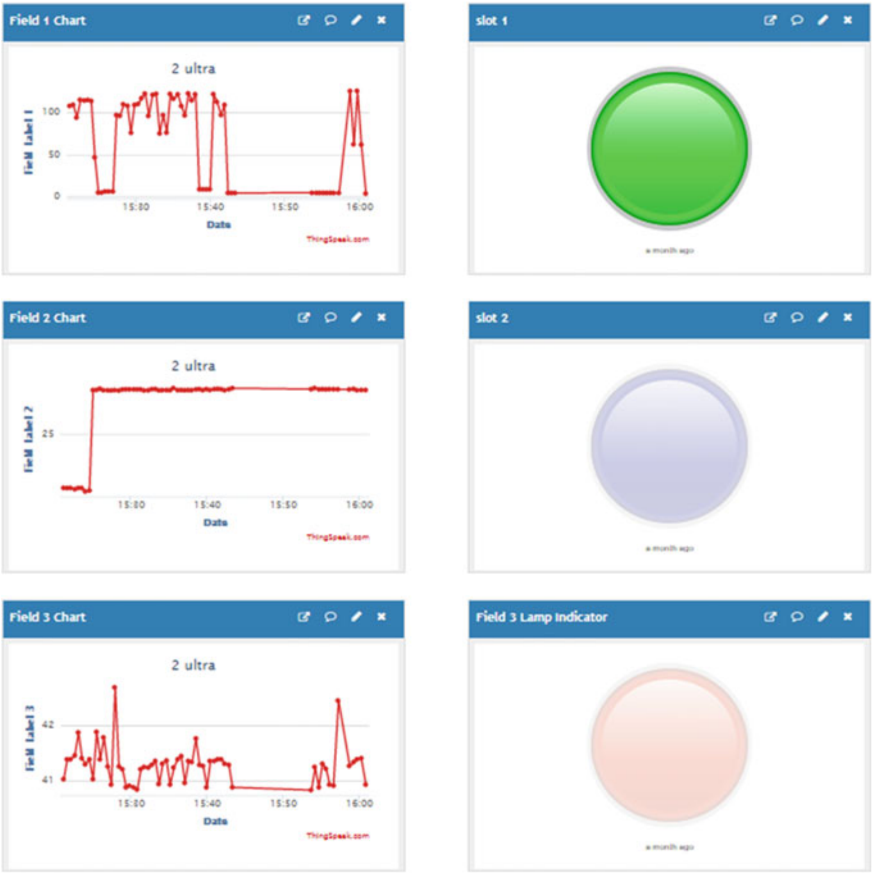


Fig. 8 ThingSpeak results

Figure 10 shows an e-mail notification that the user receives as soon as he parks his car, in this case, at the P3 parking slot.

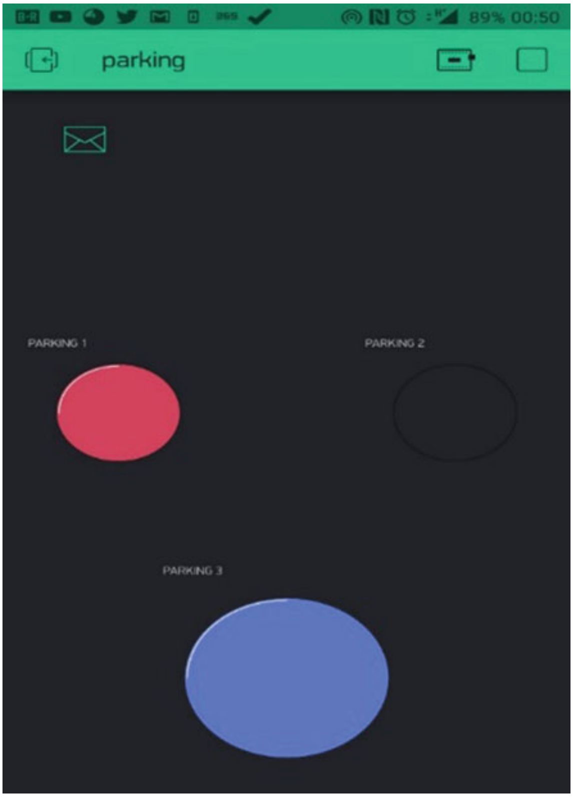
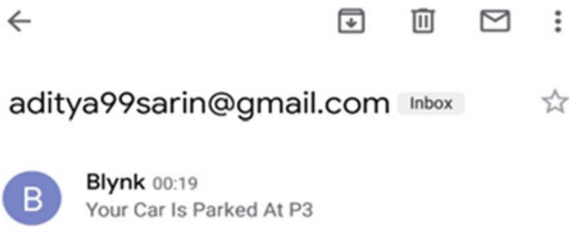


Fig. 9 Output on Blynk application


aditya99sarin@gmail.com Inbox

Blynk 00:19
Your Car Is Parked At P3

Fig. 10 User receiving e-mail

5 Conclusion

Parking has always posed an additional threat to the interminable congestion on roads and public areas leading to frustration of the driver. This model provides a solution to such a problem with the Internet of Things-based technology to build a parking management system that integrates sensors at the parking site and sends data to the online cloud platform as well as the user application. It provides an e-mail service with vehicle location thus limiting congestion by keeping track of vehicle entry and exit.

The proposed model shows that the time user takes to find his/her parking has been reduced significantly by deploying three sources by which the user can keep track of the parking availability, that is, from an online cloud platform, user application as well as the LCD indicator on-site. The proposed system not only satisfies the user demands but also forms a well-regulated chain of structured data where the security personnel can keep track of their incoming and outgoing vehicles leading to minimum inconvenience.

In the future, the same model can be made more efficient by including an application that can be used for the reservation of parking slots. Along with this, number plate recognition could be used to keep track of the vehicles that have entered the parking space.

References

1. Singh J (2016) Head of the UITP India Office, Public Transport Developments in Indian Cities, Intelligent Transport. https://www.intelligenttransport.com/transport-articles/21458/city-public-transportation-india/
2. Rane S, Dubey A, Parida T, Design of IoT based intelligent parking system using image processing algorithms. In: Proceedings of the IEEE 2017 international conference on computing methodologies and communication (ICCMC)
3. Clark J (2016) What is the internet of things?. https://www.ibm.com/blogs/internet-of-things/what-is-the-iot/
4. Correa J, Katz E, Collins P, Griss M (2008) Room-level wi-fi location tracking
5. Fatima N, Jagtap P, Natkar A, Choudhari ST, Nazish F et al (2018) IOT Based smart car parking system for smart cities. Int J Adv Res Ideas Innov Technol 4(1)
6. Anusooya G, Christy Jackson J, Sathyarajasekaran K, Kannan K (2017) RFID based smart car parking system. Int J Appl Eng Res 12(17):6559–6563. ISSN 0973-4562, ©Research India Publications
7. Hanche SC, Munot P, Bagal P, Sonawane K, Pise P (2013) Automated vehicle parking system using RFID, ISSN (PRINT) 1(2):2320–8945
8. Mahendra BM, Dr. Sonoli S, Bhat N (2017) IoT based sensor enabled smart car parking for advanced driver assistance system. In: 2017 2nd IEEE international conference on recent trends in electronics information & communication technology (RTEICT), India
9. Vakula D, Kolli YK, Low cost smart parking system for smart cities. In: Proceedings of the international conference on intelligent sustainable systems (ICISS 2017) IEEE Xplore Compliant—Part Number: CFP17M19-ART, ISBN:978-1-5386-1959-9
10. Dhumal YR, Waghmare HA, Tole AS, Shilimkar SR (2016) Android based smart car parking system. IJREEIE 5(3):1371–1374
11. Chippalkatti P, Kadam G, Ichake V (2018) I-SPARK: IoT based smart parking system. In: 2018 international conference on advances in communication and computing technology (ICACCT) Amrutvahini College of Engineering, Sangamner, Ahmednagar, India
12. Shanmugasundaram G, Vigneswari T, Abinaya R, Jane Nemisha V, Monisha M, A review on dimension based parking system. In: Proceeding of international conference on systems computation automation and networking 2019, @IEEE 978-1-7281-1524-5
13. Pandey D, Hanchate S (2018) Navigation based-intelligent parking management system using queuing theory and IOT. IEEE. 978-1-5386-5657-0/18/$31.00c
14. Hasan MO, Islam MM, Alsaawy Y, Smart parking model based on internet of things (IoT) and TensorFlow. In: IEEE 2019 7th international conference on smart computing & communications (ICSCC). 978-1-7281-1557-3/19/$31.00©2019

Machine Learning based Restaurant Revenue Prediction

G. P. Sanjana Rao, K. Aditya Shastry, S. R. Sathyashree, and Shivani Sahu

Abstract Food industry has a crucial part in enhancing the financial progress of a country. This is very true for metropolitan cities than any small towns of our country. Despite the contribution of food industry to the economy, the revenue prediction of the restaurant has been limited. The agenda of this work is basically to detect the revenue for any upcoming setting of restaurant. There are three types of restaurant which have been encountered. They are inline, food court, and mobile. In our proposed solution, we take into consideration the various features of the datasets for the prediction. The input features were ordered based on their impact on the target attribute which was the restaurant revenue. Various other pre-processing techniques like Principal Component Analysis (PCA), feature selection and label encoding have been used. Without the proper analysis of Kaggle datasets pre-processing cannot be done. Algorithms are then evaluated on the test data after being trained on the training datasets. Random Forest (RF) was found to be the best performing model for revenue prediction when compared to linear regression model. The model accuracy does make a difference before pre-processing and after pre-processing. The accuracy increases after the applied methods of pre-processing.

Keywords Prediction · Machine learning · Pre-processing · Restaurant · Revenue

G. P. Sanjana Rao · K. Aditya Shastry (✉) · S. R. Sathyashree · S. Sahu
Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: adityashastry.k@nmit.ac.in

G. P. Sanjana Rao
e-mail: sanjanasachit@gmail.com

S. R. Sathyashree
e-mail: sathyashree97@gmail.com

S. Sahu
e-mail: sahushivani001@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_35

1 Introduction

Restaurant is just not about food, it is about giving the customer satisfaction of their time spent there. In the cities and small towns, we often find many inlets not running properly despite being present in the best locality in town. The reasons are many. However, few of the rules may be followed before opening a restaurant. With its additional importance of contributing to the economy of a country, it should be prioritized before opening the restaurant. On numerous occasions, there is a disparity amid the location and the inlets of the restaurant [1].

Hence, it is a part of our day-to-day life in one way or the other. A good restaurant can lead to the employment of a few more workers. Manually trying to open a restaurant in the traditional layman fashion can be tedious job and time-consuming. It is just not about giving a lot of money for its opening. The proposed solution does give us an insight into the revenue. If it is higher then it can be opened. This solution is more efficient and cheaper. Some of the challenges faced in predicting the revenue using machine learning are availability, granularity, and clarity of dataset as follows [2, 3].

The main objectives of this work are as follows:

- Analysis of the Kaggle datasets.
- Pre-process the raw dataset by removing the unwanted variables.
- To understand the feature importance of the datasets and their requirement.
- Forecast the profits of a restaurant.

The above objectives are achieved using the following modules:

- Pre-processing: This module cleans the data in the sense as to remove the unwanted variables, fill up the missing values, resolving the inconsistencies in the data.
- Feature: This module is utilized to choose the subset of relevant attributes out of all the features present. The chosen attributes are used for further analysis.
- Random Forest: This module is a supervised learning algorithm. It creates a forest of decision trees. It constructs several decision trees and combines them to obtain a more precise and stable forecast.

The remainder of the paper is structured as follows. Section 2 discusses the related work in the area of restaurant revenue prediction. Section 3 demonstrates the proposed work. Section 4 describes the experimental setup and results. The paper ends with a conclusion and references.

2 Related Work

This section discusses some of the recent and significant works carried out for restaurant revenue prediction using machine learning. The work by Raul et al. [4] is performed on the demographic data, real estate data, and points of interest. They applied the concept of support vector machines and random forest algorithms. The prediction is the annual revenue of the restaurants which would help in determining the feasibility of the outlets. The RF algorithm is promising here because it is a flexible, efficient, and effective for a variety of applications. It provides an error estimate in which the predictor with the lowest error is preferred.

The work by Danks et al. [5] focuses more on whether the weather forecast improves sales forecast. Daily sales per restaurant are predicted. Predictive accuracy comparison is done where the level of daily sales per restaurant including holiday and weather gives little additional accuracy. Considering a single 3-h mealtime (breakfast, lunch, and dinner) with data related to holiday and weather gives higher predictive accuracy. Concepts such as linear regression with and without including the weather data are done to make a valid comparison for its predictability. Hence, it proves that features such as weather make a valid impact on the sales forecast.

In [6], the authors Holmerg & Hallden used supervised learning over unsupervised learning. It creates a more accurate revenue forecast than what can be done by comparing it with the previous revenue of a restaurant. Different datasets are utilized to forecast the revenue by implementing integration among the datasets. The implementation is done by first understanding the availability of the datasets. Attributes such as weather are kept into consideration. Data that is extracted is used for the training phase. It is important for the result. This includes both creating and selecting the good ones. Data pre-processed is where they check the distribution of weakness in the datasets. Finally, the co-relation among the datasets is formed where the task of selecting good features is useful if it is correlated with the target attribute. The dataset is subjected to normalization and standardization.

3 Proposed Work

This section demonstrates the devised technique used for restaurant revenue prediction using a machine learning approach. Figure 1 depicts this process.

3.1 Dataset Description

In this work, the training data of the Kaggle dataset consists of 137 restaurants [7]. The data column includes the open date, location, city type, and three categories of co-related data: demographic data, real estate data, and commercial data. The columns indicate the revenue that was transformed in a specified year that denotes the target of predictive analysis. Compared with other datasets which had a missing feature of revenue which itself was avoided. Each feature serves toward better clarity of the desired solution. It has a parameter of p1–p37 which is collected from third-party providers through GIS systems. It includes the population in any given area, age and gender distribution, and development scales.

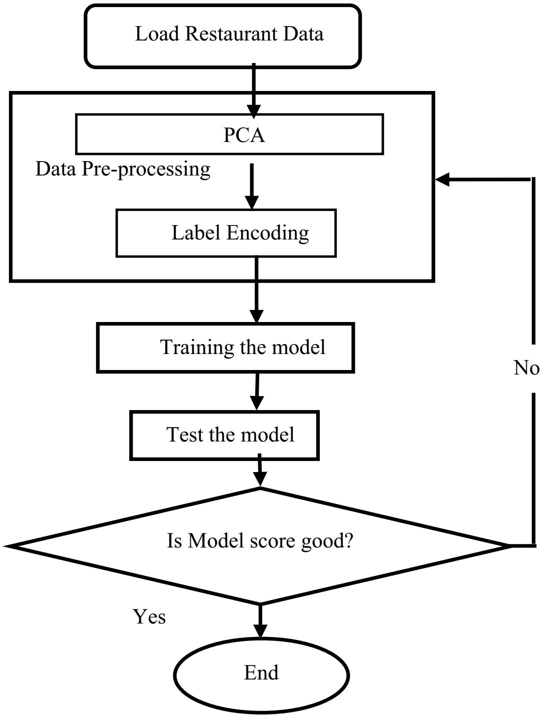


Fig. 1 Proposed ML approach for restaurant revenue prediction

3.2 *Pre-processing*

The pre-processing step involves the preparation of raw data into a format that is readable to the machine learning algorithm. In our work, we have utilized PCA and label encoding as the data pre-processing steps

(i) Principal Component Analysis (PCA):

PCA is a simple technique utilized for dimensionality reduction which defines that the number of feature variables is decreased by narrowing down the important features. It basically has three main steps [8]:

- *Step-1 Computing the covariance matrix of the data*: This involves that the features are well balanced initially among each other. In order to accomplish this data is normalized. Each feature initially is weighed equally for the computation. Covariance specifies how well correlated the variables are among each other. Two variables are positively correlated if they are dependent on each other such that increasing or decreasing one variable results in the increase or decrease of the other variable. A negative correlation indicates that changing one variable results in the modification of another variable in the opposite direction. Equation-1 provides the covariance of feature X.

$$\mathrm{Cov}(X) = 1/(n-1)\left[\left(X - x'\right)^T\left(X - x'\right)\right] \tag{1}$$

where x' denotes the vector of mean values for each feature of X, "n" is the number of observations, and T is the transpose. When a transposed matrix is multiplied by the original one, each of the features is multiplied together.

- *Step-2 Computing the Eigen Values and Eigen Vectors*: The principal components denote the eigenvalues and their magnitude denotes the eigenvectors. Higher eigenvalues with high magnitude of eigenvectors denote the data with high variance associated with the corresponding feature present in the vector space. Any movement causes a lot of variances. Low variance is achieved by vectors having small eigenvalues since the data does not change significantly when moving along the vector. The main agenda behind is it is that finding the most significant attribute and ignoring the rest.
- *Step-3 Projections onto new vectors:* By this step, the eigenvectors list has been arranged in order of importance built on their eigenvalues. For example, suppose our dataset has ten features. After computing the covariance matrix we get certain eigenvalues such as [12, 10, 8, 7, 5, 1, 0.1, 0.03, 0.005, 0.0009]. The total sum of this array accounts for 43.1359. The first 6 values give 42 and therefore 42/43.1359 = 99.68% of the total. This states that the first six holds effective than the rest four which can be neglected. Finally, we simply concatenate all eigenvectors decided to keep. These are the steps to decrease the dimensionality.

(i) Label encoding: Here, the labels are encoded into its corresponding numeric representation for transforming it into a machine-readable form. The ML algorithms then operate on these in a better manner. It is a significant step in pre-processing that works on the structured dataset in supervised learning. For instance, presuming that a height column comprises three values like short, tall, and medium. Then, after the application of the label encoding technique, tall is assigned 0, the medium is assigned 1, and 2 is assigned to short [9].

(ii) Feature selection: Every problem statement could have many numbers of features on which it is dependent on. It is crucial to recognize the features contributing to the solution. Picking the initial best of it helps in predicting [9]. We have taken the 14 features from a set of 43 of it after this process. For the feature containing the City group, it has two parts of it other and Big Cities. The data concludes that Big Cities have a higher contribution toward the prediction. Similarly, with the City type, it has been classified into three subcategories such as DT, IL (in line), FC (food court). The food court has the highest contribution among the two. It can be concluded that bigger cities and food courts contribute to larger revenue. In order to decrease the number

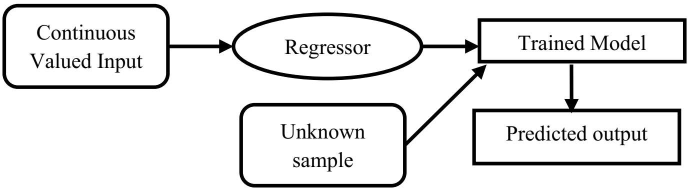


Fig. 2 Working of linear regression

of variables, the dataset was analyzed and based on its variance and impurity the variables were ranked.

3.3 Random Forest Regression Algorithm

RF algorithm represents an ensemble method. There are many decision trees in it. It uses a bagging concept where every decision tree is trained on separate data records in which sampling is done with replacement. The main objective is to combine multiple decision trees in determining the final output rather than relying on individual decision trees [10]. Regression trees are used because the variables are continuous. It separates the predictor space into regions that are unique and non-overlapping. The decision to split affects the tree's accuracy [10].

3.4 Linear Regression

In this model, we construct an equation using our own data. It is utilized to make forecasts about one variable based on the values of another variable. The variables we are predicting become the dependent variable and the variables being used to make these predictions are called independent variables. The line that best fits the curve is called the regression line. It reduces the Sum of Squared Errors (SSE) between the points that are plotted. Error is basically the variation and not the mistake [11]. Figure 2 depicts the working of linear regression.

3.5 Cross-Validation (CV)

The stability of the ML model can be validated using the CV technique. Any model cannot be fit in the training data and hope it would give the exact work or output. The indication of the generalizing ability on the dataset can be provided by the CV technique. It is basically the technique that trains the model using the subset of the dataset and then evaluates using the complementary subset of the dataset [12].

4 Result Analysis

This section describes the setup of experiments along with the associated results which are discussed. In our work, we made use of PyCharm, Pandas, NumPy, and Matplotlib libraries for development on the Windows platform.

The dataset contains variables such as Id, open date, city, city group, type, and other unclear data.

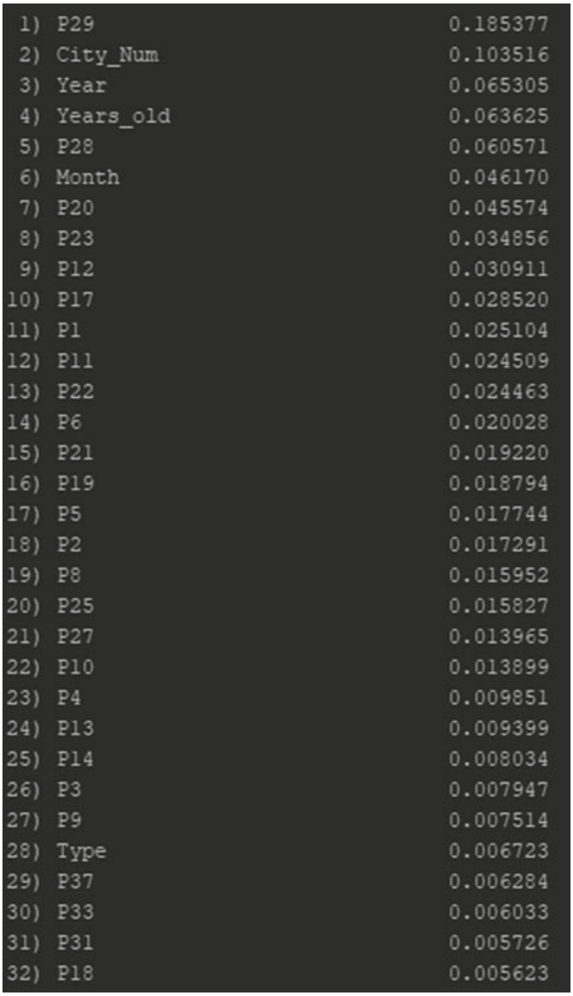
- Id: Restaurant Id.
- Open Date: opening date of the restaurant.
- City: a city that restaurants are present.
- City Group: Type of the city.
- Type: Type of Restaurants.
- P1, P2, …, P37: there are three categories of these obfuscated data. Demographic data are gathered from third-party providers. Based on its ranking a feature importance graph was plotted using Matplotlib. To understand how much difference each variable or parameters make.

Figure 3 illustrates the score obtained for each value indicating its importance in revenue prediction.

The RF and linear regression were compared using Root Mean Squared Error (RMSE) as the performance metric. We obtained an RMSE value of 1625604.12 for the random forest and 1923941.26 for the linear regression. This indicated that the random forest performed better restaurant revenue prediction than the linear regression as it had lesser RMSE value.

5 Conclusion

In conclusion, we would like to conclude that the work on restaurant revenue prediction is developed with the intention to predict the revenue in any upcoming location. Our work used the datasets provided by Kaggle. Many models with their ability to solve the specifics have been used here. RF algorithm has been the best because of its ability to handle huge and diverse datasets. A reference can be provided to aid human judgment and operational losses. It predicts the annual revenue of a new restaurant which would help the food chains determine the feasibility of a new outlet. A comparison between the RF algorithm and regression models demonstrated that the RF algorithm performed better than the regression model with respect to RMSE. In the future, this work can be extended by utilizing a larger dataset and extracting a greater number of features to provide a higher level of accuracy. A mobile application or a web interface can be developed for the same.



```
 1) P29                              0.185377
 2) City_Num                         0.103516
 3) Year                             0.065305
 4) Years_old                        0.063625
 5) P28                              0.060571
 6) Month                            0.046170
 7) P20                              0.045574
 8) P23                              0.034856
 9) P12                              0.030911
10) P17                              0.028520
11) P1                               0.025104
12) P11                              0.024509
13) P22                              0.024463
14) P6                               0.020028
15) P21                              0.019220
16) P19                              0.018794
17) P5                               0.017744
18) P2                               0.017291
19) P8                               0.015952
20) P25                              0.015827
21) P27                              0.013965
22) P10                              0.013899
23) P4                               0.009851
24) P13                              0.009399
25) P14                              0.008034
26) P3                               0.007947
27) P9                               0.007514
28) Type                             0.006723
29) P37                              0.006284
30) P33                              0.006033
31) P31                              0.005726
32) P18                              0.005623
```

Fig. 3 Variables contributing to the impact

References

1. Rarh F, Pojee D, Zulphekari S, Shah V (2017) Restaurant table reservation using time-series prediction. In: 2017 2nd international conference on communication and electronics systems (ICCES), Coimbatore, pp 153–155
2. Ma X, Tian Y, Luo C, Zhang Y (2018) Predicting future visitors of restaurants using big data. In: 2018 international conference on machine learning and cybernetics (ICMLC), Chengdu, pp 269–274
3. Hossain FMT, Hossain MI, Nawshin S (2017) Machine learning based class level prediction of restaurant reviews. In: 2017 IEEE Region 10 humanitarian technology conference (R10-HTC), Dhaka, pp 420–423
4. Raul N, Shah Y, Devganiya M (2016) Restaurant revenue prediction using machine learning, research inventy. Int J Eng Sci 6(4):91–94

5. Danks N, Martinez I, Ashouri M, Rivera P (2016) Forecasting restaurant sales using data from iChef, weather forecasts, and holiday information, National Tsing Hua University
6. Holmberg M, Halldén P (2018) Machine learning for restaurant sales forecast (Dissertation). http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-353225
7. Dataset: Restaurant Revenue Prediction (2018). https://www.kaggle.com/c/restaurant-revenue-prediction
8. Sehgal S, Singh H, Agarwal M, Bhasker V, Shantanu (2014) Data analysis using PCA. In: 2014 international conference on medical imaging, m-health and emerging communication systems (MedCom), Greater Noida, pp 45–48
9. Song F, Guo Z, Mei D (2010) Feature selection using PCA. In: 2010 international conference on system science, engineering design and manufacturing informatization, Yichang, pp 27–30
10. Paul A, Mukherjee DP, Das P, Gangopadhyay A, Chintha AR, Kundu S (2018) Improved random forest for classification. IEEE Trans Image Process 27(8):4012–4024
11. Kavitha S, Varuna S, Ramya R (2016) A comparative analysis on linear regression and support vector regression. In: 2016 online international conference on green engineering and technologies (IC-GET), Coimbatore, pp 1–5. https://doi.org/10.1109/get.2016.7916627
12. Yadav S, Shukla S (2016) Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In: 2016 IEEE 6th international conference on advanced computing (IACC), Bhimavaram, pp 78–83. https://doi.org/10.1109/iacc.2016.25

Solving Multi-objective Fixed Charged Transportation Problem Using a Modified Particle Swarm Optimization Algorithm

Gurwinder Singh and Amarinder Singh

Abstract Particle Swarm Optimization (PSO) is population-based algorithm established and enhanced to solve a wide variety of real-life problems. During the last decade, different aspects of PSO have been modified and many variants have been proposed. In this paper, a modified PSO is proposed to solve multi-objective fixed charge transportation problem wherein it optimizes the transportation cost (variable and fixed) as well as time to deliver goods from sources to destinations satisfying certain constraints. The method starts with the variable cost only and then with addition of fixed cost, iterates toward optimal Pareto pair. The simulation results show a significant performance gain by the proposed method and prove it as a competent alternative to existing methods.

Keywords Transportation problem · Optimal solution · Evolutionary technique · Swarm intelligence

1 Introduction

Transportation Problem (TP) is well-known subclass of linear programming problem which undergoes major modifications to realize the management and industrial need. This need, generally, realized in determining an optimal distribution plan to transport goods from the sources to the destinations with an objective to minimize the transportation cost. Hitchcock [1] and Koopmans [2] have standardized the primitive case of TP, usually referred as Hitchcock–Koopmans transportation problem.

When a fixed amount of cost is deferred for active route of distribution plan, the transportation problem is modeled as Fixed Charge Transportation Problem (FCTP). This fixed amount of cost, sometime may be a setup cost, renders the objective

G. Singh (✉)
Research Scholar, IKG Punjab Technical University, Jalandhar, Punjab, India
e-mail: gurwinder.singh@bbsbec.ac.in

A. Singh
BBSB Engineering College, Fatehgarh Sahib, Punjab, India
e-mail: amarinder77@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_36

function nonlinear and problem as NP hard [3]. In practice, various distribution problems can only be modeled as fixed charge transportation problems. For instance, there are invariant transport rates incurred by rail, roads, and trucks which comprise a fixed and a variable cost. This fixed amount of cost may be represented in form of renting and landing fee of property and installation cost of machines/equipment in production units, etc.

Traditionally, the FCTP was considered for single objective only but nowadays the real-life situations encourage for multi-objective criterion. While satisfying a common set of constraints, it copes with conflicting objectives. When these conflicts are resolved simultaneously the problem is known as multi-objective FCTP and it leads toward a set of compromised solutions, i.e., Pareto front. The multi-objective FCTP is stated with a system of m sources of capacities $a_i(i = 1, 2, \ldots, m)$ and n destinations of demands $b_j(i = 1, 2, \ldots, n)$. Let $c_{i,j}$ be the variable cost per unit amount and $t_{i,j}$ be the time of transportation of goods from ith source to jth destination. And $f_{i,j}$ be the fixed cost associated with route i, j. The time $t_{i,j}$ is independent of the shipping amount. The objective of problem is to determine $x_{i,j}$ i.e. the amount of units being transported from ith the source to jth destination such that the total cost and the duration of transportation is minimized. The mathematical model is stated as follows:

$$\begin{aligned}
\text{Minimize } Z &= \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} + \sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}\delta_{ij} \\
\text{Minimize } T &= \text{Max}\{t_{ij} : x_{ij} > 0\} \\
\text{subject to } \sum_{j=1}^{n} x_{ij} &\leq a_i(i = 1, 2, \cdots, m) \\
\sum_{i=1}^{m} x_{ij} &= b_j(j = 1, 2, \cdots, n) \\
x_{ij} &\geq 0(i = 1, 2, \cdots, m; j = 1, 2, \cdots, n)
\end{aligned} \tag{1}$$

where

$$\delta_{ij} = \begin{cases} 1, \text{ if } x_{ij} > 0; \; i = 1, 2, \cdots, m; \; j = 1, 2, \cdots, n \\ 0, \qquad \text{Otherwise} \end{cases} \tag{2}$$

Initially, FCTP was studied as a mixed integer programming problem and solved generally, either by exact or heuristic methods. Among others, Murty [4] has used the exact solution method to solve the FCTP which was later improved by Sadagopan and Ravindran [5]. Gray [6] has solved it by decomposing into a series of subprograms. And later, Steinberg [7], Barr et al. [8], Cabot and Erenquc [9] and Palekar et al. [10] have addressed it with some exact algorithms having a branch-and-bound-like approach. Alternatively, robers and cooper [11] applied an adjacent extreme point technique to Balinski's heuristic [12] in an effort to find a better solution. Adlakha et al. [13–15] have made some significant improvements to refine the results. Although, these heuristics have made a meaningful contribution, still these are computationally expensive and disrupted by the solution dragged at local optimum value far away from global optima. To overcome these issues, many authors have turned to meta-heuristics/evolutionary techniques such as GA, TS, PSO etc. Likewise, Gottlieb and Paulmann [16], Sun et al. [17], Raj and Rajendran [18], Lotfi and Tavakkoli-Moghaddam [19], and El-Sherbiny and Alhamali [20] have deal with FCTP in a different perspective. Recently, the multi-objective FCTP is addressed by Midya [21], Roy [22, 23], and Majumder et al. [24].

Particle Swarm Optimization (PSO) algorithm has gained substantial performance gain over other similar evolutionary techniques due to its prominence features like (a) self-organization, (b) simple algorithmic structure, (c) exploration–exploitation trade-off capability, (d) computationally less expensive, etc. Therefore, it has been effectively applied to different optimization problems.

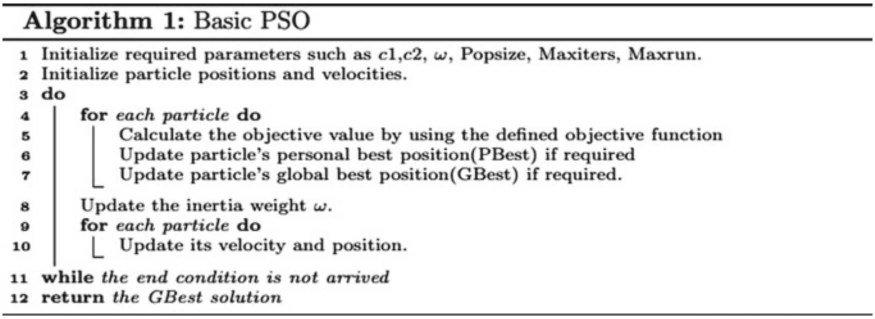
So in this motivation, a modified PSO (discussed in Sect. 3) is proposed to solve the FCTP. The proposed PSO starts with variable cost only and then adding fixed cost iterates toward an optimal solution sequentially. The main objective of this paper is to insinuate a good alternative to existing techniques on one hand and on the other hand to introduce a novel method modifying PSO. The technique also overcomes the rigid condition of analytical techniques. Thus, it provides the necessary decision support to decision-makers handling such kind of logistic problems.

2 Particle Swarm Optimization (PSO)

PSO, perceived by Eberhart and Kennedy [25], is a population based optimization algorithm simulating by the collective conduct of natural organism such as birds or insects. The basic features of PSO are employed when a swarm of randomly initialized particles probes into the solution space while retaining its best positions (local & global) moving towards a global optimum. The path of the particles is regulated by the balance of the nominated local and global best position. Mathematically, it is defined with an assumption of a -dimensional search space wherein the ith particle of the swarm is represented by $X_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ and the velocity of the particle is denoted by $V_i = (v_{i1}, v_{i2}, \ldots, v_{in})$. The best visited position of X_i and the global best position of the swarm may be denoted as $P_{i,best} = (p_{i1}, p_{i2}, \ldots, p_{in})$ and $P_{g,best} = (p_{g1}, p_{g2}, \ldots, p_{gn})$. Then the velocity and the position of the particles are updated using the following equations:

$$\begin{array}{ll} V_i(t+1) = \omega * V_i(t) + c_1 r_1 \left(P_{i,best} - X_i\right) + c_2 r_2 \left(P_{g,best} - X_i\right) \\ X_i(t+1) = \qquad\qquad X_i(t) + V_i(t+1) \end{array} \tag{3}$$

where the inertia weight , the acceleration coefficients c_1, c_2 and the random variables r_1, r_2 play a vital role to retain the accelerated rate of particles towards their *pbest* and *gbest* locations. The pseudocode of basic PSO is devised as follows:



Algorithm 1: Basic PSO

```
Initialize required parameters such as $c1$,$c2$, $\omega$, Popsize, Maxiters, Maxrun.
Initialize particle positions and velocities.
do
    for each particle do
        Calculate the objective value by using the defined objective function
        Update particle's personal best position(PBest) if required
        Update particle's global best position(GBest) if required.
    Update the inertia weight $\omega$.
    for each particle do
        Update its velocity and position.
while the end condition is not arrived
return the GBest solution
```

3 PSO for Fixed Charge Transportation Problem

The complexity of FCTP has made conventional methods inefficient or less effective with increase in size of the problem. PSO resolves this issue by using its intelligent computational feature, which emerges as each particle of swarm interacting iteratively and moving along best positions in search space towards global optima. Moreover, researchers have introduced various encoding techniques so that the discrete optimization problems are amenable to solution by PSO. These encoding techniques are used to discretize the algorithm/solution space wherein the candidate solution, representing the position of particle, is confined to only integer values. During the updation of particle's position, it is necessary to keep the particles in the feasible region.

In implementing the basic PSO on the transportation problem, the particles enter the infeasible search space because of the effects of inertia weight and random variables $r1$ and $r2$. These infeasible particles, containing negative and fractional parts, are repaired to be retained in the feasible solution space. In this paper, two additional modules, viz. Amend Negatives and Amend Fractions, have been incorporated into the PSO to solve the multi-objective fixed charge transportation problem. The convergence of some test problems towards the respective optimal solution using this procedure is also presented.

3.1 Proposed Modified PSO for FCTP

The basic PSO (Algorithm 1) has modified by incorporating three modules which are explained as follows:

In first module (Algorithm 2), the particles are initialized randomly and the Initial Basic Feasible Solution (IBFS) of the FCTP is obtained with a reduced objective function. This IBFS is a non-negative integer solution fulfilling the supply and demand constraints. This solution is then iterated with some velocity (Eq. 2) and position

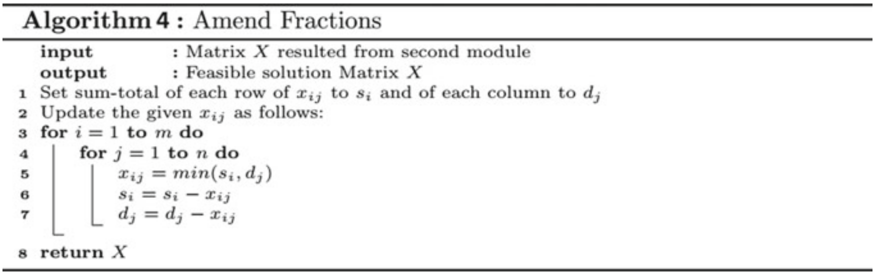
(Eq. 3) toward the optima. During intermediate steps, the solution becomes infeasible due to the presence of negative and/or fractional values in variables, which are to be repaired by utilizing the second and third modules as given below.

```
Algorithm 2: Initialization
input   : $TP$, $Supply$, $Demand$
output  : Initialized Matrix $X$
Take dimensions of $X$ as $[m, n] = size(TP)$
Assign $a = Supply$ & $b = Demand$
Set $RX = rand(m, n)$, a matrix of random numbers
Take a null matrix $X$ by using $X = zeros(m, n)$
Set $maxitr = m * n$
for $itrs = 1$ to $maxitr$ do
    Take maximum element $maxr$ of $RX$
    Locate $maxr$ in terms of $(i, j)$ in matrix $RX$
    Update $x_{ij}$ as $x_{ij} = min(a_i, b_j)$
    Update supply and demand as
    $a_i = a_i - x_{ij}$
    $b_j = b_j - x_{ij}$
    $maxr = 0$
return $X$
```

In second module (Algorithm 3), the infeasible solution containing negative values is repaired by applying the idea of Huang [26]. This procedure starts by selecting the most negative values of the first column of solution matrix and compensating that equivalent amount to some selected rows by executing Steps 8 and 9 of Algorithm 3. This repaired solution still has fractional values which are repaired using third module.

```
Algorithm 3: Amend Negatives
   input   : Initialized Matrix $X$
   output  : Negatives-Free Matrix $X$
1  do
2      if $x_{ij} < 0$ then
3          Denote the most negative element of column $j$ as $x_{hj}$.
4          Set $x_{hj}$, $x_{ij}$ as
5          $x_0 = x_{ij}$
6          $x_{hj} = x_{hj} - |x_0|$
7          $x_{ij} = 0$
8          Change row $i$ into
               $x_{ig} = \begin{cases} x_{ig} & x_{ig} = 0 \\ x_{ig} - \frac{|x_0|}{u} & x_{ig} > 0 \end{cases}$
            ($u$ is the count when $x_{ig} > 0; g = 1, 2, \cdots, m$ is met)
9          Update row $h$ into
               $x_{hg} = \begin{cases} x_{hg} & x_{ig} = 0 \\ x_{hg} + \frac{|x_0|}{u} & x_{ig} > 0 \end{cases}$
10 while ($x_{ij} < 0;\ i \in \{1, 2, ..., m\}, j \in \{1, 2, ..., n\}$)
11 return $X$
```

In third module (Algorithm 4), the solution with fractional values is repaired. The procedure, firstly, divides the matrix into integer and fraction matrix. The fractions of the matrix are resolved by reallocating the sum total of each row and column to the selected cell as mentioned in Algorithm 4. Thus, the obtained solution is feasible non-negative integer solution.



Algorithm 4 : Amend Fractions

```
input   : Matrix $X$ resulted from second module
output  : Feasible solution Matrix $X$
Set sum-total of each row of $x_{ij}$ to $s_i$ and of each column to $d_j$
Update the given $x_{ij}$ as follows:
for $i = 1$ to $m$ do
    for $j = 1$ to $n$ do
        $x_{ij} = min(s_i, d_j)$
        $s_i = s_i - x_{ij}$
        $d_j = d_j - x_{ij}$
return $X$
```

4 Results and Discussion

In order to validate the proposed PSO algorithm, an experimental design containing five test problems of different sizes was developed. For these five test problems, the cost matrices (variable cost and fixed cost) as well as supply and demand matrices were generated by using random integer numbers generator randi() in MATLAB with range as given in Table 1. The problems were balanced by adjusting the demand/supply quantities accordingly and denoted as P1, P2, P3, P4, and P5 as given in Tables 2, 3, 4, 5 and 6, respectively.

The proposed method has been implemented using the parameters as given in Table 7 to solve the FCTPs without incorporating any reduction techniques, given in existing literature, to linearize the objective function. Instead an independent approach which starts with the variable cost and then adding the fixed cost has been explored.

The proposed method, during execution, divided into four phases (I–IV) to accommodate both objectives (transportation cost as well as time). For instance, on attaining

Table 1 Data range for test problems

Variable	Range	Description
c_{ij}	$1 \leq c_{ij} \leq 100$	Variable cost
f_{ij}	$20 \leq f_{ij} \leq 100$	Fixed cost
t_{ij}	$05 \leq t_{ij} \leq 50$	Time matrix
S_i	$40 \leq S_i \leq 100$	Supply (availability)
D_j	$20 \leq D_j \leq 80$	Demand (requirement)
Where $i = 1, 2 \ldots, m$	$j = 1, 2, \ldots, n$	For m sources and n destinations

Table 2 Data of test problem P1

S_i	Variable cost (c_{ij})				Fixed cost (f_{ij})				Time (t_{ij})			
76	34	97	57	37	91	47	44	68	32	18	20	5
83	99	49	8	70	26	62	60	45	19	44	29	45
63	50	78	47	63	57	40	32	96	18	47	11	8
D_j	73	31	66	52								

Table 3 Data of test problem P2

S_i	Variable cost (c_{ij})					Fixed cost (f_{ij})					Time (t_{ij})				
49	6	65	55	23	21	46	57	55	45	58	29	35	32	34	34
87	25	82	57	46	71	95	91	60	59	92	22	31	43	20	20
97	34	46	40	68	16	31	90	39	56	45	16	13	13	32	32
D_j	40	54	43	42	54										

Table 4 Data of test problem P3

S_i	Variable cost (c_{ij})					Fixed cost (f_{ij})					Time (t_{ij})				
86	45	30	85	61	44	58	33	69	61	72	25	6	42	13	8
50	98	62	85	39	78	63	27	37	21	64	10	43	17	42	34
99	31	61	100	69	1	68	57	99	46	100	46	30	29	10	25
75	90	83	29	43	57	88	94	56	22	47	48	28	22	27	29
D_j	58	75	31	77	69										

Table 5 Data of test problem P4

S_i	Variable cost (c_{ij})						Fixed cost (f_{ij})						Time (t_{ij})					
50	77	22	69	8	93	41	45	88	89	64	42	62	14	36	29	12	31	19
95	19	31	33	16	90	6	90	56	75	32	80	86	46	48	23	29	10	22
98	53	13	40	69	74	61	57	81	24	72	98	82	49	11	34	29	43	10
87	36	80	7	63	31	10	77	92	23	75	52	52	42	6	10	46	30	26
D_j	54	29	67	70	36	74												

Table 6 Data of test problem P5

S_i	Variable cost (c_{ij})						Fixed cost (f_{ij})						Time (t_{ij})					
76	73	56	64	50	75	62	55	30	32	21	52	64	21	24	15	29	22	15
78	89	96	46	97	82	60	47	45	37	23	90	73	32	27	15	43	22	17
40	95	87	78	20	3	40	31	61	91	32	34	20	39	44	42	12	32	44
42	51	32	71	33	72	86	24	54	78	74	31	21	31	41	9	17	24	41
51	58	26	49	59	26	59	54	37	100	36	84	20	40	14	38	33	28	28
D_j	57	53	34	59	36	48												

Table 7 Parameters of PSO

Parameter	Value	Description
c_1	2	Acceleration coefficient (cognitive)
c_2	2	Acceleration coefficient (social)
r_1	Rand()	Random variable
r_2	$1 - r_1$	Random variable
ω	0.9–0.4	Time decreasing
Popsize maxiter	50	Population or swarm size
	1000	Maximum ıterations

the global best (Gbest) solution after phase-I, the corresponding cell of time matrix (t_{ij}) is marked as **x**, and the proposed algorithm enters into next phase. In this manner, after every phase, the corresponding cells of time matrix are crossed out (×) with respect to each Gbest solution. The attained optimal solutions of P1 are summarized in Table 8.

Similarly, the attained results for other test problems are summarized in Tables 9, 10 and 11 and Table 12. It has been observed from these tables that the transportation cost (Gbest) is increasing with each phase whereas the time is decreasing. Thus, it yields the Pareto pair (cost, time) for each test problem P1–P5 which is given in Table 13.

Table 8 Solution of test problem P1

Phase	I				II				III				IV			
	24 0 49	0 17 14	0 66 0	52 0 0	27 0 46	0 31 x	0 52 14	49 0 3	10 0 63	31 x x	0 66 0	52 x 0	x 73 0	31 x x	0 10 56	45 x 7
Cost	8021				8279				9272				15381			
	32 19 18	18 44 x	20 29 11	5 45 8	32 19 18	18 X X	20 29 11	5 X 8	X 19 18	18 X X	20 29 11	5 X 8	X 19 18	18 X X	20 X 11	5 X 8
Time	45				32				29				20			

Table 9 Solution of test problem P2

Phase	I					II					III					IV				
	0 40 0	0 0 54	0 43 0	38 4 0	11 0 43	24 16 0	0 0 54	10 29 4	X 42 0	15 0 39	22 18 0	0 27 27	0 X 43	X 42 0	27 0 27	6 34 0	X 11 43	43 X 0	X 42 0	X 0 54
Cost	8364					8809					9212					9408				
	29 22 16	35 31 13	32 43 13	X 24 5	34 20 32	29 22 16	35 31 13	32 X 13	X 24 5	34 20 32	29 22 16	X 31 13	32 X 13	X 24 5	X 20 32	29 22 16	X 31 13	X X 13	X 24 5	X 20 X
Time	43					35					32					31				

Table 10 Solution of test problem P3

Phase	I					II					III					IV				
	28	58	0	0	0	58	28	0	0	0	58	28	0	0	0	8	75	X	3	0
	0	17	0	33	74	0	17	0	33	0	0	X	0	50	0	50	X	0	X	0
	30	0	0	0	74	X	30	0	0	69	X	30	0	0	69	X	0	0	30	69
	0	0	31	44	0	X	0	31	44	0	X	17	31	247	0	X	0	31	44	0
Cost	9516					10855					11211					13062				
	25	6	0	0	0	25	6	42	13	8	25	6	X	13	8	25	6	X	13	8
	10	43	0	33	0	10	X	17	42	34	10	X	17	X	34	10	X	17	X	X
	X	30	0	0	69	X	30	29	10	25	X	30	29	10	25	X	X	X	10	25
	X	28	31	44	0	X	28	22	27	29	X	28	22	27	29	X	X	22	X	x
Time	43					42					34					25				

Also the plots of attained Pareto pairs for each test problem are given in Fig. 1. The transportation cost (Gbest) is depicted on horizontal axis and time on vertical axis wherein the coordinates for each pair are in blue dots. It can be observed from these plots that with increase of transportation cost, the time dropped down. And this pattern elaborates the quality of solution obtained by the proposed PSO.

5 Conclusion

In this paper, a modified PSO algorithm is insinuated to solve the multi-objective Fixed Charge Transportation Problem (FCTP). The intended technique even worked well without reducing the objective function to linear form and the procedure was carried out on the variable cost only and the fixed cost was considered at each iteration to improve the solution subsequently. Moreover, the advantages of PSO were also observed during the implementation, since it does not require the two important optimality test conditions as in the case of exact methods. The number of basic cells in case of exact method needs to be $(m + n - 1)$ and the basic cells need to be in independent positions. It is also established that the PSO works well on the variable cost matrix itself rather than considering the reduced form of the objective function as mentioned earlier. This establishes the efficient exploration of the solution space by the PSO. The efficiency of the proposed method is recognized by the trace of Pareto line iterating toward the Pareto pairs, thereby validating the capability of proposed method and its scope, as an alternative, with other evolutionary techniques.

Table 11 Solution of test problem P4

Phase	I						II						III						IV					
	0	0	0	50	0	0	0	0	0	50	0	0	0	0	0	50	0	0	0	0	0	9	0	41
	4	0	0	20	0	74	54	0	0	20	0	24	X	X	4	20	0	74	X	X	0	61	34	3
	50	29	67	70	36	74	X	29	66	0	0	0	X	29	44	0	22	0	X	29	66	0	X	0
	0	0	51	0	36	0	0	0	1	0	36	50	54	0	19	X	14	0	54	0	1	X	2	30
Cost	6889						7034						8184						11770					
	14	36	29	12	31	19	14	36	29	12	31	19	14	36	29	12	31	19	14	36	29	12	31	19
	46	48	23	29	10	22	X	X	23	29	10	22	X	X	33	29	10	22	X	X	23	29	10	22
	X	11	34	29	43	10	X	11	34	29	43	10	X	11	34	29	X	10	X	11	34	29	X	10
	42	6	10	46	30	46	42	6	10	X	30	26	42	6	10	X	30	26	X	6	10	X	30	26
Time	48						43						42						36					

Table 12 Solution of test problem P5

Phase	I						II						III					
	15	2	0	59	0	0	57	8	0	8	0	3	0	53	0	19	0	4
	0	0	34	0	0	44	0	0	34	0	0	44	15	0	34	X	0	29
	0	0	0	0	36	4	0	X	0	16	24	X	0	X	X	40	0	X
	42	0	0	0	0	0	0	7	0	35	0	0	42	X	0	0	0	X
	0	51	0	0	0	0	0	38	0	0	12	1	0	0	0	0	36	15
Cost	12468						13144						14000					
	21	24	15	29	22	15	21	24	15	29	22	15	21	24	15	29	22	15
	32	27	15	43	22	17	32	27	15	X	22	17	X	27	15	X	22	17
	39	X	42	12	32	X	39	X	X	12	32	X	X	X	X	12	X	X
	31	41	9	17	24	41	31	X	9	17	24	41	31	X	9	17	24	X
	40	14	38	33	28	28	40	14	38	33	28	28	X	14	X	X	28	28
Time	43						40						31					

Table 13 Solutions obtained by using proposed PSO

Test problems	P1		P2		P3		P4		P5	
Phases	Cost	Time	Cost	Time	Cost	Time	Cost	Time	Cost	Time
Phase-I	8021	45	8364	43	9516	43	6889	48	12468	43
Phase-II	8279	32	8809	35	10855	42	7034	43	13144	40
Phase-III	9272	29	9212	32	11211	34	8184	42	14000	31
Phase-IV	15381	20	9408	31	13062	25	11770	36	–	–

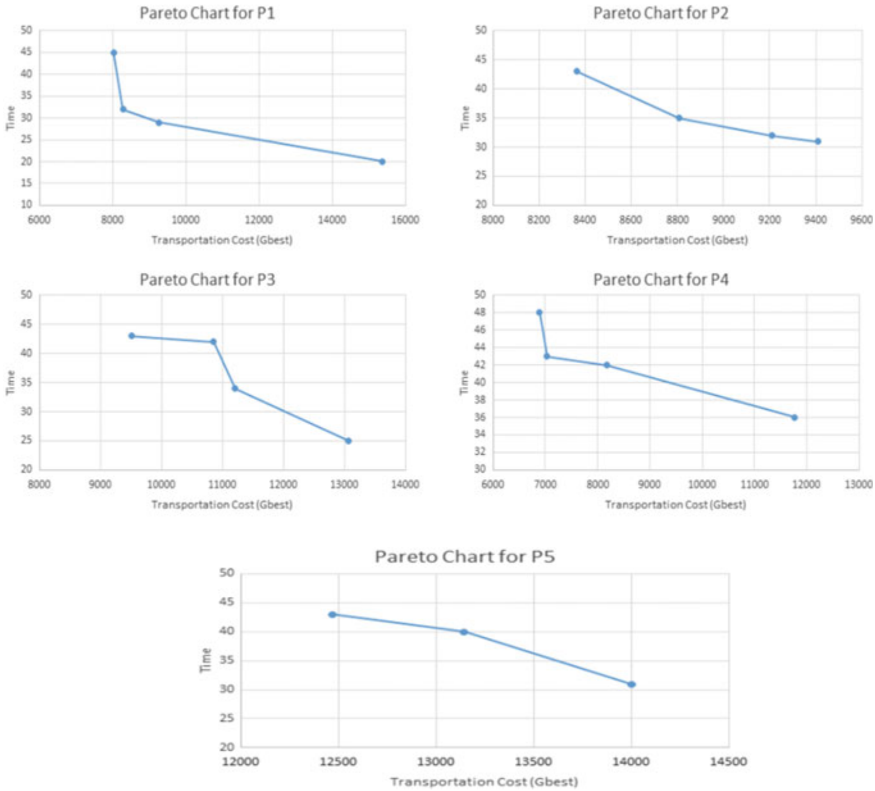


Fig.1 Pareto charts for P1–P5 by proposed PSO

References

1. Hitchcock F (1941) The distribution of a product from several sources to numerous locations. J Math Phys 20(224):230
2. Koopmans T (1947) Optimum utilization of the transport system. In: Proceedings of international statistical conferences: econometric, pp 136–145
3. Hirsch WM, Dantzig GB (1968) The fixed charge problem. Naval Res Logist Quart 15(3):413–424
4. Murty KG (1968) Solving the fixed charge problem by ranking the extreme points. Oper Res 16(2):268–279
5. Sadagopan S, Ravindran A (1982) A vertex ranking algorithm for the fixed-charge transportation problem. J Optim Theory Appl 37(2):221–230
6. Gray P (1971) Exact solution of the fixed-charge transportation problem. Oper Res 19(6):1529–1538
7. Steinberg DI (1970) The fixed charge problem. Naval Res Logist Quart 17(2):217–235
8. Barr RS, Glover F, Klingman D (1981) A new optimization method for large scale fixed charge transportation problems. Oper Res 29(3):448–463
9. Cabot AV, Erenguc SS (1984) Some branch-and-bound procedures for fixed-cost transportation problems. Naval Res Logist Quart 31(1):145–154

10. Palekar US, Karwan MH, Zionts S (1990) A branch-and-bound method for the fixed charge transportation problem. Manag Sci 36(9):1092–1105
11. Robers P, Cooper L (1976) A study of the fixed charge transportation problem. Comput Math Appl 2(2):125–135
12. Balinski ML (1961) Fixed-cost transportation problems. Naval Res Logist Quart 8(1):41–54
13. Adlakha V, Kowalski K (2003) A simple heuristic for solving small fixed-charge transportation problems. Omega 31(3):205–211
14. Adlakha V, Kowalski K, Vemuganti R (2006) Heuristic algorithms for the fixed-charge transportation problem. Opsearch 43(2):132–151
15. Adlakha V, Kowalski K, Wang S, Lev B, Shen W (2014) On approximation of the fixed charge transportation problem. Omega 43:64–70
16. Gottlieb J, Paulmann L (1998) Genetic algorithms for the fixed charge transportation problem. In: 1998 IEEE international conference on evolutionary computation proceedings, 1998. IEEE world congress on computational intelligence. IEEE, pp 330–335
17. Sun M, Aronson JE, McKeown PG, Drinka D (1998) A tabu search heuristic procedure for the fixed charge transportation problem. Eur J Oper Res 106(2–3):441–456
18. Raj KAAD, Rajendran C (2011) A hybrid genetic algorithm for solving single-stage fixed-charge transportation problems. Technol Oper Manag 2(1):1–15
19. Lotfi M, Tavakkoli-Moghaddam R (2013) A genetic algorithm using priority-based encoding with new operators for fixed charge transportation problems. Appl Soft Comput 13(5):2711–2726
20. El-Sherbiny MM, Alhamali RM (2013) A hybrid particle swarm algorithm with artificial immune learning for solving the fixed charge transportation problem. Comput Ind Eng 64(2):610–620
21. Midya S, Roy SK (2014) Solving single-sink, fixed-charge, multi-objective, multi-index stochastic transportation problem. Am J Math Manag Sci 33(4):300–314
22. Roy SK, Midya S, Vincent FY (2018) Multi-objective fixed-charge transportation problem with random rough variables. Int J Uncert Fuzziness Knowl-Based Syst 26(6):971–996
23. Roy SK, Midya S (2019) Multi-objective fixed-charge solid transportation problem with product blending under intuitionistic fuzzy environment. Appl Intell 1–15
24. Majumder S, Kundu P, Kar S, Pal T (2019) Uncertain multi-objective multi-item fixed charge solid transportation problem with budget constraint. Soft Comput 23(10):3279–3301
25. Eberhart R, Kennedy J (1995) A new optimizer using particle swarm theory. In: Proceedings of the sixth international symposium on micro machine and human science. Nagoya, Japan, pp 39–43
26. Huang H, Zhifang H (2009) Particle swarm optimization algorithm for transportation problems. Particle Swarm Optimization, Intech, Shanghai, pp 275–290

Binomial Logistic Regression Resource Optimized Routing in MANET

M. Ilango, A. V. Senthil Kumar, and Amit Dutta

Abstract MANETs consist of the node that moves continuously in a random direction. In MANET architecture, devices can move in any direction. They have recently confined imperativeness, enlisting power and memory. The hubs of the system are versatile and the topology changes quickly. In binomial logistic regression resource optimized routing in MANET (BLR-OR), disappointment methodology includes visit hub distinguished a high directing overhead and stretched out start to finish postponement and power. The proposed method used logistic regression technique to detect optimal network path. It can be obtained depending on neighbor node using coefficient value. Regression coefficient is directed by thinking about the vitality, latency, and transmission capacity of versatile hub by utilizing the coefficient incentive to choose the mobile node with high energy, latency, and bandwidth utilization for communicating packets in MANET and enhance lifetime of the network (N-L), packet delivery ratio(P-D-R) in successful way. The presentation of BLR-OR regression procedure is estimated regarding energy consumption (E-C), end-to-end delay (E-E-D).

Keywords MANET · Route · Data packets · Logistic regression · Energy · Latency · Bandwidth utilization

M. Ilango (✉)
Hindusthan College of Arts and Science, Coimbatore, India
e-mail: ilango_kn@yahoo.co.in

A. V. Senthil Kumar
Department of MCA, Hindusthan College of Arts and Science, Coimbatore, India

A. Dutta
All India Council for Technical Education, New Delhi, India

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_37

1 Introduction

Mobile Ad Hoc Network (MANET) is a gathering of portable hubs framing a system without a fixed framework. MANET consists of a routing system condition over a connection layer ad hoc network which comprises a distributed, self-shaping, self-recuperating system. In MANETs, every versatile hub contains constrained preparing rate and force, battery, stockpiling, and correspondence abilities. Regression analysis is the way toward evaluating the relationship among hub and anticipating where a specific hub has a place class. Logistic regression is a non-linear model to find an optimized route in the proposed method. Logistic regression utilizing a calculated capacity which produces "S"-shaped curve in the range between "0" and "1", making it conceivable to get a non-linear limit.

Logistic regression is a kind of regression to breakdown the connection among dependent and independent variables and then computes the node occurrence of an event. It is used to measure the binary variable; the dependent variable is binary value 1 (true) 0 (false). There are mainly two logistic regression models, namely, binomial and multinomial logistic regression. In this, author [1] fully describes link failure prediction algorithm to combine dynamic source routing protocol, each node for finding link failure node depends on received signal strength using linear regression.

Linear regression had continuous node, because of that identification of an optimal path is difficult [1]. In stepwise regression [2], routing method is applied for finding more than one independent path, and it fails to concentrate on a node by node in a network. To overcome this problem, BLR-OR is used to easily identify group or category of the dependent node.

The objective of the logistic method is defined as follows:

1. To identify the efficiency node in a computational network.
2. To choose the best categorical node from the network path and remove uncovered nodes using logistic regression.

2 Related Works

Ad Hoc On-demand Distance Vector (AODV) is the routing overhead [3] for analyzing the dependency on the connection failure likelihood in MANET. In this approach, the crash likelihood which is brought about by covered up hub issue and the effects on the connection failure probability were analyzed. A numerical examination of the hypothetical directing overhead of AODV protocol was presented according to the link failure probability. However, there is an increased maximum routing overhead but only two scenarios such as chain and square shape situations by making all hubs stationary are considered. Connection-security-based multicast routing component [4] was proposed in MANET. The stable routes were identified dependent on the determination of stable sending hubs that have high stability and connection network. The link stability was estimated dependent on the parameters, for example, got power, separation between neighboring hubs, and connection quality. However, the scalability and flexibility were not improved.

Link stability estimation mechanism [5] was proposed for performing multicast routing in MANET. The major contribution of the aforementioned approach was to provide a constant connection service in a cost-efficient manner. A novel link stability estimation framework was proposed subject to the examining of got signal quality data. In this approach, the proposed model was integrated with multicast AODV which is used for finding progressively accessible stable paths and adjusting to organize topology changes. However, the estimation of link expiration time was required for further improving the efficiency of finding more stable routes. Multi-objective OLSR for proactive routing [6] was proposed in mobile ad hoc network with energy, delay, and connection time prediction. In this approach, three objectives are considered to reduce the end-to-end delay, network lifetime maximization, and boosting P-D-R. As a result, three routing measurements were grown such as mean lining delay on every hub, energy cost on each middle point, and connection dependability on each connection. In this approach, queuing delay and E-C were anticipated by utilizing twofold exponential smoothing and the leftover connection lifetime was anticipated by utilizing a heuristic of the spreads of the association lifetimes in MANET. However, P-D-R was less and the complexity of the method was high.

A reliable routing mechanism to predict the linkl availability [7] was proposed for MANET. In this approach, a novel LBRP was proposed for mobile ad hoc network. A scientific articulation of connection accessibility for MANET was determined by utilizing probabilistic and factual registering. This approach was proposed dependent on arbitrary walk versatility model which is a ceaseless time stochastic procedure that portrays the irregular development of hubs in a two-dimensional space. However, the complexity of the approach was high. Link stability estimation [8] was proposed dependent on the connection availability changes in MANET. In this approach, a novel scheme was proposed for estimating the link stability according to connection availability changes which might be performed on the system layer. A variable measured sampling window was embraced and the strategy was proposed for evaluating the connection progress rates. Then, the routing method was proposed for adjusting its working mode dependent on the assessed connection strength. In any case, the variance of the inspecting window length was not diminished and the unpredictability of the strategy was high.

Xink et al. exhibited the ECDC [9] protocol. ECDC handles the recognizing territory/point incorporation for figuring of abundance center point. ECDC utilizes randomized turn of bunch head inside the groups for each round. In ECDC, the gathering heads are picked subject to self-assertive probabilities without contemplating the residual energy. Because of this, ECDC does not keep up energy-adjusted inside the system. RBSP [10] is an area uninformed, residual-energy-based conveyed convention. The discretionary hub booking is utilized in RBSP.

Logistic regression is a kind regression examination utilized to figure the result of ward variable dependent on at least anyone of the free factors, where the estimation of indicator variable is somewhere in the range of 0 and 1. It can be classified either as binomial or multinomial. It is generally utilized in epidemiologic, the credit appraisal for account, etc. [11]. In probability-based methods [12], every hub is allocated likelihood for retransmission relying on the topology of the system. In area-based methods, a typical transmission extend is expected and a hub will rebroadcast if just adequate new zone can be secured with the retransmission. The reliability of the network [4] is characterized as a system that works effectively (or un-interruptedly) over a given time allotment under known ecological conditions. It is additionally expected that the system works persistently from the beginning or restart of the system [13]. Zombie state node to get CABR calculation [14] to transmit packets starting with one hub then onto the next. It fully depends on regression model to reduce E-C which increases the N-L. AODV method [15] is to predict the delay dependency based on the distance between two nodes. This algorithm deals with energy, delay to optimize the path from starting place to end place.

Logistic regression on the basis of reliability analysis performed on MANET is exhibited [16] which fixed most extreme speed and fluctuating respite time of portable hub. MANETS' reliability for running on DSDV, DSR, and AODV routing protocol was figured [17] which is proposed for identifying on sending packets, transmitting speed, multistep ahead prediction, and destroying the fake packets. The logistic regression [18] helps to compute AODV, DSR, DSDV routing protocol regression method to solve reliability on MANET. This method [19] is trusted included over ad hoc based on the demand distance vector routing protocol and optimized link state routing protocol available in MANET. The performance evaluations are computed by these parameters like energy, data loss, E-E-D, and network traffic. Multipath [20], deterministic [21], non-linear [22], and stepwise regression [2] enhance the power, bandwidth, P-D-R, and delay time despite that the selection of path is automatically taken from one hop to another.

3 Proposed Method

The residual energy of versatile hub s_i is assessed by utilizing following scientific articulation:

$$S_i = \text{Starting Energy}_i - (b_i \times T_B) \tag{1}$$

While observing the condition (1) $\mathrm{b_i}$ means the transmitted number of bits while $\mathrm{T_B}$ represents the transmission power for every byte. Here, starting energy shows the beginning residual energy present in versatile hub ahead the route ID process procedure by utilizing condition (1), residual energy of a portable hub is resolved for course confirmation.

$$\mathrm{BW_1} = \mathrm{BW_{raw}} - \mathrm{BW_2} \tag{2}$$

From condition (2,) $\mathrm{BW}_{\mathrm{raw}}$ indicates the crude channel transfer speed and BW_2 alludes a used transmission limit with regard to transmitting the data packets. By utilizing the above condition, the mobile compact with higher data transmission accessibility is chosen for ideal way determination. The defer time of mobile hub is estimated by utilizing underneath scientific equation

$$DT = Q^i \times T_P^i + T_N^i \times H_N \tag{3}$$

From the condition (3), Q^i speak to the line length of portable hub "i" and "T_P^i" indicates the nearby handling time of information in versatile hub "i". Here "T_N^i" refers transmission time of nodes to be closed to hubs and H_N is the quantity of bounces between the source and objective hub in MANET.

The BLR-OR used logistic regression method for performing resource optimized routing enhanced directing dependent on the above parameters to be specific residual energy, bandwidth availability, and delay time. By estimating the above parameters, best portable hub in system is chosen for packets transmission in MANET. The BLR-OR processes the logistic regression coefficient value for every portable hub in network. So as to choose the best hubs for routing the data packet, mathematical formula is given below:

$$\text{Logistic Regression Coefficient Value(LRCV)} = \text{HRE} + \text{LBW} + \text{LDT} \tag{4}$$

From the condition (4), LRCV is resolved dependent on the hub residual energy, data transfer capacity accessibility, and delay time between the hubs. A logistic capacity is characterized by the recipe given in condition

$$x = \frac{e^y}{1 + e^y} \tag{5}$$

The proposed capacity in condition (5) can be communicated from in conditions (6) and (7).

$$x = \frac{\left(e^{\alpha + \beta_{yi}}\right)}{\left(1 + e^{\alpha + \beta_{yi}}\right)} \tag{6}$$

$$x = \frac{1}{1 + e^{-(\alpha + \beta_{yi}}} \tag{7}$$

where α, β are regression coefficients which decide the slant and logistic intercepts for which y_i is an indicator.

As appeared in Fig. 1, a MANET's structure includes various portable hubs, for example, [MN_1], [NN_2], [MN_3] … [MN_12]. Here [MN_1] needs to transmit packet to versatile hub [MN_12] though [MN_1] is a sourcing hub and [MN_12] is an objective hub and remaining residual hubs are called as middle of the road hubs in network.

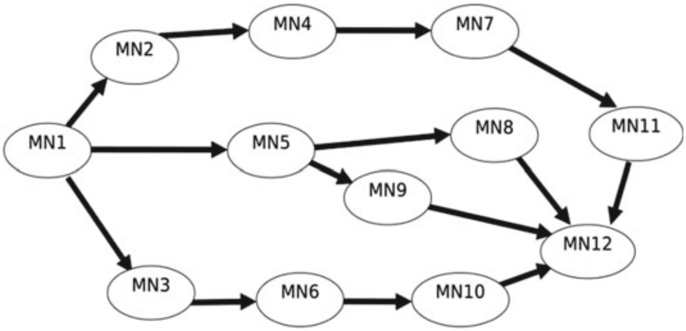
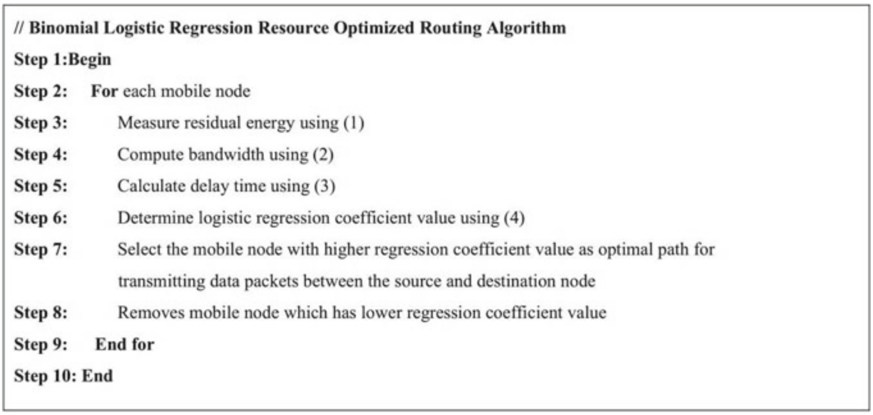


Fig. 1 Example of MANET structure

More number of route ways are accessible to transmit the media data content from MN$_1$ to MN$_{12}$ such as MN$_1$-MN$_2$-MN$_4$-MN$_7$-MN$_{11}$-MN$_{12}$,MN$_1$-MN$_5$-MN$_8$-MN$_{12}$, MN$_1$-MN$_5$-MN$_9$-MN$_{12}$, MN$_1$-MN$_3$-MN$_6$-MN$_{10}$-MN$_{12}$, etc. Be that as it may, we cannot realize which course way is propelled as the resource route path to transmit the information in a productive way from basis node MN$_1$ to objective MN$_{12}$. The algorithmic process of BLR-OR is shown below:



```
// Binomial Logistic Regression Resource Optimized Routing Algorithm
Step 1:Begin
Step 2:    For each mobile node
Step 3:        Measure residual energy using (1)
Step 4:        Compute bandwidth using (2)
Step 5:        Calculate delay time using (3)
Step 6:        Determine logistic regression coefficient value using (4)
Step 7:        Select the mobile node with higher regression coefficient value as optimal path for
               transmitting data packets between the source and destination node
Step 8:        Removes mobile node which has lower regression coefficient value
Step 9:    End for
Step 10: End
```

By utilizing the above algorithmic procedure BLR-OR picks the resource-enhanced course way for information broadcast. This improves the P-D-R with least E-C and improves the N-L in MANET.

4 Experimental Results

The BLR-OR procedure is actualized by NS-2 test system with the scope of 1200 * 1200 m size. The number of portable hubs picked for performing reenactment is 500. The consequences of BLR-OR method are contrasted beside and presented strategy Shortest path Link-Based Resource Optimized Routing (SPROR) [23] to quantify the viability of our work.

4.1 Energy Consumption

In BLR-OR method, the E-C is estimated utilizing the energy devoured by a solitary portable hub concerning all the versatile hubs in MANETs. The power exploitation rate is estimated in terms of Joules (J) and detailed as

$$\text{Energy Consumption} = \frac{\text{Energy}_{\text{DP}}}{\text{Total}_{\text{DP}}} \tag{8}$$

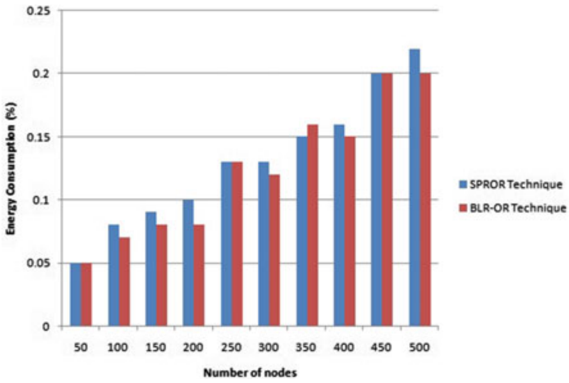
From condition (8), the E-C of routing process is acquired “$\text{Energy}_{\text{DP}}$” speaks to the proportion of power inspired for single packet and total power devoured for all the packets “Total_{DP}” in the networks.

Table 1 and Fig. 2 show the measure and effect of the E-C for packets communication dependent on various portable hubs in the scope of 50–500. From Table 1 to uncovered in Fig. 2, the E-C of proposed BLR-OR strategy is lower when contrasted and existing. Likewise, while expanding the number of hubs, the energy is also expanded utilizing both the techniques.

Table 1 Energy consumption

Number of mobile nodes	Energy consumption (J)	
	SPROR technique	BLR-OR technique
50	0.05	0.05
100	0.08	0.07
150	0.09	0.08
200	0.10	0.08
250	0.13	0.13
300	0.13	0.12
350	0.15	0.16
400	0.16	0.15
450	0.20	0.20
500	0.22	0.20

Fig. 2 Energy consumption



4.2 Average End-to-End Delay

In BLR-OR method, the average E-E-D quantifies the time expended for a packet to be transmitted over a system from basis to objective. The average E-E-D is estimated as far as milliseconds (ms) and numerically detailed as

$$\text{Average End to End Delay} = \text{Receiving data packets} - \text{sending data packets} \tag{9}$$

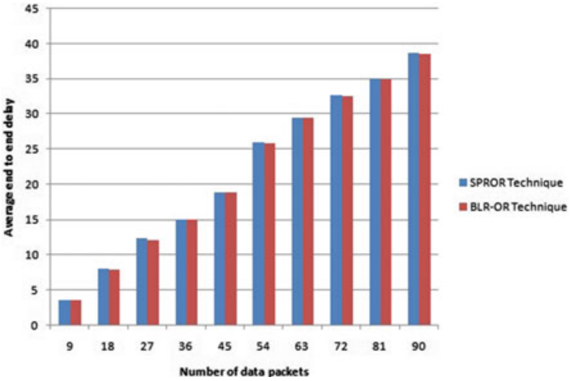
From condition (9), the average E-E-D is acquired. While the E-E-D is lesser, the strategy is supposed to be progressively proficient.

The similar examination of the average E-E-D dependent on various packets ranges from 9 to 90 is shown in Table 2 just as in Fig. 3. The proposed technique provides lower average E-E-D when contrasted and existing strategy.

Table 2 End-to-end delay

Number of packets	Average end-to-end delay (ms)	
	SPROR technique	BLR-OR technique
9	3.5	3.5
18	8.0	7.9
27	12.3	12.1
36	15.0	15.0
45	18.76	18.72
54	26.0	25.86
63	29.4	29.37
72	32.67	32.59
81	35.0	34.96
90	38.7	38.58

Fig. 3 End-to-end delay



4.3 Packet Delivery Ratio

In BLR-OR procedure, P-D-R is characterized as the proportion of a number of packets expected by the objective to the total number of packets sent. The P-D-R is quantified as far as a rate (%) and planned as follows:

$$\text{Packet Delivery Ratio} = \frac{\text{Number of data packets received}}{\text{Total number of data packets sent}} \times 100 \qquad (10)$$

From condition (10), the P-D-R is gotten while the P-D-R is higher, the strategy is said to be progressively proficient.

The consequence of P-D-R dependent on various packets in the scope of 90–900 is exposed in Table 3 and Fig. 4. Table 3 and Fig. 4 obviously express that the P-D-R utilizing proposed systems is higher when contrasted with existing technique.

Table 3 Packet delivery ratio

Number of packets	Packet delivery ratio (%)	
	SPROR technique	BLR-OR technique
9	90.15	90.18
18	90.99	91.08
27	91.26	91.31
36	92.45	92.49
45	93.29	93.34
54	94.94	95.01
63	96.38	96.42
72	97.95	97.97
81	98.40	98.47
90	98.46	98.46

Fig. 4 Packet delivery ratio

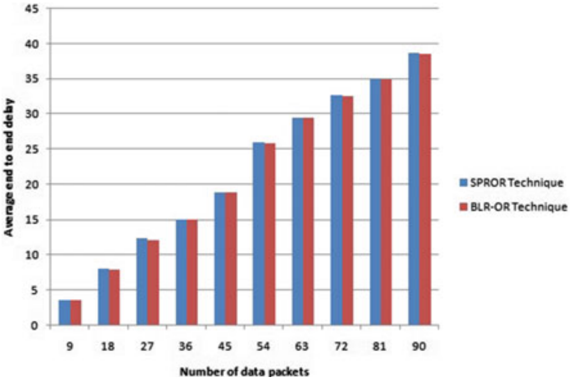


Table 4 Network lifetime

Number of nodes	Network lifetime (%)	
	SPROR technique	BLR-OR technique
50	90.39	90.43
100	91.23	91.28
150	92.80	92.84
200	93.32	93.36
250	95.26	95.33
300	96.05	96.17
350	97.50	97.54
400	97.56	97.62
450	98.32	98.35
500	99.48	99.51

4.4 Network Lifetime

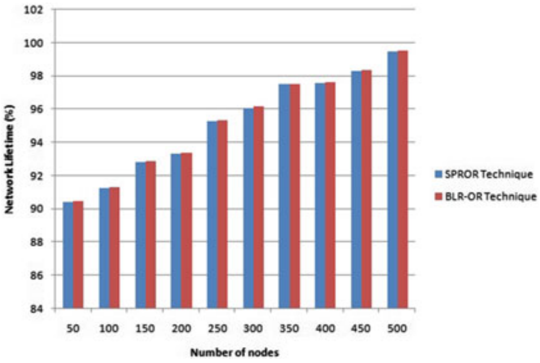
In BLR-OR procedure, the lifetime of a system is estimated by utilizing various portable hubs in MANETs. The N-L is estimated regarding rate (%) and defined as

$$\text{Network Lifetime} = \frac{\text{S}_{\text{addressed}}}{\text{Total}_{\text{S}}} \times 100 \tag{11}$$

From condition (11), the N-L is got utilizing the total number of mobile node “Total_{s}” in the system and steering addressed for the portable hub “$\text{S}_{\text{addressed}}$” in MANETs while the N-L is elevated, progressively productive the strategy is alleged to be.

The N-L effect is obtained dependent on a contradictory number of packets in the scope of 90–900 utilizing the technique is revealed in Table 4 and Fig. 5. From Table 4

Fig. 5 Network lifetime



to Fig. 5, plainly the N-L utilizing proposed procedure is higher when contrasted with existing strategy.

5 Conclusion

The proposed BLR-OR approach is utilized to locate the best routing path from source to goal. The BLR-OR regression method improves the MANET performance like energy, latency, and bandwidth. The source node can quickly establish a stable node from the basis node to all neighbor nodes. If neighboring three parameter values are satisfied then data packets are transmitted to that node. The simulation results shows in the aspect of identified best data using coefficient value of the node in a network. The performance of BLR-OR regression technique is estimated as far as E-C, E-E-D, P-D-R, and N-L contrasted and existing technique. The future work is proposed to explore and improve different protocols and reduce the resources of routing.

Reference

1. Rathiga P, Shathappan S (2017) Regrssion-based link failure prediction with fuzzy based hybrid black hole/grey hole attack detection technique. Int J Appl Eng Res 8:7459–7465
2. Ilango M, Senthil Kumar AV (2018) Stepwise regression based resource optimized routing in mobile ad hoc network. Int J Res Appl Sci Eng Technol (IJRASET) 6(II), 504–514
3. Zhang QJ, Wu MQ, Yan ZHEN, Shang CL (2010) Aodv routing overhead analysis based on link failure probability in MANET. J China Univ Posts Telecommun 17(5):109–115
4. Biradar R, Manvi S, Reddy M (2010) Link stability based multicast routing scheme in MANET. Comput Netw 54(7):1183–1196
5. Xia H, Xia S, Yu J, Jia Z, Sha EHM (2014) Applying link stability estimation mechanism to multicast routing in MANETs. J Syst Archit 60(5):467–480

6. Guo Z, Malakooti S, Sheikh S, Al-Najjar C, Malakooti B (2011) Multi-objective OLSR for proactive routing in MANET with delay, energy, and link lifetime predictions. Appl Math Model 35(3):1413–1426
7. Han Q, Bai Y, Gong L, Wu W (2011) Link availability prediction-based reliable routing for mobile ad hoc networks. IET Commun 5(16):2291–2300
8. Song Q, Ning Z, Wang S, Jamalipour A (2012) Link stability estimation based on link connectivity changes in mobile ad-hoc networks. J Netw Comput Appl 35(6):2051–2058
9. Xin G, Jiguo Y, Dongxiao Y, Guanghui W, Yuhua L (2014) Ecdc: an energy and coverage-aware distributed clustering protocol for wireless sensor networks. Comput Electr Eng (Elsevier) 40:384–398
10. Avinash M, Vijay R (2014) Random backoff sleep protocol for energy efficient coverage in wireless sensor networks. Smart Innov Syst Technol (Springer) 28(2):323–331
11. Huang F, Jiang Z, Zhang S, Gao S, Reliability evaluation of wireless sensor networks using logistic regression. In: 2010 international conference on communication and mobile computing
12. Ilango M, Senthil Kumar AV (2016) Probabilistic and link based energy efficient routing in MANET. Int J Comput Trends Technol (IJCTT) 38(1):38–45
13. Rodriguez G, Logit models for binary data. Chapter 3
14. More A (2016) Coverage aware battery regression curve node scheduling in wireless sensor node. Int J Comput Netw Commun Secur 4(7):218–226
15. Singh A, Singh T (2019) End to end delay using ADODV-artificial neural network (ANN) to improve performance of MANET. Int J Recent Technol Eng 8(1)
16. Singh MM, Mandal JK (2017) Logistic regression based reliability analysis for mobile ad hoc network with fixed maximum speed and varying pause times. J Sci Ind Res 76:18–84
17. Silva AAA, Ponder E, Zhou F, Kofuji S (2014) Grey model and polynomial regression for identifying malicious node in MANET. In: Globecom-ad hoc and sensor symposium 2014, pp 162–168
18. Sing MM, Baruah M, Mandal JK, Reliability computation of mobile ad hoc network using logistic regression. In: IEEE Conference on proceedings of the 2014
19. Venkatraman R, Ramarao MP (2012) Regression based trust model for mobile ad hoc network. IET Inf Secur 6(3):131–140
20. Ilango M, Senthil Kumar AV, Multipath strategies and link based resource optimized routing in mobile ad hoc networks. Int J Recent Technol Eng (IJRTE) 8(5)
21. Ilango M, Senthil Kumar AV (2017) Deterministic multicast link based energy optimized routing in MANET. In: 2017 IEEE international conference on electrical, computer and communication technologies (ICECCT), vol 3, pp 1102–1110
22. Ilango M, Senthil Kumar AV (2019) Non linear differential optimization for quality aware resource efficient routing in mobile ad hoc networks. Int J Eng Adv Technol (IJEAT) 9(1)
23. Ilango M, Senthil Kumar AV, Resource optimized routing using shortestpath technique in MANET. (Under review)

A Lightweight Approach for Policy-Based Messaging

P. P. Abdul Haleem

Abstract The popularity of Resource-Constrained Networks (RCNs) is increasing rapidly in value, volume, velocity, variety, and veracity. The limited battery power and verbosity of messages are the major limiting factors in the constrained wireless mobile environment. Hence, devising a standard that is adaptive to the limitations of the present-day wireless mobile environment is crucial for the emerging generations of constrained wireless mobile devices; it is also expected to accelerate the penetration of wireless mobile technologies to the underprivileged class of users. The eXtensible Markup Language (XML) is the natural standard for messaging across heterogeneous types of mobile devices. But it is not suitable in the wireless mobile environment, due to the increased storage and processing requirements needed. This work presents a lightweight Policy-Based Messaging (PBM) mechanism for trusted transmission of information that can be used especially in Resource-Constrained Networks (RCNs). The work is based on a data format which is less verbose. The data format is derived from the YAML Aint Markup Language (YAML), a lightweight data serialization language. Proposals include measures to define policy assertions and a two-level mechanism to make sure trusted transmission of information. The performance analysis indicates an advantage over the existing methods.

Keywords Mobile computing · Verbosity reduction · XML · Rewriting attacks · Resource-constrained networks · PBM

1 Introduction

Widespread usage of wireless mobile devices resulted in the popularity of pervasive computing. In such an environment, regulating and ensuring the security and trust of the transmitted data and users is an important task. A lot of existing research trying to ensure trusted transmission and restricted access in the pervasive environment is reported.

P. P. Abdul Haleem (✉)
Department of Computer Science, Farook College (Autonomous), Kozhikode, Kerala, India
e-mail: abdulhaleem@farookcollege.ac.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_38

The relevance of the proposed work stems from the fact that the size of the data format to be transmitted has a direct impact on the energy being consumed by the device and network. The feasibility of the idea can be further inferred from a number of facts: (i) transmission and reception of messages takes the biggest share of power in a wireless mobile device [1], (ii) security protocols for the constrained wireless environment tend to prefer reduced size of data and less number of packets to be transmitted/received [2], (iii) in majority of applications on wireless mobile devices the amount of information transmitted from computer to user is much larger than the amount of information transmitted from user to computer, and (iv) considering the limited input and screen of the devices, a format that requires fewer bytes for data representation, but expressive as XML, is well suited. Hence, a method for reducing verbosity has been proposed in [3]. This work is further enhanced by introducing security mechanisms in [4] and energy-conserving proposals in [5]. A schema definition called YASchema is also proposed to reduce the verbosity of the format [3, 6].

Obviously, the messages are to be protected against any kind of tampering during the transmission. PBM infrastructure, which is developed to promote a distributed and secure mechanism for messaging, can be used to strengthen any messaging scheme [7, 8].

This work addresses the issue of trusted transmission in pervasive environment. Mechanism to evolve as a lightweight policy-based messaging environment is discussed. The contributions in this work are (i) extension of the application of the schema definition discussed in [9] to encompass the PBM, and (ii) introduction of a two-level lightweight PBM mechanism making use of the YASchema/YAccount combination.

The rest of the paper is organized as follows: In Sect. 2, the state of the art of discussing. Section 3 discusses the proposed scheme. In Sect. 4, performance evaluation of the proposed scheme is explained. Section 5 concludes the paper.

2 Literature Survey

The proposed work is centered around a data formatting scheme that is less verbose than XML. In [3], a method with reduced verbosity based on YAML (TYAML) is proposed. In [4] this work is enhanced with security enhancements, and in [5, 6] the work is further refined to achieve energy-conserving security specifications. The literature survey is focused on: (i) specifications to ensure trusted exchange of information, and (ii) measures for a lightweight mechanism for PBM. The survey converges, especially to the viability of a method suitable to the constrained wireless mobile environment—to conserve energy and memory requirements.

2.1 Specifications to Ensure Trusted Exchange of Information

Many novel and innovative solutions were reported to overcome rewriting attacks in XML which include XML Digital Signature [10], WS Policy [11, 12], WS Security [13], SOAP Account [14], and WS Policy Advisor [15]. These methods provide excellent mechanisms to check rewriting attacks, but several shortcomings of these methods are reported in the literature.

XML Digital Signature has a security loophole that makes it vulnerable for modification without detection [9]. Since WS Policy and WS Security make use of XML Digital Signature, they are also affected by this issue [9].

WS Security standard is heavyweight for RCNs as it encompasses standards such as XML Digital Signature, XML Encryption, X.509 certificate, and Kerberos ticket. With larger and complex configuration files, WSE Policy Advisor reports performance degradation. It fails in the detection of signed element reordering attack [7, 14, 16, 17]. A novel method is proposed in [14, 18] to check the rewriting attack—but [19] reported a flaw in this method [19]. The security specification that is proposed in [4] is enhanced in [9] to include measures to check rewriting attacks. A modified accounting method is proposed with reduced verbosity and patches for the reported security holes. YASchema is also refined to include the WS policy assertions. The enhanced security measures introduced in [3] related to rewriting attacks include a refined XML digital security specification, inclusion of policy assertions in the schema definition, and structures to check rewriting attacks [3, 9]. The possibility of applying these techniques to propose a PBM mechanism is to be explored [9].

2.2 Measures for a Lightweight Mechanism for PBM

In PBM, messages can be packed with security policies by the sender. Receiver has to enforce these policies. This arrangement provides a distributed and secure mechanism for transmitting messages [8]. Several papers such as [11, 20, 21] discuss the method of adopting PBM. In a Policy-Based Network (PBN), policies are implemented using a high-level language.

Several standards are available for the choice of specifying policy. These standards can be broadly classified into XML-based and non-XML-based technologies. Prominent XML-based languages used in PBM are eXtensible Access Control Markup Language (XACML) [22], Security Assertion Markup Language (SAML) [23], and Enterprise Privacy Authorization Language (EPAL) [24]. XACML is an OASIS standard. It consists of a policy language and an access control decision request/response language [25]. Although XACML is accepted as a precise and integral policy description method, several difficulties are reported [18, 26] that make it undesirable for RCNs. Also it assumes that data is stored in XML documents, which sacrifices the flexibility to use other types of data storage and system implementations [27]. EPAL is a formal language for writing enterprise privacy policies according to fine-grained positive and negative authorization rights [28]. An EPAL policy is a well-formed XML-based document. In spite of these advantages, it is reported that EPAL cannot suitably handle non-data-related policies. It is also reported that EPAL and XACML are not commercially adopted [24]. SAML is an XML-based open standard for exchanging authentication and authorization data between parties [23]. In addition to the difficulties listed above, the verbosity of this XML-like format is also a hindering factor for the use of these languages in a resource-constrained environment.

Hence, it can be concluded that there is a scope for a "lightweight policy mechanism" that can overcome the verbosity barrier of existing techniques.

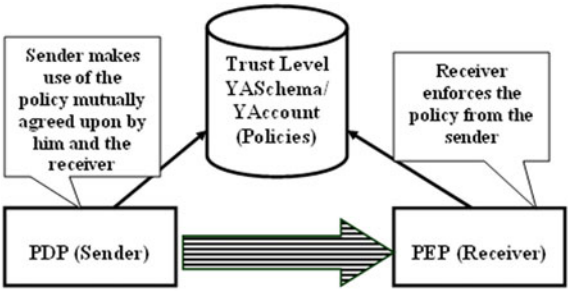
3 Proposed Work

3.1 Scheme of the Proposed Work

A PBM system has the following stages [6, 8]: (i) prepare the payload of the messages and choose the recipients, (ii) identify the constraints and rules to be applied to open the message at the recipient side and define them using a policy, (iii) send the message to the recipient by adding the policy to it, (iv) the messaging system at the recipient side retrieves the policy attached with the incoming message, and (v) the target system verifies the policy and if it's ok, message is handed over to the recipient. Otherwise, message is refused.

The essential components (Fig. 1) of a fully fledged PBM include a Policy Decision Point (PDP) and a Policy Enforcement Point (PEP). PDP interprets the policies stored in a repository, frames the policy decisions, and sends them to PEP. PEP applies and executes different policies. Two management models exist for PBM: one for outsourcing and the other for provisioning [20]. In outsourcing, data is sent to update the PDP (which in turn may update other PEPs). Provisioning is concerned with the installation of a policy by the PDP in PEP. The former method is used here for this scheme, as it suits the limited scope of the proposed scheme.

Fig. 1 Lightweight PBM mechanism


Sender makes use of the policy mutually agreed upon by him and the receiver
Trust Level YASchema/ YAccount (Policies)
Receiver enforces the policy from the sender
PDP (Sender)
PEP (Receiver)

In the proposed scheme, the message is prepared in TYAML format and policy assertions are specified in YASchema. Only the basic policy assertions and built-in features of YASchema are applied in specifying the constraints. The trust level is increased with the introduction of YAccount. The role of PDP is taken over by the sender of the message. Sender either attaches a YASchema or points to a YASchema which is to be referred by the receiver (PEP) to decide whether the receiver has the rights to access the message (as shown in Fig. 1). The receiver then cross-checks whether the message is as per the structures sent or mutually agreed upon by the sender and receiver.

3.2 Application of YASchema for Assertion of Policies

The proposed YASchema in [3] had an important role in verbosity reduction. In this section, the possibility of extending YASchema to contain the policy assertions as well is explored.

It is seen that the policy assertions are prepared in separate policy files in mechanisms such as WS Policy [11], WS Policy Assertion [6], and WS Security Policy [12]. YASchema can be used for policy assertions as well and hence eliminating the use of a separate policy file. In [9], four basic assertions [11, 23] that are commonly in use in the policy files are mapped to the YASchema. These details are included in the YASchema definition. In addition to these assertions, other assertions such as “Expires”, “QOS”, and “Acknowledgement Interval” can also be added, when required. It is possible to prevent the addition of false entries to the message by using these assertions.

When YASchema is utilized for defining the policy assertions, size of the policy file is reduced and the need for a separate policy file is eliminated.

In spite of these advantages, the flaws made in preparing policy files can result in the tampering of messages [14].

To check these issues, a modified accounting structure proposed in [9] is used.

3.3 Design of a Lightweight PBM Mechanism

Similar to the layered approach to tackle the rewriting attacks [9], a two-level lightweight PBM mechanism is proposed. In the first level, YASchema is employed as a policy assertion file and in the second level YAccount is used to check the inclusion of bogus additions to the original message.

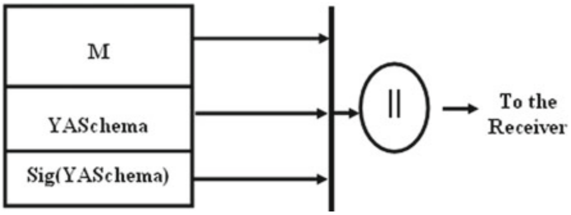


Fig. 2 Level 1—YASchema is attached with the message (Sender's Side)

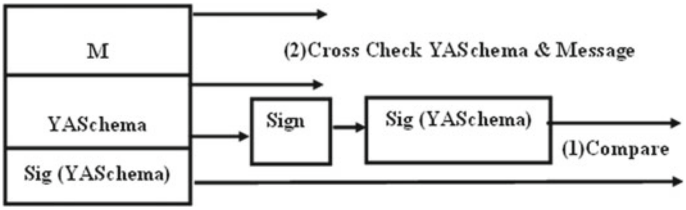


Fig. 3 Level 1—YASchema is attached with the message (Receiver's side)

3.3.1 Level 1 in This Level, the YASchema Is Used to Check the Attacks

Sender and receiver can have a common YASchema that is predefined or the sender can attach a message-specific YASchema along with the message and send that to the receiver. When a common predefined YASchema is used, the message at the receiver is verified with the YASchema. If the message has tampered while in transit, the test will fail, as the YASchema is predefined and agreed upon by the sender and receiver.

When sender attaches the YASchema along with the message, (s)he adds the digest value of YASchema to the message (Fig. 2). Receiver calculates the digest value and compares it with the digest value being sent by the sender. This will reveal any tampering made to the message in transit. If there is no mismatch, the receiver will verify the message with the YASchema to check the existence of anomalies (Fig. 3).

3.3.2 Level 2

In level 2, YAccount is used to strengthen the security of the message further (Figs. 5 and 6). Sender calculates the YAccount for the message, signs it, and sends it to the receiver along with the message (Fig. 4). Receiver calculates the YAccount of the message and compares the calculated and received values of YAccount (Fig. 5).

Policy assertions are sent with the document through these steps: (i) sender and receiver choose a trust level (level 1, level 2, or both) for message transfer, (ii) in case level 1 is chosen, sender and receiver decide whether to use a predefined YASchema or to send the YAShema corresponding to the message, (iii) if a common

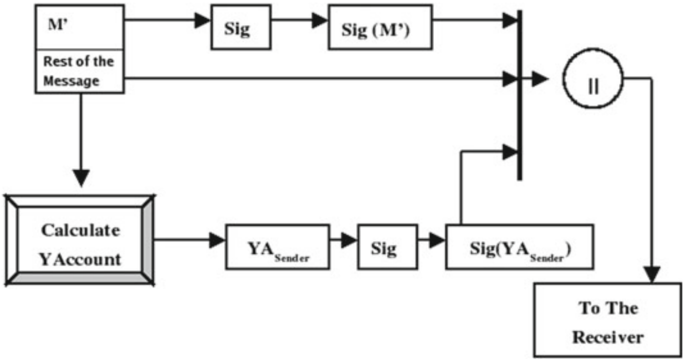


Fig. 4 YAccount derived by the sender

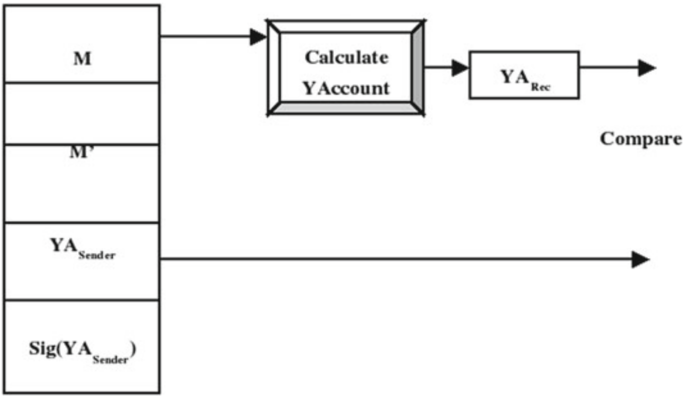


Fig. 5 YAccount derived by the receiver

YASchema is used, the message received is verified with the YASchema constructs and policy assertions. Message is rejected if there are any violations, (iv) if YASchema is also a part of the message, its digest value is added with the message before sending. Receiver calculates the digest value and compares it against the digest value it received. Message is rejected in case of a mismatch. Otherwise, receiver verifies the message as per the constructs and policy assertions mentioned in the YASchema, (v) if level 2 is chosen, YAccount received is compared against the calculated YAccount for anomalies, and (vi) all of the steps (ii to v) are done, if both trust levels are chosen.

This combined approach using YASchema and YAccount has several merits over conventional policy files: (i) reduced verbosity, (ii) simple approach in the preparation of YASchema and YAccount, in comparison with the complex nature of PBM languages such as XACML, (iii) increased trust level with the introduction of YASchema along with YAccount, and (iv) two-level approach taking care of the limitations of the RCNs.

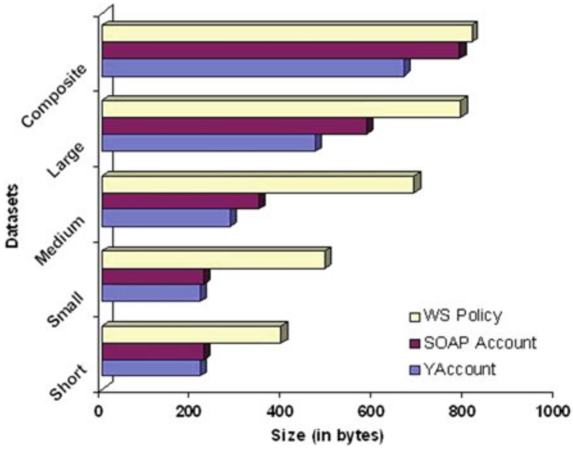


Fig. 6 Size comparison—WS policy, SOAP account, and YAccount

4 Performance Evaluation

The primary focus of the performance evaluation is on the size of the policy files to be transmitted. Test data is generated as per the categories of messages identified in [4]—short, small, medium, large, and complex.

4.1 *Size of WS Policy, SOAP Account, and YAccount Size*

A size comparison of WS Policy file, SOAP Account, and YAccount is shown in Fig. 6. It can be seen that YAccount takes the least size, and WS Policy files take the maximum size. Also when the message sizes are small, both SOAP Account YAccounts take almost equal number of bytes.

4.2 *Size of WS Policy and YASchema*

It can be seen from Fig. 7 that YASchema outperforms WS Policy files for short, small, and medium datasets. But for the remaining categories, WS Policy file outperforms YASchema.

Thus it can be concluded from the above discussion that the proposed work has performance advantages over existing XML-based standards for PBM.

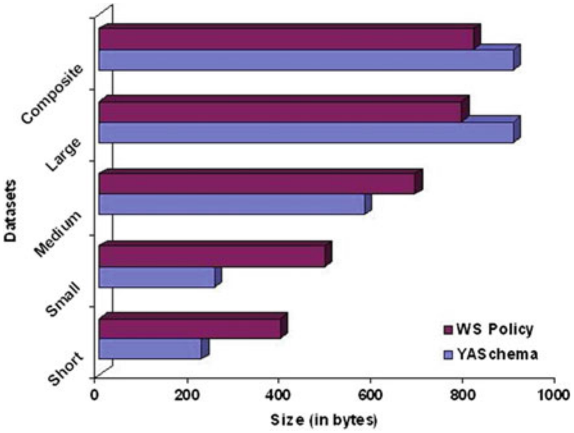


Fig. 7 Size comparison—WS policy versus YASchema

5 Conclusion

The increased penetration and importance of RCNs demands a better secure data transmission scheme suitable for the RCNs. This work presents an alternative technique for PBM to work within the limitations of network latency, bandwidth requirement, and low battery backup, commonly seen in the RCNs.

The proposed work is based on a format derived from YAML (TYAML). TYAML is less verbose than XML. The features of the proposed work include (i) applying YASchema as a policy assertion file, (ii) introduction of an accounting structure known as YAccount as a secondary tool for PBM, and (iii) a two-level lightweight PBM mechanism that consists of YASchema in the lower level and YAccount in the upper level. Results obtained indicate a performance advantage over the existing methods.

TYAML and the schema (YASchema) are to be fine-tuned to include the attribute-based access control mechanism and semantic-based security aspects. A working model of PBM has to be designed and developed to assess its functionality and problems in a real-time environment. These are suggestions for future work.

Acknowledgments The author of this paper would like to thank University Grants Commission SWRO Bangalore for sponsoring this work (1646-MRP-/14-15/KLCA028/UGC-SWRO).

References

1. Michael MP (2005) Energy awareness for mobile devices, research seminar on energy awareness. University of Helsinki
2. Ravi P, Raghunathan N (2006) Heterogeneous grid computing for energy constrained mobile device. IEEE Trans Mob Comput V:128–143

3. Haleem PPA, Sebastian MP (2009) An efficient approach for thinning of wireless mobile messages. ICICI Express Lett III:99–105
4. Haleem PPA, Sebastian MP (2008) An alternative approach for slicing down the message size and enhancing the security in wireless mobile network. Mediterr J Comput Netw V:148–149
5. Haleem PPA, Sebastian MP (2009) Optimizing message verbosity and energy consumption in secured wireless mobile networks. Int J Mob Comput Multimed Commun (IJMCMC) I:22–35
6. Abdul Haleem PP, Sebastian MP (2012) An energy conserving approach for data formatting and trusted data exchange in resource constrained networks. Knowl Inf Syst 32(3):559–587 (Springer, London). https://doi.org/10.1007/s10115-011-0450-0, Print ISN: 0219-1377
7. Benameur FAK, Fenet S (2008) XML rewriting attacks: existing solutions and their limitations. In: IADIS international conference on applied computing 2008, Algarve, Portugal, IADIS Press 2008, vol. abs/0812.4181
8. Eggenberger M, Prakash N, Matsumoto K, Thurmond D (2009) Policy based messaging framework. Lecture notes in computer science, vol 4749. Springer Berlin, Heidelberg, pp. 497–505
9. Haleem PPA, Sebastian MP (2014) An alternative approach for XML messaging. Int J Adv Res III:251–294. ISSN 2320-5407
10. Bartel M, Boyer J, Fox B, LaMacchia B, Simon B (2008) XML signature syntax and processing (second edition), W3C Recommendation. http://www.w3.org/TR/xmldsig-core/
11. Eggenberger M, Prakash N, Matsumoto K, Thurmond D (2007) Policy based messaging framework. In: ICSOC ‘07: proceedings of the 5th international conference on service-oriented computing. Springer, Berlin, Heidelberg, pp 497–505
12. Della-Libera G, Gudgin M, Hallam-Baker P, Hondo M, Granqvist H, Kaler CM, McIntosh M, Nadalin, Nagaratnam N, Philpott R, Prafullchandra, Shewchuk, Walter, Zolfonoon R (2005) Web services security policy language (ws-security policy). http://specs.xmlsoap.org/ws/2005/07/securitypolicy/ws-securitypolicy.pdf
13. Rahaman MA, Rits M, Schaad A (2006) An inline approach for secure SOAP requests and early validation. In: OWASP’06: proceedings of the OWASP Europe 2006 conference. OWASP, pp 19–33
14. Bhargavan K, Fournet C, Gordon AD, Shea GO (2005) An advisor for web services security policies. In: SWS ‘05: proceedings of the 2005 workshop on Secure web services. ACM, New York, NY, USA, pp 1–9
15. Sinham SK, Benameur A (2008) A formal solution to rewriting attacks on SOAP messages. In: SWS ‘08: proceedings of the 2008 ACM workshop on Secure web services. ACM, New York, NY, USA, pp 53–60
16. Rahaman MA, Schaad A (2007) SOAP-based secure conversation and collaboration. In: ICWS 2007: proceedings of the 2007 IEEE international conference on web services. IEEE Computer Society, pp 471–480
17. Gajek S, Liao J, Schwenk A (2007) Breaking and fixing the inline approach. In: SWS ‘07: proceedings of the 2007 ACM workshop on secure web services. ACM, New York, NY, USA, pp 37–43
18. Lang B, Zhao N, Ge K, Chen K (2008) An XACML policy generating method based on policy view. In: ICPCA 2008: proceedings of the 2008 third international conference on pervasive computing and applications. IEEE, pp 295–301
19. Grand GL, Springinsfeld F, Riguidel M (2003) Policy based management for critical infrastructure protection. In: INFORMATIK 2003: proceedings of the first international conference on communities and technologies. GI, Gesellschaft fur Informatik, Bonn, pp 67–78
20. Matthys N, Joosen W (2008) Towards policy-based management of sensor networks. In: MidSens ‘08: proceedings of the 3rd international workshop on middleware for sensor networks. ACM, New York, NY, USA, pp 13–18
21. Parducci, Lockhart, Levinson, McRae: OASIS extensible access markup language, OASIS (2005). http://www.oasis-open.org/committees/tc/home.php/wg/abbrev/xacml
22. Security assertion markup language (SAML), XML Cover Pages (2008). http://xml.coverpages.org/saml.html

23. Enterprise privacy authorization language (EPAL 1.2), W3C (2003). http://www.w3.org/Submission/2003/SUBM-EPAL-20031110/#Introduction
24. Liu AX, Chen F, Hwang J, Xie T (2008) Xengine: a fast and scalable XACML policy evaluation engine. In: SIGMETRICS '08: proceedings of the 2008 ACM SIGMETRICS international conference on measurement and modeling of computer systems. ACM, New York, NY, USA, pp 265–276
25. Mordore Intelligence, Wireless Sensors Network Market—growth, trends, and forecast (2019–2024). https://www.mordorintelligence.com/industry-reports/wireless-sensor-networks-market
26. Baquero Merino A (2014) Coast services: achieving service customization and policy-based differential access in personal information systems, Ph.D. dissertation. University of California, Irvine
27. Box D, Hondo M, Kaler C, Maruyama H, Nadalin A, Nagaratnam N, Patrick P, von Riegen, Shewchuk J (2003) Web services policy assertions language (ws-policy assertions). http://xml.coverpages.org/ws-policyassertionsV11.pdf
28. Nadalin A, Kaler C, Hallam-Baker P, Monzillo R (2004) Web services security: SOAP message security 1.0 (ws-security 2004), OASIS. http://docs.oasis-open.org/wss/2004/01/oasis-200401-wss-soap-message-security-1.0.pdf

A Lightweight Effective Randomized Caesar Cipher Algorithm for Security of Data

Vardaan Sharma, Sahil Jalwa, Abdur Rehman Siddiqi, Ishu Gupta, and Ashutosh Kumar Singh

Abstract In any kind of organization in the present scenario, raw data or meaningful data needs to be shared among different personnel; therefore, the chances of frauds or treachery are more that creates vulnerability in the working environment. Therefore, the protection of organizations' confidential and sensitive data among different levels of employees against these kinds of theft or illegal activity that violates the company security policy is prerequisite. ABE pronounced as attribute-based encryption was introduced. ABE is a parameter that plays a vital role in providing access control in a fine-grained manner for outsourcing data in a data sharing system. Moreover, CP-ABE was introduced that defines an access policy in order to cover all the attributes within the system. In this scheme, for encryption and decryption purposes, the user's private key is accompanied to the group of attributes. but due to its lack of efficiency and several other paradigms, it was proved ineffective. Hence, in order to overcome the existing issues, we have implemented a new cryptographic algorithm which is competent enough to encrypt and decrypt any type of file that contains any kind of data written in the range of ASCII values. We named this algorithm as randomized Caesar cipher algorithm.

Keywords Attribute-Based encryption · Data sharing · CP-ABE · Access policy · Access control · RSA · Randomized Caesar cipher

V. Sharma · S. Jalwa · A. R. Siddiqi · I. Gupta (✉) · A. K. Singh
Department of Computer Applications, National Institute of Technology, Kurukshetra, India
e-mail: ishugupta23@gmail.com

V. Sharma
e-mail: vardaansharma096@gmail.com

S. Jalwa
e-mail: sahiljalwa@gmail.com

A. R. Siddiqi
e-mail: abdur9808@gmail.com

A. K. Singh
e-mail: ashutosh@nitkkr.ac.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_39

1 Introduction

As we see in our day-to-day life, use of machines and personal computers is coming vital resources to save our data and to provide a large number of services that we need at regular interval of time. Similar to that all the big organizations and industries need to save and process a large amount of data in day-to-day business transactions and operations. Apart from that, they need to handle a large amount of data that is generated in the heavy amount on an everyday basis [1]. Thus, from the above, it can be guessed that this data is how much crucial for them to perform any task or to make any decision related to the organization. It will contain information of employees, their designation, salaries, and much more; moreover, it will contain information about the company name, location, their achievements and year of establishment, kind of services they are providing, product-based or service-based, etc. So, from the above, it can be inferred that data is of utmost importance and it is necessary to protect this data from any kind of vulnerability [2] [3]. It is necessary to take appropriate actions and steps to prevent and secure our data, otherwise may lead to huge business losses and cause serious problem. We have come with the idea to contribute toward security and protection by using cryptography. There are several cryptographic algorithms that have been proposed so far, performing a wide variety of security features [4]. Some of the earlier techniques that have been implemented so far like Caesar cipher, RSA algorithm, AES algorithm, DES, Triple DES, etc. [5–8]. All of the existing algorithms have contributed a lot toward security paradigms.

2 Related Work

Earlier techniques and methodology that have been proposed like Symmetric-Key Cryptography, Public-Key Cryptography in which decryption is done with the help of private key but proved unsuccessful due to certain drawbacks [9, 10]. Then an advancement over PKC known as Identity-Based Encryption introduced that mainly concentrates on identity of end user. Then after fuzzy identity-based encryption, comes into play stating that if a person credential value satisfies access policy, only then decryption will take place and many more techniques are preexisted. But CP-ABE is considered more efficient due to its spectacular functionality; here, attributes will define every user credentials and decryption process is operated by encryptor. Algorithms like Caesar cipher are particularly applied to text messages which are used in communication among individual so that privacy is maintained. This text is encoded with the help of substitution of another text in place of original text [11, 12]. But major drawback of this algorithm is that the KEY value is fixed due to which it provides least level of security and scope of this technique is very narrow [13]. Above algorithm uses a pair of large primes for constructing both public and private keys and commonly used for authentication purpose. Difficulty in decomposition of large digits (factorization of integral values) will determine its level of security. It is

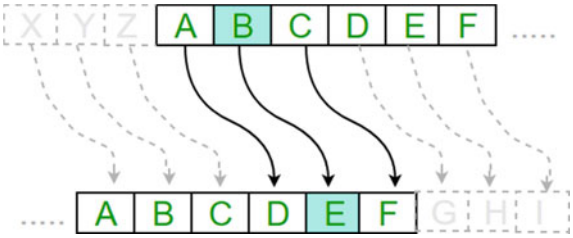


Fig. 1 Random Caesar cipher

resistant to all kinds of password attacks that it is familiar with [14, 15]. End user can take the help of probabilistic algorithm. The implementation process of RSA is a little bit complicated one.

3 Randomized Caesar Cipher Algorithms

The Caesar cipher is one of the preliminary methodologies of encryption technique. It's just a kind of substitution in plaintext, i.e., for each letter of a given text of a sentence some other alphabet will take place, i.e., replaced by some fixed number of positions in the alphabet sequence [16, 17] as shown in Fig. 1. For example, with a shift of 1, B replaces A, C would replace B, and this will continue on. The technique is named after Julius Caesar, used by him to communicate with his peers.

$$\text{Encn}\ (X) = (X + n) \bmod 26,\ n \text{ implies number of shifts}$$
$$\text{Decn}\ (X) = (X - n) \bmod 26$$

But in randomized Caesar cipher, the alphabets are not replaced by any fixed number of positions; instead, we provide or decide a range of numbers for random key generation and each time when we execute the algorithm that key is distinct but it will fall within the range that we have provided [18–22] as shown Fig. 2. This is one of the biggest advantages of our algorithm, i.e., generated Key is not static it will take different values every time we execute this algorithm.

4 Pseudocode (Randomized Caesar Cipher)

- First of all, import **random** and **time** package.
- **Random** package will give **random.int()** function and **time** package will help us to determine the encryption and decryption time of file.
- Now we generate a random key with the help of **random.int()** function. This key will differ in value for each iteration of execution.

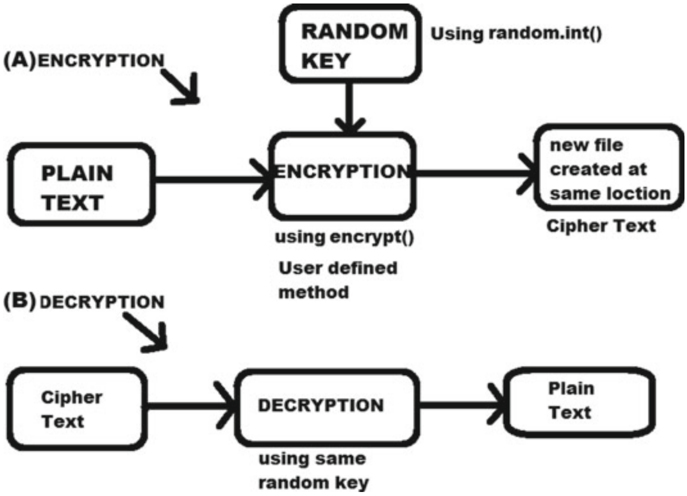


Fig. 2 Working random Caesar cipher

- We provide a range of values according to our requirement in order to generate randomized key.
- After generation of key, it will be shown that what key value is used at that particular instant of time.
- Now we create a **encrypt()** function which will encrypt the original file that is placed in that particular folder or desired path given by us.
- **encrypt()** function perform character by character and line by line encryption of entire file.
- After encryption of the entire file, a new encrypted file will be created at the same location where our original file is present.
- Now by using the time package, we will show how much time is taken by **encrypt()** function in order to encrypt the file.
- Just after that **decrypt()** function working will take place. It will perform decoding of that newly generated encrypted file in the same fashion, i.e., character by character and line by line till the end.
- Upon complete execution of **decrypt()** function time taken to decrypt, the entire file will be displayed.

5 Performance Evaluation

5.1 Experimental Setup

Experiment description of the environment in which the proposed scheme has been implemented and results have been evaluated is depicted in Table 1.

Table 1 Experimental setup

Parameters	Specifications
CPU	Intel (R) core(TM) i5-5200U CPU @ 2.20 GHz
RAM	8.00 GB
OS	Ubuntu 16.04, Windows
External library	GMP, PBC, OpenSSL
Language	C, Java, Python
Compiler	Gcc

5.2 *Experimental Results*

Figure 3 shows the encryption time taken by algorithm for encrypting the file of different sizes. One point to note here is that it is not necessary that a large file will always take more time to encrypt than a smaller size file because key is generated different every time, and we encrypt file of different sizes. Its value may vary according to the range provided. If a large key value is generated for a small size file and small key value is generated for a larger file, then there are chances that the file with larger key value takes more time to encrypt than the other.

Here in Fig. 4 we can see that file size of 15 MB takes more time to decrypt as compared to other files. So, there are chances that its shift value is very large as compared to other two files.

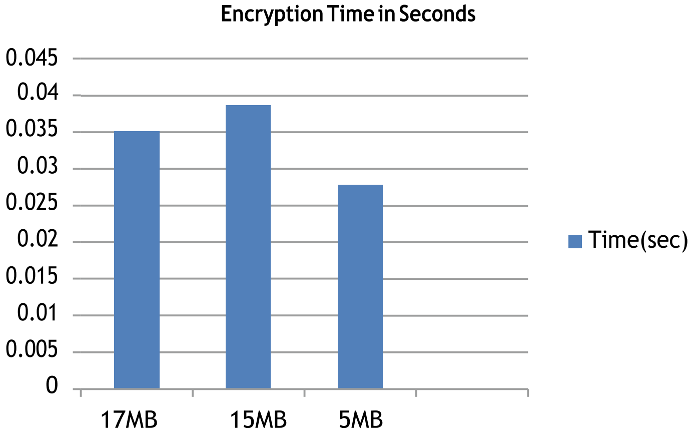


Fig. 3 Encryption time of different sizes of files

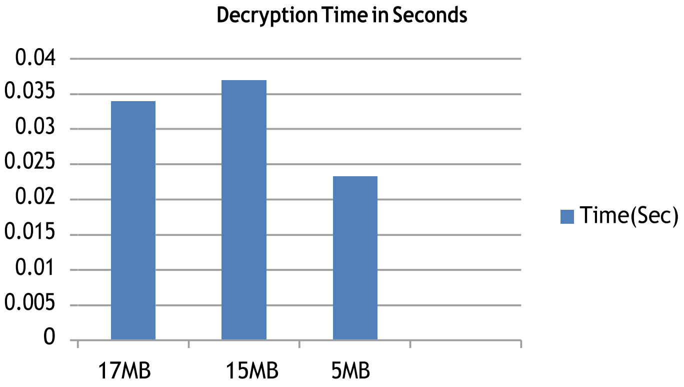


Fig. 4 Decryption time of different sizes of files

5.3 *Comparison*

5.3.1 Encryption

The proposed scheme has been compared with the existing scheme, to evaluate the performance of the proposed scheme; in Fig. 5, it is clear that time taken by the proposed scheme in encryption is lesser than the existing scheme, which proves more efficiency of proposed scheme.

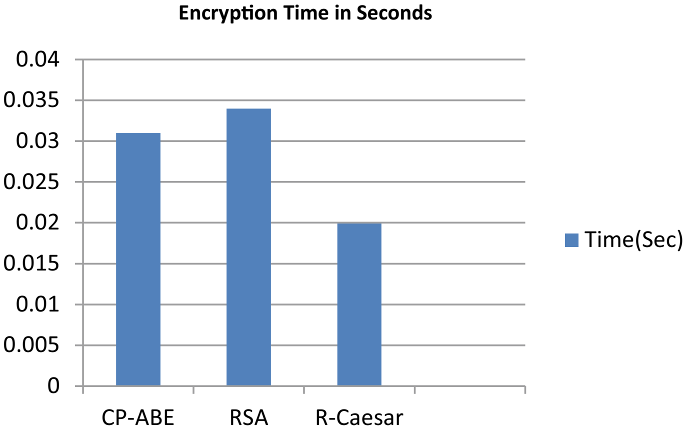


Fig. 5 Encryption time of different algorithms

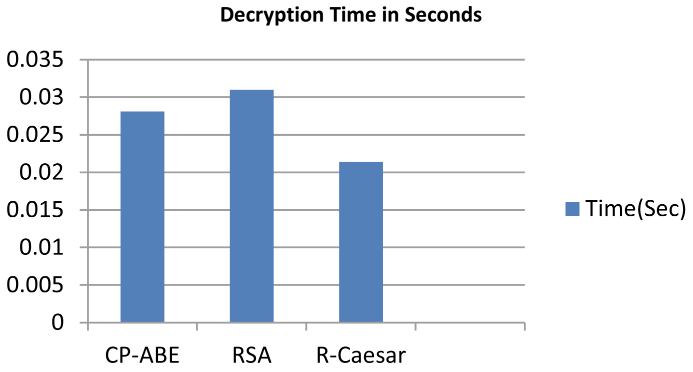


Fig. 6 Decryption time of different algorithms

5.3.2 Decryption

The proposed scheme has been compared with the existing scheme, to evaluate the performance of the proposed scheme; in Fig. 6, it is clear that time taken by our proposed scheme in decryption is lesser than the existing scheme, which proves more efficiency of the proposed scheme.

6 Conclusion

In this paper, we have taken the reference of Caesar cipher algorithm which was designed by Julius Caesar for the purpose of security measure but due to its large number of limitations (like simple storage structure) in various areas of information technology and computer science, it is no longer considered as a successful and efficient algorithm and later on also proved that it provides a minimum level of security. So, we come with an idea of providing a key randomly from a pool of keys. And for that we will define a range according to our requirement after that, the key is chosen from that range, and process of encryption and decryption takes place. The larger the range, the larger the complexity in encryption of text will increase. So, to conclude, one can achieve the goal of cryptography with the help of our algorithm. We have implemented this algorithm for the purpose to provide a high level of security and protection of any kind of malicious activity. In order to run this algorithm, one has to fulfill its requirement condition and provide the desired environment. This algorithm is purely written in Python and in order to execute, installation of PyCharm is necessary. Otherwise, the end user will not be able to run this algorithm. With the use of Python language, it makes it easier for end user to understand the code of algorithm, and also makes them able to learn the algorithm in a very short period of time. Scope of our proposed algorithm can be extended further and any kind of modification can be done easily by the person who knows

Python language. Hence, by using our algorithm one can relish the above-prescribed benefits for their own convenience without caring about compatibility environment.

References

1. Gupta I, Singh AK (2019) Dynamic Threshold based Information Leaker Identification Scheme. Inf Process Lett 147:69–73
2. Gupta I, Singh AK (2019) An integrated approach for data leaker detection in cloud environment. J Inf Sci Eng
3. Gupta I, Singh N, Singh AK (2019) Layer-based privacy and security architecture for cloud data sharing. J Commun Softw Syst (JCOMSS) 15(2)
4. Gupta I, Singh AK (2018) A probabilistic approach for guilty agent detection using bigraph after distribution of sample data. Procedia Comput Sci 125:662–668
5. Jalwa S, Siddiqi AR, Sharma V, Gupta I, Singh AK (2019) Comprehensive and comparative analysis of different files using CP-ABE. In: First international conference on advanced communication & computational technology (ICACCT)
6. Bethencourt J, Sahai A, Waters B (2007) Ciphertext-policy attribute-based encryption. In: 2007 IEEE symposium on security and privacy (SP'07). IEEE
7. Goyal V, Pandey O, Sahai A, Waters B (2006) Attribute-based encryption for fine-grained access control of encrypted data. In: Proceedings of the 13th ACM conference on computer and communications security. ACM, pp 89–98
8. Chor B, Rivest RL (1988) A knapsack-type public key cryptosystem based on arithmetic infinite fields. IEEE Trans Inf Theory 34(5):901–909
9. Shamir A (1999) Identity-based cryptosystems and signature schemes, 2nd edn. In: Workshop on the theory and application of cryptographic techniques. Springer, Berlin, Heidelberg
10. Sahai A, Waters B (2005) Fuzzy identity-based encryption. In: Annual international conference on the theory and applications of cryptographic techniques. Springer, Berlin, Heidelberg
11. Han F et al (2014) A general transformation from KP-ABE to searchable encryption. Future Gener Comput Syst 30:107–115
12. Liu P et al (2014) Efficient verifiable public key encryption with keyword search based on KP-ABE. In: 2014 ninth international conference on broadband and wireless computing, communication and applications. IEEE
13. Porwal S, Mittal S (2017) Implementation of Ciphertext policy-attribute based encryption (CP-ABE) for fine grained access control of university data. In: 2017 tenth international conference on contemporary computing (IC3). IEEE
14. Guo F et al (2014) CP-ABE with constant-size keys for lightweight devices. IEEE Trans Inf Forensics Secur 9(5):763–771
15. Bobba R, Khurana H, Prabhakaran M (2009) Attribute-sets: a practically motivated enhancement to attribute-based encryption. In: European symposium on research in computer security. Springer, Berlin, Heidelberg
16. Ning J et al (2016) White-box traceable CP-ABE for cloud storage service: how to catch people leaking their access credentials effectively. IEEE Trans Dependable Secur Comput 15(5):883–897
17. Li J et al (2017) User collision avoidance CP-ABE with efficient attribute revocation for cloud storage. IEEE Syst J 12(2):1767–1777
18. Wang S et al (2016) An efficient file hierarchy attribute-based encryption scheme in cloud computing. IEEE Trans Inf Forensics Secur 11(6):1265–1277
19. Li L et al (2017) A ciphertext-policy attribute-based encryption based on an ordered binary decision diagram. IEEE Access 5:1137–1145
20. Luo W, Ma W (2018) Efficient and secure access control scheme in the standard model for vehicular cloud computing. IEEE Access 6:40420–40428

21. Gupta I, Singh AK (2017) A probability based model for data leakage detection using bigraph. In: 7th international conference on communication and network security (ICCNS-2017). ACM, Tokyo, Japan
22. Gupta I, Singh AK (2019) A confidentiality preserving data leaker detection model for secure sharing of cloud data using integrated techniques. In: Seventh international conference on smart computing and communication systems (ICSCC), Sarawak, Malaysia, pp 1–5

RPL-Based Hybrid Hierarchical Topologies for Scalable IoT Applications

Animesh Giri and D. Annapurna

Abstract The applications built nowadays are mainly distributed, and most of them have sensors enabling them to do it. Internet of Things in the same context is solving real-world problems. Although there are challenges still like on ground scalability, efficiency, etc. and a lot of other bottlenecks, in our work here we have focused on RPL routing protocol and the potential to scale under strained networks. Some real-world application scenarios like military and agriculture build in an environment with the "strained" transmission and interference ranges, which requires the nodes to be retained as part of Destination-Oriented Directed Acyclic Graph (DODAG). The simulation study done using Contiki OS-based Cooja simulation environment on hierarchical and circular network topologies for highly scalable and strained networks shows high energy consumption and the impact on the radio duty cycle on few selected nodes of DODAG. Combining the features of hierarchical and circular network topology, we propose a hybrid hierarchical topology with multiple sinks which resembles the real-world applications. The testing and relative comparison of RPL's Objective Functions (OFs) consists of the following parameters: Power Consumption, Radio Duty Cycle, and possible topologies. The results of the simulation study of RPL protocol show that the proposed hybrid network topology results in much stable energy consumption and radio duty cycle increasing the scale and strain on the network.

Keywords RPL · Routing protocol · Scalability · Strained · Contiki · Cooja · IoT · Hierarchal network

A. Giri (✉)
Department of Computer Science and Engineering, PESIT - Bangalore South Campus,
Visvesvaraya Technological University, Belagavi, Karnataka, India
e-mail: animeshgiri@pes.edu

D. Annapurna
Department of Computer Science and Engineering, PESIT-Bangalore South Campus,
Bengaluru, Karnataka, India
e-mail: annapurnad@pes.edu

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_40

1 Introduction

In the new age of technology, every segment is touched by the advances. One of such areas is the countries' defense system that protects the country at the enemy's line. In the past decades, the operations in the military are highly affected by technology. As discussed by Yushi et al. [1] the military has used the networking equipment for various aspects like information sensing, communication, transmission, etc., and it helps them make crucial decisions at times. Based on the work that most of the military does the topology that suits them the most is a hierarchical network which can be seen as a level by the level arrangement of nodes which can have unique tasks. Therefore, to test the applications of these operations we have used the hierarchical structure itself.

Considering the impact of IoT in the domain of agriculture, today there are remote sensors in every large agricultural field that leads to the adaptation of smart farming. An example is the precision farming as discussed by Anurag et al. [2], which analyses the routing problems and shows a mesh network topology as a solution for such a scenario. Now with a lot more sensors and even more economic scalability constraints with the Internet of things [IoT], there is a need to analyze the routing techniques suitable for real-world application scenarios.

There are two categories of routing techniques broadly categorized as distance vector and link state routing.

The former finds the best path by using the minimum distance between source and destination, and that is done by a distance vector which holds all the distances from that node. In case of the latter nodes, each link is associated with weight and it's updated regularly. The weights are proportional to the cost to the destination so we choose the smallest one.

We'll be dealing with Wireless Sensor Networks (WSN). In the WSNs, the routing protocol selection is a problem. There are many protocols like, as discussed by Aijaz et al. [3], CORPL (A Routing Protocol for Cognitive Radio Enabled AMI Networks) [4], RPL (Routing Over Low-power and Lossy network), CARP (Common Address Redundancy Protocol), as mentioned by Gnawali et al. [5], CTP (Collection Tree Protocol), and few observations by Clausen et al. [6] LOADng (Lightweight On-demand Ad hoc Distance-vector Routing Protocol – Next Generation), and a few more. Due to constraints with data routing in such wireless networks, which impact the routing protocol significantly, to give options to suffice various needs, the Routing Over Low-power and Lossy network [4] was given by team which works under (IETF). This paper presents the analysis of the problems faced in the military application using the hierarchical topology scenarios. The simulations of these scenarios have been done using the Contiki OS-based Cooja simulator.

Rest of the paper consists of the following sections. Section 2 includes related work in the area of WSN routing. Section 3 contains key terms around the simulation environment and the parameters defined in the experiment. The performance quality study of the RPL under various constraints is described in Sect. 4, and the paper is concluded in Sect. 5.

2 Related Work

2.1 Problems with Hierarchical Topology

There has been a lot of research on applications of IoT like military and agriculture. We will focus on the research papers, which work with the application and analyze the problems and limitations of the current routing protocols.

Referring to the work done by Kumar's et al. [7], whereafter the simulations they have concluded that small-size hierarchical networks work as per the requirement of the military but problems arise with a high frequency of messages needed to set it up, in a large number of nodes, in high-scale hierarchical networks. Similarly, there are a lot of survey papers who have concluded the following about the RPL protocols under various scenarios.

Gaddour et al. [8] mentioned: "One of the most important issues still left open is the specification of the Objective Function" which was left out by the ROLL team. Along with that RPL don't have a single parameter to enhance its objective function. They have mentioned the limitations like Immature Security in the RPL protocol, but that aspect isn't considered at this point by the wok done by Clausen et al. [9] goes on and add, issues are there because of dual directionality of communication and the scope of loops, the DODAG formation, and therefore in multipoint-to-point route provisioning techniques.

Afonso et al. [10] analyzed the unnecessary optional resource consumptions. Kim et al. [11] raised the issues on the tradeoff between simplicity and scalability with the RPL protocol. Liu et al. [12] bring the analysis of high-level scalability networks and shows the security issues as well as the problem with the power and computation issues.

RPL Being the primary choice for the Low-Power Lossy Network (LLNs), there is a lot of opportunities to analyze similar problems and see the more suitable alternatives.

2.2 Routing Protocol for Low-Power and Lossy Network (RPL)

RPL is a new, a protocol that is designed to work on top of offering from the link layer. It is mainly used for collection-based scenarios, where every node sends messages to a point after every fixed interval of time.

The messages it sends are controlled by a protocol to distribute information over the continuously changing network topology by ICMPv6 packets, such as DAO (Destination Advertisement Object), DIS (DODAG Information Solicitation), and DIO (DODAG Information Object). The DIO packets contain knowledge about the type of objective functions, the ranking of parents node and nodes' information details, etc.

The topology of RPL protocol is emerged from Acyclic Graphs, in particular, the DAGs. It's a tree topology, based on structure connecting the nodes in the Lossy Sensor Network. It's usually bottom-up or top-down search for nodes. It does have a different way of connecting, to the parent node. A single node here can be allowed to connect to multiple parent nodes to reach the destination, so we have those called as Destination-Oriented DAGs (DODAGs); most common nodes for the destination are the sink node, which even we have used route in practice for the Internet (i.e., Gateway) that acts as root for the tree. RPL has many features; some are Automatic-configuration of features, Automatic-healing, Loop prevention and recognition, and Independence.

3 Simulation and Network Setup

This section discusses the simulation and how the network is established, corresponding to the parameter of the protocol, with mimicking some topologies in area of interest which depends upon the application needs for stable and scalable routing protocol, and efficient self-healing or modification in configuration.

3.1 Simulation Parameter

We have used Cooja Simulator to study the RPL under various topologies and scenarios; parameters are shown in Table 1. The Unit Disk Graph Medium (UDGM) has been used to simulate the lossy networks. The sender nodes send the UDP packets which contain the node ID of the sender node to the sink node used in the DODAG formation.

3.2 Simulation Network Topologies

The network topologies considered for the experiments are hierarchical and circular. In topology diagrams, the green color, yellow color node represents the sink, sender nodes, respectively. The diagrams represent the radio medium of communication using UDGM. In the two concentric circles, the green circle in the inner ring is the transmission range, whereas the outer gray ring is the interference range with other radio sensors. The percentage shown in the figure below represents the reception ratio of transmission among sink node(s) and sender nodes.

Table 1 Network simulation parameters

Test parameters	Values	
Operating system (OS)	Contiki 2.7/Ubuntu 14.04	
Objective functions (OFs)	OF0	MRHOF
Metrics	Radio Duty, Energy	
TX range/INT range	70 m for transmission (TX) and 90 m for interference (INT) 50 m for transmission (TX) and 90 m for Interference (INT)	
TX ratio/RX ratio	100% for transmission (TX) and reception (RX)	
Topologies	Tree (Sparse), Circle (Sparse), Hybrid(Sparse)	
Simulation time	900s	
Mote type	Tmote sky	
Wireless channel model	Unit disk graph medium (UDGM): distance loss	
Sink node(s)	1, 3, and 5 (udp.sink.c)	
Sender nodes	30, 40, and 60 (udp.sender.c)	

3.2.1 Hierarchical Network Topology

A hierarchical network can be shown in Fig. 1 that allows nodes to reach the sink directly or contact each other to reach the sink, especially nodes at the edges and in the node's interference range. The hierarchical network topology is built by increasing the strain on the network. This topology is designed with the intention of "straining" the transmission and interferences ranges for the communication between the nodes.

As shown in Fig. 1a, all the sender nodes are within inside the green circle of transmission and very low strain for the communicating nodes.

As shown in Fig. 1b, only a small proportion of sender nodes are located inside the transmission range, few are placed at the network boundary along with the interference range, and few are scattered outside the sink's interference range, this shows that there is a moderate increase in the strain for the communicating nodes.

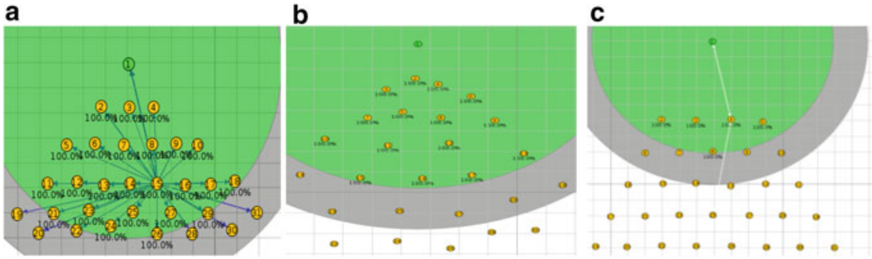


Fig. 1 **a** Low-strain hierarchical network topology. **b** Moderate-strain hierarchical network topology with single sink node and 30 sender nodes. **c** High-strain hierarchical network topology with single sink node and 30 sender nodes

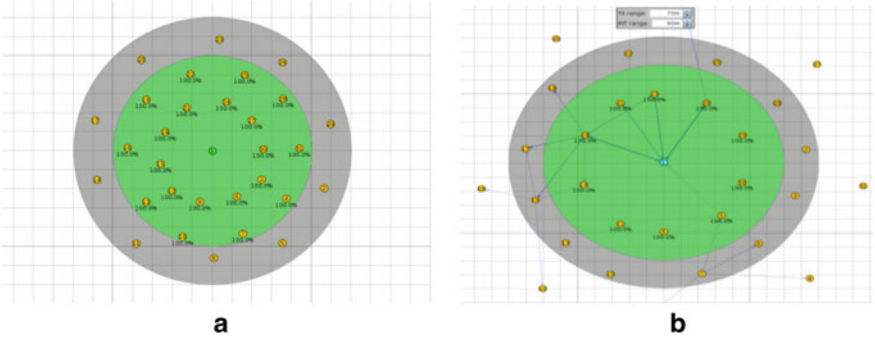


Fig. 2 Circular network topology with **a** Low strain, and **b** High strain with single sink and 30 sender nodes

As shown in Fig. 1c, the scenario in high strain shows that a very large number of nodes are placed outside the transmission circle and the interference ranges and beyond.

3.2.2 Circular Network Topology

A circular network topology can be shown in Fig. 2 that allows the placement of the nodes surrounding the sink node. The topology is built by increasing the strain on the network. To analyze the working of RPL in a sparse circular topology, we created scenarios where a single sink node is located in center and the sender nodes are in the circular form such that every node is reachable and within the transmission range. There are not much of significant changes in the circular topologies for low- and moderate-strain networks. Thus we have only considered low- and high-strain networks.

4 Simulation Results

In this section, we study the outcome of the various topologies discussed above and compare their average power consumption and average radio duty cycle, to understand the effect of these topologies on the performance of RPL protocol.

Average Power Consumption Performance—It is the electrical power consumed by each of the sensors to operate. The sensors' performance shown in graphs is relative to each node.

Average Radio Duty Cycle Performance—The Duty Cycle indicates that the fraction of time a resource is busy either listening or transmitting; the graphs show the ration of radio listen(red) and radio transmit(blue) of the nodes.

Referring to Fig. 3a–c, the observations made in the hierarchical network topology highlight that the few nodes had uneven power consumption and radio duty cycle distribution across the topology, and there is a significant increase in power consumption and radio duty cycle with the change in the scalability and strain on the network.

Referring to Fig. 3d, e, the observations made in the circular network topology highlight that the almost all nodes had the much stable power consumption and radio duty cycle distribution across the topology as compared to hierarchical network topology; the power consumption and radio duty cycle does not increase significantly with the change in the scalability and strains on the network.

The observation here, for the uneven distribution of resources under the network build with the hierarchical topology and the average power consumption metric for the circular network topology, motivates us to explore the advantages of using both hierarchical topology and the circular network topology to make a new hybrid network topology (Fig. 4).

4.1 Proposed Hybrid Network Topology

We propose the new hybrid network topology by combining the hierarchical and circular topology. In the hybrid network topology, the placement of the nodes is as per the circular topology and structured to function like the hierarchical topology. To build such a hybrid network topology, we require multiple sink nodes and multiple sender nodes. The stability of the hybrid topology increases with the increase in the sink nodes in the network. For the proposed hybrid topology, we have considered two scenarios as shown in Table 2.

We have considered two scenarios as shown in Table 2.

Refer to all in Fig. 5a, b. The observations made in the hybrid network topology highlights that the problem of extremely high power consumption by few nodes has reduced significantly. However, this problem persists in the hierarchical and circular topology. As shown in the hybrid network topology, by the usage of multiple sink nodes and appropriate placement of the sender nodes, the nodes are able to form a reliable DODAG. Thus, to conclude that the proposed hybrid network topology results in much stable behavior with the increase in the scale and strain on the network.

5 Conclusion

The performance metrics discussed in the paper suggest that the hybrid topology has a much more balanced and even distribution of load to all the nodes in the network as shown with the radio duty and average power consumption. It's a hybrid made from circular and hierarchical topology which outperforms similar hierarchical topology.

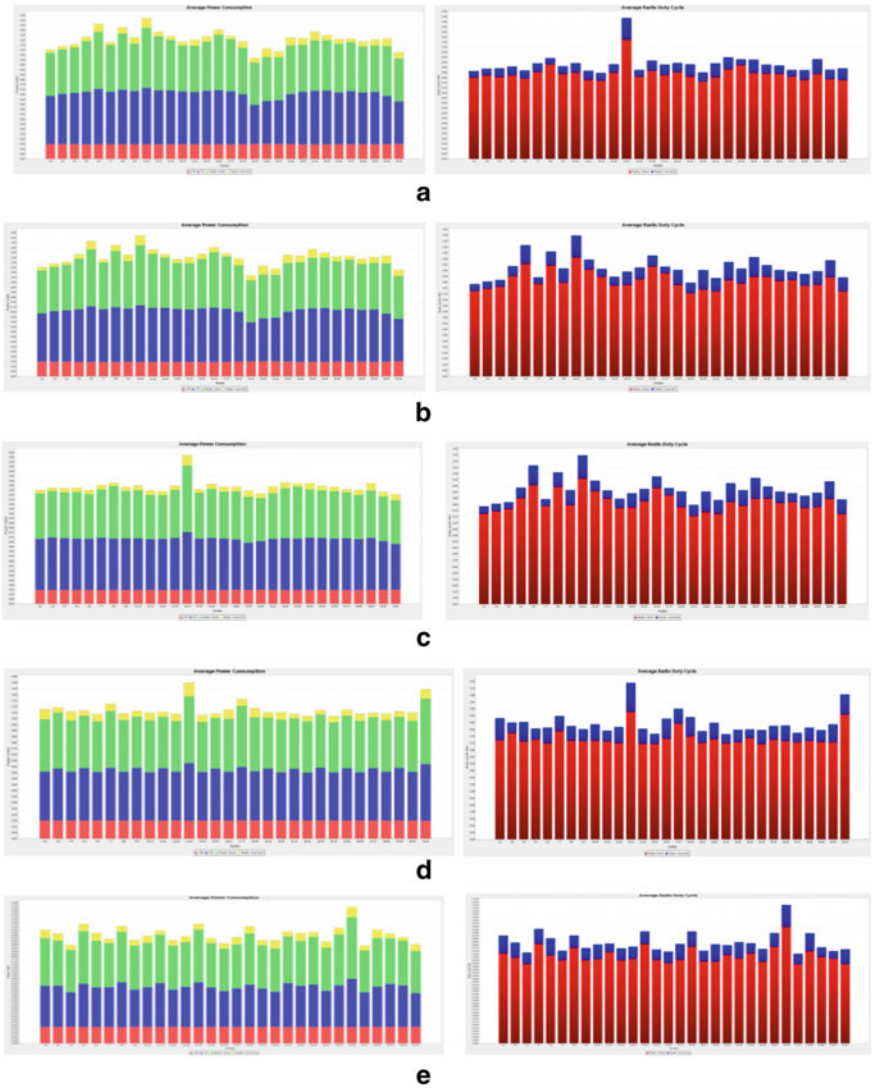


Fig. 3 **a** Average power consumption and average radio duty cycle in low-strain hierarchical network with single sink node and 30 sender nodes. **b** Average power consumption and average radio duty cycle in moderate-strain hierarchical network with single sink node and 30 sender nodes. **c** Average power consumption and average radio duty cycle in high-strain hierarchical network with single sink node and 30 sender nodes. **d** Average power consumption and average radio duty cycle in low-strain circular network with single sink node and 30 sender nodes. **e** Average power consumption and average radio duty cycle in high-strain circular network with single sink node and 30 sender nodes

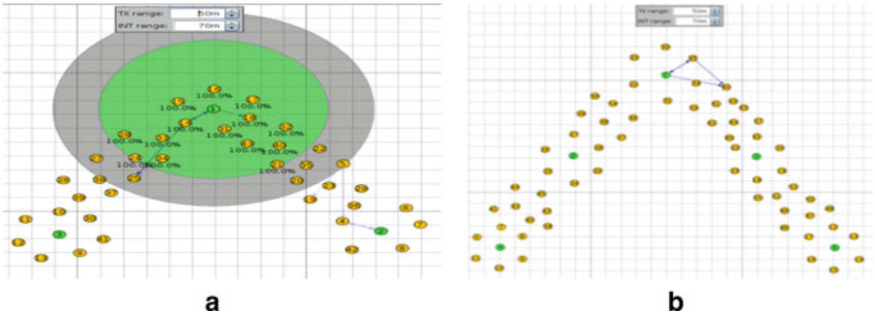


Fig. 4 Hybrid network topology with **a** Multi-sink nodes (3 sinks, 40 sender nodes), **b** multi-sink nodes (5 sinks, 60 sender nodes)

Table 2 Scenario for hybrid network topology

Hybrid network topology	No. of sink nodes	No. of sender nodes
Scenario—1	3	40
Scenario—2	5	60

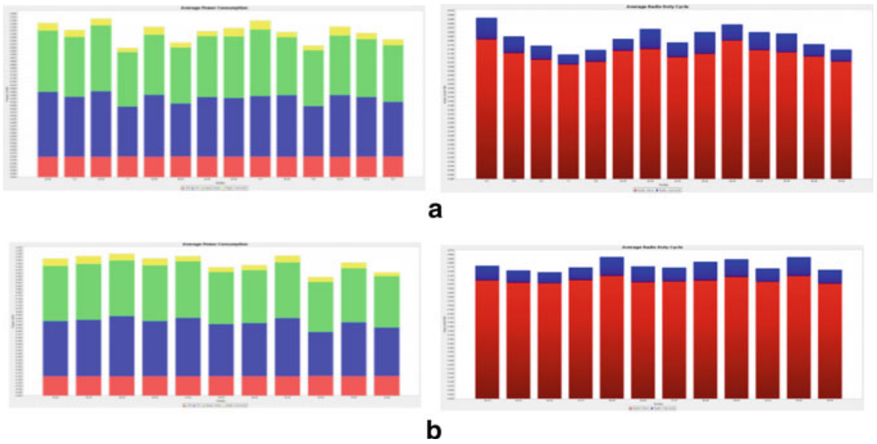


Fig. 5 **a** Average power consumption and average radio duty cycle in hybrid network topology with 3 sinks and 40 sender nodes. **b** Average power consumption and average radio duty cycle in hybrid network topology with 5 sinks and 60 sender nodes

Additionally, the simulation study demonstrates that we have better results with more sinks in the hybrid topology, placed at the right positions.

There are several directions that could be further evaluated, to observe the performance of the network. For example, all the tests done so far focus on only two parameters, which could be extended to focus on Hop Counts, ETX, Drop count, control messages, etc. along with increasing the size of the network.

In addition, future work could look deeper into the effect of the mobility of these nodes as per the need for applications. Another possibility includes the impact of different versions of trickle algorithms, which controls the intervals of packet communication. Finally, optimization of the objective function as per the condition of the network with some learning methods can help it reorganize itself better will be a target in our future research.

References

1. Yushi L, Fei J, Hui Y (2012) Study on application modes of military internet of things (MIOT). In: 2012 IEEE international conference on computer science and automation engineering (CSAE), vol 3. IEEE, pp 630–634
2. Anurag D, Roy S, Bandyopadhyay S (2008) Agro-sense: precision agriculture using sensor based wireless mesh network. In: First ITU-T kaleidoscope academic conference—innovations in NGN: future Network and services
3. Aijaz A, Su H (2015) Abdol-Hamid Aghvami CORPL: a routing protocol for cognitive radio enabled ami networks. IEEE Trans Smart Grid 6(1)
4. Valanarasu MR (2019) Smart and secure Iot and AI integration framework for hospital environment. J ISMAC 1(03):172–179
5. Gnawali O, Fonseca R, Jamieson K, Moss D, Levis P (2009) Collection tree protocol. In: Proceedings of the international conference on embedded networked sensor systems (ACM SenSys), Berkeley, CA, USA
6. Clausen T, Yi J, Herberg U (2017) Lightweight on-demandadhocdistance-vectorrouting-nextgeneration(LOADng): protocol, extension, and applicability, pp 125–140
7. Kumar N, Motia S, Jain AK (2018) Performance analysis of routing protocol for low power and lossy links (RPL) in military hierarchical networks, IAC3T
8. Gaddour O, Koubâa A (2012) RPL in a nutshell: a survey. Comput Netw 56(14):3163–3178
9. Clausen T, Herberg U, Philipp M (2011) A critical evaluation of the IPv6 routing protocol for low power and lossy networks (RPL). In: The IEEE 7th international conference on wireless and mobile computing, networking and communications (WiMob), Wuhan, 2011, pp 365–372
10. Afonso O, VazÃ£o T (2016) Low-power and lossy networks under mobility: a survey. Comput Netw 107(2):339–352
11. Kim HS, Ko J, Culler DE, Paek J (2017) Challenging the IPv6 routing protocol for low-power and lossy networks (RPL): a survey. IEEE Commun Surv Tutor 19(4):2502–2525
12. Liu X, Sheng Z, Yin C, Ali F, Roggen D (2017) Performance analysis of routing protocol for low power and lossy networks (RPL) in large scale networks. IEEE Internet Things J 4(6):2172–2185

A Quick Survey of Security and Privacy Issues in Cloud and a Proposed Data-Centric Security Model for Data Security

Abraham Ekow Dadzie and Shri Kant

Abstract Cloud is being utilized by the majority of Internet users. Businesses storing their critical information in the cloud have increased over the ages due to simplistic and attractive features the cloud possesses. In spite of cloud utilization, there are concerns raised by users regarding cloud security. Hackers are using this opportunity to steal data stored in the cloud. In this regard, researchers have proposed techniques to provide security and privacy to the cloud. Some researchers focused on securing the cloud server, while others concentrated on securing the data transmitted to the cloud. In this research, a comparative analysis is done on papers that were published between 2013 and 2019. Based on the challenges identified in those techniques, a well-secured scheme is proposed to provide extensive security to data before it is transmitted to the cloud. The proposed technique is Data-Centric Security (DCS), and it utilizes Attribute-Based Encryption (ABE) as its framework. This scheme is at the implementation stage, and we believe it will come to address the security and privacy flaws observed.

Keywords Attribute-based encryption · Data-centric security · Cloud computing

1 Introduction

According to National Institute of Standards and Technology (NIST), cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (networks, servers, storage, applications) and services that can be rapidly provisioned and released with minimal management effort or services' provider interactions. Due to the simplistic,

A. E. Dadzie (✉)
Department of Computer Science and Engineering, Sharda University, Greater Noida, India
e-mail: 2018009150.abraham@pg.sharda.ac.in

S. Kant
Research and Technology Development Center/Department of Computer Science and Engineering, Sharda University, Greater Noida, India
e-mail: shrikant.ojha@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_41

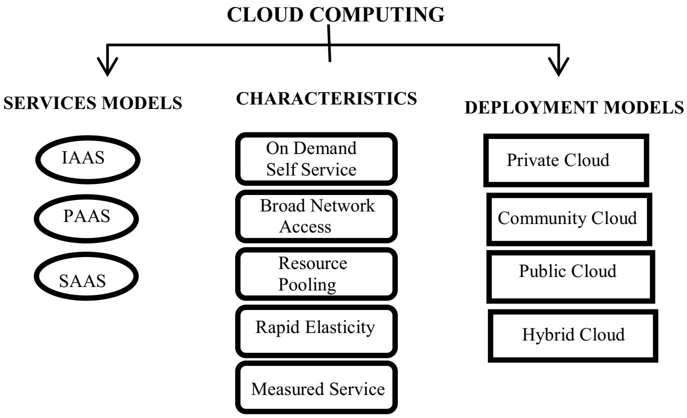


Fig. 1 Background of cloud computing

affordable, and flexible nature of the cloud, its popularity is widespread and various institutions and businesses are moving in that direction [1]. Cloud-based applications are predominantly increasing due to easy access, cheap, efficiency, and availability of computing resources [1, 2]. Cloud computing presents essential models (Deployment and Service) as well as its essential features as indicated in Fig. 1. The service models consist of Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS), while the Deployment Models include Private, Public, Hybrid, and Community [3]. Notwithstanding these numerous benefits that cloud computing offers, it also surrounds itself with environment for cybercrime attacks [4, 5].

This research delves into various techniques that have been proposed to provide security and privacy to the cloud. Comparative analysis is done on these techniques by considering features like domain of protection (cloud server or data), the three security aspects (Confidentiality, Integrity, and Availability), the model the technique works with (Infrastructure, Platform, and Software), the challenges of the techniques, etc. The motivation of this research is to educate incoming researchers about the techniques that have already been proposed and the trend to follow as well as alerting cloud service providers on a particular technique to adopt in terms of implementation on their cloud server.

1.1 Characteristics/Features of Cloud Computing

On-demand self-service. This feature is often considered as utility computing and is compared to the use of light and water. Clients have the privilege of subscribing to a service and can activate the service whenever needed and pay per use. This flexibility access gives full control to clients over their usage and spending [6].

Broad Network Access. Users can access services provided by the cloud across the globe. What cloud users need is a device compatible with Internet connectivity. Cloud providers have multiple servers located across the world which makes this feature possible.

Resource Pooling. Locations do not matter in cloud service. You can be anywhere around the globe and still receive the same and equal services provision provided there is Internet connectivity. This is because the same and equal computing resources and storage can be assigned or distributed as and when needed by the user.

Rapid elasticity. Cloud has the tendency of shrinking and expanding without negative effects on its user's information. In case of heavy load, you can request for expansion which is possible without hindering the progress of your business. The same process can be adopted to reduce cloud usage.

Measured service. Cloud users are billed per service used as it is done in the usage of electricity and water. Therefore, many cloud providers utilize pay as you go model. This model ensures clients are billed with the right amount per service used to avoid any cheat.

2 Review of Proposed Techniques from 2013 to 2019

Research analysis indicates that good amount of researches have been conducted with regard to cloud security and privacy. Several scholars have proposed a variety of techniques to mitigate the security challenges identified in cloud security. This section presents an analysis of the techniques that have been proposed by researchers in their publication.

In 2013, Giweli et al. [7] proposed a solution based on data-centric security approach. The goal of the approach is to provide data security, hence ensuring data are self-protecting, self-describing, and self-defending during their lifecycle in the cloud environments. The approach gives full responsibility to the data owner to set and manage security and privacy relating to data. This scheme is centered on Chinese Remainder Theorem (CRT), and it uses hybrid encryption (asymmetric and symmetric). This technique focuses on providing security to the data but not the cloud.

In 2014, Saravana et al. [8] focused on fusing attribute-based encryption with digital signature, hash functions, and asymmetric encryption. Their paper emphasizes on how the proposed technique is convenient for applications which require high-level security, hence minimizing access time and becoming cost-effective. Their technique when implemented might work efficiently but the system might be robust and complex since it involves many computations, thereby slowing down the system and maximizing computing power.

In 2015, Yang [9] proposed an attribute-based searchable encryption with synonym keyword search function (SK-ABSE). This new scheme activates the synonym keyword search and purposely used to provide assistance to fine-grained access control. This technique provides flexible search authorization encryption scheme since it focuses on multiple users and multiple sender application scenarios.

In 2015, Sengupta and Chinnasamy [10] proposed Hybrid DESCAST encryption algorithm to provide cloud security. Their technique addresses the challenges that emanated from DES to CAST algorithm and then combine their strength to form a hybrid DESCAST. This technique focuses on providing data security in cloud server in addition to the data which is being transferred from client to the cloud. Their proposed technique according to the experimental analysis performed well with 3G and 4G but had a challenge with 5G. Further improvement is required on the technique to perform well with big data.

In 2016, Sarojini et al. [11] proposed a technique called Enhanced Mutual Trusted Access Control Algorithm (EMTACA). The technique offers a mutual trust for both cloud providers and cloud users to deal with issues relating to cloud security. The technique integrates three key elements (trust, guarantee, and reputation) as the basis for cloud services among clients.

In 2016, Songyang and Yong [12] proposed an identity-based signature with escrow free and ID protection for cloud computing. This technique came as a solution to the key escrow problem. The brain surrounding their technique is to make use of a trusted third party to bind the partial secret and the identity of the user together; this will prevent the malicious public-key generator from mimicking the honest user's identity.

In 2016, Yuh-Min et al. [13] presented a technique in their paper known as Identity-Based Encryption with Cloud Revocation Authority. The goal of their paper is to provide solutions to the performance of IBE which had been a challenge. With this technique, all the users' secret key to the system is kept by the cloud revocation authority. The demonstration as seen in their paper suggested that the technique is semantically secure in connection with bilinear Diffie–Hellman assumption.

In 2016, Victor and Muthu [14] developed a framework referred to as Cloud Computing Adoption Framework (CCAF). This scheme was developed purposely for providing data security in the cloud. The focus of the paper was to create a system that will provide security to huge data stored in data centers. This system will also prevent malicious attacks from datacenters by providing real-time protection and if necessary also quarantine those viruses. Their framework was simulated using Business Process Modeling Notation.

In 2017, Lalitha et al. [15] proposed a technique that ensures data is stored in a server in encrypted form. In their approach, only the administrator is mandated to decrypt the data or file. The administrator also has the prerogative to block unauthorized users from getting access to the file. Each user is assigned an Internet protocol address which is attached to a particular resource; this helps the administrator in monitoring and restricting unauthorized access.

In 2017, Bodrul et al. [1] proposed a Reliable Resource Allocation Approach for cloud computing. The essence of this approach was to provide reliability in terms of allocating resources to users while minimizing cost. The focus of their paper was to address the reliability feature which was not really paid attention to in the previous papers.

In 2018, Roman et al. [16] proposed Dynamic Node-RED Based approach as an aid in developing and performing operational activities of gateways. With this approach, offloading is implemented directly as part of the IoT rapid prototyping process embedded in the software stack, based on the Node-RED.

In 2019, Yujiao et al. [17] presented a new Attribute-Based Encryption (ABE) scheme. Their scheme secures user's privacy during key issuing. This technique segregates the functionality of key generating and attribute auditing to ensure the key generation center (KGC) is not aware of user's attribute, thereby making it cumbersome for Attribute Auditing Center (AAC) to get secret key of the user. The scheme represents enterprise attributes by the procurement plans owned by the enterprise, qualifications as well as patent.

In 2019, Mohamed et al. [18] proposed a meta-heuristic placement algorithm which utilizes Ant Colony Optimization (ACO-BF). Their scheme uses Max Fit and Best Fit based on a fitness function which concurrently assesses the waste resource of both physical machines and virtual machines. Their experimental analysis indicates that the technical performance is very good as compared to Max-fit and Best-Fit heuristics, and again there is an increased improvement in terms of resource utilization of the physical machines and virtual machines. In contrast, their scheme focused on limited resources (few numbers of memory size and CPU cores) at the expense of considering processing cores, memory, data transfers, etc.

In 2019, El-Moursy et al. [19] proposed an algorithm for cloud security known as Multi-Dimensional Regression Host Utilization Algorithm. Their focus was to increase the performance of CPU, memory, and BW utilization and as well reduce the energy consumption of the cloud server. The analysis of their simulation indicates that there is improved performance with regard to service-level agreement violation and energy consumption.

2.1 Comparative Analysis of the Proposed Techniques

This is the section where the various techniques that have been proposed are put together for extensive analysis. In Table 1, the researcher made a comparative analysis in a tabular form based on the information retrieved from published papers between the years of 2013 and 2019. These papers were obtained from a recognized journal. Based on information gotten from the papers read, a detailed distinction was drawn between the techniques that have been proposed and used so far based on what was read by the researcher. In Table 1, the letters C, I, A in the results column indicates Confidentiality, Integrity, and Availability. Also I, P, S at the Service Model's column signifies Infrastructure as a Service, Platform as a Service, and Software as a Service.

Table 1 Comparative analysis

S/n	Author	Title	Technique(s)	Results			Service models			Implemented/not	Challenges/future work
				C	I	A	I	P	S		
1	Giweli [7]	Enhancing data privacy and access anonymity in cloud computing	Data-centric security (DCS) approach The approach is based on Chinese Remainder theorem (CRT)	✓				✓		Implemented	The approach does not support smartphones and tablets Same secret key is used in encrypting keywords Incapable to search encrypted data securely
2	Saravana et al. [8]	Enhanced Attribute-Based encryption for cloud computing	Attribute-Based Encryption using hash function, digital signature, and asymmetric encryption	✓	✓	✓	✓		✓	Proposed (not implemented) Proven mathematically	Encryption and decryption involve multiple and complex steps Too much time overhead
3	Yang [9]	Attribute-based data retrieval with semantic keyword search for e-health cloud	Attribute-based searchable encryption with synonym keyword search function (SK-ABSE)	✓	✓	✓	✓			Not implemented	Not favorable for cloud users who are not I.T. inclined since they need to generate trapdoor for keyword search and decrypt.
4	Sengupta and Chinnasam [10]	Contriving hybrid DESCAST algorithm for cloud security	Hybrid DESCAST	✓	✓	✓	✓			Implemented	Performance is slow with big data Cannot be applied on above 1 MB data

(continued)

Table 1 (continued)

S/n	Author	Title	Technique(s)	Results			Service models			Implemented/not	Challenges/future work
				C	I	A	I	P	S		
5	Sarojini et al. [11]	Trusted and reputed services using enhanced mutual trusted and reputed access control algorithm in cloud	Enhanced mutual trusted Access control algorithm (EMTACA)	✓		✓	✓			Simulated	Authentication and authorization were not taken care of. System might generate more uncertainty
6	Yuh-Min et al. [13]	Identity-based encryption with cloud revocation authority and its applications	Identity-based encryption	✓	✓	✓	✓		✓	Experimented	Users that miss their time will find it difficult to decrypt their data since key would be revoked.
7	Victor and Muthu [14]	Toward achieving data security with the cloud computing adoption framework	Cloud computing adoption framework (CCAF)	✓	✓		✓	✓		Simulated	Performance problem with many users and attacks
8	Songyang and Yong [12]	Efficient verification of data possession in cloud computing	Identity-based signature with escrow free and ID protection	✓	✓	✓	✓		✓	Proven mathematically but not implemented	The system leads to slightly low performance

(continued)

Table 1 (continued)

S/n	Author	Title	Technique(s)	Results			Service models			Implemented/not	Challenges/future work
				C	I	A	I	P	S		
9	Lalitha et al. [15]	Data security in cloud	Encryption and decryption	✓	✓	✓	✓			Tested but not deployed	Performance problem when number of users are more Workload on admin since he performs both encryption and decryption
10	Bodrul et al. [1]	A reliability-based resource allocation approach for cloud computing	Reliable resource allocation approach		✓	✓	✓			Simulated	No evaluation model was created to calculate the overall reliability
11	Roman et al. [16]	Offloading execution from edge to cloud: a dynamic node-RED based approach	Node-RED	✓	✓	✓	✓			Implemented	Service-level agreement guarantees were not considered

(continued)

Table 1 (continued)

S/n	Author	Title	Technique(s)	Results			Service models			Implemented/not	Challenges/future work
				C	I	A	I	P	S		
12	Yujiao et al. [17]	Efficient attribute-based encryption with privacy-preserving key generation and its application in industrial cloud	Efficient attribute-based encryption	✓	✓	✓	✓	✓		Proven mathematically	Performance of the system might be affected since it involves a lot of parameters
13	Mohamed et al. [18]	A placement architecture for a container as a service (CaaS) in a cloud environment	Ant colony optimization based on Best Fit (ACO-BF)	✓			✓			Experimented	Limited resources (memory size and CPU cores) were considered
14	Ahmed et al. [4]	Multi-dimensional regression host utilization algorithm(MDRHU) for host overload detection in cloud computing	Multi-dimensional regression host utilization algorithm	✓			✓			Simulated	The energy consumption needs to be reduced further

2.2 Research Questions

1. How do those proposed techniques differentiate from each other in terms of strength and weaknesses?
2. Which cryptographic technique can be implemented to provide extensive security and privacy to data in cloud?

2.3 Analysis

It is observed from the survey that most of the papers concentrated on providing security to the cloud server [1, 10, 14, 18, 19]. Few that focused on securing data before outsourcing used the same private key for each user, techniques had a challenge with performance or required the service of a third party in generating the secret keys [7, 8, 12, 17]. With regard to the challenges observed from the comparative analysis of the techniques studied. The researcher, therefore, proposes a cryptographic scheme to curtail those challenges. The proposed scheme is still under implementation and we hope to achieve better results when implemented and tested. The scheme is proposed to provide privacy and security to data before it is transmitted to the cloud. The scheme will concentrate on both data at rest and data in transmission. All the three goals of security (Confidentiality, Integrity, and Availability) would be sorted out when the scheme is finally implemented and tested [20].

3 Proposed Model

Data-Centric Security (DCS) approach is the proposed technique to address the security and privacy concerns in the cloud. DCS is a useful approach for safeguarding sensitive data from misuse or theft. In this approach, protection is applied to the data itself independent of where it is located. This approach gives full responsibility to data owner. A model architecture of this scheme is illustrated in Fig. 2.

Attribute-Based Encryption (ABE) is a framework our scheme is going to use to utilize the DCS concepts for data privacy and security. ABE is a recent approach that reconsiders the concept of public-key cryptography. Encrypting data using ABE is very secure and easy as compared to other algorithms since it offers flexible access control by employing controlled access structures relating to Master key, Private Key, and Ciphertext, respectively [8]. In Fig. 2, we have Data Owner, User, and the cloud. All the required security parameters are attached to the encrypted data to create a secure file container which is referred to as DCS file. Users who satisfy the access policy are authorized to access the DCS file based on the embedded policies that are set and managed exclusively by the data owner. Cloud users are the first to initiate communication. So the cloud user will request a file and an alert would be communicated to data owner. Data owner then checks details of the cloud user

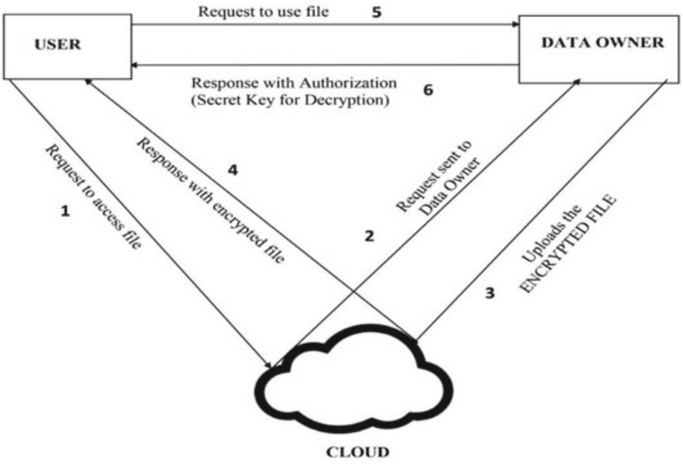

USER
DATA OWNER
Request to use file 5
Response with Authorization (Secret Key for Decryption) 6
Request to access file 1
Response with encrypted file 4
Request sent to Data Owner 2
Uploads the ENCRYPTED FILE 3
CLOUD

Fig. 2 Model architecture

who requested the file for confirmation. The file is then encrypted in addition to the attributes of the user and sent to the cloud user. Upon authorization, the cloud user will then be able to follow certain sequence or patterns to decrypt the file. File is decrypted only on users' machine.

This scheme is proposed to be used in Academic Institutions to enable students who have completed their course and return to their country or different states to request for their credentials with ease without the need to wait after completion or return to the school for their academic credentials. The system might also profit school administration and staffs of the schools in terms of exchange of confidential files without the need to walk to and fro in offices. With this scheme, cloud providers' responsibility is to store encrypted data since their service is not required in both encryption and decryption. Files are protected both at rest and in transmission.

3.1 Implementation

Hardware for Implementation. System: Pentium Dual Core, Ram: 1 GB, Hard Disk: 100 GB, Monitor: 15″ LCD or LED

Software for Implementation. Operating System: Windows (7, 8, 10), Coding Language: Python/Django, Database: Sqlite.

To see the actual working of the scheme, a system is developed as a basis for experimentation. This system has platforms that both users and data owners follow to enroll their details, request, and upload encrypted files as well. These platforms are web based and connected to the cloud. In the proposed system, users (students, staffs, etc.) will have to enroll their details into the system before they can be authenticated to request a file (transcript, certificate, etc.). Data owners upon receiving request would then encrypt the requested file with the attributes of the user and then outsource to the cloud. The authorized user on receiving an email with the sequence of the secret key then decrypts the file on his machine. Meaning, encryption, and decryption are all done on authorized user's machine. Cloud providers' responsibilities are only to store the outsourced file.

4 Discussion and Future Work

ABE is the framework to be utilized in our proposed scheme. Our proposed scheme will be cumbersome for eavesdroppers to attack since its security requirements and parameters are the sole prerogatives of the data owner. Data can only be decrypted by the authorized user since user's attributes need to satisfy the access policy. The work done so far was tested by using 20 students as a synthetic dataset. The students were permitted to register and request a file from the system concurrently. The system worked perfectly by receiving the requests and replied to the students through their personal mail as scheduled. This indicates that our scheme supports multiple users. The implementation is already in final stages to be tested on the cloud environment and evaluates its performance.

5 Conclusion

In this research, we conducted a quick comparative analysis of techniques that have been proposed to provide security and privacy to the cloud. It was realized from the analysis that while some techniques concentrated on offering security at the cloud providers' end other schemes also focused on securing data that are been transmitted to the cloud. Most of the techniques are yet to be deployed in real-life environment. From Table 1, we see the techniques that are prototyped and those that have pass-through experimental stage. Few challenges that were drawn from the analysis include the following: some techniques allowed a third party to generate and keep secret keys; other techniques also used same secret keys for all transactions; other techniques utilize much computational power while few also focused on providing security only to the cloud server without considering the data. To overcome these challenges, we have proposed a scheme which is under implementation stage that is based on Data-Centric Security and utilizes Attribute-Based Encryption as its framework. We hope and believe that our proposed scheme would address the above challenges after implementation and testing.

Competing Interest The authors declare that they have no conflict of interest.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Bodrul ABM, Mohammed Z, Anwar H (2017) A reliability-based resource allocation approach for cloud computing. In: 2017 IEEE 7th international symposium on cloud and service computing, pp 147, 249–252
2. Al-Said A, Andras P (2019) Scalability analysis comparisons of cloud-based software services. J Cloud Comput: Adv Syst Appl 8(10):1–17
3. Investopedia. www.investopedia.com/terms/c/cloud-computing.asp, Accessed 20 Nov 2019
4. Ahmed A et.al (2019) Experts reviews of a cloud forensic readiness framework for organizations. J Cloud Comput: Adv, Syst Appl 8(11):1–14
5. Gururaj R, Mohsin I, Farrukh AK (2017) A comprehensive survey on security in cloud computing. In: The 3rd international workshop on cyber security and digital investigation (CSDI 2017), vol 110, pp 465–472
6. Mahesh USC, Ambika VP (2015) Security and privacy in cloud computing: a survey. In: Proceedings of the 3rd international conference on frontiers of ıntelligent computing (FICTA), vol 2, pp 1–2
7. Giweli N, Shahrestani S, Cheung H (2013) Enhancing data privacy and access anonymity in cloud computing. Commun IBIMA 1–10
8. Saravana K, Raiva L, Balamurugan B (2014) Enhanced attribute based encryption for cloud computing. In: International conference on information and communication technologies (ICICT 2014), vol 2015, pp 689–696
9. Yang Y (2015) Attribute-based data retrieval with semantic keyword search for e-health cloud. J Cloud Comput: Adv Syst Appl 1–6
10. Hussein MK et al (2019) A placement architecture for a container as a service (CaaS) in a cloud environment. J Cloud Comput: Adv Syst Appl 8(7):1–15
11. Sarojini G, Vijayakumar A, Selvamani K (2016) Trusted and reputed services using enhanced mutual trusted and reputed access control algorithm in cloud. In: 2nd international conference on intelligent computing, communication and convergence (ICCC-2016), vol 92, pp 506–512
12. Songyang W, Yong Z (2016) Efficient verification of data possession in cloud computing. In: Conference of computational interdisciplinary sciences (CCIS 2016), pp 424–428
13. Yuh-Min T et.al (2016) Identity-based encryption with cloud revocation authority and its applications. IEEE Trans Cloud Comput 1–14
14. Victor C, Muthu R (2016) Towards achieving data security with the cloud computing adoption framework. IEEE Trans Serv Comput 9:138–151
15. Lalitha VP et.al (2017) Data security in cloud. In: International conference on energy, communications, data analytics and soft computing (ICECDS-2017), pp 3604–3608
16. Sosa R et al (2018) Offloading execution from edge to cloud: a dynamic node-RED based approach. In: 2018 IEEE international conference on cloud computing technology and science (CloudCom), pp 1–4
17. Sridhar S, Smys S (2016) A survey on cloud security issues and challenges with possible measures. In: International conference on inventive research in engineering and technology, vol 4
18. Yujiao S et al (2019) Efficient attribute-based encryption with privacy-preserving key generation and its application in industrial cloud. Hindawi Secur Commun Netw pp 1–10
19. El-Moursy AA et al (2019) Multi-dimensional regression host utilization algorithm (MDRHU) for host overload detection in cloud computing in: journal of cloud computing: advances. Syst Appl 8(8):1–17
20. Sengupta N, Chinnasamy R (2019) Contriving hybrid DESCAST algorithm for cloud security. In: Eleventh international multi-conference on information processing (IMCIP-2015). Procedia Comput Sci 54:47–56

Homo Sapiens Diabetes Mellitus Detection and Classification

Anu Agarwal, Anjay Sahoo, Indrashis Das, Siddharth S. Rautaray, and Manjusha Pandey

Abstract Diabetes mellitus can be defined as a set of deficiency disorders which is caused due to under-secretion of insulin. In other words, it results in very high blood sugar levels. Diabetes mellitus influences and is influenced by various factors. Diabetes mellitus, if remains unidentified or untreated can lead to lethal disorders like a cardiovascular disease such as heart attack, narrowing of arteries, nerve damage, kidney damage, skin conditions, depression, and many such complications. Statistics suggest that human beings are getting affected by this disease at an alarming rate. Yet, it remains unidentified and hence untreated in most cases. Hence, machine learning is introduced in the field of biomedical sciences such that these disorders can be treated at a larger scale without conducting pathological tests. The below paper solely focuses on predicting over a set of features for every human that if the person has a tendency of high blood sugar or diabetes mellitus or not. Building the classifier includes libraries like Python, Numpy, Pandas, Matplotlib, Seaborn, Scikit Learn, and Scipy.

Keywords Python · Pandas · Matplotlib · Seaborn · Scikit learn · Scipy · Machine learning · Diabetes · Data science · Django

A. Agarwal (✉) · A. Sahoo (✉) · I. Das · S. S. Rautaray · M. Pandey
School of Computer Engineering, Kalinga Institute of Industrial Technology (Deemed to Be University), Bhubaneswar, India
e-mail: anuagarwal308@gmail.com

A. Sahoo
e-mail: 1705208@kiit.ac.in

I. Das
e-mail: indrashisdas98@gmail.com

S. S. Rautaray
e-mail: siddharthfcs@kiit.ac.in

M. Pandey
e-mail: manjushafcs@kiit.ac.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_42

1 Introduction

Diabetes mellitus is a chronic disorder caused by a high blood glucose level. Mostly, diabetes is hereditary in nature. Failure of the pancreas to produce the required amount of insulin or inability of the body to use the insulin produced is the major pathologic cause. Insulin is a hormone secreted by the β cells of the pancreas used to derive energy for the body. Diabetes mellitus is of two types—Type 1 and Type 2. A person is diagnosed with Type 1 Diabetes Mellitus (T1DM) if the insulin-secreting β cells are damaged and are diagnosed with Type 2 Diabetes Mellitus (T2DM) if the produced insulin is not used accordingly by the body. The major symptoms of diabetes mellitus include increased thirst, increased hunger, blurred vision, slow healing, etc.

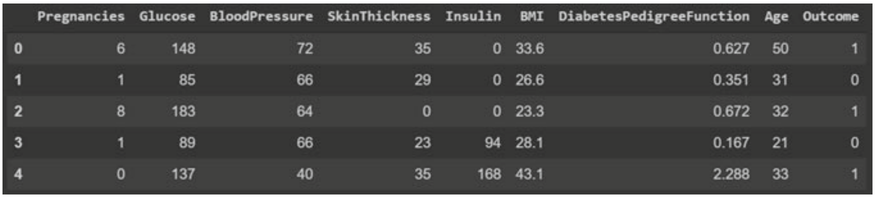
It is estimated that 415 million people are living with diabetes in the world. So it can be said that 1 in 11 of the world's adult population is diabetic. Moreover, this figure is likely to rise to 642 million by the year 2040. Type 2 diabetes is the most common type in adults. IDF (International Diabetes Federation) reports that more and more people are getting diagnosed with Type 2 Diabetes in the world. However, Type 1 diabetes is common in kids. In the year 2015, more than 542,000 children were living with Type 1 diabetes. According to the reports of the IDF(International Diabetes Federation), the top 3 countries with the highest amount of diabetic patients are China, India, and Brazil. Yet, 46% of the world population is still undiagnosed.

To reduce the death rate and prevent other fatal diseases caused due to diabetes mellitus, we need to identify the group of common masses prone to get affected by the disease. To identify that group of people mostly to get affected by this disease, we need to advance technology. Hence, data mining is the best-suited method for us. Data mining is looking for concealed, rational, and probably important patterns in huge data. It is more likely a process of coming across relationships between the data that was not predicted just by looking at the data. As the number of diabetics is huge and is increasing exponentially, it is important to develop that can predict it so that the hassle caused due to various medical tests can be avoided.

2 Basic Concepts/Technology Used

On reading the data, it could be seen that it consists of nine parameters including the target variables that were used to determine or rather predict if the patient has diabetes mellitus or not. The basic parameters included glucose level in the blood on 2 h resting, insulin level on 2 h resting, skin thickness of the patient, number of pregnancies, age, body mass index, the pedigree function of diabetes, and so on.

The data had almost 770 data points for which the data was separated into training, validation, and testing datasets. The split was done on the basis of the golden rule where a major part of the data resides under training set so that maximum patterns



	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

Fig. 1 Head of the diabetes mellitus dataset

can be captured out of it. Hence, around 60% of the data was kept as training data and 20% was kept for validating with the rest 20% to be as a testing set.

On reading the data, a few that were checked included the basic head of the data so that the columns or rather the parameters could be understood to determine the output label. Hence, below the head of the data can be seen (Fig. 1).

3 Study of Similar Projects or Technology/Literature Review

See Table 1.

4 Proposed Model/Tool

As seen and observed for insights and trends in the dataset, it could be seen that there are various deciding factors or rather parameters that to the decision-making of a patient having diabetes mellitus or not. Especially after doing the required Exploratory Data Analysis, a lot many trends could be captured that matched the logic of having diabetes mellitus. Firstly, it could be seen from the data that the higher the number of pregnancies, the more was the probability of having diabetes as shown in the figure below. Accordingly, it was also observed that there was a linear relationship between plasma glucose concentration and the probability of diabetic patients under particular discrete values. Similar was the case of diastolic blood pressure where there was a similar linear relationship with the probability of being diabetic in nature. Similarly in the case of skin thickness, it was seen that the higher the skin thickness, the more was the probability of being diabetic in nature. Also, body mass index and diabetes pedigree function were continuous variables where they were not that gaussian, rather right-skewed in nature. Hence for that, required transformations were done which included a square-root transformation and an exponential for Body Mass Index and Pedigree Function for Diabetes, respectively. Also, as a factor of age always plays in the case of diabetes, it could also be seen that most of the patients were diagnosed with diabetes mellitus at a higher age or as age of a person increases,

Table 1 Literature review

Sl. no.	Author year	Title	Proposed model	Review
01	Hasan Abbas et al. [1–3]	Predicting diabetes in healthy population through machine learning	The provided study shows the future development of type2 diabetes with the use of machine learning. The study shows a validation accuracy of 84.1%, which can be used in future for identifying greater risk of type 2I diabetes	Mellitus such as time classification could be employed to find out the performance. This technique could be used in pediatric emergencies
02	Alić et al. [4–6]	Machine learning techniques for classification of diabetes and cardiovascular diseases	The concerned paper classifies diabetes and cardiovascular diseases (CVD) using machine learning techniques with artificial neural networks (to be specific multilayer feedforward neural network) and Bayesian networks (Naive Bayesian Network). To be more calculative the mean accuracy shows the better results with ANN, indicating high possibility to get accurate results in diabetes	Artificial neural networks have higher possibility of yielding more accurate results in diabetes prediction

(continued)

Table 1 (continued)

Sl. no.	Author year	Title	Proposed model	Review
03	Lee et al. [7–9]	Identification of Type 2 diabetes risk factors using phenotypes consisting of anthropometry and triglycerides based on machine learning	The main agenda of the study was to access the association between the HW phenotypes consisting of combinations of individual anthropometric measurements and TG level. To be more appropriate, two machine learning algorithms were used to predict the power of various phenotypes, and the phenotypes were having more predictive power in women than men	This study was not able to establish a cause–effect relationship. It can be used to find out the best phenotype of Diabetes. Since the study only includes Korean women and men the finding cannot be applied as population other than that of Korea
04	Faruque et al. [10–12]	Performance analysis of machine learning techniques to predict diabetes mellitus	The concerned study was done for the early prediction of disease called diabetes mellitus caused by a group of metabolic disorder, using machine learning techniques. This technique gives us an efficient way to extract knowledge from the medical datasets acquired from the diabetic patients	They have used four different kinds of machine learning algorithm, and they are Naive Bayes, K-nearest neighbor, support vector machine, and C4.5 decision tree. The best working algorithm is found to be C4.5 decision tree
05	Sowjanya et al. [13–15]	MobDBTest: It uses mobile to predict diabetes risk by employing machine learning based system into the mobile	With the change in the lifestyle in our life are moving the degree of diabetes which is proven to be linked with number of other complications. This study is related to a mobile application based solution which uses novel machine learning techniques with 4 machine learning algorithms out of which Decision Tree is the one	Among the four algorithms employed it is found that J48 algorithm yields the best results

(continued)

Table 1 (continued)

Sl. no.	Author year	Title	Proposed model	Review
06	Benbelkacem et al. [16–18]	Random forests for diabetes diagnosis	From last few studies, it's proven that random forest is one of the best researches to get decision tree learning. In this study, we have exploited the principle of random forest and at the end of the study it's proven to be more efficient than any other method of machine learning	The performance of random forest is subjected to various tests to find out the best performance by modifying the number of trees. Then random forest function is further compared to other algorithms. This proves that random forest gives the lowest ever role
07	Priyadarshini et al. [19–21]	A novel approach to predict diabetes mellitus using modified extreme learning machine	Modified extreme learning has been used in their study to find whether the concerned patient is having diabetes or not. Backpropagation neural network and modified extreme learning machine are used to address the diabetes prediction problem	Use of feature selection improves the improvement to the model by removing the features which are basically noise on the data. Genetic algorithm can be used as it helps in optimization
08	Morton et al. [22–24]	A comparison of supervised machine learning techniques for predicting short-term in-hospital length of stay among diabetic patients	The need to predict the duration of stay in the hospital of a diabetic person needs to be calculated as it results in many hospitalizations annually. Machine Learning algorithms can be used to do so	They have employed five different machine learning algorithms. The result indicates the SVM+ is the most effective method of all. Amere advanced feature selection work can be done and often algorithms such as artificial neural networks, logistics regressions can also be used

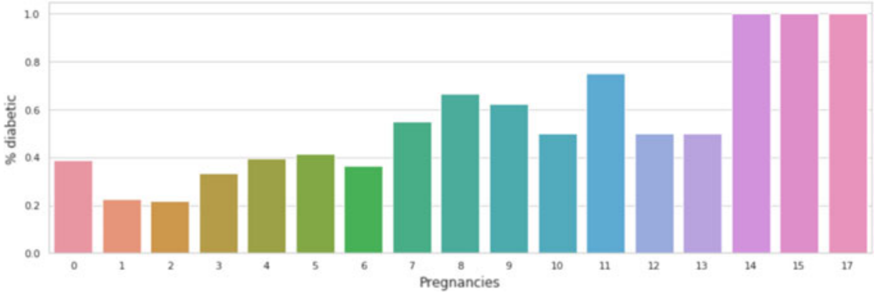


Fig. 2 Pregnancies probability of diabetes

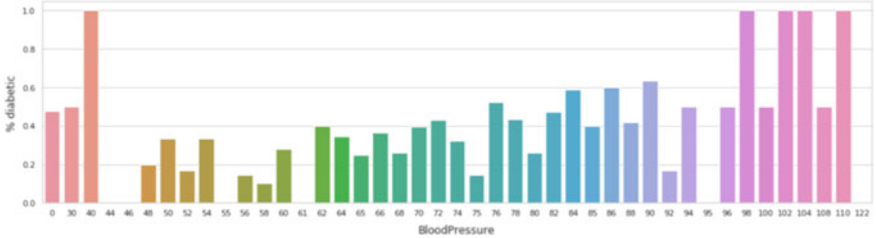


Fig. 3 Diastolic blood pressure probability of diabetes

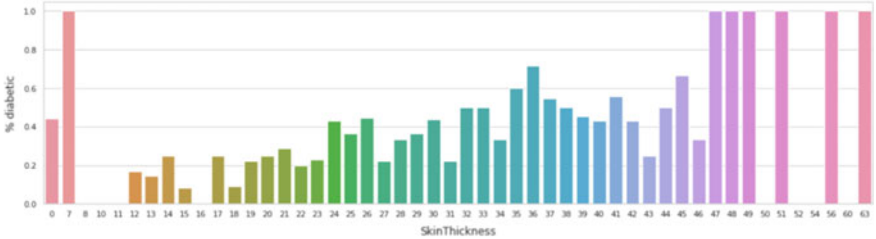


Fig. 4 Skin thickness probability of diabetes

so increases the probability of diabetes. Below are some bar plots for the different trends captured via the visualization techniques (Figs. 2, 3, 4, 5, 6, and 7).

5 Implementation and Results

It is quite clear from the problem that it is a classification supervised problem. Hence, many classification algorithms were applied over the data in order to gain a quality prediction out of it both the training and the testing data. It could be seen that after passing the features like age, blood pressure, 2 h resting insulin level, plasma

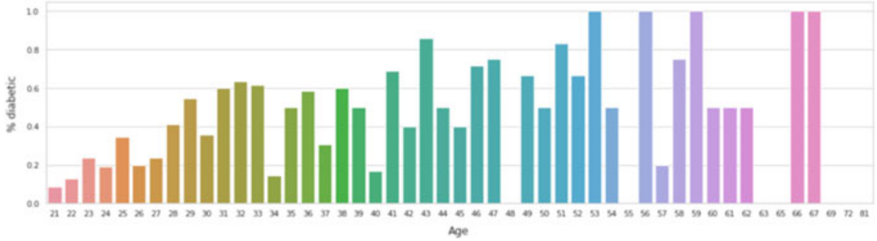


Fig. 5 Age probability of diabetes

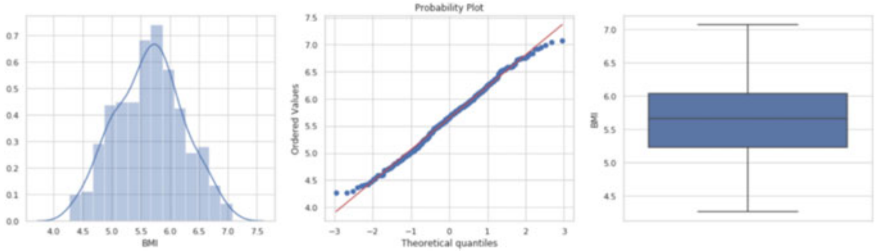


Fig. 6 Distplot, probability plot and box plot of body mass index after the required transformation

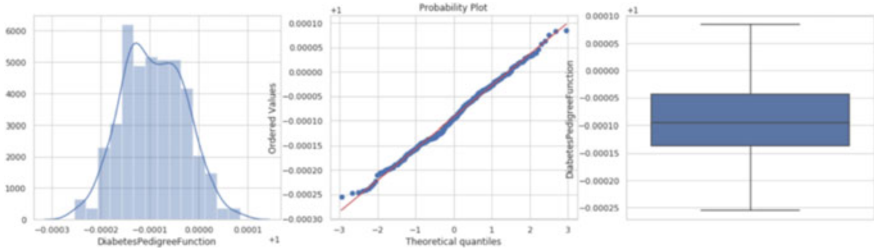
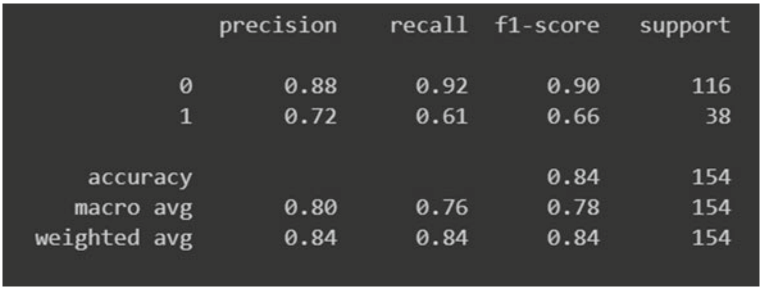


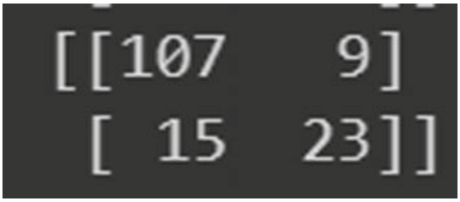
Fig. 7 Distplot, probability plot, and box plot of diabetes pedigree function after the required transformation

glucose level, skin thickness, and so on, the best accuracy given was by random forest classifier where the accuracy was around 85% after doing a hyperparameter tuning on the model. This particular bagging algorithm performed well than the other bootstrapping, bagging, and boosting algorithms that include Decision Tree Classifier, XGBoost Classifier, LightGBM Classifier, AdaBoost Classifier, and so on. The recall for 0 was around 0.92, while that for 1 was around 0.62. Below is the classification report and the confusion matrix for which the metrics are calculated (Figs. 8 and 9).



```
              precision    recall  f1-score   support

           0       0.88      0.92      0.90       116
           1       0.72      0.61      0.66        38

    accuracy                           0.84       154
   macro avg       0.80      0.76      0.78       154
weighted avg       0.84      0.84      0.84       154
```

Fig. 8 Classification report



```
[[107   9]
 [ 15  23]]
```

Fig. 9 Confusion matrix

6 Conclusion

As seen and observed, random forest classifier worked best with an accuracy of around 85–86%. Along with this, it could also be suggested that a few more features be built so that the performance of the model increases.

While such models are being built for diabetes classification, there can be many more such algorithms that can be used to build classifiers for different diseases and disorders. Not only in the field of diabetology, but such algorithms could also be applied anywhere and everywhere like in the field of oncology and cardiology as well.

Competing Interest The authors declare that they have no conflict of interest.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Abbas H, Alic L, Rios M, Abdul-Ghani M, Khalid Qaraqe (2019) Predicting diabetes in healthy population through machine learning. In: 2019 IEEE 32nd international symposium on computer-based medical systems (CBMS). IEEE
2. Mathers CD, Loncar D (2006) Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 3(11):e442

3. Tschritter O, Fritsche A, Shirkavand F, Machicao F, Haring H, Stumvoll M (2003) Assessing the shape of the glucose curve during an oral glucose tolerance test. Diabetes Care 26(4):1026–1033
4. Alić B, Gurbeta L, Badnjević A (2017) Machine learning techniques for classification of diabetes and cardiovascular diseases. In: 2017 6th mediterranean conference on embedded computing (MECO). IEEE
5. Sandhya N, Charanjeet KR (2016) A review on machine learning techniques. Int J Recent Innov Trends Comput Commun 395–399. ISSN 2321-8169
6. Ghaheri A, Shoar S, Naderan M, Hoseini SS (2015) The applications of genetic algorithms in medicine. Oman Med J 30(6):406
7. Lee BJ, Kim JY (2015) Identification of type 2 diabetes risk factors using phenotypes consisting of anthropometry and triglycerides based on machine learning. IEEE J Biomed Health Inform
8. Katzmarzyk PT, Craig CL, Gauvin L (2007) Adiposity physical fitness and incident diabetes: the physical activity longitudinal study. Diabetologia 50(3):538–544
9. Xu Z, Qi X, Dahl AK, Xu W (2013) Waist-to-height ratio is the best indicator for undiagnosed type 2 diabetes. Diabet Med 30(6):e201–e207
10. Faruque MF, Sarker IH (2019) Performance analysis of machine learning techniques to predict diabetes mellitus. In: 2019 international conference on electrical, computer and communication engineering (ECCE). IEEE
11. Platt, JC (1999) 12 fast training of support vector machines using sequential minimal optimization. In: Advances in kernel methods, pp 185–208
12. John GH, Langley P (1995) Estimating continuous distributions in Bayesian classifiers. In: Proceedings of the eleventh conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc.
13. Sowjanya K, Singhal A, Choudhary C (2015) MobDBTest: a machine learning-based system for predicting diabetes risk using mobile devices. In: 2015 IEEE international advance computing conference (IACC). IEEE
14. Kaveeshwar SA, Cornwall J (2014) The current state of diabetes mellitus in India. Aust Med J PMCID: PMC3920109, pp 45–48
15. Georga EI, Protopappas VC, Mougiakakou SG, Fotiadis DI (2013) Short-term versus long-term analysis of diabetes data: application of machine learning and data mining techniques. In: IEEE: 13th international conference on bioinformatics and bioengineering (BIBE)
16. Benbelkacem S, Atmani B (2019) Random forests for diabetes diagnosis. In: 2019 international conference on computer and information sciences (ICCIS). IEEE
17. Settouti N, Daho ME, Lazouni MA, Chikh MA (2013) Random forest in semi supervised learning co-forest. In: International workshop on systems signal processing and their applications, pp 12–15
18. Butwall M, Kumar S (2015) A data mining approach for the diagnosis of diabetes mellitus using random forest classifier. Int J Comput Appl 120:0975–8887
19. Priyadarshini R, Dash N, Mishra, R (2014) A novel approach to predict diabetes mellitus using modified extreme learning machine. In: 2014 international conference on electronics and communication systems (ICECS). IEEE
20. Pradhan M, Sahu RK (2011) Predict the onset of diabetes disease using artificial neural network (ANN). Int J Comput Sci Emerg Technol 2(2). E-ISSN: 2044-6004
21. Siva Prakash J, Rajeshalayam R (2011) Random iterative extreme learning machine for classification of electronic nose data. Int J Wisdom Based Comput 1(3):24–27
22. Morton A, Marzban E, Giannoulis G, Patel A, Aparasu R, Kakadiaris IA (2015) A comparison of supervised machine learning techniques for predicting short-term in- hospital length of stay among diabetic patients. In: 2014 13th international conference on machine learning and applications. IEEE
23. Cornall RJ, Prins J-B, Todd JA, Pressey A, DeLarato NH, Wicker LS, Peterson, LB (1991) Type 1 diabetes in mice is linked to the interleukin-l receptor and lshllty/BCG genes on chromosome 1. Nature 353(6341):262–265
24. Robinson GH, Davis LE, Leifer RP (1996) Prediction of hospital length of stay. Health Serv Res 1(3):287

Learning Platform and Smart Assistant for Students

R. Rashmi, Sharan Rudresh, V. A. Sheetal, and Dexler information Solutions Pvt Limited

Abstract Students and knowledge have always been abstract entities. Earlier, students found it tough to find the source of knowledge and now they find it difficult to gather, curate and utilise the immense amount of knowledge available from various sources. Both the scenarios seem like they are the two faces of a coin. Students face yet another problem, i.e. the problem of tracking and scheduling. More often than not we have seen that with proper guidance and materials to study, people have emerged with flying colours. If we can provide this to the students by using technology, we'll be able to help many lead a stress-free and efficient academic life. In this paper, we have detailed the problems faced by students. We look at some existing platforms which aim to address these problems. Further, we propose a smart learning assistant for curating the study materials and tracking the student's progress.

Keywords Data science · Recommendation system · Machine learning · Neural networks · Education system · Chatbot

1 Introduction

Students today have access to a lot of study resources. Learning has become a tedious process largely due to improper organisation of these resources. We have identified four problems students face while learning:

R. Rashmi (✉) · S. Rudresh · V. A. Sheetal
B.M.S College of Engineering, Bengaluru, India
e-mail: rashmibhle@gmail.com

S. Rudresh
e-mail: rudreshsharan@gmail.com

V. A. Sheetal
e-mail: sheetal.cse@bmsce.ac.in

Dexler information Solutions Pvt Limited
Dexler Information System, Domlur, Bengaluru, India
e-mail: kaustubh.p@gyanamite.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_43

1. Students face trouble while gathering necessary study materials.
   Nowadays, it is quite common to get materials from various sources.
   For example, we have books, ebooks and pdfs which are obtained from the internet. We get notes prepared by our seniors and teachers.
   We could get access to these via shared Google Drive folders, or we could receive them as email attachments, or even from file-sharing mobile apps. Some of these materials might also be redundant. Such a haphazard and scattered state of content makes learning a hassle.
2. Students face trouble in organising their study material.
   Today, a major part of study material used by students is digital—i.e. accessed by computers and mobile phones. The most common way a student organises study material is by folders—mainly a folder for each subject, and maybe folders for each chapter/unit.
   The students generally prefer to use cloud storage services like Google Drive or OneDrive—this allows access from any device and the changes are synced.
   However, merely organising by subjects and chapter names aren't as efficient and convenient to the student, as explained in the following point.
3. Students face trouble in finding what they are looking for.
   When a student thinks of opening a file to study, most of the times, the first thought is not the chapter number—it is the topic of interest. The student will have to look for the right chapter and then look for the topic within the possibly tens of files in that folder. Further, some topics might have relevant content in other chapters, which the student might miss.
4. Students face trouble in tracking their progress.
   Often we tend to misjudge the effort we need for studying. The effort could be the time required to study or the practice questions to solve. We also tend to forget our previous thoughts on what we need to do to learn a topic. Today I might decide to solve five more math questions from previous year's question papers. However, if I do not find and solve those questions immediately and postpone to the next day, I might forget or end up not moving on to the next topic.

2 Related Work

In Muhammad et al. [1], introduction of the learning path in online learning systems is one of the new trending technologies that has been adopted by a few education platforms. The main intent of this adaptation is to enhance the learning experience by optimising the relationship between the learning objective and the course content just through the learning outcomes attained by the minimum time spent on the platform and the various interactions expended during the entire course. A brief insight into the problems faced by the online learning system's adaptation is of cardinal importance which needs to be addressed [2]. Learning path construction is a difficult and tedious task as it contains many tasks like formulating, assessing learning outcomes, defining and redefining students' learning improvement and to improve with the learning resource. This paper aimed at showing the novel and open model approach towards the learning path construction and automatic path generation [3]. Virtual university has proposed a user-specific learning path for the various courses they offer. They presented a system that recommends a learning path to students that is unique for individual users based on user-specific as well as system-specific factors. Since this problem of learning path generation is yet to be proven as an NP-complete or NP-hard problem, efficient methods to tackle this problem are expected to be present. Therefore, a method to tackle this problem of learning path generation which is efficient and not computationally intensive is needed and can be developed [4]. Course offers a feature for users who wish to start a new career, which are provided with a comprehensive guide to the prerequisites and skills required to succeed in the field along with a learning path consisting of the various courses that can help them develop these skills. Learning paths have been developed for various careers including full stack developer, data scientist, marketing specialist, UI/UX designer, machine learning engineer, etc. These learning paths are developed based on information gathered from the catalog of 2000+ top university and industry taught courses [5]. Fuel education is an online platform for school students. It provides online curriculum for the students and provides equal learning opportunities for them. The education is personalised according to the students' learning pace, capability, etc. It also caters to different students' interests. Although the education is personalised to some extent, it still follows the same pattern as general education.

3 Proposed System

We propose a system that comprises the following:

1. Smart learning assistant which collates the content and caters it to students and assists in tracking their progress.
2. Interactive interface for the users to upload and consume content.

The proposed solution is hosted on the cloud with a web interface.

Our proposed solution aims at helping several schools, colleges and other educational institutions in the long run.

In the system we propose to develop, we will have a user termed administrators (admin) who shall add teaching/non-teaching staff to handle the courses. These staff assigned are termed as course handlers. The course handlers can upload the content with respect to a course or multiple courses. The content can then be tagged. Tagging is the feature which allows quick and easy access to the content. It also assists in organising and reorganising the content. Now, the content is ready for the students to view. We have a built-in feature to track the progress of the student's studies, i.e. the consumption of content on the website is automatically tracked, thereby helping to track his/her progress. Moreover, the student can inform the voice-based/text-based smart assistant of any progress outside the system. The smart assistant also helps query and search for a particular course in the content present on the platform.

3.1 Modules and Functionalities

1. Uploading content
   It is a web interface for the course handlers to upload content to the courses they handle.
   Storage and database in the backend are used to store the content and the metadata.
2. Tagging content
   It is a web interface for the users to tag the content. There are two kinds of tagging:
   a. The course handler can tag the content. This is visible and used by everyone who can access the content.
   b. The student can tag the content. This is visible only to that student.
3. Searching
   The users can search across all accessible content by name, course, topic or any tag.
4. Progress tracking
   The system tracks the consumption of content on the website.
   The students can inform the system of their progress.
5. Analysis and summariser
   The system summarises and analyses the activity of the user on the website.
6. Tests
   The system does not have a feature for the professors to hold tests and quizzes. However, the course handlers can enter the details of the test—including the date and syllabus—into the system.
   The system now knows the syllabus—by topics and chapters.
   The student can now search for content within this subset. The system can keep track of the student's progress with respect to this test.
7. Chatbot
   It is an another interface to the system.
   Chatbot can be used for searching queries and informing progress [6].

Here, the uploaded content will be stored in AWS s3 buckets [7]. The details regarding the storage and accessing them will be stored in a no-sql database as they are known for their cloud computing capabilities and storage scalability [8].

We plan to use conversational chatbots which can be interfaced with existing messaging applications to concentrate on the functionalities rather than the interface [9].

3.2 Users and Use Cases

Use a case summary for the user Admin (Fig. 1):

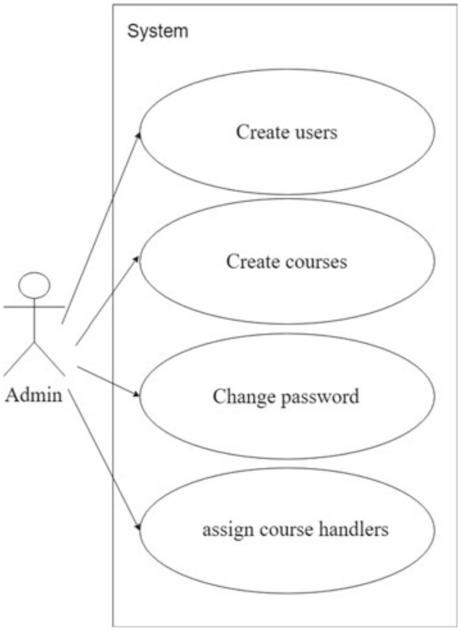


Fig. 1 Use case diagram of the user admin

1. Admin interacts with the system to create users, i.e. course handlers. He also assigns them courses.
2. Admin creates courses corresponding to a particular branch and semester.
3. Admin can also change/reset password pertaining to the course handlers.

Use case summary for the user Course handler (Fig. 2):

1. Course handler can upload materials for one or more courses.
2. He can tag content for easy access and availability.
3. He can delete unnecessary content.
4. He can modify the content for a course and can re-upload it on the platform.
5. He can enter the details of the test—including the date and syllabus—into the system, i.e. schedule the tests.

Use case summary for the user Student (Fig. 3):

1. Students can view the content available on the system and follow the courses.
2. They can tag all the content.
3. They can upload content, share and comment on the materials present.
4. The system tracks the consumption of content on the website. The students can inform the system of their progress and hence track progress.
5. They can use the chatbot to interact with the system, i.e. query the content and search for a particular chapter, search by tags, update progress, etc.
6. The students while viewing the content can add comments for reference. The student should be able to share the comments doc with his/her peers.

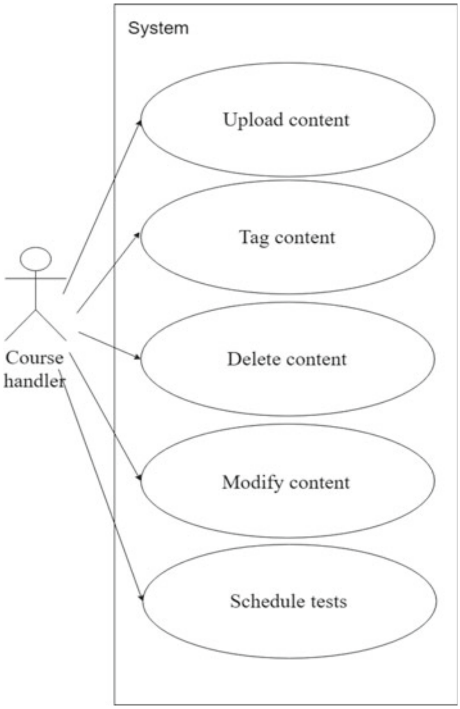


Fig. 2 Use case diagram of the user course handler

The above-mentioned modules shall be implemented as microservices [10], and each service will be used by the system. Services will be loosely coupled, and they will be implemented as python modules interfaced with flask, a micro framework architecture. There are other frameworks as well but flask proved to be efficient [11].

4 Dataflow Diagram

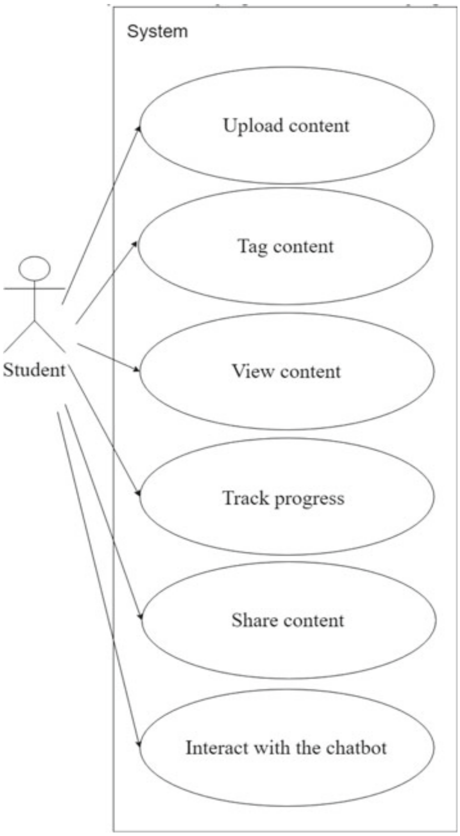
The major modules as mentioned in the previous section are as follows:

1. Uploading content.
2. Searching content.
3. Progress tracking.
4. Tests—curating relevant content.

We now explain each of these in greater detail.

The modules described earlier are abstracted into reusable services.

Fig. 3 Use case diagram of the user student



4.1 Uploading Content

The user can upload content and relevant tags along with the content. As shown in Fig. 4, the content is stored in a content data store and the metadata along with the tags are stored in a database.

4.2 Searching Content

The users can search across all the content accessible to them. Figure 5 shows the flow design. The search service returns the metadata including the location address in the data store, which is used to return the content as the output.

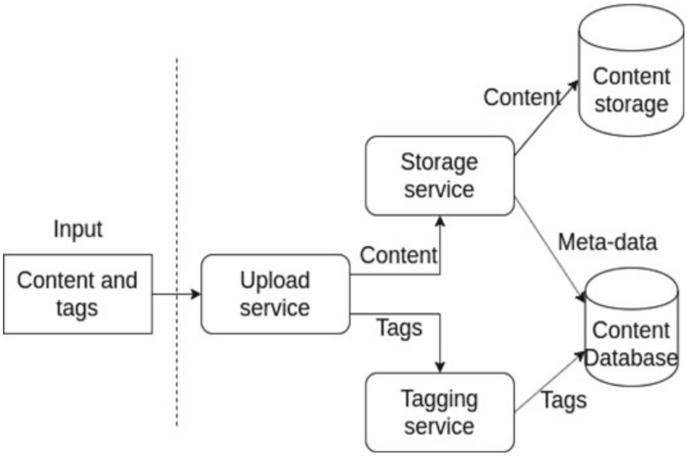


Fig. 4 Design diagram for uploading content

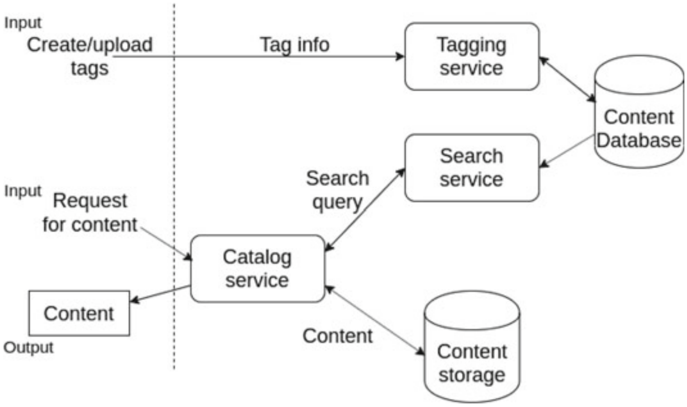


Fig. 5 Design diagram for searching

4.3 Progress Tracking

Figure 6 shows the use of chatbot as an interface to the progress tracking service. The progress tracking service may also take input from other internal services. As shown, the progress information and the analysis information are stored in the user database. Further, the chatbot is versatile. It can be used for search queries as well.

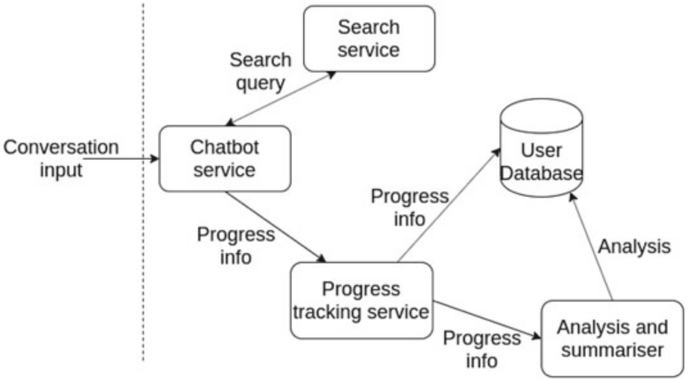
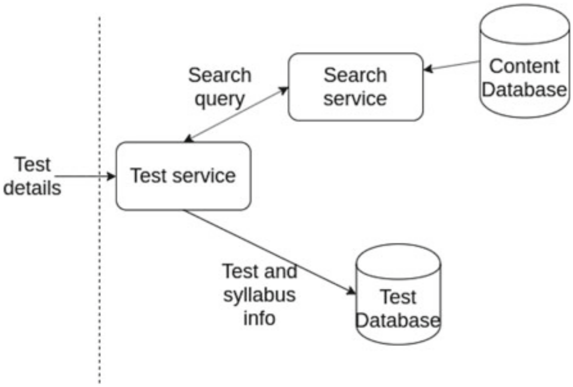


Fig. 6 Design diagram for progress tracking

Fig. 7 Design diagram for curating relevant content for tests



4.4 *Tests—Curating Relevant Content*

The test service takes the test details as input. It uses the search feature to get and curate the metadata of the relevant content. The curated metadata is stored in a test database for quick access (Fig. 7).

5 Conclusion

We have seen many problems faced by the students and by other peers who are into learning. We have proposed the above solution to help them overcome the problems. In the long run, we want to achieve an efficient solution and an evergreen platform with many more enhancements. To state a few, initially, we aim to put our system into use by giving access to colleges, schools and other educational institutions. We want to be helpful to students who are unable to attend the college physically, i.e. help in distant-based education. The college staff can upload content. Students who have enrolled themselves through correspondence to any course offered by the college can view content, upload content and get the test schedule. This will help many students by easing their struggle at many stages of education. Also, a mobile app would enhance the user experience of students. Besides timely notifications, a mobile app would open the doors to many new features such as uploading photographs as content.

References

1. Muhammad A, Zhou Q, Beydoun G, Xu D, Shen J (2019) Learning path adaptation in online learning systems
2. Yang F, Li FWB, Lau RWH (2019) An open model for learning path construction
3. Basu P, Bhattacharya S, Roy S (2019) Online recommendation of learning path for an e-learner under virtual university
4. Coursera: start-to-finish learning paths for starting a new career (2019). https://blog.coursera.org/new-coursera-start-finish-learning-paths-starting-new-career/
5. Fuel Education (2019). https://www.fueleducation.com/
6. Edmodo (2019). https://new.edmodo.com/home/
7. AWS S3 (2020). https://aws.amazon.com/s3/?nc=sn&loc=0/
8. Reasons to use no-sql database (2020). https://support.rackspace.com/how-to/reasons-to-use-a-nosql-db/
9. Chatbots (2020). https://chatbotslife.com/, https://-www.medium-com-hellohaptik-understanding-conversational-ai-what-where-how-you-can-use-bots-a5df897886f9?gi=44912e84dbd3/
10. Microservices (2020). https://martinfowler.com/articles/microservices.html/
11. Python (Microservice Architecture) MSA (2020). https://www.fullstackpython.com/microservices.html/

Eye Disease Detection Using YOLO and Ensembled GoogleNet

Saikiran Gogineni, Anjusha Pimpalshende, and Suryanarayana Goddumarri

Abstract Our research work has succeeded in integrating ensemble into GoogleNet image classification technique aiming higher accuracy and performance than existing models. The convolutional layers are apt for feature extraction from images. In GoogleNet classifier, only 1 fully connected dense layer (weak) is used for classification of class from these outputted features. But we integrated ensemble immediately after convolutional layers for purpose of better classification output. Thus, the output (features of image) of convolutional layers is passed as a separate input to both ensemble methods and fully connected layers of GoogleNet for obtaining the class of image. The final class of image is determined by the specific strategy after analyzing outputs of ensemble and GoogleNet fully connected layer. All earlier works focused on eye disease classification. Here, we have also experimented with YOLO for detection of location and class of diseases. The eye is considered as the most significant part of the body. But this most significant part of the body is easily subjected to various kinds of diseases. Early detection and prevention are needed. Our research work aimed at detecting top five common eye diseases with higher accuracy. User can upload a pic in a mobile or cloud application, and inbuilt AI algorithms will detect the type of eye disease with higher accuracy and thus offering prevention suggestions at an early stage without doctor intervention.

Keywords YOLO · ResNet101 · GoogleNet · Object detection · Bagging · Decision tree · Random forest

S. Gogineni (✉) · A. Pimpalshende · S. Goddumarri
CMR College of Engineering & Technology, Hyderabad, India
e-mail: goginenisaikiran31677@gmail.com

A. Pimpalshende
e-mail: anjushap@cmrcet.org

S. Goddumarri
e-mail: gsuryanarayana@cmrcet.org

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_44

1 Introduction

The classification of disease must be more accurate than classifying cars or dogs because as a result of classification associated medicine is suggested. Intake of wrong medicine may lead to death sometimes. So, classification, in this case, became more important. So, we decided to carry on further research in achieving higher accuracy compared to existing classifiers in the task of classification. We have chosen ensembles as a means to achieve our aim. Ensembles consist of bagging and boosting techniques. Later, we also experimented with YOLO to perform object detection in case of eye diseases. There are hundreds of eye issues that affect humans. However, every eye issue is identified with certain symptoms. The eye is considered as the most significant part of the body. It discharges its responsibilities such as looking at the world, color vision, protects from other dangers, and so on. It also protects the body from various viruses and bacterial attacks through vision. It acts as a shield by preventing us from the sun's rays. It forms a boundary separating external environments. But this most significant part of the body is easily subjected to various kinds of diseases that can do greater harm. Early detection and prevention are needed. We have considered the following five eye diseases to carry on further research work as they are most common and typically dangerous. They have to be detected in time with the use of proper technology.

1. Bulging eyes,
2. Cataracts,
3. Crossed eye,
4. Glaucoma, and
5. Uveitis.

2 Related Work

Authors [1–3] performed different image processing techniques to analyze retina therapy [4]. GoogleNet is the base of our research work. [5–8] have experimented with glaucoma and cataracts' detection individually but resulted in poor results. In contrast to this, we have experimented with the latest YOLO and SSD object detection techniques and also developed models that can detect several eye diseases in one step; [9–11] used powerful neural networks to classify images but our research work proposes to use ensembled CNNS to boost accuracy. The data has been split into training and test dataset. Because of the availability of limited images, the validation set is not considered. In addition to dataset, various eye disease images are collected from the Internet. Out of entire data, 75% is taken for training and rest 25% of data is taken for testing purpose. Image augmentation technique is adapted to increase data. PyTorch and TensorFlow frameworks are used to code.

3 Proposed Work

3.1 Detection with YOLO Algorithm on Eye Disease Dataset

Two types of works are done. As a part of the experimentation, the first part is focused on detection of eye disease in an image using YOLO. As a result of research work, the second part of paper is focused on integrating ensembles to GoogleNet architecture. Among the several object detection techniques, YOLO is preferred as a result of its capability to achieve significant accuracy and to track tiny and large objects in an image at one go. But also, the results and analysis of other several object detection techniques such as RCNN, faster RCNN, and SSD are also presented.

Object detection in an image is quite simpler and straightforward with YOLO. YOLO is definitely fast. The fast version of YOLO will run at more than 150 frames per second, and the base version of YOLO will run at nearly 45 fps. Fps stands for frames per second or relative speed of the model in processing the number of images per second. The detection process of YOLO is focused on detection of tiny resolution objects also. So, YOLO increased the input size of image from 224 × 224 (preferred for vgg) to standard 448 × 448. The architecture resizes the given image into 448 × 448 sized image and runs a mixture of convolutional, pooling layers on it and finally non-max suppression to obtain final output. The entire resized image is divided into G × G grid, and each grid cell predicts B bounding boxes, confidence for those boxes and C class probabilities if assumed there are total C classes. The YOLO architecture outputs a tensor of size G × G × (B × 5 +C). B is bounding box. Each and every bounding box has five parameters associated, namely, *x*, *y*, *w*, *h* coordinates and confidence of class. The confidence of a class is probability value that lies between 0 and 1 and ensures how confident the bounding box is, regarding the presence of an object in it. The (*x*, *y*) point represents center of the box. The height and width are predicted in association with the complete image. Finally, the confidence outputted will represent the intersection over union between the predicted box and ground truth of object. The probability (class ilobject) is calculated for every grid of image. Only one set of this probability is calculated per each grid regardless of the number of boxes each grid holds. The final layer of this architecture is responsible for the prediction of object coordinates and class of object. We generally scale the anchor box height and weight by input image height and weight so that they are always constrained to appear within zero and one. Linear activation function is used for the last layer of YOLO system that is involved in delivering the final output coordinates and class probabilities. All the intermediate layers of the system use leaky rectified linear activation function which is described below:

$$g(x) = \begin{cases} x, \text{if} x > 0 \\ 0.1 * x, \text{else} \end{cases}$$

Sum squared error is used as a metric to measure coordinate regression loss and thus to mitigate loss by backpropagating. But this sum squared metric weighs localization error and classification error equally which is not highly suitable or preferred. It pushes the class score of a grid to zero if that particular grid doesn't hold any object. This can cause high instability as it overpowers the gradient from these grids. In order to deal with this, the YOLO introduces two new terms ($\lambda_{\text{noobject}}$ and $\lambda_{\text{coordinate}}$) to increase error from (x, y, w, h) predictions and eventually mitigate loss from label predictions for predicted (x, y, w, h) bounding boxes that don't hold objects. The network architecture has 24 convolutional layers (composed of kernel weights) that are followed by 2 fully dense or connected layers. This simple architecture preferably uses 1 × 1 kernel weights followed by kernel weights of 3 × 3 in comparison with inception model of GoogleNet.

For eye disease detection problem, we have decided to work on five different diseases as mentioned above. So, the number of classes $C = 5$. We used $G = 7$ grid size and $B = 2$ number of bounding boxes or anchor boxes. The final output tensor of YOLO architecture is (7 × 7 × 15). (7 × 7×15) can be written as (7 × 7×2 × 5+5). Total there are 49 grids in an image and 15 predictions associated with each and every grid. These 15 predictions for each grid include $[p_c, b_x, b_y, b_h, b_w]$ for first anchor box + $[p_c, b_x, b_y, b_h, b_w]$ for second anchor box + $[c_1, c_2, c_3, c_4, c_5]$ the class confidence of an object in that grid. p_c will be 1 if corresponding anchor box contains object else 0 if it doesn't contain object. If p_c is zero then other elements of vector such as x, y, w, h coordinates can be simply don't care. If the particular grid contains Cataracts Eye disease symptom, then class confidence vector is [0, 1, 0, 0, 0] and if grid contains disease Crossed Eye then ground truth confidence vector will be [0, 0, 1, 0, 0] and ground truth box coordinates vector can be assumed as [1, 0.2, 0.6, 0.6, 0.8]. For each grid cell get two predicted bounding boxes. Get rid of low probability predictions. The aspect ratios and sizes of anchors are to be decided in consideration with types of objects being detected. The anchor used for prediction of pedestrian can't be used for prediction of a car. We are generally aware of the fact that humans can be fitted into vertical boxes rather than square boxes. In YOLO, the prediction is always done without any assumptions on the shape of target objects. But YOLOV2 put some constraints on anchor sizes taking into consideration of target object sizes. This has driven model by significant increase in performance.

$$\text{anchors} = [(1.8744, 2.0625),\ (0.5727,\ 0.67738),\ (3.3384, 5.4743),\ (9.7705, 9.1682),\ (7.8828, 3.5277)]$$

$$\begin{aligned}&\text{For anchor pair}(1.87446,\ 2.06253), \text{Width} = 9.07705 \times 32 = 290.46\,\text{pixels},\\ &\text{Height} = 9.1682 \times 32 = 293.38\,\text{pixels}\end{aligned}$$

In above-illustrated example, (3.3384, 5.4743) are coordinates of one of the anchor boxes. They usually represent the height and width coordinates of that particular anchor box. The above list represents five different anchor boxes. Remember that these are always chosen in consideration with output object shape. YOLOV2 generates a 13 × 13 output tensor. So, you can obtain actual values by performing multiplication with 13 by anchor box coordinates. In YOLOV2 the image 416 * 416 is divided into 13 grids where $G = 13$ instead of 7 as in YOLO. Yolov2 divides image of size into 13 * 13 grid each containing 32 pixels.

3.2 Results of Detection

The mean average precision is slightly higher in fast RCNN using ResNet backbone classifier after YOLOV2. However, YOLO also achieved comparable performance on eye disease image dataset. Flips per second are higher in YOLOV2 than any other object detection algorithm (Figs. 1, 2, 3, 4, and 5) (Table 1).

Over the past few years, deep convolutional layers are capable of performing both detection and classification tasks successfully in the domain of computer vision. Indeed, as each year passes, the performance of these classification systems is increasing significantly with increase in depth of layers. Priority is always given to performance of the model regardless of depth of layers and high computational space it holds.

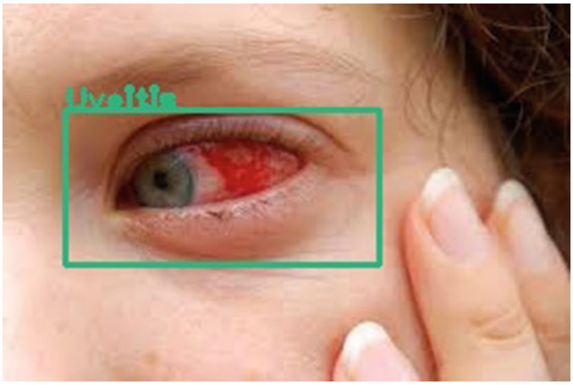


Fig. 1 YOLO algorithm detecting Uveitis disease location

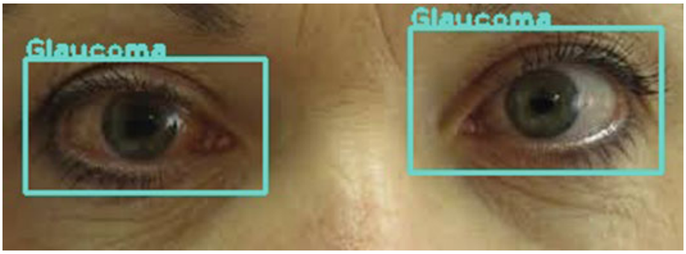


Fig. 2 YOLO algorithm detecting Glaucoma disease locations

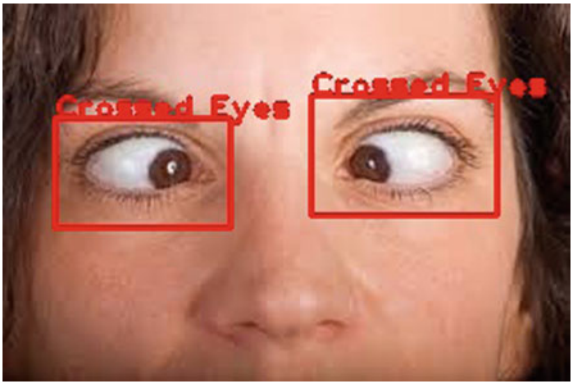


Fig. 3 YOLO algorithm detecting crossed eye disease locations

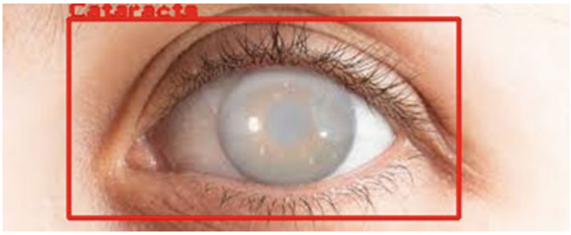


Fig. 4 YOLO algorithm detecting cataracts disease locations

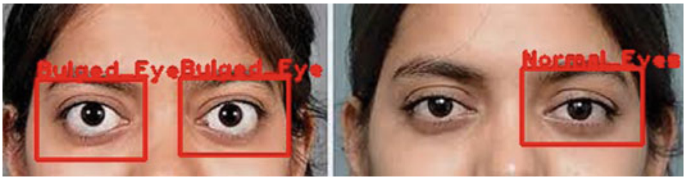


Fig. 5 YOLO algorithm detecting blurred eye disease locations

Table 1 Illustrating performance of various object detection algorithms on eye disease dataset

Algorithm	mAP	FPS
YOLO 448 × 448	73.8	45
YOLO V2 416 × 416	75	67
Fast RCNN	69.1	0.5
Fast RCNN-VGG	69.2	7
Fast RCNN-ResNet	72.8	5
SSD 300	70.7	46
SSD 500	66.6	19

GoogleNet model is an improvement over VGG. It is also known as inception model. All the earlier models suffered in choosing kernel size (whether it should be 3 * 3 or 1 * 1 or 5 * 5). This architecture suggests using all those kernel sizes. Suppose input for inception module is of shape 28 * 28 * 192. Output of 1 * 1 convolution with depth 64 is 28 * 28 * 64. Output of 3 * 3 padded convolution for same input with depth 128 is 28 * 28 * 128. Similarly, output for 5 * 5 padded convolution is 28 * 28 * 32. Pooling (stride = 1) on input generates 28 * 28 * 32. Now we need to stack or concatenate all these outputs depth-wise to obtain 28 * 28 * 256. This depth-wise (64 + 128 + 32 + 32) concatenation is the final output of inception module which will be input to the next inception module. GoogleNet is the collection of all such inception modules together. But above inception module involves high computational cost. So, a slight modification is introduced in inception module. Before performing 3 * 3 and 5 * 5 convolutions, we need to perform intermediate 1 * 1 convolution to reduce parameters. For the input 28 * 28 * 192, using 1 * 1 * 192 filter with depth 16 and further on this using 5 * 5 * 16 with depth 32 is preferred than using 5 * 5 * 192 with depth 32. Deep inception modules are usually followed by 1 fully connected layer and 1 softmax for the sake of classification.

3.2.1 ResNet

This network model is based on concept of deep residual block, proposed in the recent research paper. It uses shortcut connections to improve model performance. ResNet-k means deep residual network consisting of number of layers. For example, ResNet-50 means ResNet consisting of total of 50 layers. The issue with deep networks is that the training error increases with an increase in depth. As a result, performance decreases. But ResNet provides a mechanism through which only performance increase with increase in depth but not training error. The current layer output is passed as concatenated input to the layer ahead of next 2 layers. The phenomenon continues to happen throughout the model.

3.2.2 Ensembled GoogleNet

Here we come up with significantly more accurate convolutional neural networks. They not only achieve the state-of-the-art accuracy on eye disease classification and localization tasks, but are also applicable to other custom image classification datasets, where they achieve more accurate performance compared to existing models. During training phase, the input to our convolutional networks is a standard 229 × 229 image.

Ensemble: An ensemble is a collection of separately trained learners (might be group of decision trees or simply neural networks) whose decisions are combined in order to classify an instance. All the earlier research work has clearly proven that an ensemble is generally more powerful and correct when compared to individual learner. Bagging technique is always more correct than a single classifier. But it is sometimes much less correct compared to boosting. These techniques depend on “resampling or sampling with replacement” strategy to generate different training datasets from actual data to different learners.

Bagging: Each individual learner’s training data is obtained by drawing data points randomly from D data points with replacement, where D is the size of actual training data. Many of the instances in actual training data may occur repeatedly in the construction of resultant training data while few of them are left; one classifier is independent of another classifier.

Boosting: The aim of boosting is to actually generate a series (one after another) of dependent classifiers. Training data required for classifier of the series is obtained as per the performance of the previous learner on dataset in the series. In this method, instances that are incorrectly classified by earlier classifiers in the series are chosen more often than instances that were correctly classified. Thus, boosting is capable of producing new excellent classifiers that are better able to classify instances for which the current classifier performance is poor.

1. The bias concept measures the closeness of classifier produced by the learning algorithm to the actual target function required to map input to output.
2. The variance term measures the disagree between classifier produced by learning algorithm and actual target function. It is a measure of how their decisions differ.

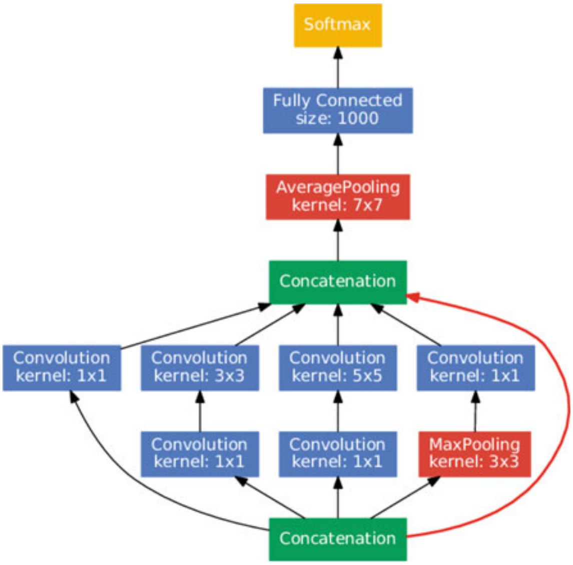
With the help of above strategy, the bagging and boosting models try to bring down both bias and variance. Many scientists feel that boosting actually attempts to decrease the miss classification rate or error in the bias term as it is focused on miss classified examples. However, the same focus can force the individual learner to produce an ensemble that differs highly from the actual single learner. Bagging can also be used to mitigate the error but highly useful in reducing the variance (Fig. 6).

3.2.3 Proposed New Image Classification Technique

The nine inception blocks of Inceptionv1 are convolutional layers that are meant for extracting features from the images. The last fully connected layer is meant for classification of class from the extracted features. We have considered the output of last inception as the required significant input (which typically contains extracted features of the image) to ensemble of classifiers. The output tensor of ninth inception module (7 * 7 * 1024) after average pool (stride = 7) is of shape $1 \times 1 \times 1024$. It is flattened simply to a vector of size $[1 \times 1024]$ or [1024]. Now we will integrate ensemble after this last inception module. Many classifiers are immediately built after this layer to achieve excellent accuracy. One of the classifiers will be GoogleNet’s fully connected layer itself. Other classifiers will be decision trees or AdaBoost or simply yet powerful another neural network.

The issue is that the input vector of size 1024 is fine for dense block of GoogleNet and another neural network classifier but not aptly suitable for decision trees because of large input space which is no more discrete but highly continuous. There must be some strategy to bring down this large input space to medium input space without

Fig. 6 GoogleNet block diagram



losing any significant information. So, we used PCA to achieve this task. The main idea of principal component analysis (PCA) is to reduce the dimensionality of training set composed of many features usually correlated with one another. But it aims at retaining or preserving the variation present in dataset to a possible extent. It is done by transforming the features of original dataset to a new feature space, known as the principal components. These principal components are orthogonal, ordered such that the retention of variation present in the original features reduces as we move down in the order. So, in this manner, the $k + 1$th principal component retains minimum variation that was present in the original features compared to kth principal component. These principal components are known as eigenvectors of covariance matrix which are usually orthogonal.

Importantly, the training set on which PCA technique is to be applied must be normalized or scaled properly. Always the results are sensitive to the normalization applied to dataset. In the dataset, normalization is done by subtracting the column mean from the respective numbers of that column.

If we consider x 1, x 2, x 3, …xn as features of dataset, these features are typically output of last inception layer of an image with n = 1024. Since we are dealing with the dataset containing 1024 features, the PCA will construct a square matrix of size 1024 * 1024. The above is the covariance matrix. Kindly remember that always var[xn] = covariance [xn, xn]. Further, we need to calculate the eigenvectors and eigenvalues of the above covariance matrix from equation $\boldsymbol{det\ (\lambda I - w) = 0}$ where I is the identity matrix. For each eigenvalue, a corresponding eigenvector is calculated using the equation $\boldsymbol{(\lambda I - w)v = 0}$***. Out of n eigenvalues, we choose some d values to reduce feature dimension.*** Covariance matrix:

$$W = \begin{pmatrix} \mathrm{var}(x1) & \mathrm{cov}(x1, x2) & \mathrm{cov}(x1, xn) \\ \mathrm{cov}(x2, x1) & \mathrm{var}(x2) & \mathrm{cov}(x2, xn) \\ \mathrm{cov}(xn, x1) & \mathrm{cov}(xn, 2) & \mathrm{var}(xn) \end{pmatrix}$$

Consider the eye disease image dataset contains N images. If there are E epochs, for each epoch there will be I iterations. For each iteration, a batch size of images is processed in GoogleNet. Instead of parallel processing we adapt serial processing.

Ninth inception layer output of GoogleNet before and after PCA.

Suppose there are N images; in Table 2, each row indicates ninth inception layers' output of GoogleNet when that particular image is given as training input. It is directly given as input to another neural network (classifier in ensemble) without PCA. However, this neural network performance will vary as it uses another random weight initialization method. But PCA is applied before it is supplied as input to the decision tree. PCA has reduced 25088-dimensional data space to much smaller dimensional space. There is loss in information but not much significant loss. Though there is a loss in information, the new set of features will be able to differentiate one category of image from the another. Here, batch-wise processing is not needed, while GoogleNet training is going on, the last convolutional layer output of all images in the last epoch is stored as separate data frame. This data frame can be input to the rest of classifiers in the ensemble. The output of last convolutional layer in the last epoch will be highly stable as it already captured a lot of features of the images as training is near to end. Classifiers working on this stable input are expected to yield much more accurate results. The class of the image is individually decided by the classifiers but the final class of the image in the test set is determined by either voting method or weighted saying method as we need to analyze all the outputs generated by all the classifiers of the ensemble. Now the issue arises in determining the class of an image in test set.

1. Voting
   For the task of eye disease classification, we opted to vote. Suppose there are k classifiers in the ensemble. Each classifier will output a particular class for an image in test set. Suppose the classifiers be GoogleNet dense block, another

Table 2 Illustrating output of GoogleNet last inception layer before and after applying PCA

Image	Feature1	Feature2	…	Feature 1024	Image	Feature1	Feature2	…	Feature 500
Image 1	f1-1	f1-2	…	f1-1024	Image 1	f1-1	f1-2	…	f1-500
Image 2	F2-1	F2-2	…	F2-1024	Image 2	F2-1	F2-2	…	F2-500
Image 3	F3-1	F3-2	…	F3-1024	Image 3	F3-1	F3-2	…	F3-500
Image 4	F4-1	F4-2	…	F4-1024	Image 4	F4-1	F4-2	…	F4-500
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
Image N	Fn-1	Fn-2	…	Fn-1024	Image N	Fn-1	Fn-2	…	Fn-500

neural network, decision tree, random forest, and AdaBoost. Their combined output is a vector of classes [c0, c1, c1, c2, c4]. C0 predicted by GoogleNets fully connected layer, c1 predicted by another neural network, c2 predicted by decision tree, and so on. The final output will be maximum occurring element of above vector. This strategy of voting is quite simple and efficient also.

2. Weighted Saying

   Here we consider the loss function of each classifier in ensemble. After the complete training is done, the loss or error of each classifier is noted down. This is used in determining the final class of test image.

$W = ½ \log (1 - K/K)$ where K is the loss made by the classifier. W is the say of the classifier or weightage of the classifier in determining the final class of an image. Generally, dense blocks and neural networks use binary or cross-entropy loss function.

If $K = 0.8$, $W = ½ \log (0.2/0.8)$, $W =$ -0.3010 If $K = 0.2$, $W = ½ \log (0.8/0.2)$, W = 0.3010. If classes predicted by classifiers in ensemble are [1, 0, 0, 1, 1] and their respective says are [0.67, 0.54, -0.75, −0.44, 0.87], it makes predictions by having each classifier classify the sample. Then, we split the classifiers into groups according to their decisions. For each group, we add up the say of every classifier inside the group. The final classification made by the ensemble as a whole is determined by the group with the largest sum.

4 Result Analysis

Ensembled GoogleNet achieved significant increase in accuracy compared to GoogleNet and ResNet architecture (Figs. 7, 8, 9, 10, 11, and 12).

As an example, consider all the images from 1 to 5 which are mentioned in the above classification results. The output of first convolutional layer of vgg16 aims at extracting certain features from an image. For better understanding purpose of feature extraction by convolutional layers, which are input to integrated ensemble, the following figures are illustrated. The output of first convolutional layer of vgg16 is of size 224 * 224 * 64. The 64 channels can be arranged as an 8 * 8 square grid which is shown below. The same feature extraction is done by GoogleNet but filters vary between VGG and GoogleNet (Figs. 13, 14, 15, 16, 17, and 18).

Features extracted at convolutional layer 1 (64 features).

An ensemble can be formed in n number of ways. It has choice of choosing individual classifiers. By choosing specific ensemble, a specific GoogleNet architecture is formed. Many such architectures are obtained and discussed below but only one can be used. The way of choosing individual classifiers will have a high notable performance impact on an ensemble. We tried and experimented with several individual classifiers in eye disease classification task. This ensemble is integrated at last possible convolutional layer of GoogleNet. The technique can be expanded to other custom datasets also.

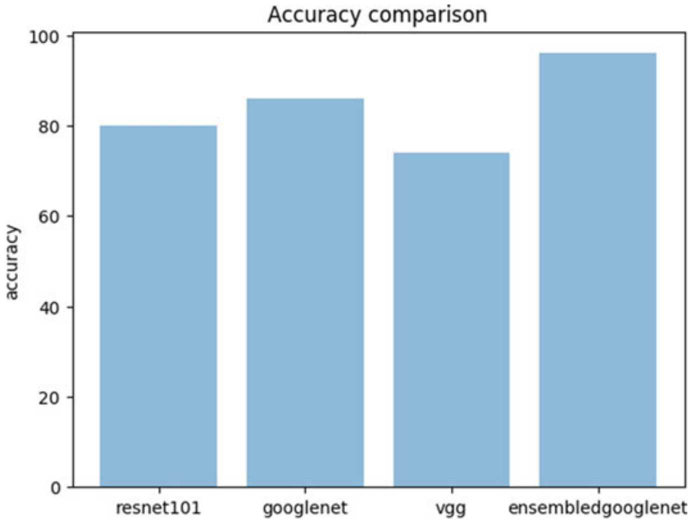


Fig. 7 Accuracy comparison among several models on eye disease image dataset

Fig. 8 Ensembled GoogleNet predicts disease type [5]

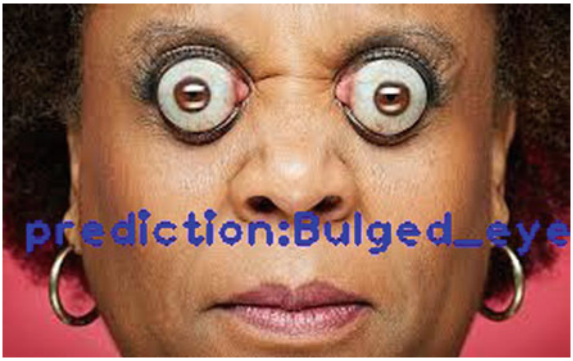
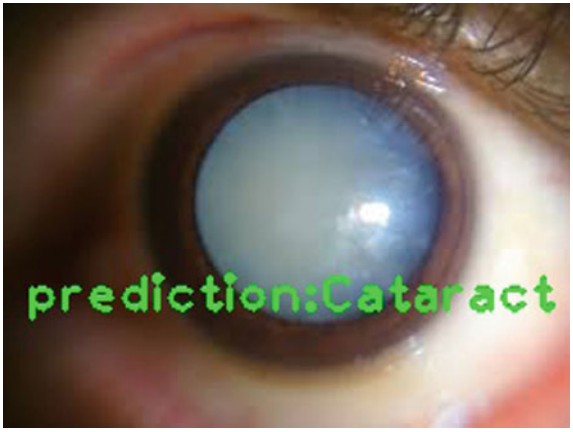


Fig. 9 Ensembled GoogleNet predicts disease type [7]



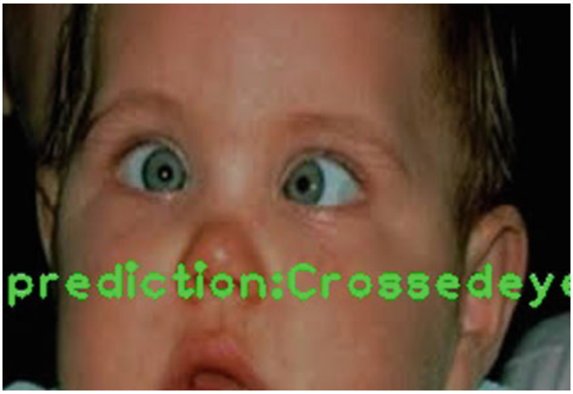


Fig. 10 Ensembled GoogleNet predicts disease type [6]

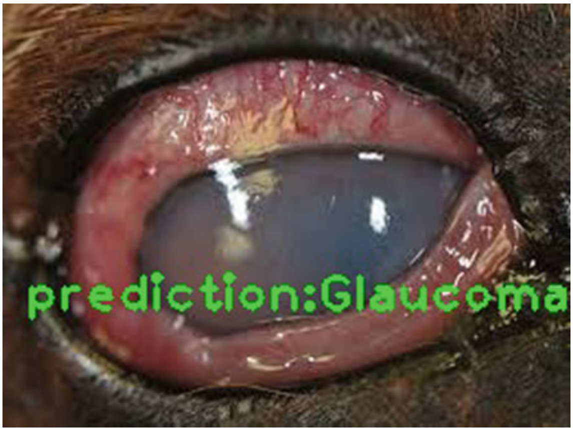


Fig. 11 Ensembled GoogleNet predicts disease type [8]

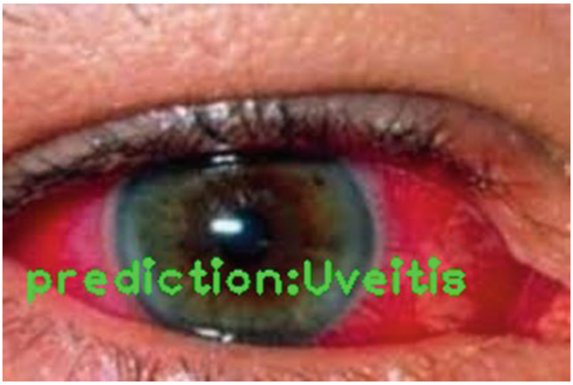


Fig. 12 Ensembled GoogleNet predicts disease type [9]

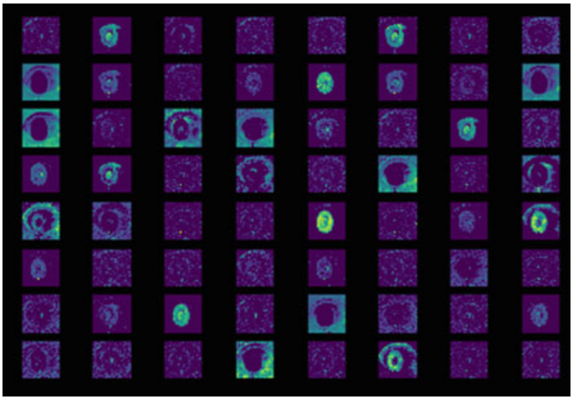


Fig. 13 Features extracted by the first convolutional layer of Vgg16 for bulged eye disease image

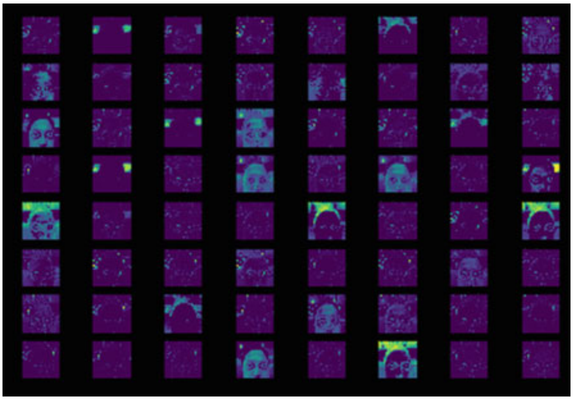


Fig. 14 Features extracted by the first convolutional layer of Vgg16 for cataracts disease

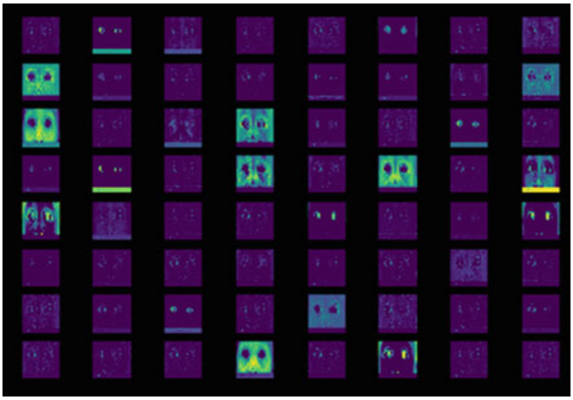


Fig. 15 Features extracted by the first convolutional layer of Vgg16 for crossed eye disease

Fig. 16 Features extracted by the first convolutional layer of Vgg16 for glaucoma disease

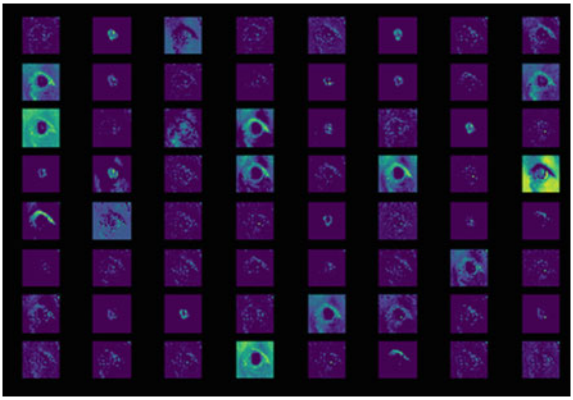


Fig. 17 Features extracted by the first convolutional layer of first convolutional layer of Vgg16 for Uveitis

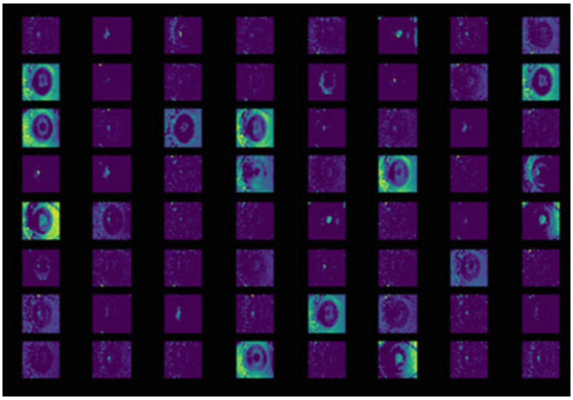
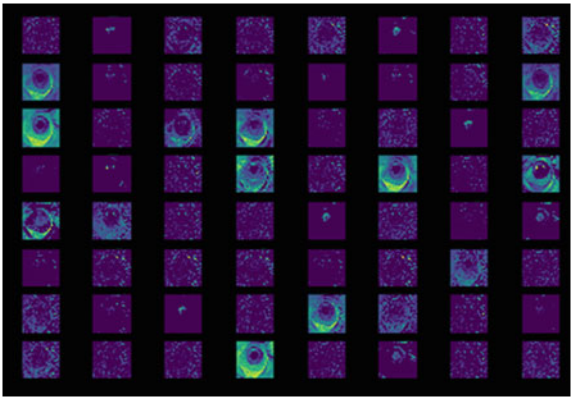


Fig. 18 Features extracted by Vgg16 for Uveitis



Ensembled GoogleNet achieved significant increase in accuracy compared to existing GoogleNet and deep RESNET101 architecture on eye disease image dataset (Fig. 19) (Table 3).

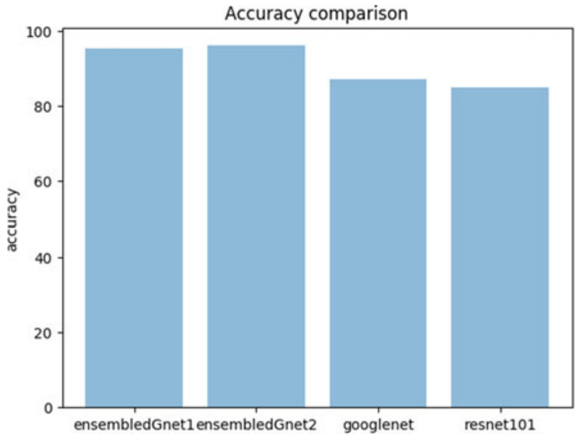


Fig. 19 Accuracy comparison among different GoogleNet ensembles, GoogleNet, and ResNet

Table 3 Describing various ensembled GoogleNet accuracies over GoogleNet and ResNet. Illustrating how an ensemble is formed by choosing its classifiers

ResNet 50 accuracy	82	Individual classifiers	Ensemble1
ResNet 101 accuracy	85.9	GoogleNet's fully connected layer	
GoogleNet accuracy on eye disease dataset	90.6	Random initialized neural network	
		Decision tree after applying pca	
		Reported ensembled GoogleNet accuracy	95.3
Individual classifiers	Ensemble2	Individual classifiers	Ensemble3
GoogleNet's fully connected layer		GoogleNet's fully connected layer	
Xavier initialized neural network		Zero initialized neural network	
Random forest		Random initialized neural network	
Reported ensembled GoogleNet accuracy	96	Reported ensembled GoogleNet accuracy	94.3
Individual classifiers	Ensemble4	Individual classifiers	Ensemble5
GoogleNet's fully connected layer		GoogleNet's fully connected layer	
Xavier initialized neural network		Zero initialized neural network	

(continued)

Table 3 (continued)

He initialized neural network		Random forest	
Decision tree with PCA			
Reported ensembled GoogleNet accuracy	96	Reported ensembled GoogleNet accuracy	94.3

5 Conclusion

The work proposed is a significant contribution to the domain of eye diseases. People can use the above YOLO experimentation and proposed ensembles as a simple cloud application to detect eye diseases in initial stages and can try to prevent them without much doctor intervention with higher accuracy. By using multiple classifiers, the accurate prediction ability of an ensemble can be much better than that of a single classifier or single model.

Competing Interest The authors declare that they have no conflict of interest.

Informed consent
Informed consent was obtained from all individual participants included in the study.

References

1. Hashim MF, Hashim SZM (2014) Diabetic retinotherapy lesion detection using region based approach. In: Malaysian software engineering conference, pp 306–310
2. Shojaeipour A, Nordin MJ (2014) Using image processing methods for diagnosis diabetic retinotherapy. In: International symposium on robotics and manufacturing automation. IEEE, pp 154–159
3. IEEE PAPER identification of different stage of diabetic retinopathy using retinal optical image
4. Going deeper with convolutions Christian Szegedy Google Inc. Wei Liu University of North Carolina, Chapel Hill Yangqing Jia Google Inc. Pierre Sermanet Google Inc. Scott Reed University of Michigan Dragomir Anguelov Google Inc. Dumitru Erhan Google Inc. Vincent Vanhoucke Google Inc. Andrew Rabinovich Google Inc
5. Salam AA, Akram MU, Wazir K, Anwar SM (2014) A review analysis on early glaucoma detection using structural features. In: International conference on imaging systems techniques. IEEE, pp 1–6
6. Yang M, Yang J-J, Zhang Q, Niu Y, Li J (2013) Classification of retinal image for automatic cataract detection. In: IEEE 15th international conference on e-health networking applications and services, pp 674–679
7. Nayak J (2013) Automated classification of normal cataract and post cataract optical eye images using SVM classifier. In: Proceedings of the world congress on engineering and computer science, pp 978–988
8. Patwari MAU, Arif MD, Chowdhury MN, Arefin A, Iman MI (2011) Detection categorization and assessment of eye cataracts using digital image processing. In: The first international conference on interdisciplinary research and development, pp 22.1–22.5
9. Raghu Raj P, Gurudatha Pai K, Shylaja SS (2007) Algorithmic approach for prediction and early detection of diseases using retinal images. pp 501–505

10. Jayaraman S, Esakkirajan T, Veerakumar, Digital image processing. Tata McGraw Hill Education Pvt. Ltd
11. Chaudhuri S, Chatterjee S, Katz N, Nelson M, Goldbaum M, Detection of blood vessels in retinal images using two-dimensional matched filters. IEEE Trans Med Imag

Comparative Analysis of MCT Load Balancing Approach in Cloud Computing Environment

Shabina Ghafir, M. Afshar Alam, and Ranjit Biswas

Abstract The efficient working of cloud computing is based on the most important aspect, which is load balancing. This paper gives a comparison and analysis of the load balancing algorithms in a cloud computing environment. The task and resource allocation is most important for efficient working of cloud computing. The load balancing is referring to task and resource allocation in an efficient manner as this [5] is related to NP-hard optimization problem. Load distribution in the cloud computing environment in an efficient manner and taking care of all the problems at the time of load balancing is most important for load balancing algorithms. The researchers have researched on various algorithms for overcoming the load balancing generating problem during the task and resource allocation phase. In this paper, the focus is on discussing the load balancing classification and their algorithms. This paper lays emphasis on heuristic algorithms and measures the performance using the CloudSim simulator.

Keywords Cloud computing · Load balancing · MCT · MET · Round robin · OLB GA · Switching

1 Introduction

Cloud computing is most popular in the current scenario. The companies are adopting cloud computing, as they are using resources without any hassle, and therefore companies' costs on computing infrastructure have reduced dramatically. When loads are mentioned, this will refer to resource-type load. The load types are (1) memory

S. Ghafir (✉) · M. Afshar Alam · R. Biswas
Jamia Hamdard, New Delhi, India
e-mail: sghafir@jamiahamdard.ac.in

M. Afshar Alam
e-mail: aalam@jamiahamdard.ac.in

R. Biswas
e-mail: ranjitbiswas@yahoo.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_45

load, (2) compute CPU load, and (3) network load. Load balancing is a process to assign and hauled the entire load to individual nodes. Load balancing is mainly related to detecting and balancing the underloaded and overloaded nodes, of all nodes. In this paper, the main focus is on load balancing classification and algorithms' performance. The load balancing algorithm objectives are to improve the performance, maintain the stability, and work on backup plan. The backup plan working is for recovering from failure losses. In this paper, the different load balancing algorithms are going to be discussed. The load balancing is utmost significant process for cloud computing. This process is related to assigning and reassigning the total load to individual nodes. The mechanism of load balancing is to identify the underloaded and overloaded nodes and balancing the load among them. This process is having an effective load for nodes, and thus the different nodes will not be overloaded and all nodes have relatively equal load. This process helps the system to make optimal resource utilization and thus improving the response time of jobs, and therefore system will be working in an efficient and effective manner.

In the environment of cloud computing, load balancing algorithm plays a decisive role. These algorithms' performance is based on various goals like to improve the performance, maintain the system stability, and when system fails to algorithm require a backup plan that helps to recover from failure loss.

The scheduling or load balancing algorithms are the solution to NP-complete problems [2, 3]. The load balancing algorithm's main objective is to distribute the load to individual nodes and take care of no node will be overloaded. These algorithms improve their performance by minimizing response time, save energy, maximize throughput, improve reliability, etc. Figure 1 shows the load balancing process.

2 Related Work

2.1 Load Balancing Algorithms

In this section, the discussion is based on several types of algorithms. The load balancing is related to task allocation, and the resource management in load balancing will play a vital role in cloud environment [2, 4]. The main objective of load balancing is fair allocation of task and resource allocation. This will improve the stability of system and help to accomplish a higher user contentedness. The load balancing algorithm classifications are based on process imitation and strategy. The process initiation type of algorithm is three: type senders, receivers, and symmetric type. According to the strategy, dynamic and static load balancing algorithms are two types.

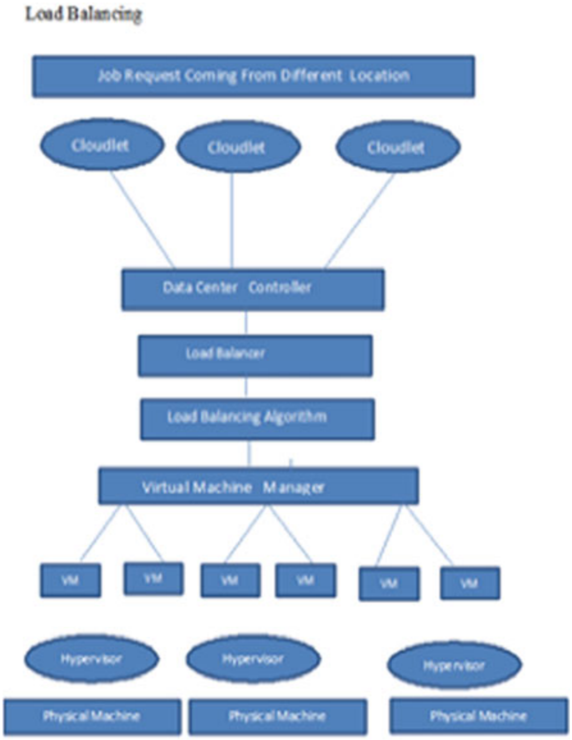


Fig. 1 Load balancing process

2.2 *Classification of Load Balancing Algorithms*

In this section, the discussion is for algorithms of load balancing and their classification. The problem of load balancing is like NP-complete enigma [1]. The researcher researches the load balancing problem solution with a multi-objective system [5]. These algorithms are basically working for minimization throughput, makespan, maximization, and energy saving. This study presents different sub-optimal or heuristic techniques used by load balancing algorithms that are presented by researchers to achieve a sub-optimal solution in cloud computing environment.

The researchers focus on improving their load balancing algorithm performance in different environments. The researcher's research load balancing algorithm for both homogeneous and heterogeneous environments. In this section, the discussion is about the load balancing algorithm [6] classification. The classification of load balancing algorithm is based on static and dynamic strategy. The dynamic strategy is further divided based on modes offline and online. In the last two decades, the researchers have been focusing on improvement of load balancing algorithms' performance in cloud computing environment. These load balancing algorithms work in different environments and strategies.

Figure 2 shows the assortment or classification of load balancing algorithms.

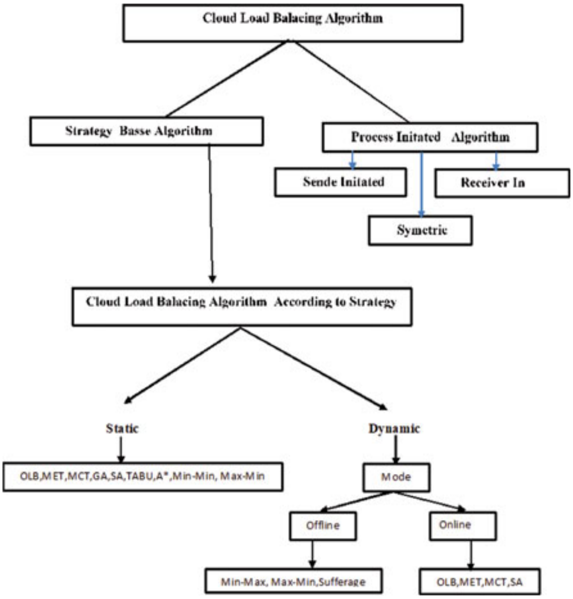

Cloud Load Balacing Algorithm
Strategy Basse Algorithm
Process Initated Algorithm
Sende Initated
Receiver In
Symetric
Cloud Load Balacing Algorithm According to Strategy
Static
Dynamic
OLB,MET,MCT,GA,SA,TABU,A*,Min-Min, Max-Min
Mode
Offline
Online
Min-Max, Max-Min,Sufferage
OLB,MET,MCT,SA

Fig. 2 Assortment or classification

2.3 Static Strategies

In the cloud computing environment when discussion is about the static strategy, working is according to two criteria at the time of beginning:

(1) Initial task arrival and
(2) Availability of physical machine.

In these types of algorithms, resource allocation will be updated according to task scheduling.

The next section in our discussion is on algorithms that fall into both strategies. These algorithms are opportunistic load balancing, MCT, MET, GA, switching, A-Star, tabu [7], min-min [8] and min-max, round-robin, ACO, honeybee foraging, biased random sampling, and active clustering algorithms.

2.4 Dynamic Strategies

This strategy is important and plays a significant role in cloud computing environment. These types of algorithms play a vital role due to distributing strategy [9]. The load will be distributed at the run time. These types of algorithms schedule the load at the run time. These algorithms work dynamically as these types of algorithms for allocation of tasks ar unusual. The VM virtual machines will be assigned and created according to input tasks. These algorithms' strategy [10, 11] is heuristics and further divided into batch mode or offline mode and online mode. The offline mode based heuristic algorithms work for the task allocation under the strategy for determining the certain execution time of huge number of tasks. Min-Min and Min-Max Suffrage algorithms fall into offline mode. In the immediate online mode, this strategy works according to user's request. The scheduler receives the user request and maps the request [12]. The online or immediate Mode algorithms are Opportunistic, SA, MET, and MCT.

3 Approaches

OLB Load Balancing: This algorithm working is based on heuristic techniques. This algorithm can work on both (static/dynamic) cloud environments. The algorithms heuristic and task allocation to virtual machines' behavior are arbitrary. The allocation of each task host based assigns to VM according to various parameters. These parameters are execution time, resource, request, and service. The OLB algorithm task execution is VM level, and this scheduling algorithm distributes the task load according to three levels. In this algorithm, the main strategy is to keep busy the node and do not deal with execution time of that task; therefore, makespan is slower OLD algorithm that is not at all proper for cloud environment. This algorithm has a lousy makespan for multiple goals considered cumulatively.

The cloud task scheduling policy [13] is the basis of the proposed algorithm and based on Load Balancing Ant Colony Optimization (LBACO) algorithm. The algorithm process is minimizing the makespan of stated task sets and balanced entire system load. This algorithm simulation is done on homogenous CloudSim [5]. This algorithm works for load balancing in a novel load balancing strategy. This strategy works to quest underloaded node to balance load from overpowered node. The strategy searching for VMS load balancing is GELS, based on GA and gravitational emulation local. In Singh et al. [1], the dynamic algorithm working environment is cloud computing. This algorithm functionality is taken from A2LB; this is known as autonomous agent-based load balancing algorithm. In Wang and Rao [14], the algorithm of load balancing is based on convex optimization theory.

MET: This algorithm is heuristic algorithm that follows both strategies, known as Minimum Execution Time or Limited Best Assignment (LBA) [15] and earlier is known as User Directed Assignment (UDA) by Armstrong et al. In this algorithm, the scheduler distributes the load as per the lowermost execution time from calculated Expected Time to Compute (ETC) table to VM, so as the system to perform all of tasks within execution time [16].

MCT: Researcher proposed an algorithm, known as minimum allocation time heuristic [17] techniques that worked on static and dynamic load balancing strategies. This algorithm of balancing the load is based on task ready-to-execute time and task expected execution time, and the task allocation is according to least completion time.

GA Algorithm: GA-based load balancing approach is proposed by Dasgupta [17] in 2013; it is used for minimizing the makespan. GA is depending on individual chromosome and population; this is related to fitness values. This algorithm design encoded binary strings and where chromosomes feel a random singular point crossover, pretense 0.05 as mutation probability.

Tabu Search (TS): Larumbe and Sanso [7, 18] proposed an accomplish TS heuristic to place the cloud data center in various locations and parallel variant master–slave model and algorithm proposed by Tsai and Rodrigues. The main goals are to improve the performance in terms of the network, reducing the CO_2 emissions. This algorithm is for optimizing the resource utilization cost.

A-Star Search: This algorithm is based on graphic searching algorithm. The A-Star algorithm functionality is based on BFS and DFS. This algorithm provides the benefit of DFS and BFS. This algorithm is based on VM and task allocation lists. The task list is based on priority queues, and VM list is based on processing efficiency of all resources. AlShawi et al. [19] proposed an algorithm [6]; researcher has to apply the blending of fuzzy and A-star algorithm.

Switching Algorithm: The algorithm vital function is to migration task for balancing the load into cloud computing. Shao et al. [5] introduced the algorithm that has accomplished the fault tolerance. This algorithm is a simple manner switching the tasks and VM according to policy that is decided by algorithm criteria.

Opportunistic Load Balancing Algorithm: This algorithm is based on the framework of the system. This algorithm works to keep every node busy. The execution time is completed [20]. The drawback of this algorithm is that node keeps busy after the job completion.

Ant Colony Optimization Algorithm: This algorithm falls under the classification of static and results shown best, and the functionality of algorithm is like actions of ants and seeking an optimal path in collecting their food [12]. The load balancing is efficient as it distributes the workload among nodes in efficient, and optimal job scheduling is achieved.

Honeybee Foraging Algorithm: This algorithm is based on honeybees' behavior and their approach to collecting honey. This algorithm dynamic strategy and behavior is similar to honeybees showing better result in response time and waiting time of the virtual machine and become reduced. In terms of increasing the resources, throughput is decreased.

BRS Algorithm: The algorithm researcher used Random Sampling Method. Load balancing is accomplished through all the nodes in the system. This dynamic algorithm becomes corrupted when load expands in the cloud environment.

Active Clustering Algorithm: The algorithm falls under the type of dynamic strategy based algorithm where nodes are grouping together. In this algorithm, similar nodes are grouped together [15]. The performance is poor when there is an increase in a variety of nodes.

Table 1 shows the comparison of load balancing algorithms [21].

Table 1 A comparison table for load balancing algorithms [22]

Load balancing algorithm	Parameters	Analysis	Simulation environment
Round Robin [21]	Completion time	This compares with FCFS result better from FCFS	Cloudsim cloud analyst
OLB	Calculate makespan	–	–
MET	Expected time to compile	–	–
Min-Min [21]	Execution time	The Min-Min is better from RR, FCFS	Workflowsim
Max-Mn [21]	Execution time	The Max-Min is better from RR. FCFS	Workflowsim
The efficient and enhanced algorithm [23]	Response time	–	Cloudanalyst
Ant Colony Optimization [3]	Completion time	The ant colony is better from GA and PSO	Cloudsim
Heterogeneous load balancing algorithm [1]	Memory, storage, MIPS	–	–
Honey Bee galvanized load balancing formula (HBBLB) [22]	Makespan, response time	–	Cloudsim
Static load balancing algorithm [9]	ORT (Overall response lime) and MIPS	Static load balancing algorithm performance, from Throttled load balancing algorithm	Cloudanalyst

4 Result Analysis

In this paper, the performance of few algorithms in CloudSim environment is analyzed. In this algorithm analysis, static task list of total of 800 tasks in the list is considered. The VM list is 200 and range from 20 to 200 in intervals of 20. Figure 3 shows that MCT load balancing algorithm is the best one among the compared five algorithms [1]. The chart is showing MCT algorithm makespan, and energy consumption is least or minimized.

5 Conclusion

In this paper, load balancing algorithms in cloud computing have been studied and analyzed. Here, the discussion is based on various types of load balancing algorithms and analyzed the performance of these various types of algorithms. The performance

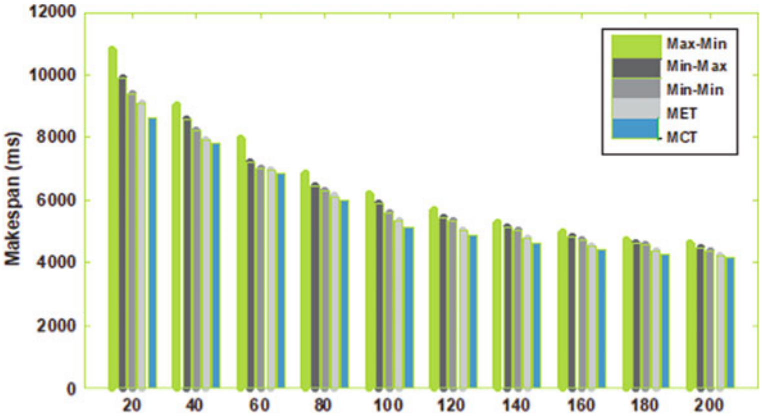


Fig. 3 Virtual machine

of MCT load balancing algorithm is the best one among the compared five load balancing algorithms.

References

1. Singh A, Bedi R, Gupta S (2014) Design and implementation of an efficient scheduling algorithm for load balancing in cloud computing. Int J Emerg Trends Technol Comput Sci (IJETTCS) 3(1)
2. Al-Ayyoub M, Daraghmeh M, Jararweh Y, Althebyan Q (2016) Towards improving resource management in cloud systems using a multi-agent framework. Int J Cloud Comput 5(1–2):112–133
3. Buyya R, Calhetros RN, Son J, Dastjerdi AV, Yoon Y (2014) Software-defined cloud computing: architectural elements and open challenges. In: International conference on advances in computing, communication and informatics (ICACCI). IEEE
4. Sethi S, Anupama S, Jena KS (2012) Efficient load balancing in cloud computing using fuzzy logic. IOSR J Eng (IOSRJEN) 2(7):65-71
5. Dam S, Mandal G, Dasgupta K, Dutta P (2015) Genetic algorithm and gravitational emulation based hybrid load balancing strategy in cloud computing. In: Third IEEE international conference on computer, communication, control and information technology (C3IT), pp 1–7
6. Alharbi F, Rabigh KSA (2012) Simple scheduling algorithm with load balancing for grid computing. Asian Trans Comput 2(2):8–15
7. Glover F, Laguna M (2013) Tabu search. Springer, New York, pp 3261–3362
8. Hassan N, Fernando X (2019) Interference mitigation and dynamic user association for load balancing in heterogeneous networks. IEEE Trans Veh Technol 68(8):7578–7592
9. Agarwal AK, Raj A (2015) A new static load balancing algorithm in cloud computing. Int J Comput Appl 132(2):0975–8887
10. Garg S, Dwivedi RK, Chauhan H (2015) Efficient utilization of virtual machines in cloud computing using synchronized throttled load balancing. In: 1st IEEE international conference on next generation computing technologies (NGCT), pp 77–80
11. Hadji M, Zeghlache D (2012) Minimum cost maximum flow algorithm for dynamic resource allocation in clouds. In: IEEE 5th international conference on cloud computing (CLOUD), pp 876–882

12. Chien NK, Son NH, Loc HD (2016) Load balancing algorithm based on estimating finish time of services in cloud computing. In: 2016 18th IEEE international conference on advanced communication technology (ICACT), pp 228–233
13. Li X, Mao Y, Xiao X, Zhuang Y (2014) An improved max-min task- scheduling algorithm for elastic cloud. In: IEEE international symposium on computer, consumer and control (IS3C), pp 340–343
14. Hsiao HC, Chung HY, Shen H, Chao YC (2013) Load rebalancing for distributed file systems in clouds. IEEE Trans Parallel Distrib Syst 24(5):951–962
15. Kanakala V, RaviTeja V, Reddy K, Karthik K (2015) Performance analysis of load balancing techniques in cloud computing environment. In: IEEE international conference on electrical, computer and communication technologies (ICECCT), pp 1–6
16. Kim SI, Kim HT, Kang GS, Kim JK (2013) Using DVFS and task scheduling algorithms for a hard real-time heterogeneous multicore processor environment, In: Proceedings of the 2013 workshop on energy efficient high performance parallel and distributed computing. ACM, pp 23–30
17. Dasgupta K, Mandal B, Dutta P, Mandal JK, Dam S (2013) A genetic algorithm (GA) based load balancing strategy for cloud computing. Procedia Technol 10:340–347
18. Larumbe F, Sanso B (2013) At tabu search algorithm for the location of datacenters and software components in green cloud computing networks. IEEE Trans Cloud Comput 1(1):22–35
19. AlShawi IS, Yan L, Pan W, Luo B (2012) Lifetime enhancement in wireless sensor networks using fuzzy approach and a-star algorithm. IEEE Sens J 12(10):3010–3018
20. Deng Y, Lau RW (2014) Dynamic load balancing in distributed virtual environments using heat diffusion. ACM Trans Multimed Comput Commun Appl (TOMM) 10(2):16
21. Kashyap D, Viradiya J (2014) A survey of various load balancing algorithms in cloud computing. Int J Sci Technol Res 3(11)
22. Rathore M, Rai S, Saluja N (2015) Load balancing of virtual machine using honey bee galvanizing algorithm in cloud. Int J Comput Sci Inf Technol (IJCSIT) 6(4)
23. Sharma T, Banga VK (2013) Proposed efficient and enhanced algorithm in cloud computing. Int J Eng Res Technol (IJERT) 2(2)

A Comparative Study of Text Classification and Missing Word Prediction Using BERT and ULMFiT

Praveenkumar Katwe, Aditya Khamparia, Kali Prasad Vittala, and Ojas Srivastava

Abstract We perform a comparative study on the two types of emerging NLP models, ULMFiT and BERT. To gain insights on the suitability of these models to industry-relevant tasks, we use Text classification and Missing word prediction and emphasize how these two tasks can cover most of the prime industry use cases. We systematically frame the performance of the above two models by using selective metrics and train them with various configurations and inputs. This paper is intended to assist the industry researchers on the pros and cons of fine-tuning the industry data with these two pre-trained language models for obtaining the best possible state-of-the-art results.

Keywords Text classification · NLP · ULMFiT · BERT · Missing word prediction

1 Introduction

Classification is a pervasively faced computing problem that encircles many diverse applications. It is defined as a task of assigning of objects or labeling of objects into one of the available predefined set of categories. In NLP, it is used for *text classification* where a text or part of the text is classified with a label or group name. Here are few industrial applications of text classification: "Categorizing customer issues using support communication," "Understanding the state of a project from

P. Katwe (✉) · K. P. Vittala · O. Srivastava
Informatica Business Solutions, Bangalore, India
e-mail: pkumark@informatica.com; Praveenkumar.41800775@lpu.co.in

K. P. Vittala
e-mail: kprasad@informatica.com

O. Srivastava
e-mail: osrivastava@informatica.com

P. Katwe · A. Khamparia
Department of Computer Science, Lovely Professional University, Jalandhar, India
e-mail: Aditya.17862@lpu.co.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_46

ongoing communication," "spam detection," etc. The performance of a model in this task identifies the model's capability to label a piece of text.

Missing Word prediction is an NLP task that has given a partial sentence, where the missing words would be identified in order to complete the sentence and understand its intended meaning. This task is not among the most popular tasks; however, it has many wide use cases in the industry, especially in the Data Quality and Data Governance domain like 'correcting incomplete data'. The performance of a model on this task helps to understand how well the model has related individual words with entire sentence structures.

Together the above two tasks help us gauge a model's suitability for industrial applications. The paper describes the methodology of dataset preparation, language model creation, task-based fine-tuning, and metrics selection. Next, we expound on the implementation method and libraries used for this paper. With the results obtained across multiple iterations and combinations, with the selected metrics, we conclude the pros and cons of the two models. The code and preprocessed datasets can be found at https://github.com/oj-srivastava/BERTvsULMFit.

2 Related Work

This research entails the study of ULMFiT and BERT models and the transfer-based learning techniques they employ. The following are brief details of the internal working of the two language models.

ULMFiT: The ULMFiT [1] is a model proposed predominantly for text classification. This model is based on ASGD Weight-Dropped LSTM [2]. It contains three phases where each phase is a stack of multiple layers. The first phase is the global domain language model training, with a word embedding layer, followed by three stacked layers of LSTM and a SoftMax layer in the end. The second phase is target task-based fine-tuning where the embeddings from the previous phase are passed over a set of weights to suit current dataset through a process called freezing. This is further fine-tuned using the discriminative fine-tuning and slanted triangular learning rates to learn task-specific features. In the third phase, the target task classifier is fine-tuned by gradual unfreezing and slanted triangular learning rates to preserve contextual representation. It contains three stacked layers of LSTM followed by a ReLu layer. ULMFiT uses a technique called concat pooling, where mean and max representation from the pool are concatenated to the last hidden state.

BERT: BERT uses an encoder to understand the language model and a decoder with attention, and to gain the needed weights from the encoder to correctly make predictions. The primary advantage of BERT from its predecessors is its bidirectional behavior. To achieve bidirectionality, BERT uses the following two tasks—*Masked Language Model*: Performs random masking of the tokens from the input and predicts the original word's vocabulary id based only on its context. The MLM technique allows the representation to freeze the left and the right context, which allows the language model to pre-train a deep bidirectional transformer [3]. *Next*

Sequence Prediction: Adds bidirectionality to allow finding the relationships between sentences.

3 Method

3.1 Dataset

Dataset selection: In this article, we are using the BBC news categories dataset [4]. It contains 2225 news articles published on bbc.co.uk website during the time duration 2004–2005. The set of 2225 articles are categorized into 5 labels. (business, entertainment, politics, sport, tech). We have used the raw format files so that we can customize the preprocessing procedure for the data. We categorized the dataset into three groups. Large dataset containing 1319 training data and 330 validation data. Medium dataset containing 792 articles for training data and 198 articles for validation. Small dataset containing 527 articles for training and 132 for validation.

3.2 Dataset Preprocessing

Before using the data for the study, we process the data as depicted in Fig. 1.

Dataset cleaning: In this phase, we clean the data to make sure it does not contain irrelevant symbols not expected to be part of the vocabulary. As we are using BERT uncased, we ensure that the entire corpus is converted to lower case.

Character cleaning: All the special characters are removed from the dataset, retaining only characters between A-Z, a-z.

Case normalization: All the characters in the text are changed to lower case to ensure all the text have the same case.

Stop word removal: In this step, we remove the stop words to reduce the noise.

Tokenization: In this process, the text is broken down into pieces such as identifiable words. These pieces are called tokens [5].

Data trimming: In this process, the extra spaces and textual noise are removed from the data.

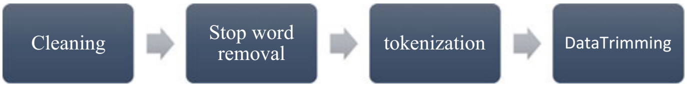


Fig. 1 Data preprocessing stages

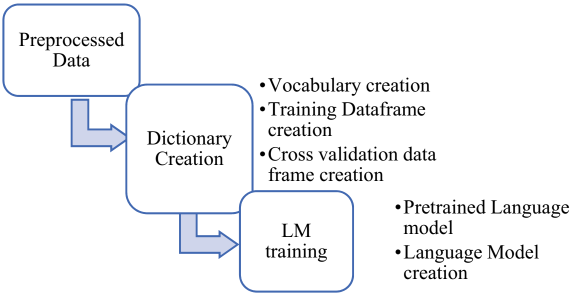

Preprocessed Data
Dictionary Creation
•Vocabulary creation
•Training Dataframe creation
•Cross validation data frame creation
LM training
•Pretrained Language model
•Language Model creation

Fig. 2 Language model creation stages

3.3 Language Model Creation

In this phase, the preprocessed data is used to train the language models. The process can be understood in Fig. 2.

This step contains two phases:

Dictionary/Vocabulary creation—A vocabulary is created from the preprocessed data obtained in the previous step. This vocabulary is used as an index to create the training and test datasets which can be passed to the language model trainer methods.

Language model training—This is the step where we use the state-of-the-art architectures of BERT or ULMFiT to train models on our datasets. The language model thus created in this phase can be saved to storage which can be directly used for performing the task-based fine-tuning.

3.4 Task-Based Fine-tuning

In this phase, the trained language models are tweaked to NLP tasks to be performed using them. This phase exposes multiple global and task-based parameters, variations of which can be used to test the underlying models' capabilities. As pointed out before this research focuses on two NLP tasks:

Text classification fine-tuning: This task involves using the language model to train a text classifier. The parameters involved here are Dataset size, accuracy thresholds, batch size.

Missing word prediction (MLM) fine-tuning: This task involves training the model to find missing words in a sentence. The parameters involved would be Dataset size, epochs for fine-tuning runs, and segment lengths.

Apart from the above task-specific parameters, we also consider the Global parameters: Parameters whose multiple variations were used in this study—Max learning rate, Number of language model epochs.

3.5 Metrics Selection

The metrics selection is based on the approach explained in the SuperGLUE benchmark [6].

Accuracy: Accuracy is the degree of correctness.

AUC ROC valuation: Area under curve under the Receiver Operating Characteristic (ROC) summarizes the trade-off between the true-positive rate and false-positive rate for a predictive model using different probability thresholds [7].

$$\mathrm{A} = \int_{-\infty}^{\infty} \mathrm{TPR}\big(\mathrm{FPR}^{-1}(\mathrm{x})\big)\mathrm{dx} \tag{1}$$

Where TPR(T) is the true-positive rate (y-axis in ROC graph) and FPR(T) is the false-positive rate (x-axis in ROC graph)

Recall: It is the ability of a model to find all the relevant cases within a dataset [8].

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \tag{2}$$

Precision: The ability of a model to find relevant cases without false-positive results.

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \tag{3}$$

F1 Ratio: It is the is the harmonic mean of recall and precision

$$\text{F1} = \frac{2 * \text{Recall} * \text{precision}}{\text{recall} + \text{precision}} \tag{4}$$

F-Beta: It is the weighted harmonic mean computed using the weighted average of the inverse of the values of precision and recall.

$$\mathrm{F}_{\text{Beta}} = \left(1 + \beta^2\right) \cdot \frac{\text{precision} \cdot \text{recall}}{\left(\beta^2 \cdot \text{precision}\right) + \text{recall}} \tag{5}$$

4 Implementation

The data is read from the same excel sources using the pandas' library for loading and preprocessing. These data frames are then passed to the models for training and validation.

BERT implementation: We used the "bert-base-uncased" [3] model from the pytorch_pretrained_bert package. Even though BERT is not present in the standard fastai packages, to keep our comparative study uniform and unaffected by library implementations, we created our own BertDataBunch class on top of TextDataBunch from fastai.text to call our implementations of fastai_tokenizer and fastai_bert_vocab. The fastai_bert_vocab uses the BERT vocabulary to create a dictionary which is then used by the fastai_tokenizer to tokenize the dataset in the format that BERT model expects. In the tokenizer, we also implement our own tokenization logic to preprocess data (cleaning, removing stop words, tokenizing based on vocabulary, and removing white spaces). We use BertForSequenceClassification from pytorch_pretrained_bert.modeling with nn.CrossEntropyLoss for text classification and BertForMaskedLM for Missing word prediction [3, 9].

ULMFiT implementation: We use the fastai TextLMDataBunch with AWD_LSTM architecture for Missing word prediction and TextClasDataBunch with AWD_LSTM architecture for text classifier. We train the language_model_learner and fine-tune the same to text_classifier_learner for the two NLP tasks [1, 10].

Both implementations use the Learner class from fastai package to create the models and train/fine-tune them using the fit_one_cycle method which is fastai implementation of one cycle training approach. Both models use fastai's implementation of transformers. The metrics are calculated using the fastai.metrics and sklearn.metrics libraries.

5 Results

We ran different combinations of global and task-based parameters and hyperparameters and calculated the previously mentioned metrics to measure models' performance on the two NLP tasks.

Text classification Accuracy: As shown in Fig. 3 the accuracy values for ULMFiT for the three datasets vary in a small range; while for BERT model, the largest dataset has reached the optimum accuracy value in a minimum number of epochs itself (Fig. 4).

Text classification F1 Score: While the f1 score values for BERT model optimize after enough runs, with the smallest score for the smallest dataset and tantamount values for the medium and largest dataset, the ULMFiT model values for the three datasets still fluctuate and are expected to optimize after much higher epochs. It is important to note that the values for F1 score in ULMFiT for the largest and smallest dataset are almost equal, even for the minimum epochs.

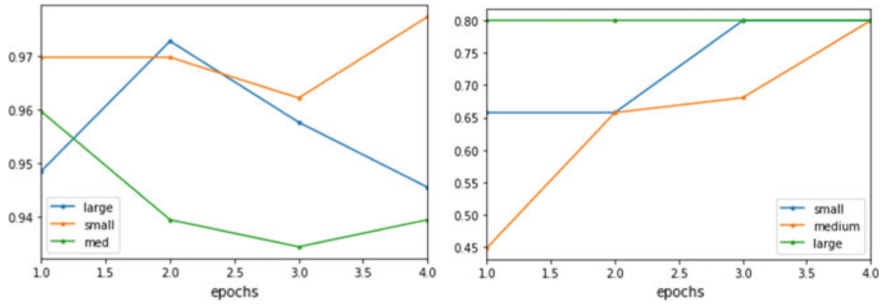


Fig. 3 (left) ULMFiT accuracy for text classification. (right) Bert accuracy for text classification

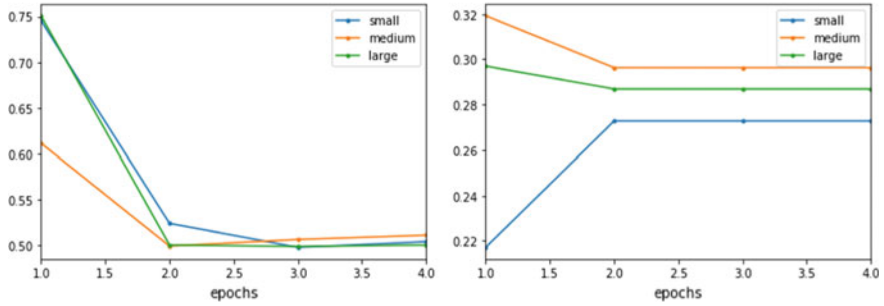


Fig. 4 (left) ULMFiT F1 for text classification. (right) Bert F1 for text classification

Text classification AUC_ROC: The Area under the curve for ROC function varies more in ULMFiT than BERT. For both the models, the values converge as we increase the epochs denoting the sensitivity to threshold values in the model at lower epochs. This sensitivity is much higher in ULMFiT (Fig. 5).

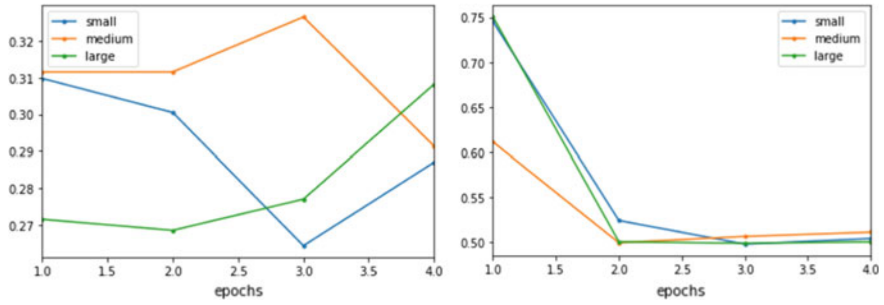


Fig. 5 (left) ULMFiT AUC_ROC for text classification. (right) BERT AUC_ROC for text classification

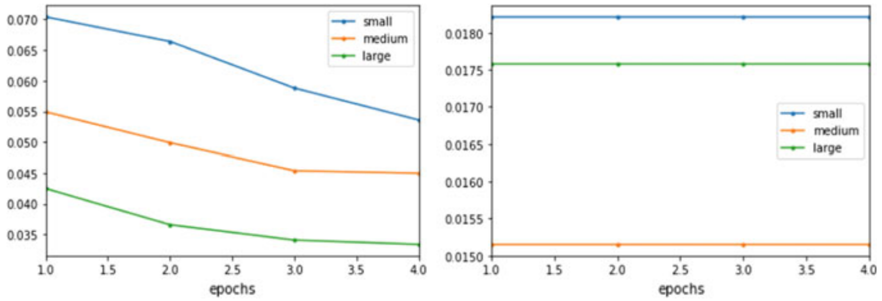


Fig. 6 (left) ULMFiT accuracy for MLM. (right) BERT accuracy for MLM

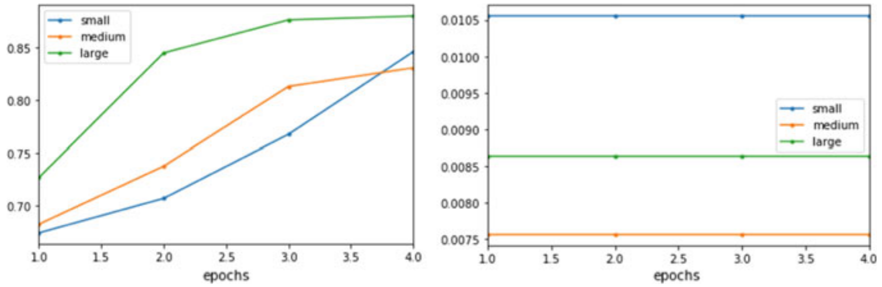


Fig. 7 (left) ULMFiT F-Beta for MLM. (right) Bert F-Beta for MLM

MLM Accuracy: The accuracy values for MLM in ULMFiT start at different values for the three datasets converging with the increase in epochs. Increasing epochs for fine-tuning the BERT model has no effect on MLM accuracy (Fig. 6).

MLM F-Beta: The F-Beta for ULMFiT, in the beginning, is much higher for the largest dataset than the other datasets and converges with increasing epochs. The BERT F-beta for MLM stays in a close range, and is not affected by changing epochs (Fig. 7).

6 Conclusion

In this study, we trained and fine-tuned BERT and ULMFit models on the same combinations of datasets and training configurations to perform Missing word prediction and Text classification. Their performance on these tasks was judged using selective metrics and their suitability to various industrial use-cases derived.

It was observed that ULMFiT, though performing better under standard computational and dataset provisions, gets superseded by BERT in case of larger datasets and a higher number of epochs. So, BERT has proven to be a much more viable option in case of industry use cases where an abundance of data and computational power to apply are available. BERT is also the better option when we have to perform a text classification task with tough to distinguish classes. On the other hand, ULMFiT would give an equivalent performance across varying sizes of dataset given enough training cycles. F-beta for MLM shows ULM fit's capability to generate good results with an even smaller dataset and BERT's dependence on dataset size to optimize its performance. Industrial use cases would often have the capability to provide higher training cycles and computational power. At the same time, the availability of large labeled datasets for training is not readily available. In these scenarios, ULMFiT would be a better model to adopt a language set and perform Natural language tasks on it.

7 Future Scope

We intend to further proceed with this research by including other innovative models like RoBERTa [11] and ALBERT [12]. Currently, the state of the art in NLP is captured using a model's performance over ideal training datasets and configurations [6]. These rankings do not consider limitations in an actual business use-case. This study can be further extended to rank and suit the various state-of-the-art models to different industrial use cases by taking into account the limitations to datasets, computational power, and optimizations in implementation. We can use additional tasks and configurations to categorize multiple industrial scenarios and with the metrics used to provide an extensive report on how each model adapts to these tasks and what trade-offs one need to make in order to select the most suitable architecture for a certain use case.

Acknowledgments We would like to thank Informatica Business Solutions Bangalore for proving the facilities needed for this research. We also extend our gratitude to Lovely Professional University mentors for providing valuable guidance throughout the research work.

References

1. Howard J, Ruder S (2018) Universal language model fine-tuning for text classification. arXiv preprint arXiv:1801.06146
2. Merity S, Keskar NS, Socher R (2017) Regularizing and optimizing LSTM language models. arXiv preprint arXiv:1708.02182
3. Devlin J et al (2018) Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805
4. Greene D, Cunningham P (2006) Practical solutions to the problem of diagonal dominance in kernel document clustering. In: Proceedings of the 23rd international conference on machine learning. ACM
5. Webster JJ, Kit C (1992) Tokenization as the initial phase in NLP. In: COLING 1992 Volume 4: The 15th international conference on computational linguistics

6. Wang A et al (2019) Superglue: a stickier benchmark for general-purpose language understanding systems. arXiv preprint arXiv:1905.00537
7. Narkhede S (2018) Understanding AUC-ROC curve. Towards Data Sci 26
8. Beyond accuracy: precision and recall. https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c
9. Fastai integration with BERT: multi-label text classification identifying toxicity in texts: https://medium.com/@abhikjha/fastai-integration-with-bert-a0a66b1cecbe
10. Tutorial for Text classification using ULMFiT and fastai library in python https://www.analyticsvidhya.com/blog/2018/11/tutorial-text-classification-ULMFiT-fastai-library/
11. Liu Y et al. (2019) Roberta: a robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692
12. Lan Z et al (2019) Albert: a lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942

Data Formats and Its Research Challenges in IoT: A Survey

Sandeep Mahanthappa and B. R. Chandavarkar

Abstract In the current context, the increase in data generated by the heterogeneous IoT sensors results in several challenges to the IoT applications' developers. One of the significant problems is handling nonstandardized zeta bytes of heterogeneous IoT data. Further, there is a lack of understanding of data generated by the sensors and incompatibility in representing the data. The objective of this survey paper is to introduce IoT sensors with their areas of applications, the different data formats, representing the data in an interchangeable format, and handling data as a stream. Further, this paper also presents challenges and research opening in IoT data.

Keywords IoT data · Data stream · Open data format (O-DF)

1 Introduction

From much research, it is evident that the growth of IoT devices is happening at a very high pace and reaching more than 75 billion[1] in this decade. According to a new IDC forecast,[2] the data generated by these devices is expected to reach around 80 zeta byte by 2025. IoT data is the output of a device or a process associated with an application, which is a physical quantity from its environment. Sensors in an IoT environment are employed to sense the event and capture sensed value.

The IoT services can be realized to the full extent only when the huge data generated by multiple devices in the IoT environment needs to be understood, collected, and analyzed to generate the required knowledge for the applications. In this paper, an effort is made to identify different types of sensors used in IoT, data format generated

[1] https://www.statista.com/statistics/471264/iot-number-of-connected-devicesworldwide.

[2] https://www.idc.com/getdoc.jsp?containerId=prUS45213219.

S. Mahanthappa (✉) · B. R. Chandavarkar
Department of Computer Science and Engineering, Wireless Information Networking Group (WiNG), National Institute of Technology Karnataka, Surathkal, Mangalore, India
e-mail: maxniroop@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_47

by them, different data representation methods for analysis, and research challenges in IoT data.

The rest of the paper is structured as below. Section 2 contains details of major IoT sensors and their application areas. Section 3 describes about IoT data formats. Section 4 specifies IoT data taxonomy and data stream. Section 5 talks about different data reduction methods. Section 6 lists challenges with IoT data and research openings. Final Sect. 7 concludes the paper.

2 Sensors and Their Areas of Applications

The section concentrates on different sensor details and their areas of application in IoT.[3],[4],[5]

- Accelerometer: detects the change in gravitational acceleration usually used for detecting vibration, tilt, acceleration.
- Proximity sensor: used for detecting the presence or absence of an object without contact, usually used in retail shops, reversing the vehicle, and parking sensor.
- PIR motion: a passive infrared sensor to detect the presence of humans or animals. The sensor detects the physical movement of the device/human in a given area.
- Level sensor: used to find the level or amount of liquid in an open or closed environment.
- Water quality sensor: for measuring the quality of water and monitoring ion levels in it.
- Transducers: measure parameters like temperature, humidity, pressure, wind direction, speed, illumination intensity, sound intensity, chemical concentration, pollution level, vital body function [1–3].

Some categories of sensors are combined to understand the IoT scenario characters properly; for example, the ambient sensor includes temperature, pressure, and humidity. Motion sensors acquire data from the movement of the object, which involves accelerometer and gyroscope. Electric sensors measure the energy consumption of an environment. Electrocardiogram (ECG), heartbeat measurement, and breathing sensors are part of biosensors. RFID and NFC are employed for entity identification. GPS and magnetometer are combinedly used for position detection [4].

IoT has become an inseparable space in everyday life, smart homes, wearable devices, connected buildings, smart health care, smart industries, connected roads, and so on. The growth in the numbers of sensors and connected smart objects have contributed to data increase at a very high pace. More data means new challenges

[3] https://internetofthingsagenda.techtarget.com/definition/sensor-data.

[4] https://iot4beginners.com/commonly-used-sensors-in-the-internet-of-things-iotdevices-and-their-application/.

[5] https://dzone.com/articles/sensorwithspecificationiniot.

need to be attended; the primitive one like preparing the data to harness knowledge from it. The top 3 priority challenges in data preparation, as identified by the author Sean [5], are a high volume of data, complexity, and interoperability.

According to authors Erick et al. [6], the attributes of the data are type, owner, location, and read value. In continuation, author Anna Gerber [7] and authors Hans et al. [8] specify that data sensed by an IoT device is a mixture of structured, semi-structured, and unstructured data.

The structured data is represented according to some model or schema, and it can easily be associated with traditional RDBMS. Structured data is represented as tabular representation, like a spreadsheet where each cell is explicitly defined and referred. Most of the computing systems like bank transactions and computer log make use of structured data. IoT sensors represent the data like temperature, humidity, pressure, and other as structured data. Structured data can be easily formatted, queried, and processed to use in decision-making.

The unstructured data does not follow any logical schema or any predefined data model for representation, so the traditional methods used for understanding and processing can not work for this data, for example, text, speech, image, and video.

The semi-structured data is the hybrid of structured and unstructured data and share the characters of both. Email is one of the good examples of semi-structured data where fields are predefined, but the content of the body and attachment is unstructured.

Data like discrete sensor readings, metadata about a device, files for image and video are part of the heterogeneous data generated. As the IoT device is built with low storage capabilities, the high volume of data acquired by them needs to be transmitted using communication protocols for further processing and storing. IoT should deal with a high volume of data and also give importance to the following issues:

- Handling heterogeneous data,
- Preparing data for the analysis by transforming,
- Aggregating, integrating, and keeping track of data origin,
- Preserving integrity and privacy of the data,
- Choosing storage that can balance between performance, reliability, flexibility, and cost. [7].

Apart from the above issues, author Fredric [9] specifies that the high volume of IoT data is processed at a high rate. The processing of the data should exercise closer to the event environment to avoid delay and loss of data.

3 Major IoT Sensor Data Formats

The major data formats generated by IoT sensors and applications are Text, Binary, XML, CSV, JSON, and RFID. The data in IoT depends on the type of sensor and the developer's interest. The sensor is connected with an application that demands less detailed data; IoT uses simple data formats like text and binary; for example, room temperature automation. Whereas for sensors connected to smart devices and applications the requirement is greater details in data; IoT tends to choose encoded data formats like XML, JSON, and CSV, for example, PTC things Worx, Arrowhead, OpenIoT [10]. IoT data includes device status, metadata about the device, and captured data. The data generated by IoT is not uniform, so a single representation of data for all the applications is difficult [7].

3.1 Text

Text data is the human-readable sequence of characters other than non-character encoded data such as graphic images, audio, and video. The IoT sensor captures the data from their environment and represents the data in the text format. Examples of data sensed by a temperature sensor on the floor, ceiling, and bedside of a hotel room provide the output as single line textual data with device identification, location of the device, environment, and read temperature data [11, 12].

deviceID: “aee62681aa9b”, “location”: “floor”, “room”: 205, “temp”: 21
deviceID: “792d3a3ef366”, “location”: “ceiling”, “room”: 205, “temp”: 25
deviceID: “b7c96bd32435”, “location”: “bedside”, “room”: 205, “temp”: 24.

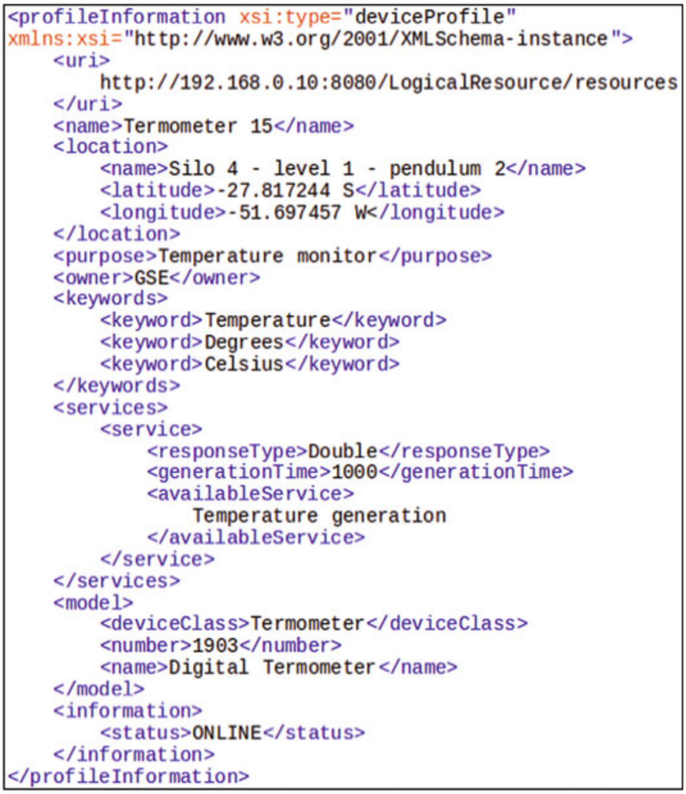
3.2 XML

Extensible Markup Language is a meta markup language, and is one of the preferred data formats on the world wide web. Cross-domain application IoT deployment faces the constraint of inter-domain data format. XML is one such language that solves the issue to some extent. XML is the human-readable representation of device information and sensed data. XML-based description of sensors and measurement process and encoding could be done by SensorML [12–14].

The example in Fig. 1 specifies an XML file format showing the web resource identity, name of the sensor, location of the sensor, the purpose of the device, owner, keywords associated, services like response type, frequency of data, device class, and status of the device.

3.3 CSV

Comma-separated values file is a text file where data values are delimited by comma, or represented as excel sheet values for easy access of data for processing. Each line in the CSV file is termed as one record which specifies sensed data as one sample. Each record will have values separated by delimiter as the comma.

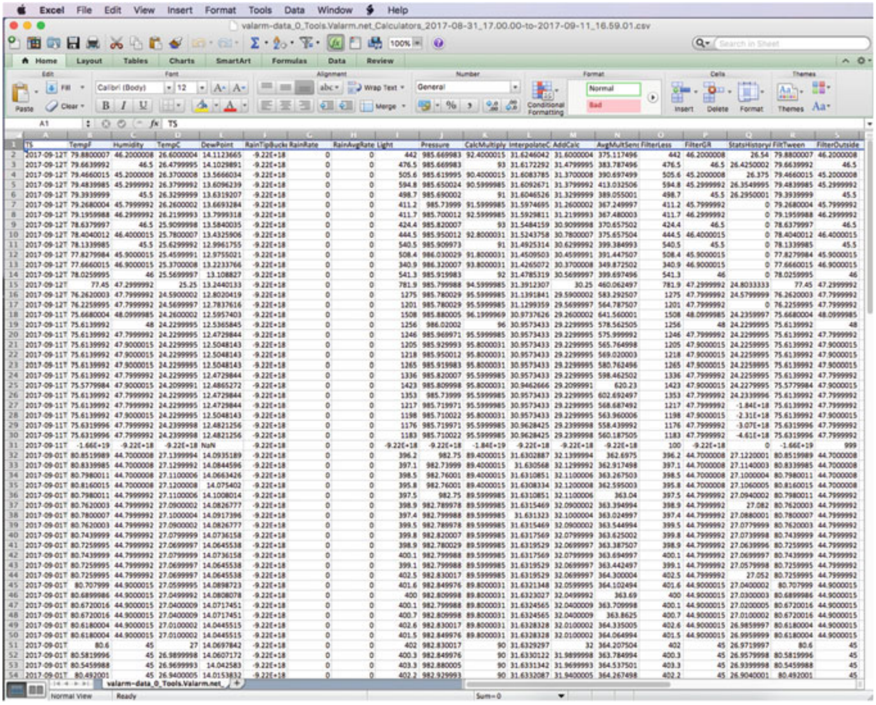


```xml
<profileInformation xsi:type="deviceProfile"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
    <uri>
        http://192.168.0.10:8080/LogicalResource/resources
    </uri>
    <name>Termometer 15</name>
    <location>
        <name>Silo 4 - level 1 - pendulum 2</name>
        <latitude>-27.817244 S</latitude>
        <longitude>-51.697457 W</longitude>
    </location>
    <purpose>Temperature monitor</purpose>
    <owner>GSE</owner>
    <keywords>
        <keyword>Temperature</keyword>
        <keyword>Degrees</keyword>
        <keyword>Celsius</keyword>
    </keywords>
    <services>
        <service>
            <responseType>Double</responseType>
            <generationTime>1000</generationTime>
            <availableService>
                Temperature generation
            </availableService>
        </service>
    </services>
    <model>
        <deviceClass>Termometer</deviceClass>
        <number>1903</number>
        <name>Digital Termometer</name>
    </model>
    <information>
        <status>ONLINE</status>
    </information>
</profileInformation>
```

Fig. 1 Example of IoT data in XML format [25]

Many IoT and other applications support the file format. For example, Tree measurements of data file with 4 records and fields are as follows: index, circumference (in), height (ft), volume (ft^3) [15]. The example in Fig. 2 is a CSV file from Valarm website displaying the records of the floodwater level monitoring sensors in an excel sheet [17].

"Index", "Girth (in)", "Height (ft)", "Volume (ft^3)"
1, 8.3, 70, 10.3
2, 8.6, 65, 10.3
3, 8.8, 63, 10.2
4, 10.5, 72, 16.4
Example of CSV file.


Excel File Edit View Insert Format Tools Data Window Help
valarm-data_0_Tools.Valarm.net_Calculators_2017-08-31_17.00.00-to-2017-09-11_16.59.01.csv
Home Layout Tables Charts SmartArt Formulas Data Review
Calibri (Body) 12
General
Normal
Bad
A1 TS
valarm-data_0_Tools.Valarm.net_
Normal View Ready
Sum=0

TS	TempF	Humidity	TempC	DewPoint	RainTipBucke	RainRate	RainAvgRate	Light	Pressure	CalcMultiply	InterpolateC	AddCalc	AvgMultSen	FilterLess	FilterGR	StatsHistory	FiltTween	FilterOutside
2017-09-12T	79.8800007	46.2000008	26.6000004	14.1123665	-9.22E+18	0	0	442	985.669983	92.4000015	31.6246042	31.6000004	375.117496	442	46.2000008	26.54	79.8800007	46.2000008
2017-09-12T	79.6639992	46.5	26.4799995	14.1029891	-9.22E+18	0	0	476.5	985.669983	93	31.6172292	31.4799995	383.787496	476.5	46.5	26.4250002	79.6639992	46.5
2017-09-12T	79.4660015	45.2000008	26.3700008	13.5666034	-9.22E+18	0	0	505.6	985.619995	90.4000015	31.6083785	31.3700008	390.697499	505.6	45.2000008	26.375	79.4660015	45.2000008
2017-09-12T	79.4839985	45.2999992	26.3799992	13.6096239	-9.22E+18	0	0	594.8	985.650024	90.5999985	31.6092671	31.3799992	413.032506	594.8	45.2999992	26.3549995	79.4839985	45.2999992
2017-09-12T	79.3939999	45.5	26.3299999	13.6319207	-9.22E+18	0	0	498.7	985.690002	91	31.6046526	31.3299999	389.055001	498.7	45.5	26.2950001	79.3939999	45.5
2017-09-12T	79.2680004	45.7999992	26.2600002	13.6693284	-9.22E+18	0	0	411.2	985.73999	91.5999985	31.5974695	31.2600002	367.249997	411.2	45.7999992	0	79.2680004	45.7999992
2017-09-12T	79.1959988	46.2999992	26.2199993	13.7999318	-9.22E+18	0	0	411.7	985.700012	92.5999985	31.5929811	31.2199993	367.480003	411.7	46.2999992	0	79.1959988	46.2999992
2017-09-12T	78.6379997	46.5	25.9099998	13.5840035	-9.22E+18	0	0	424.4	985.820007	93	31.5484159	30.9099998	370.657502	424.4	46.5	0	78.6379997	46.5
2017-09-12T	78.4040012	46.4000015	25.7800007	13.4325906	-9.22E+18	0	0	444.5	985.950012	92.8000031	31.5243758	30.7800007	375.657504	444.5	46.4000015	0	78.4040012	46.4000015
2017-09-12T	78.1339985	45.5	25.6299992	12.9961755	-9.22E+18	0	0	540.5	985.909973	91	31.4925314	30.6299992	399.384993	540.5	45.5	0	78.1339985	45.5
2017-09-12T	77.8279984	45.9000015	25.4599991	12.9755021	-9.22E+18	0	0	508.4	986.030029	91.8000031	31.4509503	30.4599991	391.447507	508.4	45.9000015	0	77.8279984	45.9000015
2017-09-12T	77.6660015	46.9000015	25.3700008	13.2233766	-9.22E+18	0	0	340.9	986.320007	93.8000031	31.4265072	30.3700008	349.872502	340.9	46.9000015	0	77.6660015	46.9000015
2017-09-12T	78.0259995	46	25.5699997	13.108827	-9.22E+18	0	0	541.3	985.919983	92	31.4785319	30.5699997	399.697496	541.3	46	0	78.0259995	46
2017-09-12T	77.45	47.2999992	25.25	13.2440133	-9.22E+18	0	0	781.9	985.799988	94.5999985	31.3912307	30.25	460.062497	781.9	47.2999992	24.8033333	77.45	47.2999992
2017-09-12T	76.2620003	47.7999992	24.5900002	12.8020419	-9.22E+18	0	0	1275	985.780029	95.5999985	31.1391841	29.5900002	583.292507	1275	47.7999992	24.5799999	76.2620003	47.7999992
2017-09-12T	76.2259995	47.7999992	24.5699997	12.7837616	-9.22E+18	0	0	1201	985.780029	95.5999985	31.1299359	29.5699997	564.787507	1201	47.7999992	0	76.2259995	47.7999992
2017-09-11T	75.6680004	48.0999985	24.2600002	12.5957403	-9.22E+18	0	0	1508	985.880005	96.1999969	30.9737626	29.2600002	641.560001	1508	48.0999985	24.2359997	75.6680004	48.0999985
2017-09-11T	75.6139992	48	24.2299995	12.5365845	-9.22E+18	0	0	1256	986.02002	96	30.9573433	29.2299995	578.562505	1256	48	24.2299995	75.6139992	48
2017-09-11T	75.6139992	47.7999992	24.2299995	12.4729844	-9.22E+18	0	0	1246	985.969971	95.5999985	30.9573433	29.2299995	575.999992	1246	47.7999992	24.2299995	75.6139992	47.7999992
2017-09-11T	75.6139992	47.9000015	24.2299995	12.5048143	-9.22E+18	0	0	1205	985.929993	95.8000031	30.9573433	29.2299995	565.764998	1205	47.9000015	24.2259995	75.6139992	47.9000015
2017-09-11T	75.6139992	47.9000015	24.2299995	12.5048143	-9.22E+18	0	0	1218	985.950012	95.8000031	30.9573433	29.2299995	569.020003	1218	47.9000015	24.2259995	75.6139992	47.9000015
2017-09-11T	75.6139992	47.9000015	24.2299995	12.5048143	-9.22E+18	0	0	1265	985.919983	95.8000031	30.9573433	29.2299995	580.762496	1265	47.9000015	24.2259995	75.6139992	47.9000015
2017-09-11T	75.6139992	47.7999992	24.2299995	12.4729844	-9.22E+18	0	0	1336	985.820007	95.5999985	30.9573433	29.2299995	598.462502	1336	47.7999992	24.2259995	75.6139992	47.7999992
2017-09-11T	75.5779984	47.9000015	24.2099991	12.4865272	-9.22E+18	0	0	1423	985.809998	95.8000031	30.9462666	29.2099991	620.23	1423	47.9000015	24.2279995	75.5779984	47.9000015
2017-09-11T	75.6139992	47.7999992	24.2299995	12.4729844	-9.22E+18	0	0	1353	985.73999	95.5999985	30.9573433	29.2299995	602.692497	1353	47.7999992	24.2339996	75.6139992	47.7999992
2017-09-11T	75.6139992	47.7999992	24.2299995	12.4729844	-9.22E+18	0	0	1217	985.719971	95.5999985	30.9573433	29.2299995	568.687492	1217	47.7999992	-1.84E+18	75.6139992	47.7999992
2017-09-11T	75.6139992	47.9000015	24.2299995	12.5048143	-9.22E+18	0	0	1198	985.710022	95.8000031	30.9573433	29.2299995	563.960006	1198	47.9000015	-2.31E+18	75.6139992	47.9000015
2017-09-11T	75.6319996	47.7999992	24.2399998	12.4821256	-9.22E+18	0	0	1176	985.719971	95.5999985	30.9628425	29.2399998	558.439992	1176	47.7999992	-3.07E+18	75.6319996	47.7999992
2017-09-11T	75.6319996	47.7999992	24.2399998	12.4821256	-9.22E+18	0	0	1183	985.710022	95.5999985	30.9628425	29.2399998	560.187505	1183	47.7999992	-4.61E+18	75.6319996	47.7999992
2017-09-11T	-1.66E+19	-9.22E+18	-9.22E+18	NaN	-9.22E+18	0	0	-9.22E+18	-9.22E+18	-1.84E+19	-9.22E+18	-9.22E+18	-9.22E+18	100	-9.22E+18	0	-1.66E+19	999
2017-09-01T	80.8519989	44.7000008	27.1399994	14.0935189	-9.22E+18	0	0	396.2	982.75	89.4000015	31.6302887	32.1399994	362.6975	396.2	44.7000008	27.1220001	80.8519989	44.7000008
2017-09-01T	80.8339985	44.7000008	27.1299992	14.0844596	-9.22E+18	0	0	397.1	982.73999	89.4000015	31.630568	32.1299992	362.917498	397.1	44.7000008	27.1140003	80.8339985	44.7000008
2017-09-01T	80.7980011	44.7000008	27.1100006	14.0663426	-9.22E+18	0	0	398.5	982.76001	89.4000015	31.6310851	32.1100006	363.267503	398.5	44.7000008	27.1000004	80.7980011	44.7000008
2017-09-01T	80.8160015	44.7000008	27.1200008	14.075402	-9.22E+18	0	0	395.8	982.76001	89.4000015	31.6308334	32.1200008	362.595003	395.8	44.7000008	27.1060005	80.8160015	44.7000008
2017-09-01T	80.7980011	44.7999992	27.1100006	14.1008014	-9.22E+18	0	0	397.5	982.75	89.5999985	31.6310851	32.1100006	363.04	397.5	44.7999992	27.0940002	80.7980011	44.7999992
2017-09-01T	80.7620003	44.7999992	27.0900002	14.0826777	-9.22E+18	0	0	398.9	982.789978	89.5999985	31.6315469	32.0900002	363.394994	398.9	44.7999992	27.082	80.7620003	44.7999992
2017-09-01T	80.7800007	44.7999992	27.1000004	14.0917396	-9.22E+18	0	0	397.4	982.799988	89.5999985	31.631323	32.1000004	363.024997	397.4	44.7999992	27.0880001	80.7800007	44.7999992
2017-09-01T	80.7620003	44.7999992	27.0900002	14.0826777	-9.22E+18	0	0	399.5	982.789978	89.5999985	31.6315469	32.0900002	363.544994	399.5	44.7999992	27.0779999	80.7620003	44.7999992
2017-09-01T	80.7439999	44.7999992	27.0799999	14.0736158	-9.22E+18	0	0	399.8	982.820007	89.5999985	31.6317569	32.0799999	363.625002	399.8	44.7999992	27.0739998	80.7439999	44.7999992
2017-09-01T	80.7259995	44.7999992	27.0699997	14.0645538	-9.22E+18	0	0	398.9	982.780029	89.5999985	31.6319529	32.0699997	363.387507	398.9	44.7999992	27.0639996	80.7259995	44.7999992
2017-09-01T	80.7439999	44.7999992	27.0799999	14.0736158	-9.22E+18	0	0	400.1	982.799988	89.5999985	31.6317569	32.0799999	363.694997	400.1	44.7999992	27.0699997	80.7439999	44.7999992
2017-09-01T	80.7259995	44.7999992	27.0699997	14.0645538	-9.22E+18	0	0	399.1	982.799988	89.5999985	31.6319529	32.0699997	363.442497	399.1	44.7999992	27.0579998	80.7259995	44.7999992
2017-09-01T	80.7259995	44.7999992	27.0699997	14.0645538	-9.22E+18	0	0	402.5	982.830017	89.5999985	31.6319529	32.0699997	364.300004	402.5	44.7999992	27.052	80.7259995	44.7999992
2017-09-01T	80.707999	44.9000015	27.0599995	14.0898723	-9.22E+18	0	0	401.6	982.849976	89.8000031	31.6321348	32.0599995	364.102494	401.6	44.9000015	27.0400002	80.707999	44.9000015
2017-09-01T	80.6899986	44.9000015	27.0499992	14.0808078	-9.22E+18	0	0	400	982.809998	89.8000031	31.6323027	32.0499992	363.69	400	44.9000015	27.0300003	80.6899986	44.9000015
2017-09-01T	80.6720016	44.9000015	27.0400009	14.0717451	-9.22E+18	0	0	400.1	982.799988	89.8000031	31.6324565	32.0400009	363.709998	400.1	44.9000015	27.0200005	80.6720016	44.9000015
2017-09-01T	80.6720016	44.9000015	27.0400009	14.0717451	-9.22E+18	0	0	400.7	982.809998	89.8000031	31.6324565	32.0400009	363.8625	400.7	44.9000015	27.0100002	80.6720016	44.9000015
2017-09-01T	80.6180004	44.9000015	27.0100002	14.0445515	-9.22E+18	0	0	402.6	982.830017	89.8000031	31.6328328	32.0100002	364.335005	402.6	44.9000015	26.9859997	80.6180004	44.9000015
2017-09-01T	80.6180004	44.9000015	27.0100002	14.0445515	-9.22E+18	0	0	401.5	982.849976	89.8000031	31.6328328	32.0100002	364.064994	401.5	44.9000015	26.9959999	80.6180004	44.9000015
2017-09-01T	80.6	45	27	14.0697842	-9.22E+18	0	0	402	982.830017	90	31.6329297	32	364.207504	402	45	26.9719997	80.6	45
2017-09-01T	80.5819996	45	26.9899998	14.0607172	-9.22E+18	0	0	400.3	982.849976	90	31.6330122	31.9899998	363.784994	400.3	45	26.9579998	80.5819996	45
2017-09-01T	80.5459988	45	26.9699993	14.042583	-9.22E+18	0	0	403.3	982.880005	90	31.6331342	31.9699993	364.537501	403.3	45	26.9399998	80.5459988	45
2017-09-01T	80.492001	45	26.9400005	14.0153832	-9.22E+18	0	0	402.2	982.929993	90	31.6332087	31.9400005	364.267498	402.2	45	26.9040001	80.492001	45

Fig. 2 Example of IoT data in CSV format [26]

3.4 *JSON*

Javascript Object Notation is a lightweight data interchange format. JSON is a comprehensive hierarchical data format supported by many modern applications. Even though it can represent complex data as an object still it is in human-readable format.

JSON is a preferred data format in IoT compared to XML, as JSON is schemaless, JSON supports strings, numbers, boolean, objects, arrays, and a null value. XML increases file size from its header information [16–18]. Below is the example JSON data format of a device with its attributes name and captured value.

"deviceid": "iot123", "temp": 54.98, "humidity": 32.43, "coords": "latitude": 47.615694, "longitude": −122.3359976

3.5 *Rfid*

Radio Frequency Identification System (RFID) helps to identify the objects with tags automatically. The following is the example of RFID tag data, from defense with each section carrying a specific number of bits, the sum bits should be equal to the size of the Tag: Header (8 bits), Filter (4 bits), CAGE Code as ASCII* (48 bits), Serial Number (36 bits). Data is b00811001111 b0040000 t048 2S194 n03612345678901. The final hexadecimal data representation with prefix and suffix of RFID tag are represented in the following 96-bit format [19]:

{XA^ RFW,H ^ FDCF02032533139342DFDC1C35 ^ FS^}

RFID systems are adopted by large companies and have contributed to publishing nID standards and industrial open standard specifications. RFID data stream includes data on RF tags (transponders), RF tag reader (transceivers), electronic product code, which can contain product info and manufacturer number. RFID data consists of tag-ID, reader-ID, timestamp; this information is insufficient, incomplete, and high volume. Advancement has been done to RFID technology "networked RFID"-connecting isolated RFID systems and software via the Internet contributed by AutoID labs, EPC global at GS1.

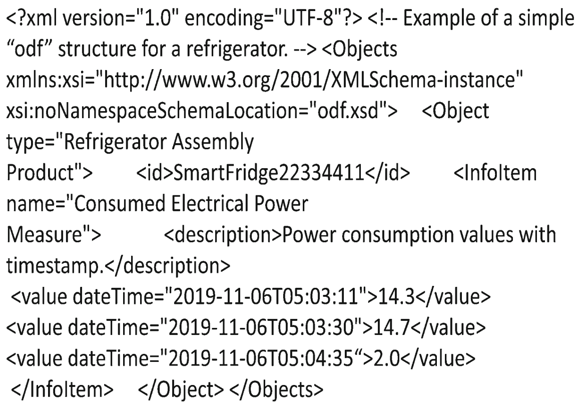
The main requirements for good data format are that it should be easily represented by a resource constraint device, power consumption due to data communication should reduce, and it should be interoperable. XML encoding is the most used data interchange format in most of the applications, but XML adds header information to the data, which increases the size of the data, and the data represented as string values. Even though JSON looks promising compared to XML, it is not space efficient encoding. Most of the data are ASCII encoded with lots of white spaces. The field labels are repeated for every occurrence of data. JSON's simple type of representation does not match with IoT programming types. So a new data encoding format must be explored to overcome this [17, 20].

3.6 Interoperable Format or New Open Data Format (O-DF)

The new open data format is developed to fill the gap of interoperability concerning IoT. The data sources should be able to generate data and provide access to other devices and applications, which securely request them. A URL publishes the data generated and sends or requests other devices when they abide by open management infrastructure. The data represented in a standard O-DF can be universally understood and exchanged by systems that manage IoT related data.

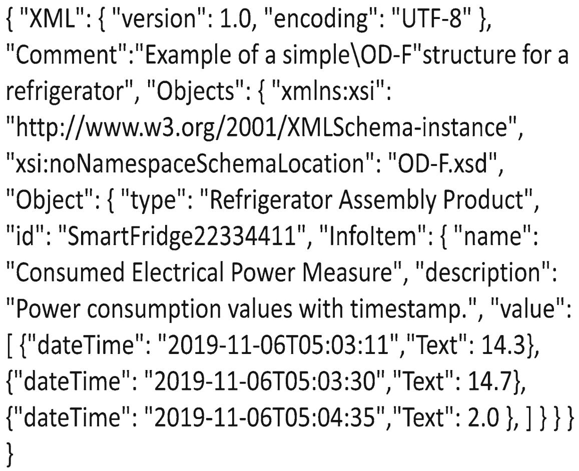
The XML schema used for specifying open format data. The information is created in a similar way of creating objects and properties in object-oriented programming. So this format is generic enough to represent an object and related information for information exchange. The O-DF format will have object hierarchy, where each object will have sub-objects, and sub-object could be device id or other information about the device. The hierarchy can have many levels depending on the details of the information. The example provides measurement values for Refrigerator power consumption in O-DF using XML in Fig. 3 and JSON in Fig. 4 [21].

Fig. 3 OD-F code using XML-sensor data from refrigerator



```
<?xml version="1.0" encoding="UTF-8"?> <!-- Example of a simple
“odf” structure for a refrigerator. --> <Objects
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="odf.xsd">    <Object
type="Refrigerator Assembly
Product">       <id>SmartFridge22334411</id>       <InfoItem
name="Consumed Electrical Power
Measure">           <description>Power consumption values with
timestamp.</description>
 <value dateTime="2019-11-06T05:03:11">14.3</value>
<value dateTime="2019-11-06T05:03:30">14.7</value>
<value dateTime="2019-11-06T05:04:35“>2.0</value>
 </InfoItem>    </Object> </Objects>
```

Fig. 4 OD-F code using JSON-sensor data from refrigerator



```
{ "XML": { "version": 1.0, "encoding": "UTF-8" },
"Comment":"Example of a simple\OD-F"structure for a
refrigerator", "Objects": { "xmlns:xsi":
"http://www.w3.org/2001/XMLSchema-instance",
"xsi:noNamespaceSchemaLocation": "OD-F.xsd",
"Object": { "type": "Refrigerator Assembly Product",
"id": "SmartFridge22334411", "InfoItem": { "name":
"Consumed Electrical Power Measure", "description":
"Power consumption values with timestamp.", "value":
[ {"dateTime": "2019-11-06T05:03:11","Text": 14.3},
{"dateTime": "2019-11-06T05:03:30","Text": 14.7},
{"dateTime": "2019-11-06T05:04:35","Text": 2.0 }, ] } } }
}
```

4 IoT Data Taxonomy and Data Stream

The rapid generation of IoT data is contributing to Big data analytics. The authors, Qin et al. [22] have made an effort to provide information about IoT with a data-centric approach. The authors have come up with three IoT data categories considering its central characters as Data generation, Data quality, and Data interoperability. This representation and their specific characters are represented as data taxonomy.

4.1 Data Taxonomy

- Generation of data: depends on factors as, at what rate samples are generated, coping up with a high amount of data generated, the dynamism of data, and a wide variety of data at a very large rate. Data is dynamic with device generating data of different locations and time; data is fragile (fast-changing) and may be at irregular intervals.
- Data quality: the quality of data depends on uncertainty due to different sources, missing reading, device identification problems, and accuracy. Different consumers perceive data produced by different sensors differently, so ambiguity is unavoidable. Data reading missing, multiple sensors reading data at different accuracy, and loss during transmission give rise to incompetency in data.
- Data interoperability: to produce a good response by the IoT system for an event sensed by the sensor requires data from multiple sources in that environment. Combined data need cooperation between devices; a failure in this situation can result in incompleteness. Processing IoT data by the device themselves, the devices need to understand data generated by humans; this is achieved by adding semantics to the data.

4.2 Data Objects and Data Stream

Data objects are a multidimensional attribute vector within a continuous, categorical, or mixed attribute space. The data stream is a huge sequence of data objects.

Data stream processing The sensor data processing technique involves data aggregation, data compression, modeling, and online querying. Queries can be aggregated to avoid high power consumption. The initial query is executed to produce an intermediate result that can be processed further. Many queries can join to get accurate results. The validity of the top received values is ensured by making mathematical constraints. The quality of the data can improve by adopting error-tolerant methods. As per the demand of application, a query can be continuously processed on sensor data and computation, and storage can be optimized by learning. Query processing over stream data will need to focus on uncertainty, ambiguity, and inconsistency, and it should also address velocity and heterogeneity. A new type of query can be adapted to select the sources which can overcome data incompleteness (which sensor needs to be considered).

Stream mining is performed to extract useful information employing clustering, classification, outlier, and frequent item set mining [9, 22, 23].

5 Data Reduction

The continuous growth of diverse data sources and transmission data has become redundant at the storage and analysis. This has also contributed to network bandwidth problem, storage, and throughput at the cloud level. Data reduction is one such solution to overcome the above problem. There are many improved data reduction techniques, for example, single-tier where data is reduced at the gateway, two-tier where reduction methods are employed at gateway and cloud or sensor and base station.

A much advanced and efficient method is proposed by authors Waleed et al. [27], it is in-network data filtering and fusion, a double layer method. The method employs filtering and fusion at network edge before sending it to the cloud. Consider an IoT application to monitor temperature, humidity, and wind speed, at time instance t greater than 0 all the sensors sense the data and send them to base station or gateway. The proposed method removes the redundant data first: for example, if there is no change in sensor data from the previous transmission data to the cloud is avoided. Even when there is a change in data deviation between actual reading and the estimated value calculated by Kalman filter [24], either data is discarded or sent to the fusion layer only when the deviation is greater compare to predefined maximum absolute error.

6 IoT Data Challenges and Research Opening

Data challenges

- Handling a very high sampling rate and low sampling rate without losing vital information.
- Heterogeneous, noise, varied size of the data need to be cleaned and represented in a uniform format.
- New format of data with reduced size.
- Incident management: identification of traffic patterns in legitimate and illegitimate users.
- Data authenticity and integrity at every level IoT ecosystem.
- Real-time implementation stream mining needs to be done.
- Correct feature extraction from multiple heterogeneous data streaming sources is not solved.
- Need efficient data compression and multiplexing technique.

- Ontology-based IoT semantic event processing.
- Defining unified ontology concept information exchange language which could be used in multiple application domains [5–8].

Research opening

- The increase in data volume has given rise to redundant and inconsistent data. In a resource constraint network transmission of data, to reduce inconsistency and redundancy is time-consuming and challenging.
- The frequency of update in the status of a device and its read values is a high example, 1,00,000 samples per second, or it will be very low, in both the cases the data is very vital. We do not have an efficient mechanism to collect and store this data. Loss of data will lead to substandard knowledge generation from analysis.
- IoT data is increasingly distributive and mobile, and a novel mechanism needs to be designed to manage the distributed and mobile data effectively.
- Development of mechanism to automatically represent sensor data in an interoperable data format so that there will be seamless interaction between heterogeneous IoT devices and networks.
- The complete realization of IoT is possible only when the data is understood and processed by a device without human interaction. Semantic enrichment of sensor data is one good research area, and enriching the data as per the requirement is also an area to be explored. Improving the expressiveness in event processing by effective enrichment technique.
- To overcome the shortcomings of current IoT data formats, Concise Binary Object Representation (CBOR), Efficient XML Interchange (EXI), and Protobuf encoding methods need to be explored [8, 10, 20, 22].

7 Conclusion

The massive data generated by IoT sensors and applications can be made useful by understanding data format correctly and representing data in an interchange format. The success of IoT depends on the seamless exchange of data between heterogeneous devices and cross-domain network. In this direction, we have referred many resources, to identify major IoT sensors, generated data formats, and found that data interchange format is the primary requirement for interaction between devices and applications. A new open data format is provided as a solution to the current problem, which can be explored further. To minimize the power consumption of devices and to improve data processing, data stream processing can be adopted. Data reduction methods can be adopted to overcome problems like redundant data, bandwidth, throughput, and storage.

Competing Interest The authors declare that they have no conflict of interest.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Tongay KN (2016) Sensor data computing as a service in internet of things. In: 2016 Symposium on colossal data analysis and networking (CDAN), pp 1–4
2. Sharma R (2017) Top 15 sensor types being used most by IoT application development companies. https://www.finoit.com/blog/top15sensortypesusediot/
3. Rogojanu T, Ghita M, Stanciu V, Ciobanu R, Marin R, Pop F, Dobre C (2018) Netiot: a versatile IoT platform integrating sensors and applications, pp 1–6
4. Cleber M, de Morais DS, Kelner J (2019) An iot sensor and scenario survey for data researchers 25:4. https://www.sciencedirect.com/science/article/pii/S1084804516000606
5. Kandel S.: The top 3 challenges of preparing IoT data (2019). https://www.iotnow.com/2019/02/08/92826top3challengespreparingiotdata
6. Novak E, Mladenić D, Kenda K, Kažič B (2019) Streaming data fusion for the internet of things, p 1955. https://internetofthingsagenda.techtarget.com/definition/sensordata
7. Gerber A (2018) Making sense of IoT data. https://developer.ibm.com/tutorials/IoTlp301iotmanagedata
8. Hanes D, Salgueiro G, Grossetete P, Barton R, Henry J (2018) IoT fundamentals: networking technologies, protocols, and use cases for the internet of things
9. Fredric combaneyre (2015) Understanding data streams in IoT. https://sas.com/enie/whitepapers/understandingdatastreamsiniot107491
10. Kenda K, Kažič B, Novak E, Mladenić D (2019) Streaming data fusion for the internet of things: taxonomies and open challenges, pp 796–809, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6514969/
11. Google cloud newsletter (2019) Overview of internet of things. https://cloud.google.com/solutions/iotoverviewtypesofinformation/
12. Al Hashami Z (2015) Xml files types and their differences and denominators with plain txt file. https://www.longdom.org/openaccess/xmlfilestypesandtheirdifferencesanddenominatorswithplaintxtfile2376130X1000130.pdf
13. Google cloud newsletter (2014) Sensor model language (sensorml). https://www.opengeospatial.org/standards/sensorml/
14. Su X, Zhang H, Riekki J, Keränen A, Nurminen JK, Du L (2014) Connecting IoT sensors to knowledge based systems by transforming SenML to RDF, pp 215–222. https://www.sciencedirect.com/science/article/pii/S1877050914006176
15. People.sc.fsc.edu: Csv file. https://people.sc.fsu.edu/jburkardt/-data/csv/csv.html/
16. Edward (2018) Water quality monitoring for governments and smart cities with IoT sensors. https://www.valarm.net/wp-ontent/uploads/2018/02/Tools.Valarm.netIndustrialIoTRemoteMonitoringFloodWaterLevelsWellsSensorDataCSVExport4.png
17. Abd El-Aziz AK (2014) JSON encryption. In: International conference on computer communication and informatics (ICCCI-2014). https://ieeexplore.ieee.org/abstract/document/6921719
18. Butcher M (2019) Rest without JSON: the future of IoT protocols. https://dzone.com/articles/jsonhttpandthefutureofiotprotocols/
19. AWS (2019) Awsiot developer guide. https://docs.aws.amazon.com/iot/latest/developerguide/iotdg.pdfiot-sql-json
20. Web page I: RFID faq and tutorial. http://www.idautomation.com/barcodefaq/rfid/ReadingRFID
21. Farkas K, Pödör Z, Mezei G, Somogy M (2018) Data interoperability across IoT domains, pp 61–65. http://www.naun.org/main/NAUN/computers/2018/a182007-052.pdf
22. The open group: Open data format (o-df), an open group internet of things (IoT) standard—introduction. http://www.opengroup.org/iot/odf/p1.htm
23. Qin Y, Sheng QZ, Falkner NJ, Dustdar S, Wang H, Vasilakos AV (2016) When things matter: a survey data centric internet of things, pp 137–153. https://www.sciencedirect.com/science/article/pii/S1084804516000606

24. Elsaleh T, Bermudez-Edo M, Enshaeifar S, Acton ST, Rezvani R, Barnaghi P (2019) IoT-stream: a lightweight ontology for internet of things data streams. http://iot.ee.surrey.ac.uk/iotcrawler/ontology/iotstream/
25. Lunardi WT, de Matos E, Tiburski R, Amaral LA, Marczak S, Hessel F (2015) Context-based search engine for industrial IoT: Discovery, search, selection, and usage of devices. In: 2015 IEEE 20th conference on emerging technologies factory automation (ETFA). https://ieeexplore.ieee.org/document/7301477
26. Wikipedia (2019) Comma-separated values). https://en.wikipedia.org/wiki/Commaseparatedvalues
27. Waleed M, Ismael MG, Zahary A (2019) An in-networking double layered data reduction for internet of things (IoT), p 795. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6412591/ffnsectitle
28. Li Q, Li R, Ji K, Dai W (2015) Kalman filter and its application, pp 74–77

Software Fault Prediction Using Cross-Validation

Yeresime Suresh

Abstract Software faults are dangerous. Software systems are often essential to a business operation or organization, and failures in such systems cause disruption of some goal-directed activity (mission critical). Faults in safety-critical systems may result in death, loss of property, or environmental harm. Run-time faults are the most damaging as they are not always detectable during the testing process. Detecting faults before they occur gives the designers a brief inner view of the possible failure and their frequency of appearance. This helps in focused testing and saves time during the software development. Prediction models have the ability to differentiate between various patterns. This article showcases the effectiveness of cross-validation in design and development of a neural network for software fault detection.

Keywords Artificial neural network · Fault · Metrics · CK metric

1 Introduction

Over the decade, the trend and need for quality software have enormously increased. As a matter of time, it is very much essential to relate issues related to testing. In software development, reducing the cost and enhancing the overall capability of the software depend on measuring fault-proneness. “Majority of faults in a software system are found in a few of its components” [1]. The *Pareto* rule says that roughly “20% of a software system is responsible for 80% of its errors, costs, and rework.” Papaccio [1] agrees that

> The major implication of this distribution is that software verification and validation activities should focus on identifying and eliminating the specific high-risk problems to be encountered by a software project, rather than spreading their available early-problem-elimination effort uniformly across trivial and severe problems [1].

Y. Suresh (✉)
Department of Computer Science & Engineering, Ballari Institute of Technology & Management, Ballari, India
e-mail: dr.suresh@bitm.edu.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_48

Planning testing activities at an early stage will be quite effective provided estimating potential faultiness of software component is of higher possibility. For instance, software components of a subset can be emphasized upon(e.g., the potentially “most troublesome” 20%). Miserably, faulty software at once cannot be computed. Further, software metrics provide an estimate of quantitative description of program attributes.

The remainder of the article is structured as follows: Section 2 provides insight on the recent work carried out in the area of finding defects. Section 3 introduces the approach followed in this paper in finding accuracy of fault prediction using cross-validation. Section 4 lists out the details of the software metrics, dataset, and the evaluation criteria used in our proposed approach for fault prediction. Section 5 highlights the results obtained and its significance and Sect. 6 accomplishes the paper and presents the further work plan.

2 Literature Survey

The following segment presents a brief survey of the activity done in design of a network for fault detection. Numerous models have been proposed by various authors to find out the impact of fault prediction. Table 1 shows the relevant literature survey.

Table 1 Literature survey on ANN models used for fault prediction

Author	Proposed methodology	Data set	Accuracy (%)
Santosh et al. [2]	Generic method better than neural networks	camel 1.4	143.47% Completeness
		camel 1.2	69.10% Completeness
Rodrigo et al. [3]	K-means, Euclidean distance	CM1 (KM-E)	83.33%
	K-means, Manhattan distance	CM1 (EM)	81.33%
	Particle Swarm Optimization (PSO),	CM1 (EM)	70.29%
	PSO, Manhattan distance	CM1 (PSO-E)	89.96%
	Expectation maximization algorithm	CM1 (PSO-M)	71.29%

(continued)

Table 1 (continued)

Author	Proposed methodology	Data set	Accuracy (%)
Bassey et al. [4]	NASA KC1, 3 sets of predictor [m1, m2, m3]	m1	97% Training accuracy
		m2	90% Training accuracy
		m3	99% Training accuracy
Jin et al. [5]	Decision tree	PC1	91.14%
	SFPM	CM1	91.17%
	SFPM	KC1	87.74%
	LR	KC3	93.42%
Sandhu et al. [6]	Decision tree, K-means clustering	CM1	100% on tenfold cross validation

3 Methodology

This section highlights the algorithms and approaches used in this article for fault prediction. In Machine Learning, algorithms enhance their ability to automatically perform task by learning through experience [7].

The characteristics of a biological neural network are used in designing an efficient information processing system known as Artificial Neural Network (ANN). It consists of numerous densely connected processing elements, commonly referred to as neurons [8]. Synapse mechanism is used to connect one neuron with another. The information about the input signal is depicted with weights for each connected link. Further, this information is used as input to a neuron to process a particular problem. An individual neuron represents its own internal state and is commonly known as activation level of a neuron, which is the representation of the function of inputs the neuron receives. In literature survey, it is observed that there are numerous activation functions that can be applied over net input [9]. Sigmoid activation function is most commonly used in neural network. ANN models are more generally identified by the three fundamental units it possesses, viz.,

- Synaptic (neuron to neuron) inter-connections in the model.
- Updating weights, which connect the neurons during training phase.
- Various activation functions used at different layers of the network.

The processing in neural network begins by designing the network and then relating the technique to be used to prepare the network using the existing dataset. Neural network designs are divided into two categories:

1. Recurrent feed forward network (RFNN).
2. Feedback networks that have no repetitive loops in the network path.

The most common architecture of neural network which is used in software fault prediction is the Backpropagation algorithm. The training algorithm of backpropagation involves the following phases:

1. Feed forward network (along with initialization of weights),
2. Backpropagation of error, and
3. Updating the weights and bias in the network.

3.1 Cross-Validation

Neural network is implemented using Keras [10]. Cross-validation is a technique employed to overcome model over-fitting, when the dataset contains a small set of training samples.

To overcome this drawback, in this paper K-fold cross-validation (with 10 folds) has been applied. In this approach, cross-validation (CV) is performed k different times, at each instance using a dissimilar batch of the data into training and validation data. Further, the average of the error rate is considered to evaluate the network. The obvious edge of using K-Fold CV is that all the instances of the input dataset are at last used for training and testing purpose. Error is estimated as the average error rate (λ is the tuning parameter.):

$$CV(\lambda) = \frac{1}{k}\sum_{k=1}^{k} E_k(\lambda) \tag{1}$$

Steps followed in tenfold cross-validation are as follows:

1. The samples or instances are randomly shuffled.
2. Split the instances into 10 batches.
3. For each distinct batch:
   - Consider the batch as test data.
   - Consider the remaining batches as training data.
   - Fit a model on the training set and evaluate it on the test set
   - Evaluate model using the performance evaluation parameters.
4. Significantly, each sample in the dataset is associated to an individual set and remains in the same for the complete duration of the process. This implies, each sample is given a chance to be associated with one set at a time and is further used to train the remaining k-1 folds.

4 Dataset and Metrics

In this section, a concise view of the dataset, software metrics, and evaluation metrics used for analysis is presented.

4.1 Dataset

The camel-1.6 dataset is used for the experimental evaluation of our model, which has been accumulated from the work of Menzies et al. [11]. The dataset totally consists of 38 proprietary open source as well academic projects of varying versions (92). This work notes about 20 object-oriented metrics and the amount of faults found in the testing phase and after release of the software.

The camel-1.6 dataset has 965 total number of modules, with 188 faulty modules [14], each with a given number of faults as shown in Table 2. In our setup, we only classify whether a module is faulty. Therefore, the count of faults in the module is not considered.

4.2 Used Metrics

The camel-1.6 dataset uses the Chidamber and Kemerer object-oriented metrics [13]. These metrics were designed to measure the complexity of an object-oriented modeled design. Though the dataset uses twenty metrics, here only the CK metrics suite is considered. These 6 metrics are calculated for each class [13].

4.3 Performance Metrics for Evaluation of Classifiers

The exploratory evaluation of the proposed model is done using several performance metrics, such as probability of Recall, Accuracy, F-measure, and Precision.

The following set of evaluation measures are being used to evaluate the model [14, 15]:

Table 2 Confusion matrix for camel 1.6 dataset [14], before cross-validation

	Not-faulty	Faulty
Not-faulty	777	0
Faulty	188	0

Table 3 Confusion matrix

	False (Predicted)	True (Predicted)
False (Actual)	True negative	False positive
True (Actual)	False negative	True positive

1. Recall is computed as the relation of classes correctly classified as faulty to the entire faulty classes.

$$PD = recall = \frac{TP}{TP + FN} \tag{2}$$

2. Accuracy is defined as the “proportion of predicted fault prone modules that are inspected out of all modules” [14, 15].

$$Accuracy = \frac{TN + TP}{TN + FP + FN + TP} \tag{3}$$

3. Precision is the “degree to which the repeated measurements under unchanged conditions show the same results” [14, 15].

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

4. The F-measure can be interpreted as a “weighted harmonic mean of the precision and recall” [14, 15].

$$F - measure = \frac{2 * precision * recall}{precision + recall} \tag{5}$$

The performance parameters for analyzing the predictor model can be determined based on the confusion matrix [15] as shown in Table 3. In Table 3, a “contingency table” is calculated after N*M CV [14].

5 Implementation and Results

The ANN is implemented using Keras [10] running on top of Theano (to calculate gradients). The Keras Sequential model is set up as shown in Fig. 1.

It can be concluded that the architecture (in Fig. 1) performed the best for the camel-1.6 dataset. In this article, a neural network with three hidden layers is considered. Table 4 gives the details of the ANN architecture with cross-validation approach used in this article.

Fig. 1 Keras sequential model for cross-validation

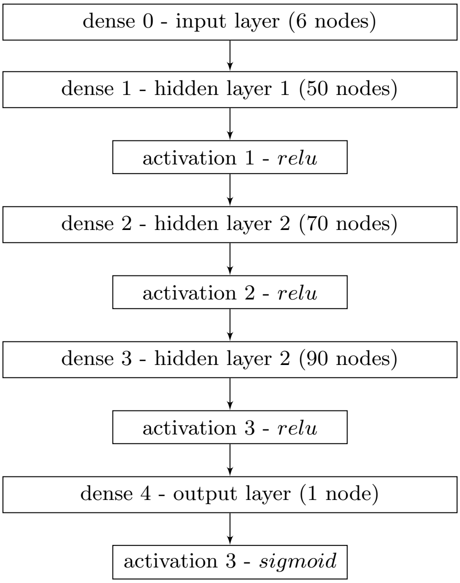

dense 0 - input layer (6 nodes)
dense 1 - hidden layer 1 (50 nodes)
activation 1 - *relu*
dense 2 - hidden layer 2 (70 nodes)
activation 2 - *relu*
dense 3 - hidden layer 2 (90 nodes)
activation 3 - *relu*
dense 4 - output layer (1 node)
activation 3 - *sigmoid*

Table 4 Components of proposed model for cross-validation

Architecture	Parameter value
Layers	5
Input layer nodes	6
Hidden nodes	1st Hidden layer-50 nodes 2nd Hidden layer-70 nodes 3rd Hidden layer-90 nodes
Output node	1
Transfer function	*relu*, *sigmoid*
Algorithm	Back-propagation, 10 fold CV

The input layer has 6 nodes, one for each CK Metric class. The $1st$ hidden layer has 50 nodes, the $2nd$ hidden layer has 70 nodes and the $3rd$ hidden layer has 90 nodes. The implementation is a binary classification.

The input layer makes use of linear activation function. The activation function used is *ReLU*, i.e., rectified linear unit function for the hidden layers and *sigmoid*, i.e, the exponential function for the output layer. The *ReLU* activation function's output ranges from -1 to 1 and is better suited for the normalized input from the normalization layer that has a mean close to 0, and a standard deviation close to 1. The *ReLU* function is also known to be faster than *tanh*, i.e., the hyperbolic tangent function [12]. The *sigmoid* function is suited for a binary classification, because it produces output probabilities in the range [0, 1]. The optimizer used was *Adagrad* and the loss considered was *binarycrossentropy*. The result obtained on tenfold cross-validation is shown in Table 5.

Table 5 Confusion matrix for camel 1.6 [14, 15], using tenfold cross-validation

	False (Predicted)	True (Predicted)
False (Actual)	77	5
True (Actual)	5	10

1. tenfold CV is employed. All the 10 folds are validated with the training data and testing data.
2. It is observed that the mean accuracy over the 10 folds produced for the training data itself is about 97% in most cases.
3. Maximum mean accuracy over the 10 folds for the testing data is 88.84% using adamax optimization and binary cross entropy loss.
4. Accuracy is found to be 89% after applying tenfold cross-validation. Before it was 80%.
5. Precision obtained on the above fold was 85.897% with a recall of 0.4210 and an F-Measure of 0.5651.
6. A recall value of 0.4210 and precision (as much as less FN or zero) is an indication that model prediction is good. Recall value is found to be 66.6%.

6 Conclusion

Fault prediction models help researchers and practitioners in improving the quality of the software and reduce the cost involved in delivering the software product. It is observed in literature that many models have been proposed in computing the fault prediction accuracy rate using various neural network models. But very little emphasis is provided in proving the effectiveness of use of cross-validation for fault prediction model. In this paper, an approach for software fault prediction was analyzed using 10 fold cross-validation technique.

Our results are of an indication that, cross-validation is a better scheme when coupled with prediction models. Further, the work can be carried out to improve the accuracy of the models by using hybrid approaches as well as meta-heuristic search techniques.

References

1. Boehm BW, Papaccio PN (1988) Understanding and controlling software costs. IEEE Trans Soft Eng 14(10):1462–1477
2. Rathore SS, Kuamr S (2015) Comparative analysis of neural network and genetic programming for number of software faults prediction. In: Proceedings of national conference on recent advances in electronics & computer engineering, pp 328–3327, Feb 2015, IIT Roorkee

3. Coelho RA, dos RN Guimarães F, Esmin AA (2014) Applying swarm ensemble clustering technique for fault prediction using software metrics. In: Proceedings of 13th international conference on machine learning and applications, pp 356–361, February 2014
4. Isong B, Ifeoma O, Mbodila M (2016) Supplementing object-oriented software change impact analysis with fault-proneness prediction. In: Proceedings of 15th IEEE/ACIS international conference on computer and information science, pp 1–6, August 2016
5. Jin C, Jin SW, Ye1 JM (2012) ANN-based metric selection for software fault-prone prediction model. IET Soft 6(6):479–487
6. Sandhu PS, Goel R, Brar AS, Kaur J, Anand S (2010) A model for early prediction of faults in software systems. In: 2nd International conference on computer and automation engineering, pp 281–285, February 2010
7. Mitchell TM (1997) Machine learning. McGraw Hill
8. Warren M, Walter P (1943) A logical calculus of ideas immanent in nervous activity. Bull Math Biophys 5(4):115–133
9. Broomhead DS, David L (1988) Multi-variable functional interpolation and adaptive networks. Complex Syst 2(3):321–355
10. Keras, a high-level neural networks library, written in Python. https://keras.io/
11. Menzies T, Krishna R, Pryor D (2016) The promise repository of empirical software engineering data. Department of Computer Science, North Carolina State University. http://openscience.us/repo
12. Krizhevsky A, Sutskever I, Hinton GE ImageNet classification with deep convolutional neural networks. http://www.cs.toronto.edu/~fritz/absps/imagenet.pdf
13. Chidamber SR, Kemerer CF (1994) A metrics suite for object-oriented design. IEEE Trans Soft Eng 20(6):476–493
14. Suresh Y (2015) Software fault prediction and test data generation using artificial intelligent techniques, NIT Rourkela, August 2015
15. Cagatay C (2012) Performance evaluation metrics for software fault prediction studies. Acta Polytech Hungarica 9(4):193–206

Implementation of Recommendation System and Technology for Villages Using Machine Learning and IoT

B. Achyuth and S. Manasa

Abstract With the advancement in smart city technologies, there is a need for developing the rural part of the country. The smart village is equipped with the integration of online recommendation online system, security and automation are aimed in this paper. There is a need for a system that defines smart security of home/go-downs with the smart face recognition technology for intruders and a fire sensor to detect the fire and smoke, and in case of any emergency the water is being sprinkled. The villages in our country need to have a modern garbage system to segregate the dry and wet waste as well as a device that checks for moisture content in soil and sprinkle water when required. Additionally, the proposed system also contains a heartbeat sensor and a pulse sensor to have a quick checkup of the local people; hence, implementing precision agriculture and health monitoring facilities for smart village. The whole system is made to work using renewable energy that is sunlight and the owner is alerted through emails and messages in case of any emergency through Twilio app and ThingSpeak. E-learning is the new medium for learning which has grown in the last decade. The Internet has evolved and has been developed rapidly in an exponential manner. With the rapid development of this new era of Internet, it has given rise to many applications and has proven its potential to impact billions all over the world. One such domain is the usage of Internet to create an impact in the educational domain. With the introduction of massive open online courses now education has been made widely accessible to each and every one. Here we discuss the recommendation system of E-learning platform and the major problem faced by E-learning, i.e. dropouts. The main intention of this paper toward recommendation system is to develop an integrated application that combines the effects of the recommendation system and dropout prediction. Using this smart education system, poor kids can get online support where they don't have proper guidance and teachers can

B. Achyuth (✉)
#21 Satyanarayana Layout, 2nd Stage, 3rd a Cross, Mahalakshmi Layout, Bangalore 560086, India
e-mail: Balaji.Achyuth@in.bosch.com

S. Manasa
#15/a, 8th Main, 4th Block Rajajinagar, Bangalore 560010, India
e-mail: manasa.suresh388@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_49

use this portal. We have our best to implement the application by researching various research papers.

Keywords Smart security system · Smart education system · Face recognition · Renewable energy · Precision agriculture · Health monitoring · Self-reliant · ThingSpeak · Twilio · E-learning · Recommendation systems

1 Introduction

India is a country with lots of villages where more than 70% of population lives. Most of this population depends on agricultural income. To most of these places, basic amenities are not available which are very much necessary. Due to these day-to-day struggles with dropping agricultural income forces youth to leave villages and migrate to urban areas in search of employment and better standard of living. In a long-term perspective, this is a devastating situation. When the backbone of our country crumbles, migration of village population will lead to the desertion of cultivable land. This will lead to a situation where our country's agricultural produce will be far less compared to consumption which is a very bad situation for the country's growth. Due to migration urban population will grow exponentially and it will be very hard for the government to handle such huge influx of people. This alters the stable eco-system which will lead to problems such as creation of slums, traffic, lack of infrastructure to sustain the population. Therefore, the government should focus on empowering village population and provide them with basic amenities and infrastructure like better education, continuous supply of electricity, etc.

MOOC has become the new way of e-learning. There is a meteoric rise in the number of students taking up courses in the MOOC platforms. The number of participants is in thousands. People taking up the MOOC courses are on the rise because of ease-of-access content, learning from anywhere, and motivation to earn a certificate from prestigious institutions. Many colleges like MIT(Massachusetts Institute of Technology), Stanford University offer free content upon enrolment of the course. The students who are pursuing courses at different colleges with graduation from a different degree can take up the course they wish to pursue. The student must only pay for the exam certification. This gives MOOC an edge over its competition. There are hundreds of websites which offer MOOC courses prominent ones are Coursera, EDx, Udemy, NPTEL. This becomes difficult for the user to choose from. With the help of recommended system, the system recommends the course based on the previous academic performance of the course the user had enrolled. But as the number of participants increases the dropout rate is also increasing. This alarming trend can be observed in all platforms. So, the dropout analysis must be conducted in order to find out the reasons for the dropout.

2 Related Work

The recommendation systems use various techniques to achieve their objectives. These techniques also involve certain use of mathematical concepts such as cosine similarity and correlation among other things. One such concept is the web-based monitoring recommendation. In this type of system, the researchers aim to retrieve the information using various mining techniques. The drawback of this method is the user may find it intrusive in nature. In addition to this content-based recommendation may not always yield the best result as the system will not consider the previous performance of the user in a course. This may impact the results the system has to achieve [1].

The above-mentioned methodology learns the user's preferences by using Web-Usage mining technique. But the learner's ability will not be considered for the recommendation of the courses which plays an important role in learner's academic performance in the course. Item Response Theory is a technique which considers the difficulty of the user. With the aid of maximum likelihood estimation (MLE) the learner's difficulty can be assessed and suitable suggestions can be made [2].

The methodology discussed uses two types of feedback. The feedback which the user gives is explicit feedback, and feedback obtained from the user's actions is called implicit feedback. In this research, there is a computerized agent which takes explicit feedback from the user. The challenges from these approaches are that the explicit feedback given by the users may not always be accurate and implicit feedback might also be varying according to the schedule of the user which makes the system to not function efficiently [3].

Various E-learning platforms offer personalization to the user based on the user's performance in the respective course. These techniques involve collaborative filtering amongst others. This research does involve methods such as content filtering which helps in recommendation of courses based on the choice of the newly registered user. The method does not also involve dropout prediction which is an important feature that helps to assess the student's performance as well as take corrective actions that lead to dropout [4].

The discussion forums in MOOC courses are also taken for prediction of dropout. Other factors on which it may depend include the number of hours of video watched, downloaded and weekly submission of assignments [5].

But midway through the course, there is increase in the numbers. Some approaches only consider the first two weeks of the course enrolment for data analysis. This approach is not accurate for prediction of dropout since the increase in numbers must also be accounted for analysis. Our approach of prediction takes consideration of data from the beginning of the course till the completion of course. [6].

By using machine learning model, techniques such as behaviour features and multi-view semi-supervised learning can be used. In this approach, the features are derived from each type of learning records. This method can be used in addition to the multi-view semi-supervised learning method. Classification is used to differentiate between two types of participants, i.e. dropout and at-risk categories. support vector machine (SVM) must be trained and modelled for each week to obtain suitable results. But regression methods like GridSearchCV trains and evaluates a set of models which differ in parameters and select the best-fit model among the tests [7, 8].

Instrumentation and Measurement Technology Conference was mainly on Internet of Things (IoT) which explains how devices can be connected remotely through the idea of Internet [9].

ISPDC 2016 shows the security-related acceptance where the user can communicate to the house-hold devices for the security alerts and where an email or a text message is being sent to the house owner from where he can monitor using FPGA [10].

IJSER June 2016 saw the design of smart security, solid waste management use of renewable energy, and facilities regarding agriculture [11].

International Conference on recent trends in engineering aimed to investigate the Smart Village project in a village community. It intended to address the major problems faced by the community of farmers, identified the Smart Village factors, and put forward a well-detailed plan for the Smart Village implementation [12].

3 Proposed Work

3.1 Security and Automation

The overall implementation consists of four different parts integrated using Arduino and Raspberry Pi. It has four divisions. (1) Waste management (2) Precision agriculture (3) Health monitoring (4) Go-down/Home security. Different sensors are installed at different locations in the go-down and in and around agricultural fields, and a solar panel is used to drive the DC motor through rechargeable battery. This battery can also be used for various other purposes such as lightning the green house at night. **Home/Go-Down Security**—There is real-time streaming using the RPI camera in which facial detection and recognition are done. This is based on the Haar cascade algorithm of machine learning. When there is an intruder present, if a human face detected in the frame is unknown then an email is sent to the owner. Smoke sensor and gas sensor are also integrated to this system as a fire emergency response system which senses the gas level and the smoke level above a certain threshold value which actuates the relay which in-turn turns on the DC pump which is powered through rechargeable battery which is charged by using solar panel. **Health Monitoring**—A pulse sensor and a heartbeat sensor are used in case of emergency or if a patient needs to know the pulse rate; a message and an email of the pulse rate data values are being sent to their respective mobile numbers and their email id's accordingly. **Precision Agriculture**—A moisture sensor checks the moisture content of the soil and if the moisture level is less than the set threshold level, then the relay is actuated, the DC motor is turned on and water is pumped. **Waste Management**—An ultrasonic sensor is used to detect whether the dustbin is full. A rain sensor is used

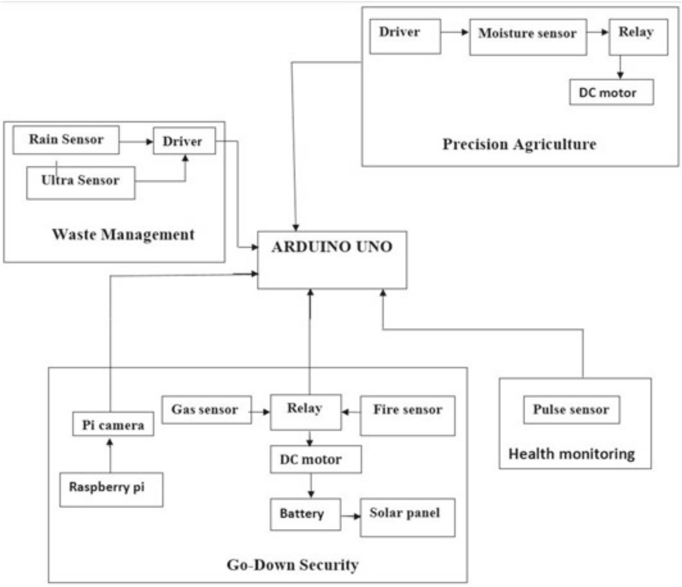


Fig. 1 Block diagram

to detect the presence of wet waste. Alert messages are informed to the house owner when required (Fig. 1).

3.2 *Smart Education System*

See Table 1.

A. *Modules*

Users—new users wanting to enrol for online courses or users already registered in the system. Prediction system is used by companies for predicting and analyzing dropouts for particular courses they are interested in studying. Recommendation system is used for recommending courses to users based on their interactions with previous courses and their preferences. In general, two types of users are involved: 1. New User—Content filtering 2. Registered User—Collaborative filtering (Tables 2, 3 and 4).

B. *Methodology*

Predicting whether a student will be certified or not is a great feature of the course hosting websites. For building a machine learning model, we take the following features based on PCA analysis. The features are '*nchapters*', '*nevents*',

Table 1 Data flow chart of recommendation system

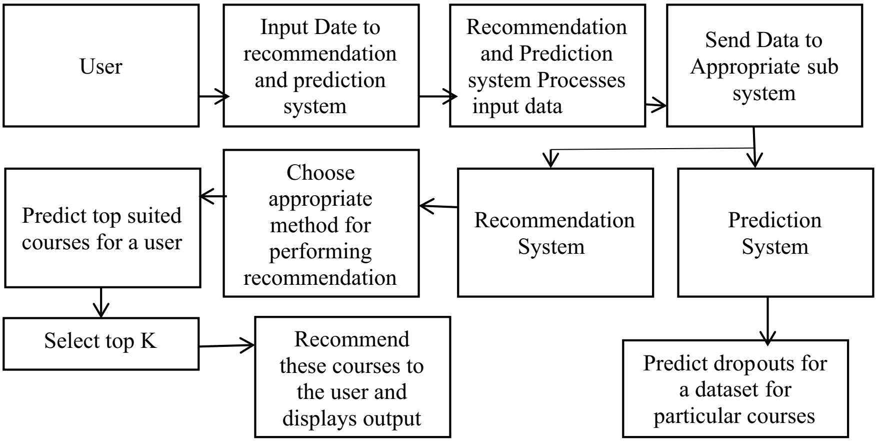
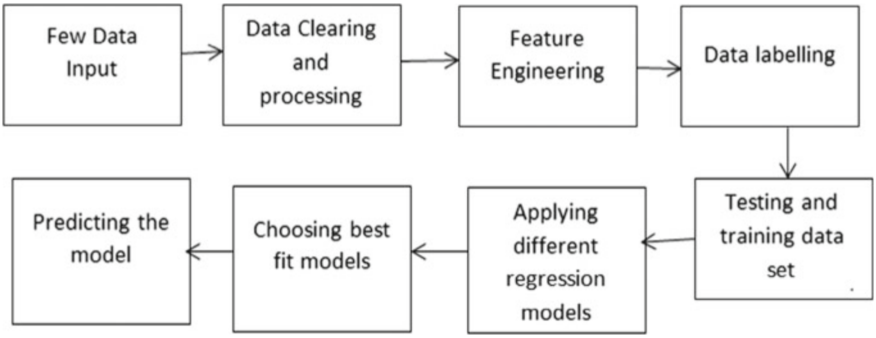
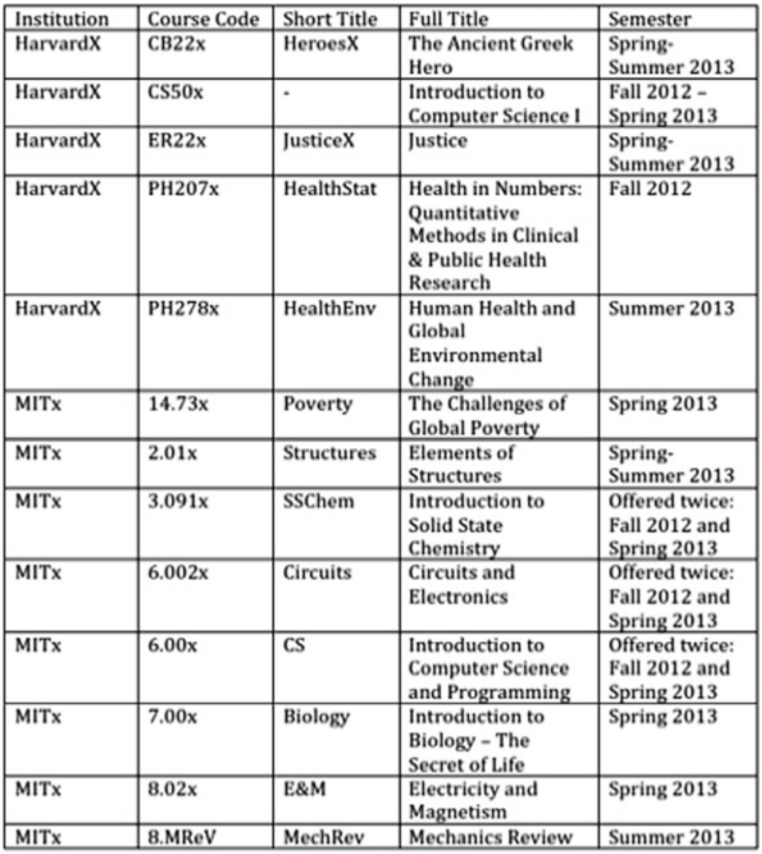

User
Input Date to recommendation and prediction system
Recommendation and Prediction system Processes input data
Send Data to Appropriate sub system
Recommendation System
Prediction System
Choose appropriate method for performing recommendation
Predict top suited courses for a user
Select top K
Recommend these courses to the user and displays output
Predict dropouts for a dataset for particular courses

Table 2 Data flow chart of dropout prediction


Few Data Input
Data Clearing and processing
Feature Engineering
Data labelling
Testing and training data set
Applying different regression models
Choosing best fit models
Predicting the model

'*nforum_posts*', '*ndays_act*'. The algorithm used for building the model is *RandomForestClassifier*. Testing and training data in this step, using TRAIN_TEST_SPLIT library, we are splitting the data into 70% as training data and 30% as testing data and fitting it into proper variables. Choosing the best-fit model—comparing the outputs from the different regression models and choosing the best-fit model. For our dataset *grid search, CV model using lasso regression* has been considered as it has the highest accuracy of 98% and overfitting was handled properly. *Grid Search CV using lasso regression. Grid search* is a method to estimate hyperparameters. The advantage of using this model is it shrinks some of the parameters (β) to exactly zero which will help in interpreting the regression model. A tuning parameter λ is added to control the strength of penalty being added (if λ is 0 the no parameters are being eliminated). As λ increases bias will increase and when λ decreases bias will also decrease (Tables 5 and 6).

Table 3 Course details



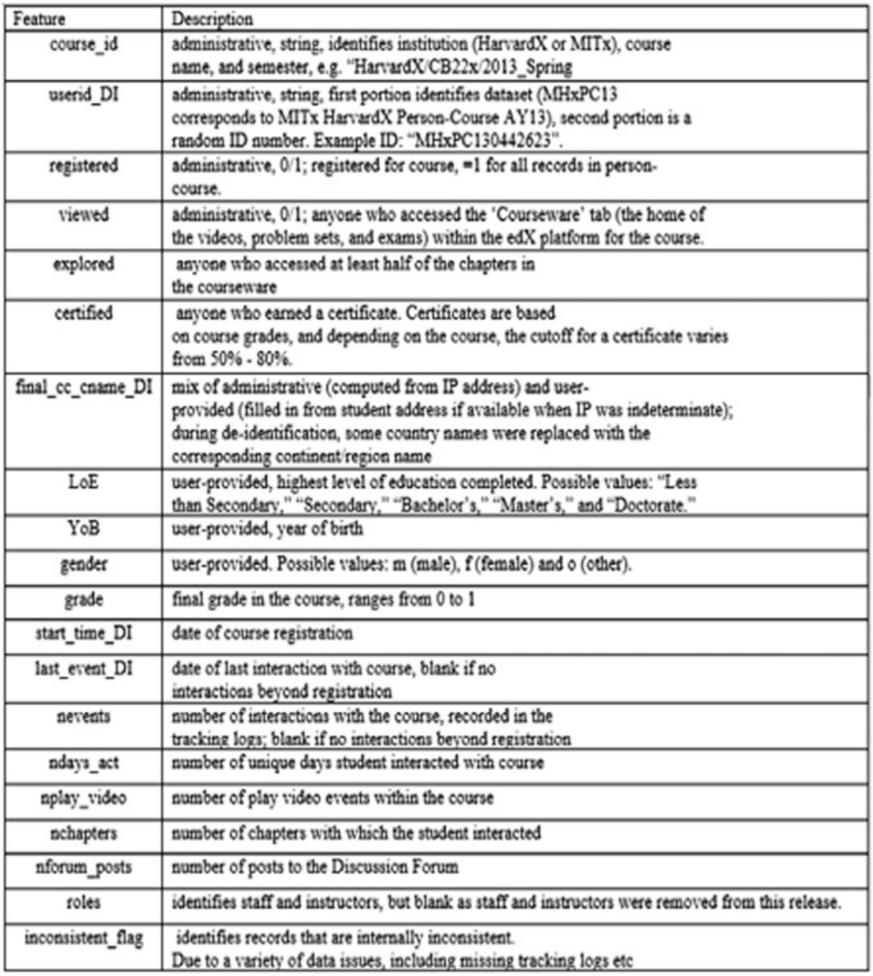
Institution	Course Code	Short Title	Full Title	Semester
HarvardX	CB22x	HeroesX	The Ancient Greek Hero	Spring-Summer 2013
HarvardX	CS50x	-	Introduction to Computer Science I	Fall 2012 – Spring 2013
HarvardX	ER22x	JusticeX	Justice	Spring-Summer 2013
HarvardX	PH207x	HealthStat	Health in Numbers: Quantitative Methods in Clinical & Public Health Research	Fall 2012
HarvardX	PH278x	HealthEnv	Human Health and Global Environmental Change	Summer 2013
MITx	14.73x	Poverty	The Challenges of Global Poverty	Spring 2013
MITx	2.01x	Structures	Elements of Structures	Spring-Summer 2013
MITx	3.091x	SSChem	Introduction to Solid State Chemistry	Offered twice: Fall 2012 and Spring 2013
MITx	6.002x	Circuits	Circuits and Electronics	Offered twice: Fall 2012 and Spring 2013
MITx	6.00x	CS	Introduction to Computer Science and Programming	Offered twice: Fall 2012 and Spring 2013
MITx	7.00x	Biology	Introduction to Biology – The Secret of Life	Spring 2013
MITx	8.02x	E&M	Electricity and Magnetism	Spring 2013
MITx	8.MReV	MechRev	Mechanics Review	Summer 2013

C. *Feature engineering*

Dataset—We have considered a MOOC dataset which consists of 516 × 28 entries. The data has been gathered from MIT and HARVARD university online courses.

Feature selection process: For this process, all the feature labels are considered and correlation matrix has been plotted using the library knows as sns. With Heatmap Fig. 2, which is based on darkness of the colour, we can analyze the features having relation with dropout column.

Table 4 Features of the dataset



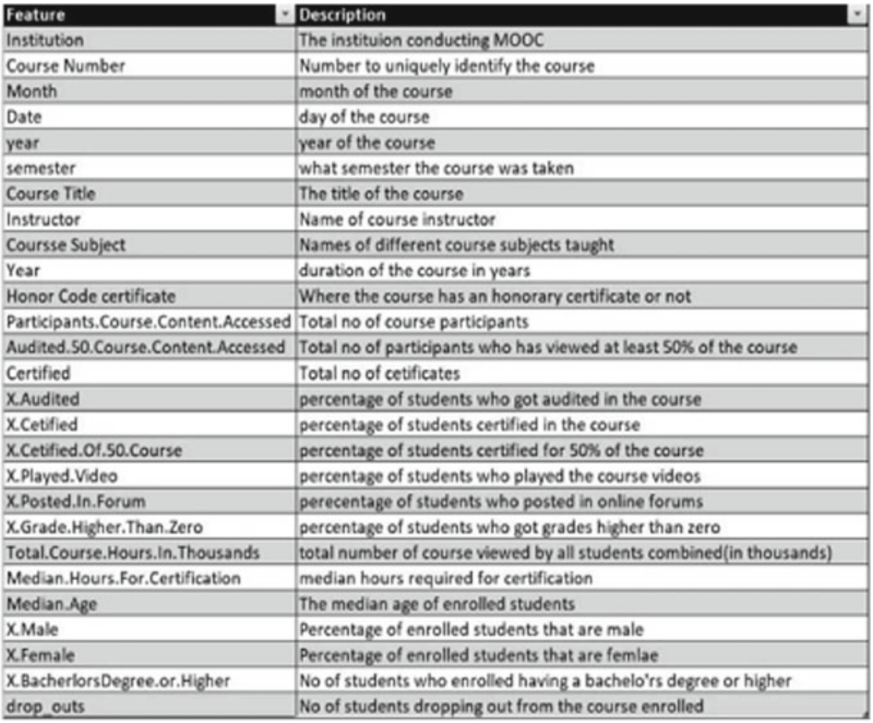
Feature	Description
course_id	administrative, string, identifies institution (HarvardX or MITx), course name, and semester, e.g. "HarvardX/CB22x/2013_Spring
userid_DI	administrative, string, first portion identifies dataset (MHxPC13 corresponds to MITx HarvardX Person-Course AY13), second portion is a random ID number. Example ID: "MHxPC130442623".
registered	administrative, 0/1; registered for course, =1 for all records in person-course.
viewed	administrative, 0/1; anyone who accessed the 'Courseware' tab (the home of the videos, problem sets, and exams) within the edX platform for the course.
explored	anyone who accessed at least half of the chapters in the courseware
certified	anyone who earned a certificate. Certificates are based on course grades, and depending on the course, the cutoff for a certificate varies from 50% - 80%.
final_cc_cname_DI	mix of administrative (computed from IP address) and user-provided (filled in from student address if available when IP was indeterminate); during de-identification, some country names were replaced with the corresponding continent/region name
LoE	user-provided, highest level of education completed. Possible values: "Less than Secondary," "Secondary," "Bachelor's," "Master's," and "Doctorate."
YoB	user-provided, year of birth
gender	user-provided. Possible values: m (male), f (female) and o (other).
grade	final grade in the course, ranges from 0 to 1
start_time_DI	date of course registration
last_event_DI	date of last interaction with course, blank if no interactions beyond registration
nevents	number of interactions with the course, recorded in the tracking logs; blank if no interactions beyond registration
ndays_act	number of unique days student interacted with course
nplay_video	number of play video events within the course
nchapters	number of chapters with which the student interacted
nforum_posts	number of posts to the Discussion Forum
roles	identifies staff and instructors, but blank as staff and instructors were removed from this release.
inconsistent_flag	identifies records that are internally inconsistent. Due to a variety of data issues, including missing tracking logs etc

4 Algorithm

A. *Content-based recommendation*

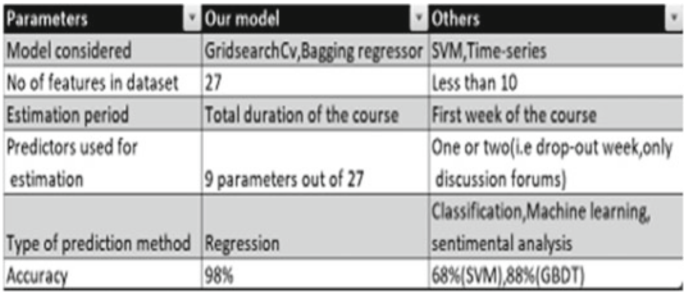
Content-based filtering is an effective technique of recommendation system. This recommendation technique is based on the title of the course. For example, if the course 'World History' is the course selected by the user, then the words 'History' and 'World' are taken into consideration. Based on the number of occurrences it is assigned a weight. If the word 'HISTORY' has more weight than the word 'WORLD' then the title containing the word 'HISTORY' is searched. If there is a course title as 'INDIAN HISTORY' then this is recommended to the user. The technique used for

Table 5 Feature description



Feature	Description
Institution	The instituion conducting MOOC
Course Number	Number to uniquely identify the course
Month	month of the course
Date	day of the course
year	year of the course
semester	what semester the course was taken
Course Title	The title of the course
Instructor	Name of course instructor
Coursse Subject	Names of different course subjects taught
Year	duration of the course in years
Honor Code certificate	Where the course has an honorary certificate or not
Participants.Course.Content.Accessed	Total no of course participants
Audited.50.Course.Content.Accessed	Total no of participants who has viewed at least 50% of the course
Certified	Total no of cetificates
X.Audited	percentage of students who got audited in the course
X.Cetified	percentage of students certified in the course
X.Cetified.Of.50.Course	percentage of students certified for 50% of the course
X.Played.Video	percentage of students who played the course videos
X.Posted.In.Forum	perecentage of students who posted in online forums
X.Grade.Higher.Than.Zero	percentage of students who got grades higher than zero
Total.Course.Hours.In.Thousands	total number of course viewed by all students combined(in thousands)
Median.Hours.For.Certification	median hours required for certification
Median.Age	The median age of enrolled students
X.Male	Percentage of enrolled students that are male
X.Female	Percentage of enrolled students that are femlae
X.BacherlorsDegree.or.Higher	No of students who enrolled having a bachelo'rs degree or higher
drop_outs	No of students dropping out from the course enrolled

Table 6 Comparison of similar models



Parameters	Our model	Others
Model considered	GridsearchCv,Bagging regressor	SVM,Time-series
No of features in dataset	27	Less than 10
Estimation period	Total duration of the course	First week of the course
Predictors used for estimation	9 parameters out of 27	One or two(i.e drop-out week,only discussion forums)
Type of prediction method	Regression	Classification,Machine learning, sentimental analysis
Accuracy	98%	68%(SVM),88%(GBDT)

this purpose is TF-IDF. Term frequency (Tf) measures how frequently a term occurs in a document. It measures number of times a word is repeated in a document, as the length of document varies so we divide it by length of document. TF(t) = (Number of times term t appears in a document)/(Total number of terms in the document). IDF measures how important a term is. In TF all terms are considered with equal importance. We need to scale up only important words. Though words like' 'of' and 'is' are repeated many times these should not be considered.

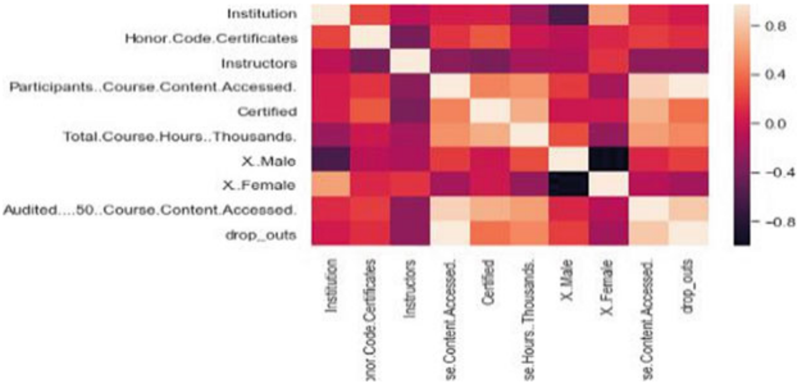


Fig. 2 Heat map

B. *Collaborative filtering*

There are 2 types of collaborative filtering techniques User–User and Item–Item. Both assume that users have given some ratings for items (explicit) or have interacted with the items from which their likings can be predicted (implicit). In User–User we find an input user has given his likings for certain items and when we find another user similar to him, we recommend courses he liked to input user and vice versa. In Item–Item we consider an input User given his likings and we find Items similar to his liked items using data from other users and recommend those items to the input user. In our model we have used implicit Item–Item based filtering for providing recommendations where items refer to courses users have registered. Step 1—forming ratings given for interaction of term ‘nchapters’ and ‘ndays_act’. We use a crude logic looking at the variance of data obtained through formulated ratings depending on user interactions with the courses. Step 2—create a user item matrix and normalize the previously formulated ratings.

A—Items from User for whom we want to recommend B—Other Items. Step 4—predicting ratings—We find weighted averages for each item and the top k items with highest weights are recommended. Weighted average is calculated as Weighted average = for each (similarity * value)/sum of similarity Step 5—recommended the top k courses with highest predicted values.

5 Visualization

Figure 3 shows the dropouts by year depicted using box plot which tells us the median, max, minimum and range of dropouts.

Figure 4 Dropouts versus Total Course Hours (scatter plot).

Figure 5 Dropouts versus Participants (scatter plot).

Fig. 3 Range of outputs

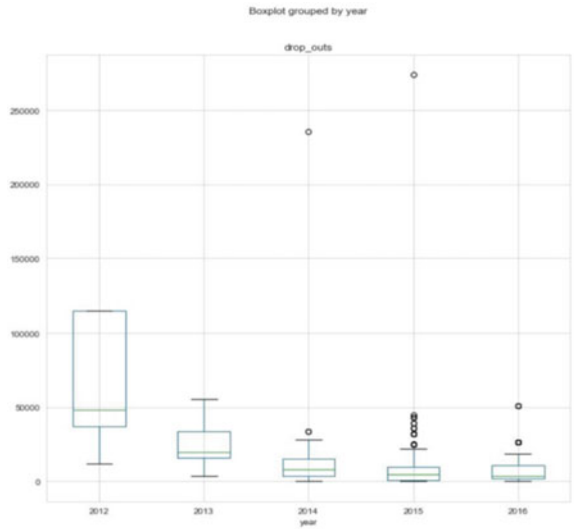


Fig. 4 Dropouts versus total course hours

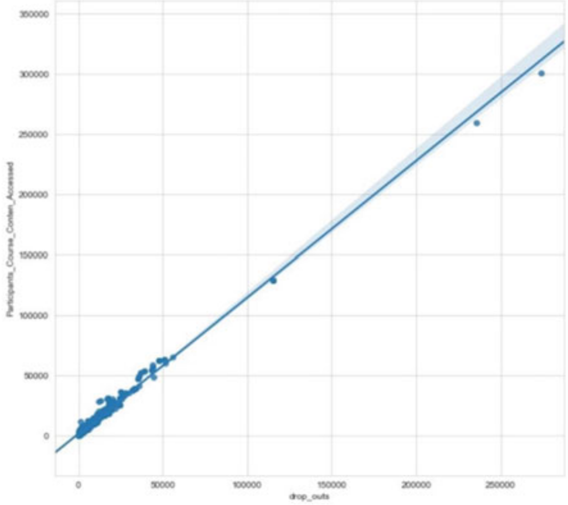


Figure 6 shows the dropouts versus year as bar graphs. Most of the dropouts occur in the year 2014–2015.

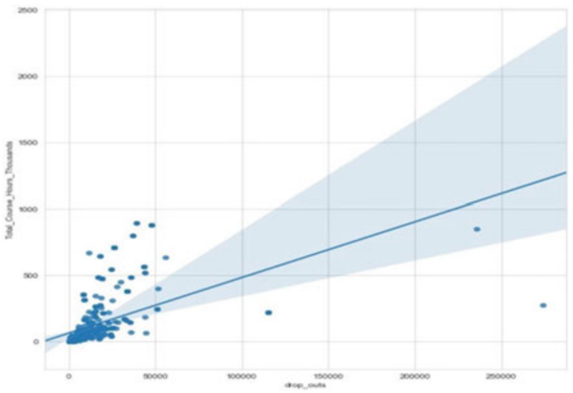


Fig. 5 Dropouts versus participants

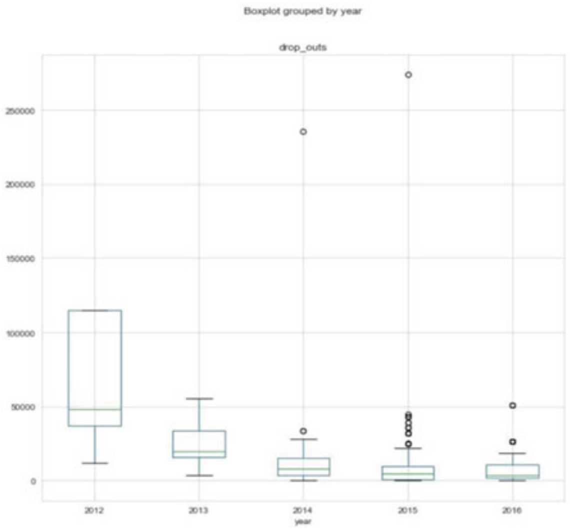


Fig. 6 Dropouts versus. year as bar graphs

6 Results Analysis

The hardware components used for the system setup are Raspberry Pi module, Arduino board, solar panel, relay, ultrasonic and rain sensor, smoke sensor, GSM module, moisture sensor, PIR sensor and web camera along with dc pump and rechargeable battery are shown in figure The hardware components are integrated with an open-source API called ThingSpeak. Raspberry Pi is integrated with a cloud application named Twilio.

Figure 7 displays hardware description. Figure 8 displays the different alert messages sent to the owner through the ThingSpeak application. All the data values of the sensors are updated to ThingSpeak for every 30 s. ThingSpeak is an open-source API in which output data values versus time graphs are plotted for different sensors.

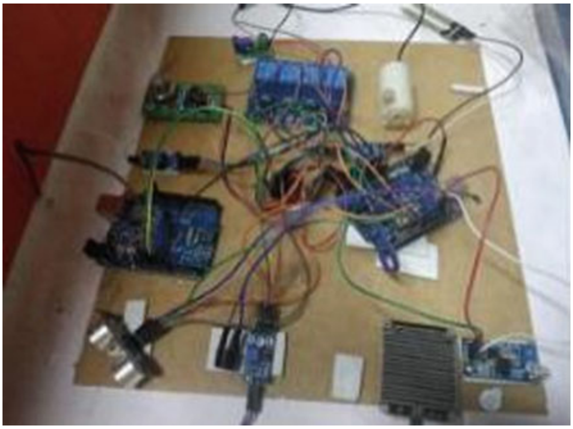


Fig. 7 Hardware description

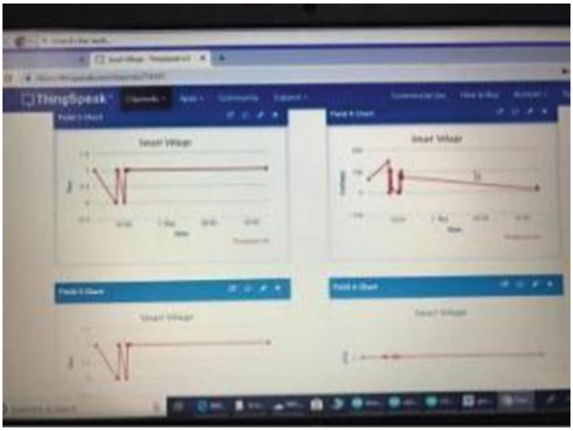
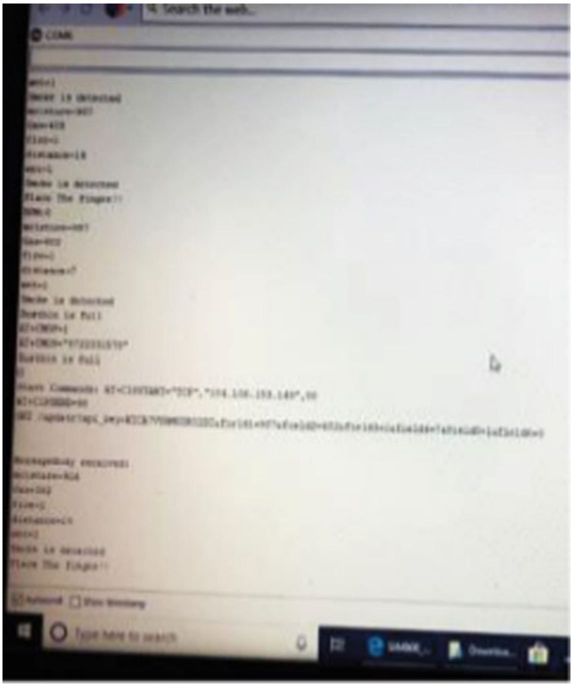


Fig. 8 Peak application

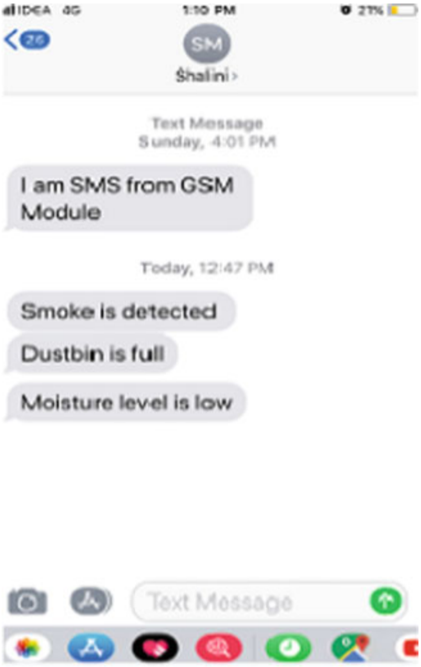
Through ThingSpeak we can act upon the data obtained according to the requirements as shown in figure. Figure 9 shows the output of different sensors displayed in the console. Output values of different sensors integrated with Arduino are displayed in the serial monitor as shown in figure. Figure 10 shows the Alert messages sent through GSM modem. The alert message sent to the villager's phone about various sensors outputs are shown in figure. Figure 11 shows the accuracy of student certification module for Recommendation System. Figure 12 shows the Accuracy rate of various models used for dropout prediction.

Fig. 9 Different sensor output



```
wet=1
Smoke is detected
moisture=987
Gas=419
fire=1
distance=18
wet=1
Smoke is detected
Place The Finger!!
BPM=0
moisture=987
Gas=412
fire=1
distance=7
wet=1
Smoke is detected
Dustbin is full
AT+CMGF=1
AT+CMGS="[illegible]"
Dustbin is full

Start Commands: AT+CIPSTART="TCP","184.106.153.149",80
AT+CIPSEND=[illegible]
GET /update?api_key=[illegible]&field1=[illegible]&field2=[illegible]&field3=[illegible]&field4=[illegible]&field5=[illegible]&field6=[illegible]

[illegible] received:
moisture=[illegible]
Gas=[illegible]
fire=1
distance=[illegible]
wet=1
Smoke is detected
Place The Finger!!
```

Fig. 10 Alert message



Sunday, 4:01 PM

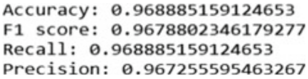
I am SMS from GSM Module

Today, 12:47 PM

Smoke is detected

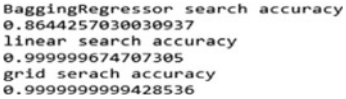
Dustbin is full

Moisture level is low



```
Accuracy: 0.968885159124653
F1 score: 0.9678802346179277
Recall: 0.968885159124653
Precision: 0.967255595463267
```

Fig. 11 Outputs



```
BaggingRegressor search accuracy
0.8644257030030937
linear search accuracy
0.999999674707305
grid serach accuracy
0.9999999999428536
```

Fig. 12 Accuracy rate

7 Conclusion and Future Scope

In this paper, we have recommended courses based on the type of user and predicted the possible number of dropouts based on various factors. This paper clearly explains the process which has been followed. Different types of recommendation systems are researched and the collaborative filtering which takes into account of the user's past performance is promising which can be developed more efficiently. Comparison between various models are done and the best model which is chosen is grid search method using lasso regression as it handles all the hyperparameter combinations and takes care of the overestimating problem. The accuracy obtain in this model is 98%. Our model helps the online web content providers with an estimate of how many dropouts may occur in future who have taken up a particular course. On further data analysis, it also estimates reasons for dropouts and in what way web content providers can improve their content and modify courses to minimize dropouts.

This paper represents the design of low-cost security system for villages. Finally, these two are related as both of them work towards development of lifestyle: one in the educational field and other in safety living conditions. The health monitoring can be improved further by making the doctors available to patients directly by storing the details of the patients in the doctor's database. This enables for precise analysis. A mobile app can be developed showing all the favourable conditions such as moisture, humidity, salt content, etc., which can be monitored by the farmer as and when required. Even the salt level and other factors can be detected and sprinkled if necessary.

References

1. Khribi MK, Jemni M, Nasraoui O (2008) Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval. In: 2008 Eighth IEEE international conference on advanced learning technologies, pp 241–245. IEEE
2. Chen C-M, Lee H-M, Chen Y-H (2005) Personalized e-learning system using item response theory. Comput Educa 44(3):237–255
3. Farzan R, Brusilovsky P (2006) Social navigation support in a course recommendation system. In: International conference on adaptive hypermedia and adaptive web-based systems. Springer, Berlin, pp 91–100

4. Klašnja-Milićević A, Vesin B, Ivanović M, Bu-dimac Z (2011) E-Learning personalization based on hybrid recommendation strategy and learning style identification. Comput Educ 56(3):885–89
5. Wen M, Yang D, Rose C (2014) Sentiment analysis in MOOC discussion forums: what does it tell us? In: Educational data mining
6. Liang J, Li C, Zheng L (2016) Machine learning application in MOOCs: dropout prediction. In: 2016 11th International conference on Computer Science & Education (ICCSE). IEEE, pp 52–57
7. Onah DF, Sinclair J, Boyatt R (2014) Dropout rates of massive open online courses: behavioural patterns. In: Proceedings of EDULEARN14, pp 5825–5834
8. Kloft M, Stiehler F, Zheng Z, Pinkwart N (2014) Predicting MOOC dropout over weeks using machine learning methods. In: Proceedings of the EMNLP 2014 workshop on analysis of large scale social interaction in MOOCs, pp 60–65
9. In: 2012 IEEE International Instrumentation and measurement technology conference (I2MTC)
10. In: 2016 15th international symposium on the parallel and distributed computing (ISPDC)
11. International Journal of Scientific & Engineering Research, vol 7(6) (2016)
12. http://data.conferenceworld.in/PGMCOE/P929-938.pdf (International Conference on recent trends in engineering, science and management)

IoT Based Inventory Management System with Recipe Recommendation Using Collaborative Filtering

Atharva S. Devasthali, Adinath J. Chaudhari, Someshkumar S. Bhutada, Snehal R. Doshi, and Vaishali P. Suryawanshi

Abstract The internet is a huge pool of data. It consists of various websites that provide numerous recipes and it becomes difficult for a person to manually search for a recipe based on the available ingredients daily. Inventory management is a difficult task because it is not feasible to keep track of every food ingredient. The objective of this paper is to provide an effective solution to manage inventory of the user using the Internet of Things (IoT) as well as recommending a recipe to the user using the available inventory. A recipe scoring algorithm that follows a collaborative filtering approach is proposed to score the recipes and the recipes that yield a higher score will be recommended to the user. This reduces the time and efforts of the user which makes the system more efficient and user-centric.

Keywords Web scraping · Recipe recommendation · Inventory · Raspberry Pi · Recipe scoring algorithm · Collaborative filtering

A. S. Devasthali · A. J. Chaudhari · S. S. Bhutada · S. R. Doshi
MITCOE, Pune, India
e-mail: atharvadevasthali22@gmail.com

A. J. Chaudhari
e-mail: adinathchaudhari@gmail.com

S. S. Bhutada
e-mail: someshbhutada67@gmail.com

S. R. Doshi
e-mail: snehal98doshi@gmail.com

V. P. Suryawanshi (✉)
MITWPU, Pune, India
e-mail: vaishali.suryawanshi@mitwpu.edu.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_50

1 Introduction

Earlier, people used different sources such as recipe books, magazines, newspapers, etc. for searching for new recipes as well as finding a particular recipe. With the advancement in internet facility and substantial growth in digitalization, recipes are now available at fingertips. Still, the majority of people face a problem while selecting a recipe due to the large volume of recipe data wherein the availability of the ingredients is questionable.

The proposed system is an application of Internet of Things (IoT) and Machine Learning technologies for the general population to make their life easy. This application saves time and effort by managing the inventory of user and helps in deciding which food item to cook on a particular day.

Initially, all possible recipes as per the user's preferences, as well as which are feasible to cook depending upon the availability of ingredients, are suggested. Unavailability of ingredients will trigger the addition of that ingredient into the shopping list. All available ingredients are listed in the inventory list of a verified user at that particular moment. Inventory of the user is continuously monitored using load sensors and the changes in the ingredients are notified to the application for making the required changes in the inventory. The user has different choices such as making a selection from the recommended recipes, adding a recipe of their choice, add or remove a favorite recipe.

The recipe database, in the proposed system, is created by web scraping of different websites [1–3] and then the data is pre-processed to remove unnecessary information.

2 Related Work

Lakshmi Narayan et al. [4] have implemented a load cell-based inventory management system. The load cell provides real-time tracking of the ingredients present in the container and notifies the user when the content level goes below a predefined threshold point. One of the approaches proposed by Praveen et al. [5] uses the K—Nearest Neighbors algorithm for recommending the appropriate recipes wherein the user ingredients are taken as input and are processed with the help of the collected dataset. In this algorithm, suitable classes are defined by utilizing the training dataset. This algorithm requires the input ingredients to be converted into a vector format. Depending on these ingredients matched recipes are recommended. Another method put forward by Ueda et al. [6] utilizes the user's food preferences and the number of ingredients to recommend recipes. In this method, food preferences are automatically identified based on the user's browsing and cooking history. Each recipe is scored based on a scoring algorithm after finding the favorite ingredients. Lo et al. [7] have used the SVM algorithm to judge if a recipe is good or not. SVM evaluates every recipe and identifies the outliers. After excluding the outliers, the nutrient content is checked to satisfy the required threshold.

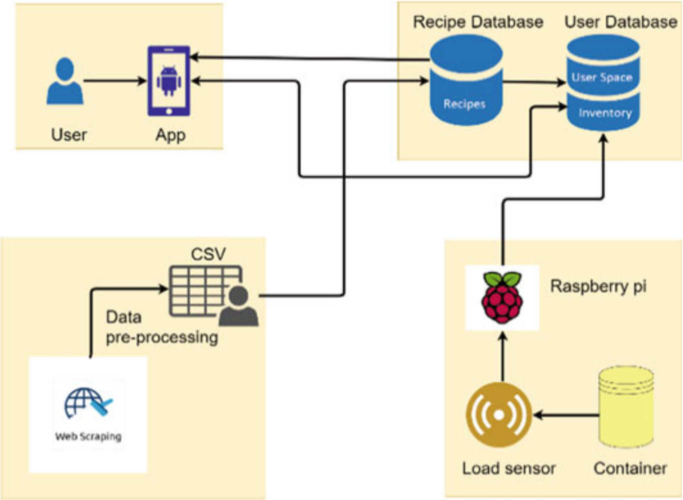


Fig. 1 System Architecture Diagram

3 Methodology

3.1 Architecture

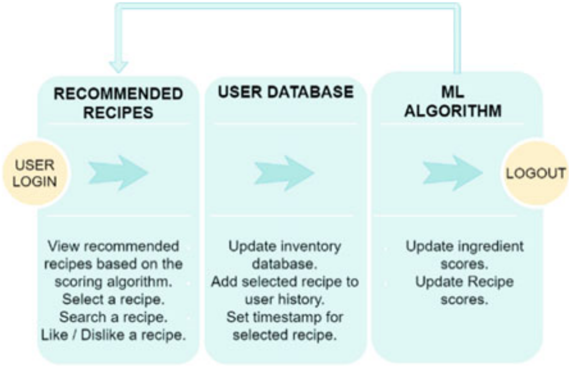
Figure 1 shows the architecture of the proposed system using the Internet of Things (IoT). Initially, raw data is extracted, cleaned and then stored in a structured format in .csv file on cloud database. The current inventory of the user consists of the available ingredients. This forms the input to our recommendation system. With the help of the recipe scoring algorithm, the score is calculated for each recipe and the recipes having higher scores will be recommended to the user. The output will be further displayed to the user on the android application.

3.2 Data Collection and Preprocessing

In the proposed system, data cleaning and preprocessing are done with the help of web scraping. Web scraping is used to extract various recipes from recipe websites [1–3] by using Python libraries—Selenium and bs4 (Beautiful Soup). The extracted data is cleaned by removing the unnecessary symbols, stop words, etc. and converted it into a structured format which was then stored in the database in .csv file because these files are easier to import in any database and help in organizing large volume of data [8].

Real-time data is collected in the system using a 64-bit microprocessor, Raspberry Pi 3 Model B by sensing the weight of the food ingredients with the help of strain

Fig. 2 Flowchart of System


RECOMMENDED RECIPES
USER DATABASE
ML ALGORITHM
USER LOGIN
LOGOUT
View recommended recipes based on the scoring algorithm. Select a recipe. Search a recipe. Like / Dislike a recipe.
Update inventory database. Add selected recipe to user history. Set timestamp for selected recipe.
Update ingredient scores. Update Recipe scores.

gauge load sensors [9]. This data is then stored in the cloud database which forms the inventory of the user.

3.3 Working

The proposed system is developed in Python as it supports machine learning, web scraping and raspberry pi.

Figure 2 shows the working model of the system. User registers to the system and then they will be assigned a personal inventory having a unique user id. They can add or remove the ingredients from the inventory. The load sensors keep updating the inventory continuously. Users can dislike a particular ingredient by using a checkbox provided in the application. Depending upon the user preferences such as likes/dislikes, allergies, and history as well as the availability of ingredients in the inventory [10, 11], recipes will be recommended. After the selection of a recipe by the user, scores of the ingredients used in that recipe will be updated. Then, the recipe score will be updated. Depending upon these scores, recipes will be recommended. After each selection of a recipe, a timestamp will be allocated to that recipe so that it will not be recommended again for the next 7 days. Users also have the option of searching a recipe of their own choice. As a result, recommended recipes will be displayed on the android app.

3.4 Recipe Scoring Algorithm

Collaborative filtering is a method generally used by recommender systems that considers user preferences such as likes, dislikes, etc. to provide a filter for user preference information [10, 12]. *Recipe Scoring Algorithm* is based on collaborative filtering approach for finding scores of each recipe. Further, the algorithm checks for available ingredients and recommends the recipes accordingly.

Input:

1. Inventory list
2. User preferences (likes, dislikes, allergies) and history

Output:

Recommended recipes

```
Algorithm Recipe_Score {
      inv_ingredient = {id, name, amt, wt}
      recipe_title = {id, iseq_no, sseq_no, name}
      recipe_ingredient = {iseq_no, name, amt, wt}
      procedure = {sseq_no, steps}
      I = $\sum$ recipe_ingredient
      R = $\sum$ recipe_title
      Inv = $\sum$ inv_ingredient
      P = $\sum$ procedure
      timestamp = 0, day = 0

      while(!R){
            if(timestamp($R_j$) != 0){
                  day($R_j$) += 1
                  timestamp($R_j$) = 1 $-\lfloor \frac{day(R_j)}{7} \rfloor$
            }
            if($R_j$ == disliked || timestamp == 1)
                  continue
            if $\forall$ I($R_j$) $\exists$ Inv then,
                  Recommendation_list $\leftarrow R_j$
      }
      display(Recommendation_list)
      if($R_j$ $\exists$ Recommendation_list && $R_j$ == selected){
            $day(R_j) = 0$
            $timestamp(R_j) = 1$
            $score(\forall I(R_j)) \mathrel{+}= 1$
            $score(R_j) = \frac{\sum_{i \in R} score(I_i)}{\sum_{i \in R} I_i}$
            $display(R_j, I_j, P_j)$
        }
    }
```

4 Experiment

The proposed system consists of 12,000 recipes that are stored on the cloud database. The *recipe scoring algorithm* is applied to this huge data by considering user preferences and available ingredients. Top 20 recipes [13] are selected by the system according to recipe scores. A graph of recommended recipes which is based on available ingredients is plotted using matplotlib library in python from these recipes.

5 Result

In the proposed system, the recommendation of recipes is based upon the recipe scoring algorithm which is based on assigning scores to each recipe, depending upon the scores and user preferences. The existing systems primarily focus on the classification of ingredients and identifying matching recipes. This system gives better results as the preferences are given to the user rather than to the recipe ingredients. Every time after the selection of a recipe, ingredients score and recipes score are updated. The system learns about the user and recommends recipes accordingly.

Figure 3 represents recipes vs recipe score graph. It consists of top 20 recipes and the recommendation is based on recipe scores and availability of ingredients in the inventory. Thus, even if a particular recipe has a higher score, but the required ingredients are not available in the inventory it will not be recommended. Similarly, if any recipe has a lower score, but all the ingredients are available in the inventory it will be recommended to the user.

Figure 4 represents days versus recipe score graph. It shows a transition of the recipe scores over a period of 7 days. This diagram helps us to understand the user's preferences in more depth.

Fig. 3 Recipe Recommendation

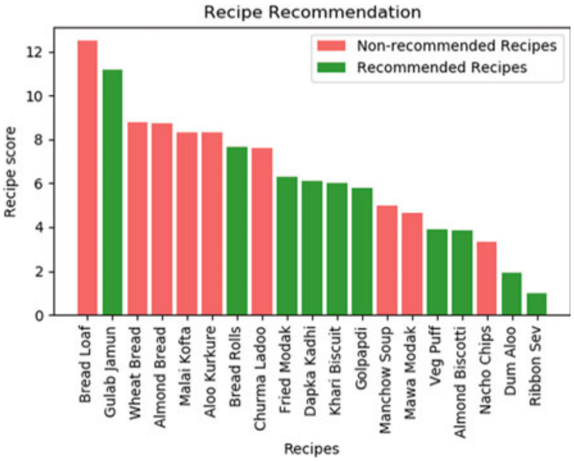
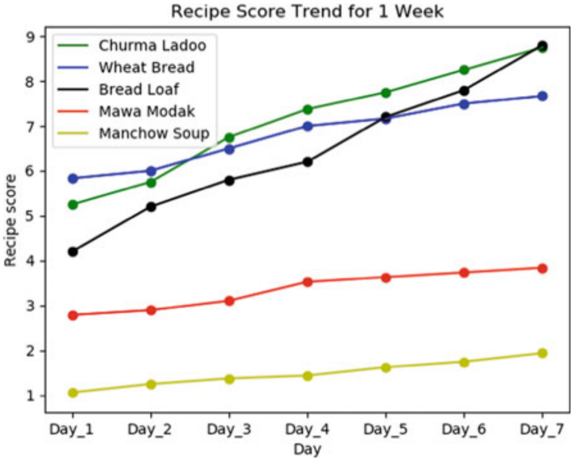


Fig. 4 Recipe Score Trend for 1 Week



6 Conclusion

The proposed system provides an efficient way of handling the user inventory with the help of load cells and raspberry pi. A large volume of recipe data is collected by using web scraping. The *recipe scoring algorithm* is then applied to this collected data to recommend recipes to the user. This algorithm is based on a collaborative filtering approach considering the user's preferences, making it more user-centric. Thus, it reduces the time and efforts required for searching a recipe. Further, load sensors with higher reading capacity can be used to manage large inventories.

References

1. Allrecipes| Food, Friends, and recipe inspiration. Allrecipes. https://www.allrecipes.com/
2. Sanjeev Kapoor Recipes. https://www.sanjeevkapoor.com/
3. TarlaDalal Recipes. https://www.tarladalal.com/
4. Lakshmi Narayan SP, Kavinkartik E, Prabhu E (2019) IoT based food inventory tracking system. Commun Comput Inf Sci 968:41–52
5. Praveen S, Prithivi Raj MV, Poovarasan R, Thiruvenkadam V, Kavinkumar M (1999) Discovery of recipes based on ingredients using machine learning. IRJET 06(02) (2019). Foster I, Kesselman C The grid: blueprint for a new computing infrastructure. Morgan Kaufmann, San Francisco
6. Ueda M, Asanuma S, Miyawaki Y, Nakajima S (2014) Recipe recommendation method by considering the user's preference and ingredient quantity of target recipe
7. Lo YW, Zhao Q, Ting YH, Chen RC (2015) Automatic generation and recommendation of recipes based on outlier analysis. In: 2015 IEEE 7th international conference on awareness science and technology (iCAST), Qinhuangdao, pp 216–221
8. Glez-Peña D et al (2014) Web scraping technologies in an API World. Brief Bioinform 15(5):788–797. Crossref. https://doi.org/10.1093/bib/bbt026
9. Desai H, Guruvayurappan D, Merchant M, Somaiya S, Mundra H (2017) IoT based grocery monitoring system. In: 2017 fourteenth international conference on wireless and optical communications networks (WOCN), Mumbai, pp 1–4. https://doi.org/10.1109/wocn.2017.8065839
10. Vivek MB et al (2018) Machine learning based food recipe recommendation system. In: DS Guru et al (ed) Proceedings of international conference on cognition and recognition, vol 14. Springer, Singapore, pp 11–19. Crossref. https://doi.org/10.1007/978-981-10-5146-3_2
11. Desai MS, Patil MP, Shinde MP, Sayyed MA, Bhosale R Recipe recommendation based on ingredients using machine learning. 8(3). https://doi.org/10.17148/ijarcce.2019.8313
12. Hameed MA et al (2012) Collaborative filtering based recommendation system: a survey. 4(05):859–876
13. All the recipes: scraping the top 20 recipes of allrecipes. NYC Data Science Academy Blog. https://nycdatascience.com/blog/student-works/recipes scraping-top-20-recipes-allrecipes/

Survey Paper on Smart Veggie Billing System

T. V. Niteesh, B. Y. Lohith, Y. M. Gopalakrishna, R. Ashok Kumar, and J. Nagaraj

Abstract The purpose of this paper is to automate the process of the billing system of vegetables. Raspberry PI is the heart of this project which monitors all the components in the system. Initially, we capture the different vegetable images and train the AI model with images. These images are used to recognize the vegetables taken by the customer. The camera captures the selected item image and AI model based on convolution neural network recognizes this image using an image recognition technique. Image recognition is not only the technique to find the vegetable, but also using the resistance of the vegetable we can recognize the vegetable. Resistance method of recognizing the vegetable is not suitable because it may cause damage to the vegetables. So we choose an image recognition method. The weight of the item is also measured using the sensor which removes the human intervention in weighing the item. Recognized item and its weight are going to display on the screen. Customer can add or delete the items using user interface. Finally, a bill is created based on the selected items of the customer.

1 Introduction

In today's world shopping in supermarkets is a time-consuming task only because of long queues at the weighing and billing counters. Reduction of this idle time is a major task. Our project aims at providing an intelligent system which can overcome the above drawbacks and provides a time-saving shopping experience.

Raspberry PI is the main controller of our smart basket which controls and coordinates all the jobs of our system. Raspberry PI camera connects to the Raspberry Pi board via a short ribbon cable (supplied). Raspberry pi camera is used to capture

T. V. Niteesh (✉) · B. Y. Lohith · Y. M. Gopalakrishna · R. A. Kumar · J. Nagaraj
Bengaluru, India
e-mail: nituv222@gmail.com

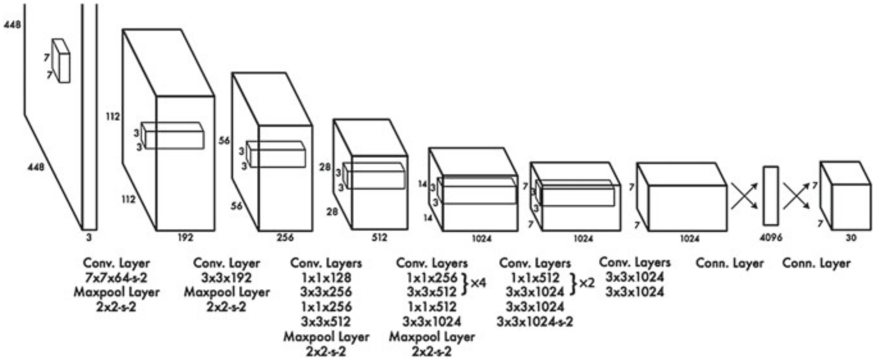
B. Y. Lohith
e-mail: 07lohithby@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_51

the images initially to create the images dataset. It is also used for capturing the customer-selected vegetable image for item recognition.

2 Yolo Network Architecture Diagram

See figure below.



3 Literature Survey

1. Ragesh et al. [1] have proposed the methodology of detecting the vegetables like potato, tomato, etc., using the YOLO model. YOLO model is an abbreviation for You Only Look Once model consists of single CNN which predicts many bounding boxes and class probabilities for them. Detecting the object with YOLO model is a regression problem. Initially an image is divided into SXS grids. Bounding boxes are created using each grid in the image, and are also used to determine the confidence score and probability of the detected object belonging to a particular class. Formula for determining the prediction is SXS (BX5 + C) tensor [8]. They also worked on measuring the weights of the vegetables using Load Cell, a weight measuring device based on the idea of Wheatstone bridge.
2. Athauda et al. [2] have developed a UHF RFID based shopping basket system. This system allows tracing and processing of the shopping data which is of low cost, robust and passive. Here shopping trolleys are equipped with UHF antenna, and shopping products contain UHF RFID tags that contain unique identity codes. This method helps to discover new ways to improve the shopping experience for the customers. This proposed system joins all the components such as polarized antennas, reader and user interface. It aims at acting through

a low-power Bluetooth and it uses the power which is one-fourth of ZigBee based smart trolley system. This shopping trolley is better than the self-checkout counter systems.

3. Pavan Kumar [3] developed a system that reduces the time spent in shopping at supermarkets. Usually, customer faces a lot of difficulties in the billing counter in which waiting is one of them. The proposed system here includes various components such as Barcode scanner and the touchscreen display. These are used to scan the items and display the item information such as cost of each product and total bill of the purchased products. During peak seasons, the number of customers in the shopping mall is very high and the waiting time in the billing system is more compared to other days. So the main aim is to reducing the waiting time. Here Raspberry Pi, Raspberry Pi touch screen display, barcode scanner and button were integrated into the shopping cart. Customers have to scan the items and can put them into the trolley system. And if the customers wish to remove the items placed in the cart, push button is provided in the cart to do the task.
4. Karjol et al. [4] have introduced a very sensible and profitable smart shopping cart by using IoT devices. This framework can suitably be applied in Walmart like supermarkets which reduces work. The cart is furnished with barcode scanner, camera, weight sensor and a computer for processing and also a display screen. The central base has all the data stored of all products, and is able to communicate with all the products using WIFI network. Customer logs in using his/her login ID and is linked with the cart ID; further, they can scan all the products and put into the cart. Weight sensors are equipped in order to avoid any dishonesty. Once the customer finishes shopping they can move to the shopping cart to pay the bill. Also cart-to-cart communication is enabled along with the supermarket management to analyze shoppers' behaviours.
5. Rupanagudi et al. [5] developed a low-cost method to locate products and carry the products to the billing counter. They created a smart trolley system by using web camera and video processing. This method overcomes the time wasted in the identification of items in the shopping mall. The cart consists of four battery-operated motors, plastic wheels and camera which are attached to FPGA. Here they divided the products into two categories and assigned them with colour light orange and blue, and also dark orange colour indicates the trolley to turn right or left. The trolley control system consists of five states, namely,, move, idle, turn left/right, products which are passed by and pickup of the products. This system also helps in carrying the selected items to the billing system in the shopping mall.
6. Sakai et al. [6] have provided insights about the recognition of the vegetable category using DNN [deep neural network]. Deep learning or DNN has connected internal layers available for feature detection and mainly for representation learning. DNN uses input values in the pixel level and through most suitable learning it acquires its characteristics and identifies it. They have used convolution neural network for category recognition of vegetables. A complete toolkit provided by the Caffe is used for training, testing, fine-tuning and for model

deployment. Initially, they have used 160 images of 8 different vegetables like images of tomato, carrot, etc. They tested the model with 40 images (5 images of each vegetables). They performed very much high learning iterations up to 3 million and obtained the learning rate of 99.14% and recognition rate with whooping 97.8%.

7. Megalingam et al. [7] have published this paper providing the information about the development of android application for smart shopping purposes. They mainly focused on providing the locations of different items within the building and automating billing after the purchase having the items having the RFID tags. This is a Java-based application designed using Microsoft android studio. Initially, customer is asked to connect to the trolley which is Bluetooth enabled. Bluetooth module HC-05 provides this service. After the customer is connected to the cart the system provides the item location. The cart movement can also be controlled using various patterns provided by this application. For the billing purpose separate window is designed, which is aided with RFID scanner. Items are scanned and added to the cart. The bill keeps updating until the customer finishes the shopping. Customer can also remove the items, again by scanning it before the item removal.
8. Chanda et al. [8] proposed the new way of automated billing system for the items with RFID tag. The framework of this idea comprises a trolley aided with RFID reader. A screen present on the trolley acts as user interface which allows the user to confirm the item: weight, cost and its expiry date. The trolley is connected to a mobile application through ESP 8266 WIFI module, which aids the customer to get the information about the items on the mobile also. Heart of the framework is the microcontroller which controls the complete framework. Customer selects the particular product, scans using RFID scanner, updates the bill in application and finally pays the bill online using the safe payment gateway associated with the above application. They summarized with that the line product may become little expensive but expense is much low.
9. Dhiraj Thote et al. [9] have worked in order to resolve the long queues for billing in supermarkets. They have used technologies like microcontrollers, database management. Introduced TECHNOBASK that is attached to the RFID reader, which is used to grab the information from the RFID card, is uniquely associated with each product. Microcontroller 1805 is to connect the components and every operation is controlled by it. When a power switch of the cart is turned on the RFID reader with getting turned on to read the RFID card and when the product is added into the cart the RFID card gets scanned and the product gets added into the product list, price of product gets added, subtotal gets displayed on screen and this information gets stored in the database. If the same RFID card number is that got already scanned previously is scanned for one more time then it is considered as the product got removed from card and the price amount gets deducted from the subtotal. Finally, at the cash counter the bill data is transferred to the main computer using Zigbee module and final bill gets generated.
10. Liu et al. [10] have put forward the stochastic gradient algorithm in order to represent the image in optimal linear representation. Since gradient process

coverages at local maximum a stochastic component is added to it. The nearest neighbour classifier is used to design the recognition function. Recognizing objects from the high-dimensional images like 2D requires large memory storage and high computations. They representing the images in linear fashion is the main goal. So in this, they have reduced the high-dimensional images like n*m (orthogonal m dimensional subspace) to one-dimensional image like n*1. The optimal linear representation or optimal subspace is formulated using a stochastic optimization process, which maximizes a needed performance function of overall subspaces. They have designed a simulated annealing algorithm which is based on MCMC and formulated an optimization problem by using Grassmann manifold.

11. Nandanwar et al. [11] have proposed a modern scale called "portable weight meter". The approach shows a grain weight of the portable scale and the price of the chosen grain, both dense and light (portable). Arduino "pro mini" professional card Arduino can save data and perform logical operations. The information refers to the price of different cereals that can be stored in the EEPROM. This article presents a new weighing technique with Arduino details. MATLAB is a mathematical software tool for estimating graphics, a fundamental. PROTEUS is a circuit development software designed to develop and monitor accurate performance. IDE is an integrated tool to register Arduino family controllers. Digital weight weighing can be calculated with great precision. This can be achieved with a control program that takes the user to a more sophisticated exit. Then enter the current market. Also, check if cereal prices are changing or not when you choose the right path. Subsequently, the grain and its weight are automatically loaded on the screen. The accessories are integrated into the glass container, which means they are compact and portable. This dish is used to measure the weight of grain. The Arduino process performance is considered sufficient to load a system. This topic is configured to reduce the disadvantages of a regular weighting system and improve the overall quality of the system.
12. Lekha and Rajeshwari [12] have designed an intelligent shopping cart by using the Bolt ESP8266, barcode scanner and LCD display. Here every trolley is fixed with barcode readers and every item is attached with the barcode tag. Whenever the customer places the item into the basket, the barcode scanner will scan the item barcode. The barcode reader will read the product-related data from the bolt ESP8266, since the bolt ESP8266 has been stored initially in item details of the supermarket. The subtotal bill gets updated and gets displayed on the LCD display. Finally, when the shopping gets completes the final bill with purchased details gets transferred from bolt ESP8266 to the billing counter. Bolt has capacity to connect with many sensors and helps in controlling them. Barcode reader converts the optimal impulses into electrical and then it decodes the barcode. The main reason for selecting the bolt ESP8266 as controller is that it is easy to monitor.
13. Prasiddhi and Gawali [13] developed a smart shopping cart which is an impending device to be seen at supermarkets to reduce the queue for billing and

estimation of bill by costumers. The project is made using accessible components hence it is affordable. LCD display, push button to delete and buzzer are user convenient. Quick scanning of product and real-time display of billing makes process quicker. The gadget looks to be compatible with cart, providing with robust technology which consists of high efficiency and low cost and can be easily adaptable, and using the localization algorithm was proposed to large scale cart "fast map algorithm" based on IoT and this algorithm helps to find particular product and location in cart. Components include which work efficiently microcontroller of ATMmega16, RF receiver, RFID tag and RFID reader. The scope of progress work is wheels of cart of energy harvesting which helps the efforts of the customer to move from one place to another in cart with charging battery.

14. Phanikrishna et al. [14] developed a knowledge extraction based on contour tracking vectors and using deep learning to recognition of images or objects when training the model with different pattern of images which is an impending device to be seen at any malls and offices to reduce the manual data collecting to find images of objects using the spectrum or syntactic pattern for recognition technique. The collected data size increasing with digital image acquisition and there is a need for developing faster, reliable which mostly human intelligence for recognizing objects. The inputs for neural network are only in 0's and 1's or lexicographic combinations symbols. The future enhancement is for recognition of objects with more accurate patterns of geometric objects and patterns with hexagonal lattice for curvilinear properties to find objects accuracy.

4 Conclusion

This research work mainly focuses on the reduction of queue at a counter in a shop. This proposed system displays the overall cost of the product which is kept inside the shopping cart. Customers can easily pay the amount directly without standing in the queue at the billing counter. It reduces the scanning of the products in queue and saving the time of each customer. Here, the shopping system is made more simple and fully automated. Also, the system has a feature to delete the scanned products in the shopping cart and to further enhance the shopping experience of the customer.

Competing Interest The authors declare that they have no conflict of interest.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Ragesh N, Giridhar B, Lingeshwaran D, Siddharth P, Peeyush KP (2019) Deep learning based automated billing cart. In: 2019 International conference on communication and signal processing (ICCSP)
2. Athauda T, Marin JCL, Lee J, Karmakar N (2018) Robust low-cost passive UHF RFID based smart shopping trolley. IEEE J Radio Freq Identif 1–1
3. Pawan Kumar V, Reddy SC (2018) Smart shopping cart. In: 2018 International conference on circuits and systems in digital enterprise technology (ICCSDET)
4. Karjol S, Holla AK, Abhilash CB (2018) An IOT based smart shopping cart for smart shopping. Cogn Comput Inf Process 373–385
5. Rupanagudi SR, Jabeen F, Vaishnav Ram Savarin KR, Adinarayana S, Bharadwaj VK, Karishma R, Bhat VG (2015) A novel video processing based cost effective smart trolley system for supermarkets using FPGA. In: 2015 International conference on communication, information & computing technology (ICCICT)
6. Sakai Y, Oda T, Ikeda M, Barolli L (2016) A vegetable category recognition system using deep neural network. In: 2016 10th international conference on innovative mobile and internet services in ubiquitous computing (IMIS), Fukuoka, pp 189–192
7. Megalingam RK, Vishnu S, Sekhar S, Sasikumar V, Sreekumar S, Nair TR (2019) Design and implementation of an android application for smart shopping. In: 2019 International conference on communication and signal processing (ICCSP)
8. Chadha R, Kakkar S, Aggarwal G (2019) Automated shopping and billing system using radio-frequency identification. In: 2019 9th International conference on cloud computing, data science & engineering (Confluence)
9. Thote D, Parsewar S, Welekar A, Sheikh N, Dhakate R, Sheikh R (2019) Automatic shopping basket technobask. In: 2019 5th International conference on advanced computing & communication systems (ICACCS). https://doi.org/10.1109/icaccs.2019.8728309
10. Liu Xiuwen, Srivastava A, Gallivan K (2004) Optimal linear representations of images for object recognition. IEEE Trans Pattern Anal Mach Intell 26(5):662–666. https://doi.org/10.1109/tpami.2004.1273986
11. Nandanwar VG, Ankushe RS (2017) Portable weight measuring instrument. In: 2017 International conference on recent trends in electrical, electronics and computing technologies (ICRTEECT)
12. Lekhaa TR, Rajeshwari S, Sequeira JA, Akshayaa S (2019) Intelligent shopping cart using bolt Esp8266 based on internet of things. In: 2019 5th International conference on advanced computing & communication systems (ICACCS)
13. Prasiddhi K, Gawali DH (2017) Innovative shopping cart for smart cities. In: 2017 2nd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT)
14. Phani Krishna C, Reddy AVN (2016) Contour tracking based knowledge extraction and object recognition using deep learning neural networks. In: 2016 2nd International conference on next generation computing technologies (NGCT)

An Optimized Approach for Virtual Machine Live Migration in Cloud Computing Environment

Ambika Gupta, Priti Dimri, and R. M. Bhatt

Abstract For efficient management of the cloud, there is a requirement of the virtualization in the emergent stage of cloud computing where resources such as memory, servers, virtual machines are shared across the World Wide Web. For getting or availing other facilities such as load balancing, auto-scaling, and fault tolerance, etc., in cloud computing live migration of virtual machine is required. Migrating virtual machines from one node to another without suspending the VMs is an important feature of cloud computing so that users do not have to deal with any kind of service downtime. In this research paper, a comparative study has been done for various methods of migrating the data from one node server to another node server. After that it also recognizes the optimization of the live virtual machine in migration techniques. Therefore, an optimized approach has been identified that will be beneficial in a live migration of virtual machines without affecting the node servers in a cloud computing environment.

Keywords Virtualization · Virtual machine manager · Downtime · Load balancing

A. Gupta (✉)
GLA University, Mathura, India
e-mail: ambika.gupta@gla.ac.in

UTU, Dehradun, India

P. Dimri
GBPEC, Pauri Garhwal, Pauri Garhwal, Uttarakhand, India
e-mail: pdimri1@gmail.com

R. M. Bhatt
IMMT and Agro Sciences, Dehradun, India
e-mail: rmbhatt@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_52

1 Introduction

In cloud computing, there are millions of time-sharing mechanisms for accessing various resources. To provide the same configuration there is an existing technology, i.e., virtualization, which allows the various guest operating systems on a single physical machine which is managed by a virtual machine manager (VMM). Each Virtual Machine manages a separate operating system which is being further managed by a single hypervisor. For maintaining integrity in the system migration is required across various nodes involved in virtualization technology [1].

But that possible shuffling of load across node servers will be live as possible as this means without affecting any virtual machine the complete shifting of the data should happen. There are lot many problems such as imbalance of load, faulty node server, etc., due to which the expensive mainframe computers could not be fully utilized. To overcome these problems, virtualization as a form of technology was introduced in 1960s by IBM to provide the maximum and efficient utilization of various resources. In this technique, there is a virtual machine manager (VMM) also known as hypervisor which is mounted between operating system and the hardware. The hypervisor is responsible to manage and control virtual machines from a single platform environment [2].

We can abstract the logical layer completely only by the hardware virtualization in which few components are required to execute on various operating systems.

Some of the popular examples of hardware virtualization are as follows:

1.1 Vmware ESX

ESX is also known as Elastic Sky X in which service console has to control and manage VMware's enterprise server virtualization kernel. The basic objective of it is to provide a management interface for the host and a lot of management agents, and other third party software agents are installed on the service console [3].

1.2 VMware ESXi

ESXi is also known as Elastic sky X Integrated into which there is no service Console to control and manage VMware's enterprise server virtualization kernel. All the VMware related managers and other third-party monitoring agents can directly run on the VM kernel [3].

1.3 Kernel-Based Virtual Machine (KVM)

It is a virtualization module which allows the kernel to work as a hypervisor in the Linux kernel as shown in Fig. 1.

There are various internal features in Kernel-Based Virtual Machines:

Firstly, set up the VM's address space in which there is a guest kernel which consists of a file system and block devices and drivers, etc.

It generates I/O requests to the host on guest's behalf and handles various events.

There is a Linux kernel situated which consists of a file system and block devices and drivers, etc.

The important benefit of virtualization is to provide better resource utilization on a single server by having many virtual machines in parallel, and we can get rid of fault tolerance and create an executable environment in isolation, etc.

The research paper is organized as follows:

In Sect. 2, there is a description of related work on virtualization technology and virtual machine manager.

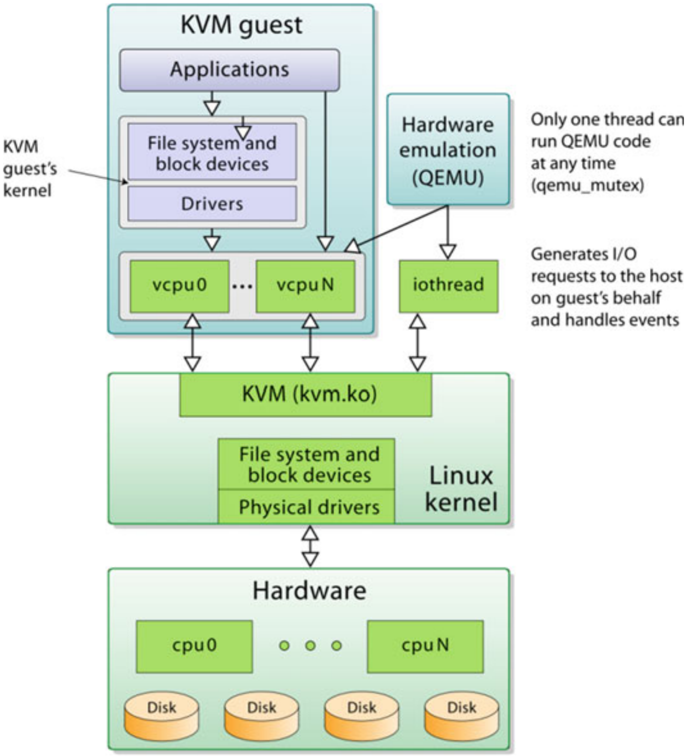

KVM guest
Applications
KVM guest's kernel
File system and block devices
Drivers
vcpu0
vcpu N
Hardware emulation (QEMU)
Only one thread can run QEMU code at any time (qemu_mutex)
iothread
Generates I/O requests to the host on guest's behalf and handles events
KVM (kvm.ko)
File system and block devices
Physical drivers
Linux kernel
Hardware
cpu0
cpu N
Disk
Disk
Disk
Disk

Fig. 1 A high-level structure of the KVM virtualization environment [4]

In Sect. 3, various techniques and methodologies about migration in virtual machine environment have been given.

Section 4 consists of discussion on concluding various techniques and which technique can be applied in the current cloud computing environment.

2 Literature Review

In the era of cloud computing environment migrating the Live Virtual Machines depending on virtualization technology is indispensable.

This section elaborates the virtualization technology. Then, we will discuss the process of live VM migration and its main approaches.

2.1 Virtualization

For reaching up-to the mark of utilization of shared resources and getting the cost-effective solutions a virtualization technology is used, which provides every guest isolated operating system for execution and an appearance that they are running on actual hardware environment. There is an abstraction that is a software layer just above the physical machine which can be managed by virtual machine manager (VMM) known as hypervisor. The interaction between the physical machine and virtual machines is shown in Fig. 2. The virtual machine manager is of two types:

(a) Bare metal or Type 1, the hypervisor runs directly on the hardware.
(b) Hosted or Type 2, the hypervisor running on a host operating system.

To fulfill the current scenario's demand live virtual machine migration is necessary. For the aforesaid issue, there are various parameters which require certain features such as fault tolerance.

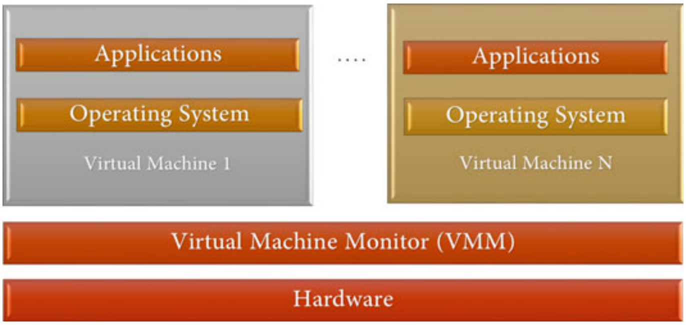


Fig. 2 Virtualization [5]

Any kind of fault can occur in the system where the 24 × 7 availability is guaranteed. We only have to ensure certain reliable conditions that need to be handled [6].

To avail the data, which need to be distributed evenly just to balance the load among the number of servers we have to ensure that no such node server is heavily loaded. For this, rigorous monitoring is required to maintain the equilibrium among various nodes [7–9].

For better manageability of load among the number of servers (data centers units), there is a requirement to transfer the accurate state of the virtual machine from the source node to the destination node.

2.2 Live VM Migration Techniques

There are various techniques available to migrate the data from one data center to another data center without affecting the client requirement; these approaches are shown in Fig. 3. For calculative performance there are the following parameters [9] to apply live virtualization technique for migration:

- Preparation Time: Time is taken to transfer the virtual machine state from source node server to destination node server.
- Resume Time: Time taken between the restarting of the virtual machine's running state and migration end.
- Pages transferred: The number of virtual memory pages transferred.
- Down Time: Time taken to stop the running of virtual machine for transferring the state of the processor.
- Total Migration Time: Time taken from the start of the migration to the end of the migration.

Pre-copy [11] is the default migration technique in which memory state of virtual machines is transferred from source node to destination node server. In fact, all Virtual machines are running on the source node.

There are two associated phases with this pre-copy migration approach:

The first phase is called a warm-up phase in which destination node server is selected. After that there is a stop and copy phase which allows transferring the virtual machine state to destination node server from host node server. Finally, the source node server is to be destroyed and the execution is to be started on destination node server [12].

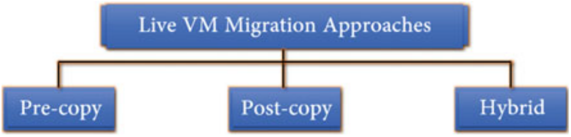


Fig. 3 Live virtual machine migration approaches [10]

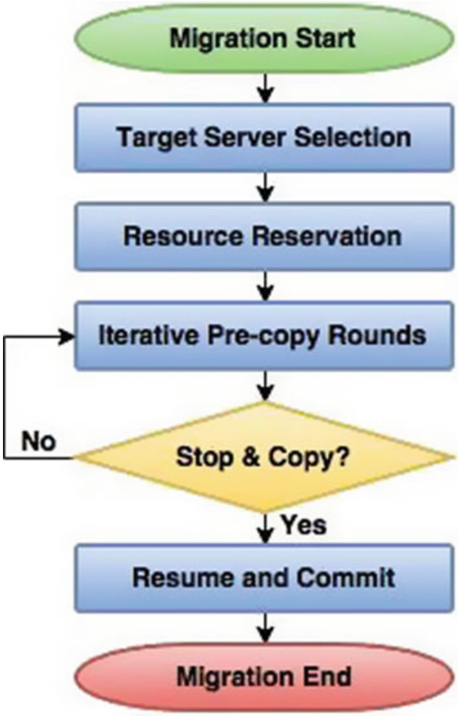


Fig. 4 Pre-copy live migration technique

As shown in Fig. 4, once the migration is started we need to select the destination server and certain resources need to be reserved, After that various pre-copy rounds need to followed and if this iteration is completed successfully then we have to resume from the destination server, If this does not happen then we need to backtrack to perform various pre-copy rounds, and then finally the migration ends and the processing can be started from the destination server node.

Post-copy is the other approach to do migration to minimize the network page fault ratio from the original host node server. There is a method followed as shown in Fig. 4, which is an adaptive pre-paging technique. There are 3 associated phases in the post-copy memory migration technique:

- Post-Copy through demand paging

In this phase, the entire number of pages is transferred at a single time and if the user is requesting the same page but when it is not available then it will behave like a page fault from source node server to the network.

- Post-Copy through Active Pushing

In this phase, the virtual machine continues its execution on the destination node server while the active push simply tells us the transferring of the pages which are going to perform page fault on the destination node server.

- Post-copy through Pre-Paging

It is better to identify the virtual machine's memory access pattern that is why it is a bit difficult to get rid of fault tolerance. Therefore, it is essential to find out the occurrence of faulty pages.

As shown in Fig. 5, there is a source node and one destination node at which directly entire number of pages has been transferred after that the state of virtual machine has been transferred. Later transferring of virtual machine state due to which it is known as post-copy migration technique which follows the page fault and replacement of pages in detail occurs[13].

Hybrid virtual machine migration [14] combines the entire properties of pre-copy and post-copy live migration methods. There are various associated phases:

- Preparation phase: We need to prepare the destination node server and also the system resources needed at the target host.
- Bounded pre-copy rounds phase: Check the limit that will refine the exact bounded pre-copy so that the working set can be transferred from source node server to destination node server.
- Virtual machine resume Phase: At the destination server node, it is required to start the transferred state of virtual machines.

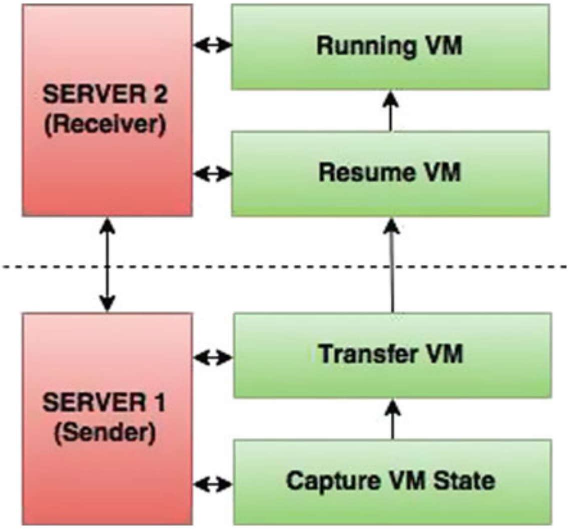


Fig. 5 Post-copy live migration technique

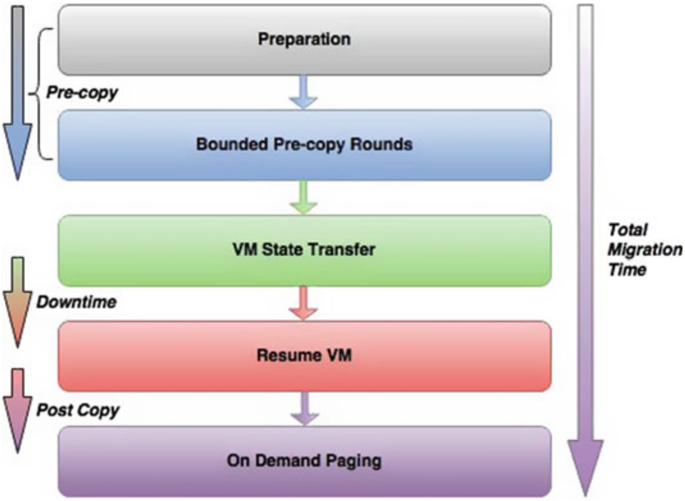


Fig. 6 Hybrid-copy live migration technique

- On demand paging phase: It is the phase where directly the request of read and write gets completed just after the completion of every migration step. As shown in Fig. 6, there is a pre-copy stage, after which the entire virtual machine state transfers and later it needs to follow the post-copy migration steps and finally the migration needs to complete in the hybrid kind of technology.

2.3 Deduplication

In terms of getting cost-effective solutions, there are some other techniques to reduce the consumption of bandwidth where we can use the concept of deduplication.

Zhang et al. [15] have given a method where processed memory pages can be saved in temporary memory and which gets reloaded just to reduce the additional consumption of bandwidth. And the proposed technique is known as migration with data deduplication which can be abbreviated as MDD which is helpful in virtual machine migration, where reduction of network consumption can pertain.

3 Proposed Methodology

There are few constraints and issues in the aforesaid techniques of live migration such as migration overhead for transferring the entire memory pages and virtual machine state, etc. To get rid of this overhead, it requires further optimization of transfer of

Table 1 Optimized approach for migration

Step:1	Start the migration policy
Step: 2	Choosing of Destination Server
Step: 3	Maintaining each processing on a temporary node server (tracker Node)
Step: 4	Transfer entire memory pages to the destination node server keeping a record to tracker node as well
Step: 5	If any request from the user further arises then that can be processed through tracker
Step: 6	After the entire memory pages are transferred, virtual machine state can be transferred to the destination node
Step: 7	Migration terminates

memory pages and other information. Further, there is a need to further optimize the transfer of memory and CPU state.

The steps of the proposed algorithm have been shown in Table 1.

The abovementioned algorithm is an optimized approach to get rid of page fault and other related issues that can be directly handled by tracker node which is a temporary node that takes care of the entire load on the source and destination node server.

4 Conclusion

In the virtualization technology, if any fault arises then it is the necessary condition to migrate the load from a particular node server to another node server without any downfall in the performance of servers. Therefore, in the aforesaid scheme Live VM migration is an important phenomenon that helps us to manage the data centers for getting smooth processing between data centers. This paper compares the existing techniques for live migration such as pre-copy migration, post-copy migration, and a hybrid-copy migration. This also supports the proposed idea for an efficient technique based on the concept of live VM migration in terms of CPU utilization, memory usage, and effective bandwidth, etc. This will help to reduce the page replacement overhead and meanwhile, when the migration process works, the entire working can be handled by a temporary node server in between the source and target server in the migration process.

References

1. Rose R (2004) Survey of system virtualization techniques, Oregon State University. http://www.ece.cmu.edu/~ece845/docs/rose-virtualization.pdf
2. Bugnion E, Devine S, Rosenblum M, Sugerman J, Wang EY (2012) Bringing virtualization to the x86 Architecture with the original VMware workstation. ACM Trans Comput Syst ACM Ref Format Bugnion 30(4):1–51

3. vSphere ESXi Bare-Metal Hypervisor. http://www.vmware.com/products/esxi-and-esx.html. Accessed 04 Nov 2016
4. Huynh K, Hajnoczi S (2010) KVM/QEMU Storage stack performance discussion (PDF). ibm.com. Linux Plumbers Conference. Accessed 3 Jan 2015
5. Kapil D, Pilli ES, Joshi RC (2012) Live virtual machine migration techniques: survey and research challenges. In: IEEE
6. Bala A, Chana I (2012) Fault tolerance—challenges, techniques and implementation in cloud computing. Int J Comput Sci Issues 9(1):288–293
7. Kikuchi S, Matsumoto Y (2011) Performance modeling of concurrent live migration operations in cloud computing systems using prism probabilistic model checker. IEEE 4th international conference on cloud computing. IEEE, DC, pp 49–56
8. Xu F, Liu F, Liu L, Jin H, Li B, Li B (2014) iAware: making live migration of virtual machines interference-aware in the cloud. IEEE Trans Comput 63(12):3012–3025
9. Yin F, Liu W, Song J (2014) Live virtual machine migration with optimized three-stage memory copy. Future information technology. Springer, Berlin, pp 69–75
10. Patel PD, Karamta M, Bhavsar MD, Potdar MB (2014) Live virtual machine migration techniques in cloud computing a survey. IJCA 86(16)
11. Noshy M, Ibrahim A, Ali HA (2018) Optimization of live virtual machine migration in cloud computing: a survey and future directions. J Netw Comput Appl. https://doi.org/10.1016/j.jnca.2018.03.002
12. Ray S, De Sarkar A (2012) Execution analysis of load balancing algorithms in cloud computing environment. IJCCSA 2(5)
13. Zhang X, Huo Z, Ma J, Meng D (2010) Exploiting data deduplication to accelerate live virtual machine migration. In: Proceedings of the 2010 IEEE international conference on cluster computing, CLUSTER'10. IEEE Computer Society, pp. 88–96. https://doi.org/10.1109/cluster.2010.17
14. Ahmad RW, Gani A, Hamid SH, Shiraz M, Xia F, Madani SA (2015) Virtual machine migration in cloud data centers: a review, taxonomy, and open research issues. Springer, pp. 2473–2515
15. Ibrahim KZ, Hofmeyr S, Iancu C, Roman E (2011) Optimized pre-copy live migration for memory intensive applications. In: Proceedings of 2011 international conference for high performance computing, networking, storage and analysis, SC'11. ACM, pp. 40:1–40:11. https://doi.org/10.1145/2063384.2063437

Digital Image Retrieval Based on Selective Conceptual Based Features for Important Documents

Premanand Ghadekar, Sushmita Kaneri, Adarsh Undre, and Atul Jagtap

Abstract Due to the increased level of digitalization, data exchange and collection has grown to a greater extent. This bulk data needs categorization else data collected will be meaningless. The exchange of multimedia data has also increased due to the availability of the Internet. Images collected need a proper categorization that can help in fetching the required data. Government documents need to be uniquely identified from the set of plenty of data to help in different kinds of proceedings. It could help to fetch images only of the required government document. Due to the increased size of data as well as an increased quantity of data, different document image identification techniques are time-consuming. The proposed model quickly identifies specific government documents from a chunk of images with ease. Various filtering categories are applied to ease the process of categorization. To speed up the process of categorization, selective efficient features are shortlisted which contributes majorly toward substantiation of a government document.

Keywords Edge detection · Image processing · Feature extraction · Concept-based image retrieval system

P. Ghadekar (✉) · S. Kaneri · A. Undre · A. Jagtap
Department of Information Technology, Vishwakarma Institute of Technology, Pune, India
e-mail: ppghadekar@gmail.com

S. Kaneri
e-mail: sushmita.kaneri17@vit.edu

A. Undre
e-mail: adarsh.undre16@vit.edu

A. Jagtap
e-mail: atul.jagtap16@vit.edu

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_53

1 Introduction

Content-based image retrieval is known as query by image content. It means that search analyzes the contents of an image rather than the descriptions associated with the image like tags, descriptions, links, keywords, etc. It checks for the context like image colors, shapes, textures, or any other derivation of information from the image itself. We get results on the basis of the content of images. This technique helps to remove the dependency of annotation quality and completeness which is created in case of searches relying on the metadata of the digital image. The scope of queries is limited when searched by keywords to the set of predetermined criteria which is also less reliable irrespective of the content.

Example: Suppose we have a database of images, we want all images having red color in it. Hence, we fire query with search string as red, and then we get all images that have a red color in them.

Concept-Based Image Retrieval is based on the image description and the structure of the image. This is an image-specific technique. This technique is useful in searching for some kind of documents. Documents like Aadhar Card, Pan Card, Driving Licence, Credit Card, etc. This document has a fixed structure. Here we don't fire queries on the basis of the content of an image. Concept-Based Image Retrieval technique finds the meaning of query based on that it gives the result.

Example: Suppose we have a database having images of documents. Moreover, we want all images of Aadhar Card. We know that the Aadhar card is rectangular in shape, it has a photo, it has a QR code. But, as we know here we don't fire query on the basis of the content of an image.

Here we directly fire a query 'Aadhar card', then it provides all images of Aadhar Card.

OCR

This paper [1] is based on the functionality of Optical Character Recognition (OCR) and speech synthesizer. They have proposed and developed one android mobile application. This application can be used for performing image to speech conversions using OCR. The OCR takes the image as the input, gets a text from that image, and then converts it into speech. OCR technology is used in this paper for character recognition.

This paper [2] is based on the comparative study of Open Source OCR tool Tesseract with other commercial OCR tool Transym OCR. They have considered vehicle number plate as input. This paper also explains about Optical Character Recognition (OCR) method, History of Open Source OCR tool Tesseract, architecture, and experiment result of OCR. In this paper, they have also compared OCR tools based on various parameters.

2 Literature Survey

A. *Related Work*

This paper [3] proposes a technique to retrieve scanned documents using fuzzy graph-based document retrieval. The document is structured using Attributed Relational Graphs. Every node in the graph is resultant of a fuzzy membership degree for each region in the document. Fuzzy membership degree is calculated on the basis of low-level properties such as texture, shape, color, etc. In this, documents are specific to Coran. Document segmentation used reduces the extraction of features. Time Complexity increases due to the creation of a tree based on a fuzzy membership degree.

This is a survey paper [4] gathering various techniques for content-based image retrieval. It proposes the idea of active learning which can increase interactive search. It also proposes a system to allow the user to ask multi-modal queries/questions and also give multi-modal input for the generated results on the query. Proposed methods are not implemented.

This paper [5] aims at solving the Perspective Distortion issue faced while Camera-Based Document Image Retrieval. It proposes a method called Locally Likely Arrangement Hashing (LLAH) which is affine invariant instead of the perspective invariant which makes it more adjustable. It also reduces the time complexity to O(n) and reduces retrieval time too. This system doesn't support query-based on textual input but on image input.

This paper [6] explains the process of latent Dirichilet scheme for keyword extraction in a document.

This paper [7, 8] discusses the important facets of image retraction techniques with different schemes color management, Gabor transform, edge gradient, and HSV. And it focused on the text and edge extraction based on the performance analysis parameters such as accuracy, error rate, sensitivity, and accuracy.

This paper [9] introduced a new scheme for classification in convolutional neural networks with different five subclasses based on the cosine similarity function.

3 Proposed Methodology

A. *Proposed Architecture*

Consider a typical scenario of an image retrieval system consisting of client–server architecture and some other layers supporting the system. The proposed architecture in this paper is of 5 tiers consisting of client, handler, preprocessor, feature extractor, and analyzer node module. The analyzer and feature extractor layers are modifiable and can be updated according to the image category. The details of each layer are discussed below.

1. Client—Client includes any user probing for specific document images. The proposed model deals with the official Indian Government Document. User needs to input the type of document images he/she wishes to fetch. This textual query triggers the classification process on the local data of the user. Thus, the user sends a GET request to the server node which is mediated by Handler. In return, the server processes the images and fetches the required images which are classified based on the input query by the user.

2. Handler—Handler works as a mediator between Client and Server. The handler is a part of the server node which resides at the server end. It handles every request made by the client and response given by the server node. Request routed through the handler is in a textual form consisting of keywords identifying the document image to be searched. While every response is an image or set of images matching the query fired by the user. Requests and Responses are in the HTTPS form. Handler also forwards the local images for the server node to process on.

3. Preprocessor—Preprocessor is used to do initial processing on the input images. It is a part of the server node. The preprocessor helps to standardize the input images which can make further steps easier. In preprocessing, images are first converted from color to grayscale images. These converted images are then cropped to the document edges irrespective of their orientation. This helps to target the required area neglecting other unimportant portions in an image.

4. Feature Extractor—Feature plays a very important role in the area of image processing. Features define the behavior of the image. It is a simple image pattern, based on which we can describe what we see in the image. Features are the unique signatures of the given image or unique properties that define an image. Feature Extraction is useful to get important features of an image and this feature can be used for the classification of images. Feature extractor node helps to extract important features that can be used to differentiate various documents.

5. Analyzer Node—Extracted features from the feature extractor are considered for the classification process. Analyzer node is a part of the server node. Analyzer node helps to classify the images based on the combinations of the various features. Details regarding the same are discussed below. After classification, the analyzer returns an image or set of images to the handler as a response to the user query.

B. *Working*

The client sends a request query on the server which signifies the type of document to be fetched via the web application interface. The server, apart from listening to request extract features from the HTTPS requests received from the client. Handler extracts the Input images from local images of the user along with the keywords in the user query. The handler then forwards these images for the classification process to the next layer called the preprocessor. Preprocessor performs processing such as converting color images to grayscale and cropping them to focus on the desired area in the image to the document edges. These preprocessed images are then fed to the feature extractor layer. In this layer, various features like emblem, logo, photograph, color information, QR code, etc., are looked for and extracted. Once these features are calculated, they are passed to the analyzer model for prediction. If the model classifies the images as an interest of query, then it is returned to the handler and then the response is sent to the Client (Fig. 1).

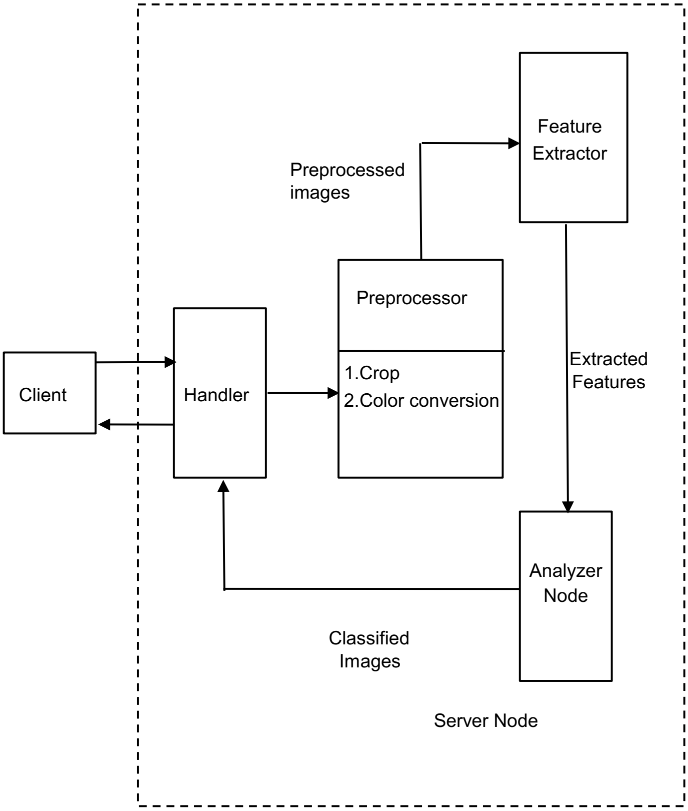


Fig. 1 Proposed model for government document image retrieval

Algorithm 1 Preprocessing and Classification algorithm

```
Extract query from request sent to server for preprocessing
keywords ← extract keywords from query
local_images ← Array of images from user collection
i ← 0
while i < len(local_images) do
    local_images = color2gray(local_images[i])
    local_images = crop_edges(local_images[i])
    i ← i+1
end while
Create feature vector $F'_n$ ($f_1$, $f_2$, ... $f_n$) using images extracted from request where $f_1$, $f_2$...$f_n$ are important features determined by observability from the document images.
    Feed $F_N$ to the analyzer $A_N$ and predict the classification of the image
        pred ← predict($A_N$, $F_N$)
        if pred == query_interest class then
            Server response-> image
        else
            Neglect image
        endif
    endif
```

4 Preprocessor

The preprocessor firstly converts a color image to a grayscale image. This helps to reduce complexity from 3D pixel value that is Red, Green, Blue color intensities to 1D value. Mathematical model [10] for the same is as follows which provides weighted average with R for Red intensity value, G for Green intensity value and B for Blue intensity value:

$$= 0.3R + 0.59G + 0.11B \tag{1}$$

$$= 0.2126R + 0.7152G + 0.0722B \tag{2}$$

$$= 0.299R + 0.587G + 0.114B \tag{3}$$

Following techniques were considered to select the target area from the image:

Edges
This provides points at its extrema which are computed with respect to its image gradient in the direction of the gradient. It provides important features in the analysis of the image and also provides significant changes in the locality.

Contours
Contour is a curve joining all the continuous points along the boundaries. It produces an enclosed edge. This helps to identify the shape and ensure object detection. When contours are formed using edges then we need to connect the obtained edges such that a closed contour is achieved [11]. Gets a measure of boundary strength at each pixel by considering the maximum response over orientations where mPb is Multiscale Pami Boundaries Pb [12], x and y are spatial coordinates and theta is the angle made by the division of diameter:

$$mPb(x, y) = max_{theta}\{mPb(x, y, theta)\} \tag{4}$$

Perspective Transformation
Transformation is done considering the 4*4 matrix. Steps involved are the translation of data such that the observer is at the origin, the observer's coordinate system is then transformed by three number of rotations. This helps to align the view vector to the global negative z-axis. Also up vector is aligned to the global y-axis. Followed by this, the z-axis is then flipped. This is achieved by negating the z-coordinate.

5 Feature Extractor and Analyzer

Feature extractor takes the preprocessed image as an input and lists out important features of images. It considered features like emblem, logo, photograph, color information, QR code, etc. Feature extractor considers a different combination of features depending on the type of image. Type of images are like frontside image, backside image, half image, etc.

Consider Aadhar card images, there are various combinations of features according to the type of Image. The following table shows features with respect to the type of image.

For the detection of various features proposed algorithm used various techniques. For Green and Orange color detection, the algorithm considers RGB values of color. Logo and emblem are detected by using template matching. For QR code and Photograph, detection algorithm used OpenCV techniques.

Using Table 1, the analyzer node can analyze whether the given image is of Aadhar card or not. According to the type of image, the analyzer checks whether the image contains corresponding features or not. If the image contains all required features then the algorithm detects that image is matched with the input query.

Table 1 Features of Aadhar Card

Features	Frontside	Backside	HalfImage
Emblem	✓		✓
Logo	✓	✓	
Photograph	✓		✓
Color (Orange and Green)	✓	✓	✓
QR code	✓		✓

6 Experimentation and Evaluation

The following snapshots showcase the results of every stage in the proposed model.

Step 1: Preprocessor (Figs. 2 and 3).

Step 2: Extractor and Analyzer (Figs. 4, 5, 6 and 7).

Results were formulated based on the photo gallery images of mobile. The input images consisted of total 500 images out of which 237 were images of important documents (Aadhar Card) while 263 were other images. The proposed model correctly classified 237 images as Aadhar Card images while 4 other images were also classified as Aadhar card images and 259 were classified as other images.

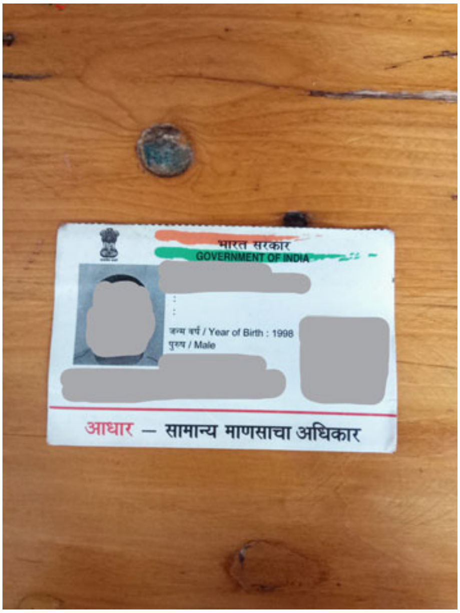

भारत सरकार
GOVERNMENT OF INDIA
जन्म वर्ष / Year of Birth : 1998
पुरुष / Male
आधार – सामान्य माणसाचा अधिकार

Fig. 2 Input image to the preprocessor

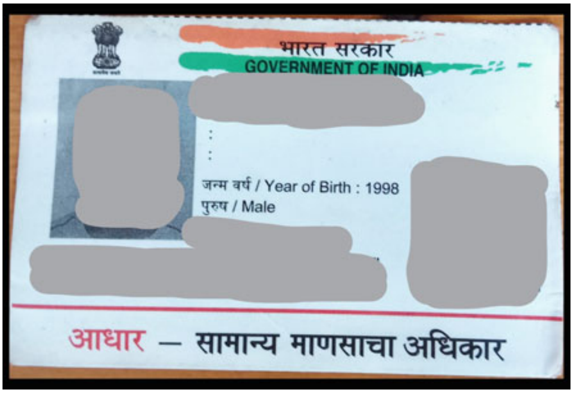

भारत सरकार
GOVERNMENT OF INDIA
जन्म वर्ष / Year of Birth : 1998
पुरुष / Male
आधार – सामान्य माणसाचा अधिकार

Fig. 3 Output cropped image of the preprocessor

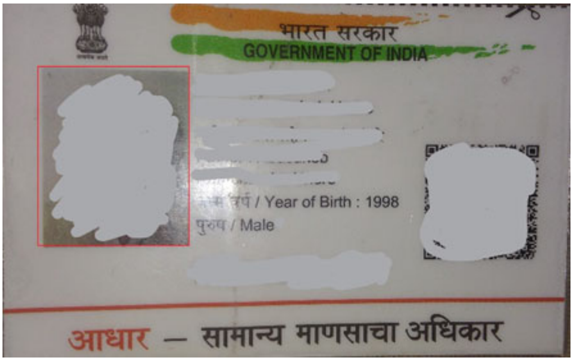


Fig. 4 Detect photo in the image

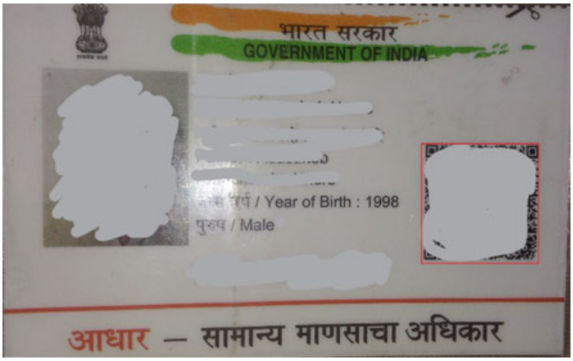


Fig. 5 Detect QR code in the image

A. *Results*

See Table 2.

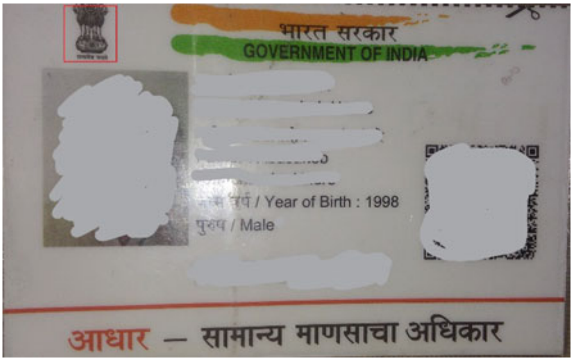


Fig. 6 Detect emblem in image

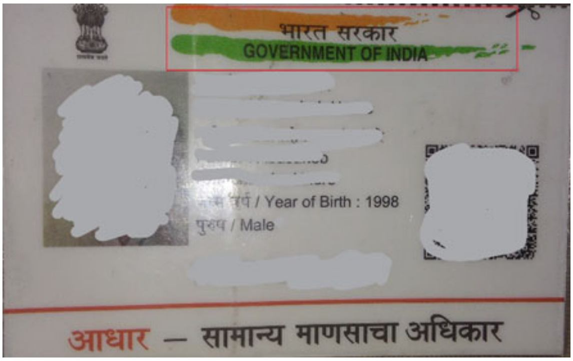


Fig. 7 Detect color in the image

Table 2 Confusion matrix

Results		Actual interest images		Total
		Positive	Negative	
Predicted interest images	Positive	237	0	237
	Negative	4	259	263
Total		241	259	500

7 Conclusion

The proposed model is used to identify and detect government document images. The proposed model quickly identifies specific government documents from a chunk of images with ease. Various filtering categories are applied to ease the process of categorization. The process of categorization is speeding up by short listing selective efficient features that contribute majorly toward the substantiation of a government document. Out of 500 input images, 241 images were of interest area. From these 241 images, the proposed model successfully identified 237 images as the query result. The proposed model provides better results.

8 Future Scope

In the proposed model, feature extraction and template matching techniques are used. Later on, this technique can be extended to other ML algorithms and Artificial Neural Networks. Interpretable Machine Learning Techniques can be used to justify truth and validation for extracted features which contributed to classifying the image as an interesting area with respect to the query. Misclassified samples can be used for rectifying the model by using continuous learning.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Mithe R, Indalkar S, Divekar N (2013) Optical character recognition. Int J Recent Technol Eng (IJRTE) 2(1):72–75
2. Patel C, Patel A, Patel D (2012) Optical character recognition by open-source OCR tool tesseract: A case study. Int J Comput Appl 55(10):50–56
3. Chaieb R, Kalti K, Essoukri Ben Amara N (2015) Interactive content-based document retrieval using fuzzy attributed relational graph matching. In: 13th International conference on document analysis and recognition (ICDAR), Tunis, 2015, pp. 921–925
4. Thomee B, Lew MS (2012) Int J Multimed Info Retr 1:71
5. Nakai T, Kise K, Iwamura M (2006) Use of affine invariants in locally likely arrangement hashing for camera-based document image retrieval. In: H. Bunke, A.L. Spitz (eds) Document analysis systems VII, DAS 2006. Lecture Notes in Computer Science, vol. 3872. Springer, Berlin
6. Cha MS, Kim SY, Ha JH, Lee MJ, Choi YJ (2015) CBDIR: fast and effective content-based document Information Retrieval system. In: 2015 IEEE/ACIS 14th International conference on computer and information science (ICIS), Las Vegas, NV, pp 203–208
7. Hou D, Wang X, Liu J (2010) A content-based retrieval algorithm for document image database. In: 2010 International Conference on Multimedia Technology, Ningbo, pp 1–5
8. Kaur M, Sohi N (2016) A novel technique for content-based image retrieval using color, texture and edge features. In: 2016 International conference on communication and electronics systems (ICCES), Coimbatore, pp 1–7
9. Rian Z, Christanti V, Hendryli J (2019) Content-based image retrieval using convolutional neural networks. In: 2019 IEEE international conference on signals and systems (ICSigSys), Bandung, Indonesia, pp 1–7
10. https://profs.info.uaic.ro/~ancai/DIP/lab/Lab_2_DIP.pdf
11. Maire MR (2009) Contour detection and image segmentation. The University of California, Berkeley
12. Martin DR, Fowlkes CC, Malik J (2004) Learning to detect natural image boundaries using local brightness, color and texture cues. PAMI

A Novel Repair and Maintenance Mechanism for 'Integrated Circuits' of Ubiquitous IoT Devices by Performing Virtual IC Inspection Based on 'Light Field Technology'

Vijay A. Kanade

Abstract The proposed research discloses a novel mechanism for repairing and maintaining the electronic circuits during an event of failure without actually dismantling or opening-up the device circuit boards. The mechanism uses light field technology, wherein a 3D interactive view of the short circuit component is projected so that the technician can directly interact with the 3D display and in turn fix the wiring in the actual device circuit without physically interacting with the circuit. As the technician interacts with the 3D display, the light source within the internal circuitry generates multiple light beams that fluctuate with varying intensities and apply an optical force on the damaged circuit components so that the circuit components align and re-arrange themselves. We have used Spatial Light Modulator (SLM) for this purpose since the SLM optimizes and focuses light on the target (damaged) circuit component and can drive the component mechanically with greater magnitude. The optimization of SLM is carried out by using the Genetic Algorithm (GA). The simulation results for the optimized SLM based on GA are presented in the research proposal.

Keywords Integrated Circuit (IC) · Light field technology · Ubiquitous IoT devices · Virtual IC inspection · Spatial Light Modulator (SLM) · Genetic Algorithm (GA)

1 Introduction

Electronic circuits have become an integral part of the IoT world we live in today. Today's electronic age highlights a wide range of ubiquitous gadgets that are used across geographies in diverse application areas, right from clothes to smart tattoos, implants to accessories. The integrated circuits employed in all such smart devices are highly sophisticated these days, however, the circuits do pose a threat of physical damage. In some cases, even an extremely small crack disrupts the current flow

V. A. Kanade (✉)
Senior Patent Associate, Intellectual Property Research, Pune, India
e-mail: kanade.science@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_54

through the circuit which may eventually lead to device crackdown or failure. The conventional electronics have traditionally been fixed by soldering process or manual fixation. However, repairing and handling the advanced electronics, i.e., ICs at the nanoscale level by using traditional methods seems problematic. Therefore, fixing such ICs, battery electrodes, or any electrical components when they break down remains a critical challenge.

The research proposal solves the above problem. The proposal discloses a novel repair and maintenance mechanism for ICs of ubiquitous IoT devices by allowing virtual IC inspection without opening-up the circuitry of the device. The proposed method operates on light field technology.

2 Survey Overview

Consider a short circuit scenario, i.e., dysfunctional electronics that have damaged the phone circuitry. The short circuit is an important cause for faults in IC wiring, leading to device failure. Short circuits occur due to loose wire connections, faulty appliance wiring, etc. In the case of loose wire connections, it allows the live and neutral wires to touch and cause a short circuit. Faulty appliance wiring implies when you plug-in an appliance into a wall socket, the wiring of the appliance becomes an extension of the circuit. Now, in case the appliance wiring is faulty, it leads to further circuit problems in the socket circuit.

In all the above cases, the solution adopted by the community today is a yearly electrical inspection by a technician or a certified professional. Here, the technician identifies any critical issue by examining the appliance or socket circuits and resolving them as per the need. Hence, the repair and maintenance of the electrical appliances, etc. are essentially performed on a time period basis. Further, loose connection problems are dealt with by the technicians by opening the circuitry and tightening the bad connections. However, this usually does not help, since it is not always possible to open the device or appliance circuitry for repairing purposes due to the damage it can cause to the nanosized circuitry of the IoT devices these days, or lack of feasibility of opening the circuitry of isolated IoT devices, etc.

Hence, there seems to be an immediate need for an advanced method for operating the circuit's wiring without opening-up the device—since, with the IoT boom, the devices are getting smaller and smaller in size. This means the internal electronic circuit is getting reduced to the nanoscale level.

3 Modules

The disclosed research operates on two basic modules which are as specified below:

3.1 Light Field Capture and Reconstruction of 3D View [2]

This module is supported by an internal intense light source and a nanosized camera module that is installed on the lower portion of the top panel of the gadget (such as a smartphone) that is facing the circuitry of the smartphone, wherein the ICs are lying on the circuit boards below the light source. The light source generates multiple light beams and illuminates the circuitry, and the adjacent camera captures the light field of the underlying circuitry and sends the signal to the light field camera situated on the top portion of the smartphone for projecting a 3D interactive display of the internal circuitry that needs repair or maintenance. Here, the user or technician can interact with the 3D view of the circuitry projected by the phone and trigger internal circuitry to re-organize based on the corresponding action on the 3D scene.

3.2 Light-Driven Circuitry [5]

Further to the light source (i.e., below the light source), a spatial light modulator is installed for focusing the light beams on the integrated phone circuitry to be repaired, wherein the Spatial Light Modulator (SLM) focuses the light beams on the circuit wires that have been disoriented or overlapped or have a bad connection. The projected 3D scene and the internal light source are time synched, i.e., as and when the technician interacts with the 3D scene of the IC projected by the light field camera of the smartphone, the intensity of the internal light beams that pass through the SLM varies. As the intensity varies, the light exerts optical force to do mechanical work on the target integrated circuit requiring repair (i.e. re-wiring the overlapped wire or correcting the bad connection).

4 Working Principle

On a fundamental level, 360° viewable 3D display is known in the art [3]. Here, properties of the optical configuration are exploited for providing integrated input sensing over the 3D view along with the simultaneous updation and re-creation of the 3D scene [1].

Now, light-driven circuitry works on a basic understanding of the fact that light carries momentum in the propagation direction that is directly proportional to its energy [4]. When an intense light beam passes through an object such as a bead or microsphere, the beam refracts, bends and changes the propagation direction. This, in turn, alters the initial momentum of the light beam. Now, according to Newton's third law, the system conserves total momentum as the light beam passes through the bead. Hence, the object (i.e., bead/microsphere) under consideration undergoes equal and opposite momentum change. This eventually leads to reactive force to the incident optical force acting on the bead.

Figure 1 discloses the mechanical work done by the light beam, wherein the microsphere is displaced due to the light beam traveling through the microsphere. The figure highlights the transfer of photon momentum to the microsphere.

Now, consider another scenario, wherein two light beams hit the micro-sized sphere or bead. One light beam hits the microsphere at the center and the second light beam at the microsphere sidelines.

In Fig. 2, the total forces experienced by the microsphere, or bead include two components: (1) a scattering force and (2) a gradient force.

The scattering force is the force developed by the light beam when an incident light beam gets scattered by the microsphere surface. The very scattering effect gives

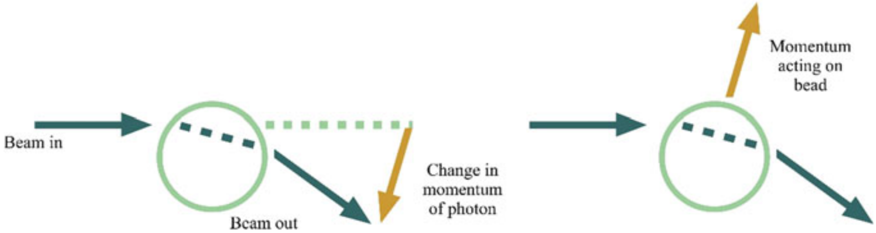


Fig. 1 Light beam applying optical force on the microsphere [6]

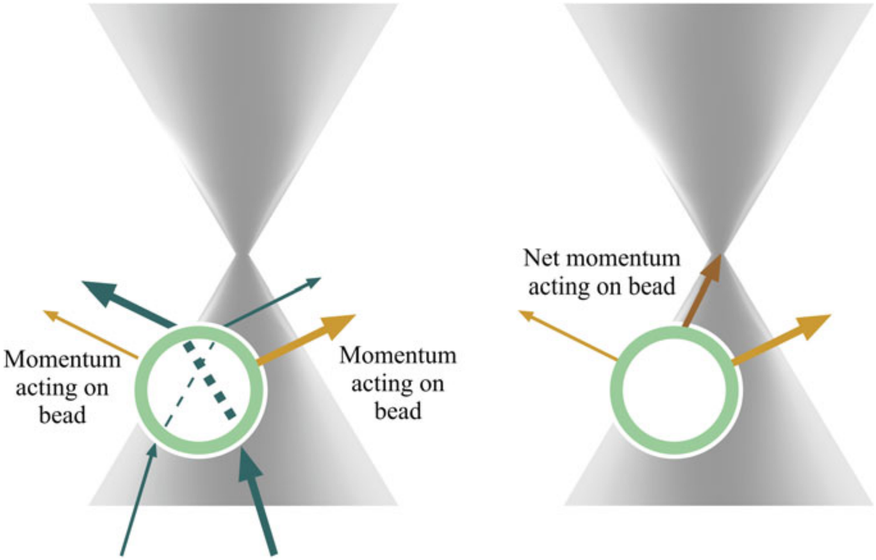


Fig. 2 Displaced bead due to two incident light beams [6]

rise to a net momentum transfer from the photons of the light beam to the microsphere and causes the microsphere to move toward the direction of the beam.

Further, the gradient force is the force that acts due to the light intensity profile of the beam. This force is an attractive force and draws the microsphere toward the greater light intensity region. Further, in scenarios wherein a highly intense light beam is used, the force acts as a restoring force that pulls the microsphere into the center of the focal plane.

Here, two light beams pass through the microsphere. Now, the light beam hitting the center of the bead carries more number of photons than the light beam hitting the bead from the sidelines. Hence, a larger force is generated by the beam hitting the center of the microsphere than the light beam hitting the bead on its sidelines. However, the latter light beam does exert some sort of force on the bead. This causes the bead to move in the direction that is a resultant of the forces applied by the two light beams, i.e. the focal plane in this case as observed in Fig. 2.

The disclosed research utilizes the above principle for its operation.

5 Methodology

This research study is conducted for a smartphone gadget. Consider a case where one of the smartphone circuitry has undergone a short circuit due to accidental wire overlap. The wire overlap has occurred due to device mishandling, and hence due to eventual short circuit, the device has turned off. Now, in order to perform the repairing task of the overlapped wires, the smartphone is plugged-into an electrical connection. This electrical connection switches on the internal light source and the nanosized camera that is provided within the smartphone circuitry. The light source focuses on the internal circuitry of the phone. Here, the focusing area of the internal light is the circuit area having the overlapped wires. The nanosized camera captures the light field of the circuit and sends the corresponding signal to the light field camera situated on the top panel of the smartphone. The light field camera projects the 3D view for the captured light field data. This 3D view provides the interactable surface to the technician. Hence, the technician starts interacting with the projected 3D scene wherever repair is needed. Now, this light field camera is provided with the additional functionality of capturing the infrared radiation of a certain wavelength (i.e., 10–12 μ). A normal human body radiates IR of about 10 μ hence the light field camera captures the IR radiation of the technician's hand. The captured IR data is communicated to the internal light source since the two entities operate in time-synched manner. Thus, as the professional interacts with the damaged or overlapped wire connection, the light field camera captures IR of technician's hand, sends signal to the internal light source and the internal light source, in turn, generates multiple light beams that fluctuate with varying intensities based on the kind of interaction the technician has with the 3D scene. Further, the intensity of light beams is optimized by the usage of SLM, wherein SLM operates at the pixel level and allows the light beams to focus on the target center, i.e. overlapped wire in this case. Here, for

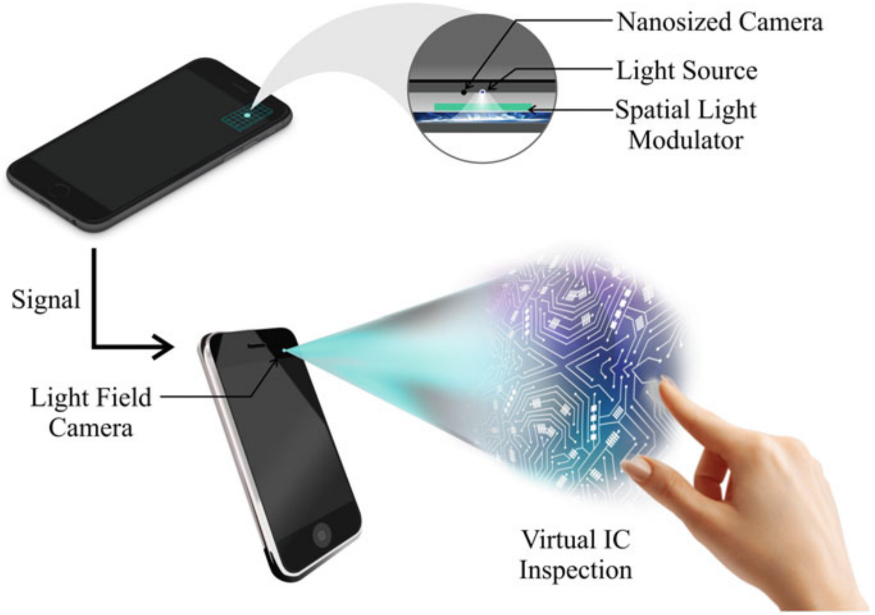


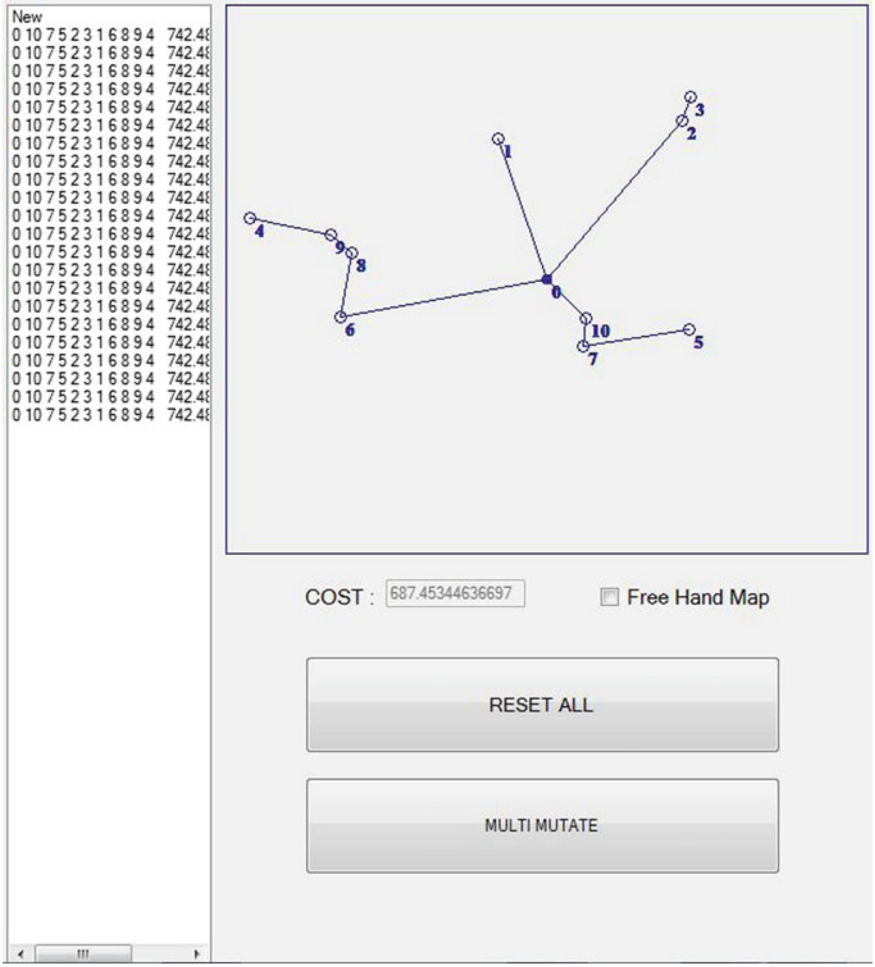
Fig. 3 Repair and maintenance of faulty 'integrated circuits' of the smartphone by performing virtual IC inspection based on light field technology

the optimization of the SLM, we have employed a genetic algorithm. The genetic algorithm ensures that the light passing through the SLM is of such high intensity that it exerts a quantifiable optical force on the target overlapped wires. Thus, as the interaction of the technician builds on, the light source shows the continuous fluctuation in the intensity of light beams and correspondingly acts on the overlapped wires. Now, as the technician disentangles the overlapped wires in the virtual 3D scene, the internal light replicates this action of the technician and thus the overlap of circuit wires is resolved. Therefore, the circuit starts functioning normally as the reason for a short circuit of the phone is resolved, through the light-driven circuitry without opening-up the smartphone and dismantling the circuit board of the device.

Figure 3 discloses the diagrammatic representation of the above methodology, wherein the smartphone circuitry is repaired via virtual IC inspection and operation performed by the professional technician.

6 Simulation Results

The simulation results for the optimization of the SLM by utilizing a genetic algorithm are elaborated below. Initially, a Genetic Algorithm (GA) generates an output map for the circuitry under the field of view of the nanosized camera. Here, a map



```
New
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
0 10 7 5 2 3 1 6 8 9 4  742.4
```

COST : 687.45344636697

□ Free Hand Map

Snapshot 1 Genetic algorithm-based UI for SLM optimization

is developed in such a way that the target area of the circuitry (i.e., for example, overlapped wires) is marked at the center of the map at 0 (0, 0) based on the [x, y] coordinates. The task of the employed GA is to focus light beams on the center 0, such that it exerts a substantial mechanical force on the circuit. Further, it is important to note that GA is specifically designed for the SLM, wherein various pixel dots corresponding to the SLM can be turned on/off based on the optimized result of GA (Snapshot 1).

In step one, GA generates the circuit map with the target area requiring optical focus marked at the center. The GA undergoes random selection of various pixel dots on the SLM configuration such that the pixel dots are configured in such a way that the phase of light passing through these pixel dots is altered in a particular manner.

Such pixel dot arrangement focuses light beams with maximum intensity at the center 0. Further, in the second step 'initial population' for randomly selected pixel dots is generated. The GA then calculates the fitness function for all the solutions of the initial population. Here, the fitness function is calculated as minimal of the total sum of the distance of various selected pixel dots from the center of the GA map. In the next step, GA iteratively performs crossover and mutation so as to yield an optimal result for the continuously evolving generations (i.e., population), wherein the fitness function for the child generations is optimum in comparison to the parent generation. Here, if the fitness function for the newly generated child generation is lesser to parent, then the child generation is retained and the parent generation is killed. And if the fitness function of the child generation is greater than the parent, then the child generation is killed and the parent is retained. In this way, the GA continues to evolve until the optimal result is produced for the SLM.

Here, it is important to understand the meaning of an 'optimized result'. The optimized result implies a minimal cumulative distance of all the pixel dots from center 0 along with the number of connection links originating from the center and terminating at different pixel dots. All the pixels lying on one connection link would be oriented to a certain degree such that the multiple light beams passing through them undergo a defined phase change (e.g. 90° phase change). The above discussed optimized result for the pixel dots determines the pixel alignment and configuration so that the ultimate maximum light focuses at the center 0.

The pixel arrangement thus achieved by the SLM allows the light beams to focus on the target center 0, wherein the focused light beams apply quantifiable magnitude of optical force on the center such that the overlapped wires resolve their position and return to their original structural construct. Thus, GA facilitates disentanglement and re-arrangement of the circuit such that the circuit board is repaired effortlessly and the device eventually starts functioning normally.

Further, the snapshot of the simulated GA for optimized SLM with maximum light intensity at the center 0 is shown below. The light intensity ensures that the maximum optical force is applied at the target center 0 so that the overlapped wires in the circuit are disentangled and resolved to a normal state.

Above disclosed is the simulation generated map which localizes center at 0 based on [x, y] coordinates and the pixel dots [x, y] of the SLM. Here, the center denotes the point in the damaged circuit where wires are observed to be overlapped. In optimization of SLM, the pixel dots (11) are oriented and configured in such a way that that the light beams received from the rear light source undergo phase change by some degree at each pixel so that maximum light intensity is focused at the center 0. This makes the center more susceptible to the optical force due to light beam(s) intensity and thus the electrical components at the center can be moved mechanically to re-arrange itself into its normal circuit configuration. This implies that the overlapped wires at the center disentangle themselves to form a usual circuit wiring. In the above case, 11-pixel dots were localized by the GA software, out of which 4-pixel sets were chosen for optimization of SLM since the map outputs four connection links from the center to the terminating pixel dot(s). This implies that each pixel in the pixel set (i.e. connection link) would be aligned and configured

Table 1 SLM optimization results

Randomly selected pixels	Optimized fitness function	Phase change (Φ) junctions (pixel sets)
9	639.817 units	Φ (3)
12	936.789 units	Φ (4)
7	715.481 units	Φ (3)
10	767.997 units	Φ (4)
8	700.781 units	Φ (3)

in such a way (at certain degree) that the light beams passing through these pixels would hit the center 0 at maximum magnitude as the light undergoes phase change at 4 different junctions on the SLM (i.e. each connection link configures all its pixel in one orientation). Consider one of the connection links above, for example (0–5). Now, for this particular connection link, all the underlying pixels (10, 7, and 5) would be oriented in a way that would allow multiple light beams originating from the light source to undergo a fixed and defined phase change (e.g., phase change of 45°) as light beams pass through these pixels. This would ensure that the maximum optical force is exerted at the center 0 by multiple light beams passing through the localized pixels.

Further, for the disclosed scenario, the fitness function turns out to be minimal for the GA optimized SLM that applies significant optical force at the center 0. Here, the fitness function turns out to be 687.453 units as seen at the center of the UI of the above snapshot. Hence, the GA-based optimization of SLM discloses an effective mechanism to apply appropriate optical force on the target circuit components.

Similarly, the optimization results for few more circuit repair scenarios developed and simulated by using Visual Studio 2008 platform and C# language are as summarized in Table 1.

7 Conclusion

The proposed research allows technicians to perform repair and maintenance activity on the IoT device circuitry by performing IC inspection and operation on virtual IC. The repair mechanism presents a novel feature of operating on the re-constructed 3D volume of the internal device circuitry and replicating it on the actual internal device circuit to resolve the circuit misconnections, overlap or any wiring problems, etc. Further, the proposal validates the usage of a genetic algorithm for optimizing the SLM in order to focus the multiple light beams on the problematic circuitry so that the light beams can perform mechanical work on the target IC. Thus, the paper discloses the first-of-its-kind technique that allows the IoT device circuitry to be repaired without ripping open the circuit board of the device.

Acknowledgments I would like to extend my sincere gratitude to Dr. A. S. Kanade for his relentless support during my research work.

References

1. Butler A et al Vermeer: direct interaction with a 360° viewable 3D display. In: UIST'11, 16–19 October 2011, Santa Barbara, CA, USA
2. Yamaguchi M, Higashida R (2016) 3D touchable holographic light-field display. Appl Opt 55(3):A178–A183
3. Shabani S Interactive 3D displays. In: Distributed systems seminar, ETH Zurich, Switzerland
4. Schirber M (2017) Focus: light pushes and pulls. Phys Rev Lett 23 January 2017
5. Bourzac K (2008) First light-driven nanomachine, 02 Dec 2008, MIT Technology Review
6. Optical Tweezers. https://lumicks.com/optical-tweezers-working-principle/

Evolutionary Optimization of Spatial Light Modulator for Advanced Wavefront Control in an Optically Addressable ‘Electric See-Through Skin’

Vijay A. Kanade

Abstract The research proposal discloses an electronic see-through skin that can be worn or wrapped around any object (i.e., human body, hands, legs, brain, etc.) and would allow to see-through the object irrespective of its physicality. The see-through skin modulates the light rays passing through the scattering opaque medium by utilizing the “Spatial Light Modulator (SLM).” The proposal elaborates on the optimization of the SLM for wavefront control of the light rays as they pass through the optically adaptive electric see-through skin. The optimized SLM helps to see-through any opaque object by resolving the shape of an object that is hidden behind the opaque object and is not in a direct view. The research proposal presents the simulation results of the evolutionary algorithm employed for optimizing the SLM.

Keywords Evolutionary computation · Spatial Light Modulator (SLM) · Wavefront · Light rays · Electronic see-through skin · Genetic Algorithm (GA)

1 Introduction

A transparent glass provides a medium that allows the light rays to pass through the glass surface with minimal distortion and disturbance and hence allows us to see-through the glass surface. However, when we try to look through another surface that does not provide such medium as in an opaque object or even a frosted glass for that matter, what we observe is a blurred image of an object laying on the other side of the frosted glass or the opaque object. This is due to the scattering phenomenon. Frosted glass serves as a scattering medium that scatters the light rays coming from the hidden object and falling on its surface, hence what we see is a vague structure instead of a carved out surface. In case of an opaque surface, we are unable to see-through the surface because the surface reflects and refracts all the light rays coming from the object and falling over it. Hence, we only see the surface itself and are

V. A. Kanade (✉)
Senior Patent Associate, Intellectual Property Research, Pune, India
e-mail: kanade.science@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_55

unable to see-through the surface. Thus, in both the above cases, we are not able to observe the scenario unfolding on the other side of the surface.

At the Weizmann Institute, however, scientists have revealed a research that allows us to see-through any opaque object by being able to resolve the shape of an object that is hidden behind the opaque object and is not in the direct view. The light (from the hidden object) after falling on the opaque obstacle gets scattered in all directions. This scattering gives rise to “white noise.” The researchers at Weizmann Institute have designed a camera that gathers these white noise bits and clubs them together, hence modulating the light path of the rays falling on the opaque object. This allows the light particles to pass through the opaque object and develop a see-through surface [4].

The research is more focused on developing a camera that can see-through any opaque object. The camera recreates an image of the hidden scene blocked by that object. However, there seems to be less research in terms of developing a full-fledged see-through device that allows us to invade any physical obstacle. Such a device could be useful in medical imaging, wherein the doctors or any layman could check and resolve any form of an anomaly occurring in the bodily organs or even brain. The device could allow us to observe in real time the internal body mechanics in greater detail [7]. This could serve as one of the solutions to identify and reason out the cause of neurological diseases, cancers, etc., which are difficult to track in their early stages.

The research proposal discloses an “Electronic See-Through Skin” that can be worn or wrapped around any object (i.e., human body, hands, legs, brain, etc.) and would allow us to see-through the object, however dense it may be.

2 Background

See-through skin is developed and optimized on the basis of two components that are well known in the art: (1) Genetic algorithms and (2) the Operative mechanism of SLMs. These aspects are briefly elaborated in the following section.

2.1 Genetic Algorithms

Genetic algorithms are optimization methods that evolve and adapt to a set of constraints over multiple iterations. For each successive “generation,” random parameter modifications are applied to the best performing object based on the predetermined optimization function. In cases where the end goal is well defined, this evolutionary process can easily outperform humans at determining optimal parameters.

We have used the “Genetic algorithms” for optimizing the SLM that control the wavefront of the light rays that pass through it and help in aligning the scattered light rays coming from the object that is not under direct view [1].

2.2 Spatial Light Modulator

Spatial light modulators are computer-controlled objects that modulate the spatial distribution of light. They are used both for transmission and reflection applications. They are usually an array of cells or pixels (similar to a checkerboard with controllable white/black square locations) that can individually induce a phase delay and/or change the reflection or transmission intensity of the incoming light.

In the above example, we can see that the distorted wavefront passes through the SLM (cross-sectional view). The gray cells in this diagram have a higher index of refraction, effectively slowing down the wavefront locally. Specifically, this shows that if the shape of the wavefront is known in advance, one can use an SLM to tune it to the desired shape. In the diagram above, it is observed that a flat wavefront is recovered by using the right pixel combination [2] (Fig. 1).

In the case of intensity modulating SLMs, the gray pixels are made in a way that partially or totally block the light, instead of simply slowing it down.

We have optimized the effect produced by SLMs in our research by utilizing Genetic algorithms (Fig. 2).

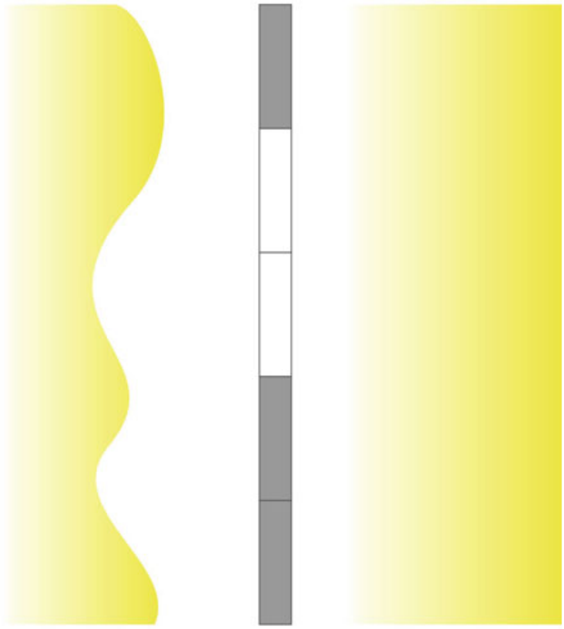
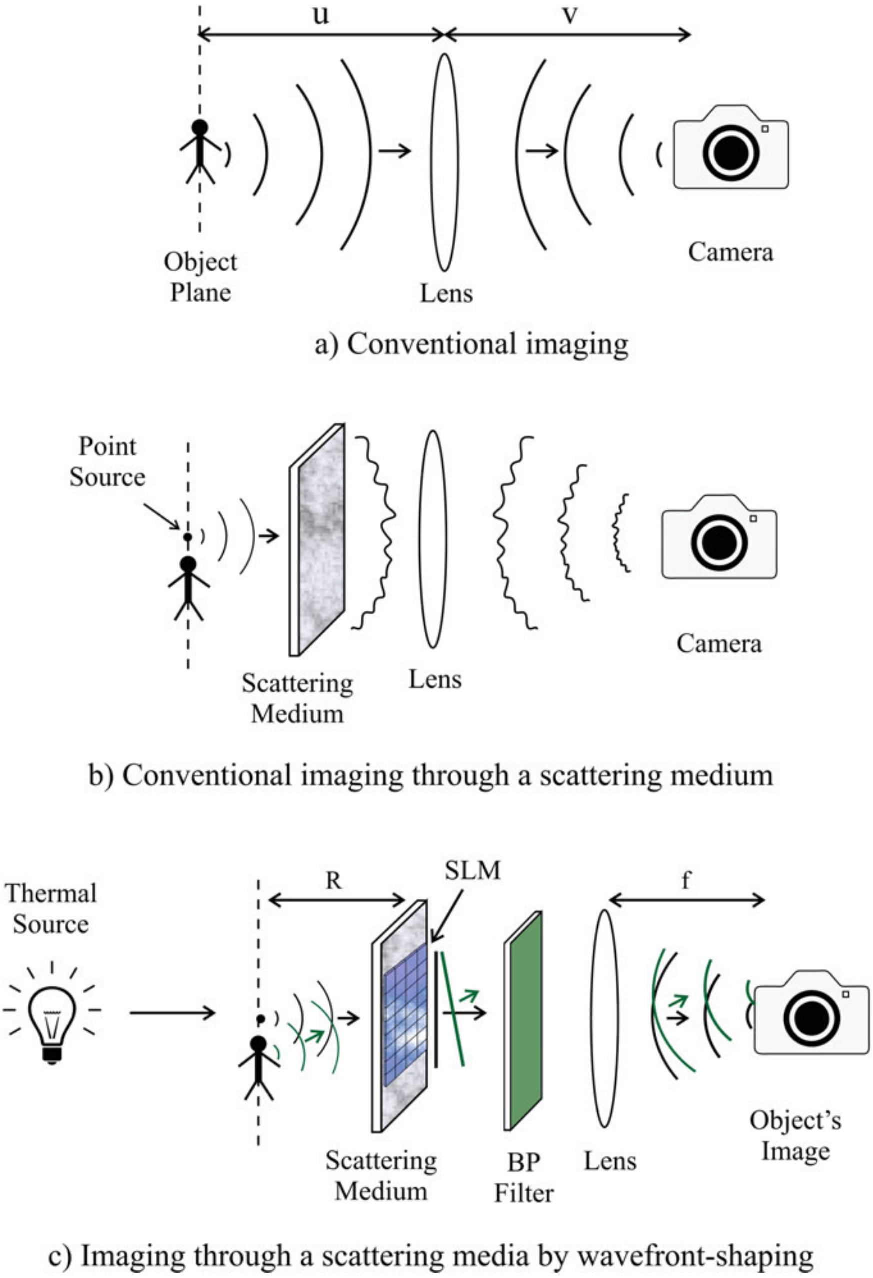


Fig. 1 SLM (cross-sectional view)



a) Conventional imaging

b) Conventional imaging through a scattering medium

c) Imaging through a scattering media by wavefront-shaping

Fig. 2 SLM Operation [3, 6]

3 SLM Operation

Disclosed is the operative mechanism of SLM. In the figure given below, (a) represents a simple imaging system, (b) after adding the scattering medium. As you can see, the scattering medium distorts the original wavefront in such a way that the perceived image is indistinguishable from noise. Just like most surfaces around you, this scattering glass is covered in microstructures that diffuse light and appears to be of a single color instead of making it perfectly transparent or reflective. In (c), an SLM is added to allow wavefront shaping correction. That element is added to attempt to "cancel" the effect of the scatterer by precisely choosing how each cell interacts with light. Since this correction can only be valid for a small bandwidth, a bandpass filter is added.

So to exactly know what pixel distribution is required in order to correct the recreated image for a scatterer, the genetic algorithm is used [5]. It is used to "learn" the best pixel configuration for the SLM by evolving over multiple iterations and building on its past experience. First, we start with a point source at a distance R from the SLM.

From this, a single speckle in the output image is selected. Now, the task given to the genetic algorithm is then to simply try to maximize the intensity around at the chosen point. It first starts with a random pattern, but after few generations, the speckle intensity ends up being enhanced.

Initially, speckle is chosen in the output image. Here, the output image is the map (graph) generated by the GA software that is capable of plotting various pixel points representing the SLM along with the corresponding mapping of the speckle at the center of the map (having [x, y] coordinates [0, 0]). The genetic algorithm then randomly selects various pixels on the SLM for optimizing the intensity produced at the speckle. Further, the "initial population" is generated by the GA. The GA then calculates the light intensity at the speckle based on a cost parameter. Here, the cost parameter is defined as the cumulative distance of various selected pixel points from the speckle (center) plotted on the GA map. Further, the GA undergoes iterative crossover and mutation steps in order to output an optimized result for the newly generated offspring set in comparison to the initially generated population set. Here, optimized distance along with the number of connection links originating from the speckle and terminating at various pixel points decide the pixel arrangement and their corresponding configuration in order to achieve maximum light intensity at the speckle—implying, based on the final output (map) generated by the GA, two parameters decide the type of pixel configuration for that particular SLM: (1) Optimized distance between the speckle and various pixel points and (2) Number of connection links between the speckle and pixel points.

Once that is done, any object in the vicinity of the optimized point source will be imaged live (Fig. 3).

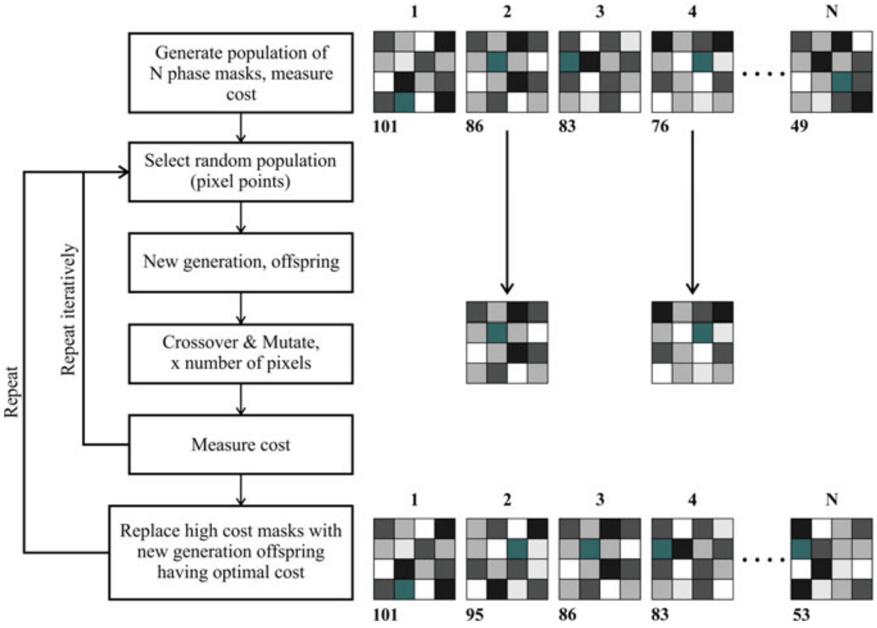

Generate population of N phase masks, measure cost
Select random population (pixel points)
New generation, offspring
Crossover & Mutate, x number of pixels
Measure cost
Replace high cost masks with new generation offspring having optimal cost
Repeat
Repeat iteratively
1
2
3
4
N
101
86
83
76
49
1
2
3
4
N
101
95
86
83
53

Fig. 3 Block diagram showing the steps of GA for SLM

4 Electronic See-Through Skin

Further, it is important to zero-in on the structural framework of the see-through skin. The see-through skin is made up of transparent structure on its upper portion of the membrane. Just below the transparent section, the SLM is embodied that performs the task of modulating the wavefront of the light rays coming from the object hidden from the direct view. Further, just beneath the SLM, the see-through skin is provided with a light source, so that the elements lying below the see-through skin are made visible.

The significant part of this innovation is that it does not use any lasers or X-rays, but employs natural incoherent light source to perform the imaging and currently available hardware. Without any high-end complications in pricy hardware and complex procedures, the proposed technology seems fit to pave the way for research in many fields like medical imaging, etc., without much difficulty (Fig. 4).

In Fig. 4, the electronic see-through skin is wrapped around the leg. The upper portion of the membrane is made transparent to observe the elements below the skin from top of the see-through skin. Now, the light source at the lowest portion of the membrane is turned on. As the conventional incoherent light source is turned on, it travels through the natural human leg skin and beyond. After multiple reflections and

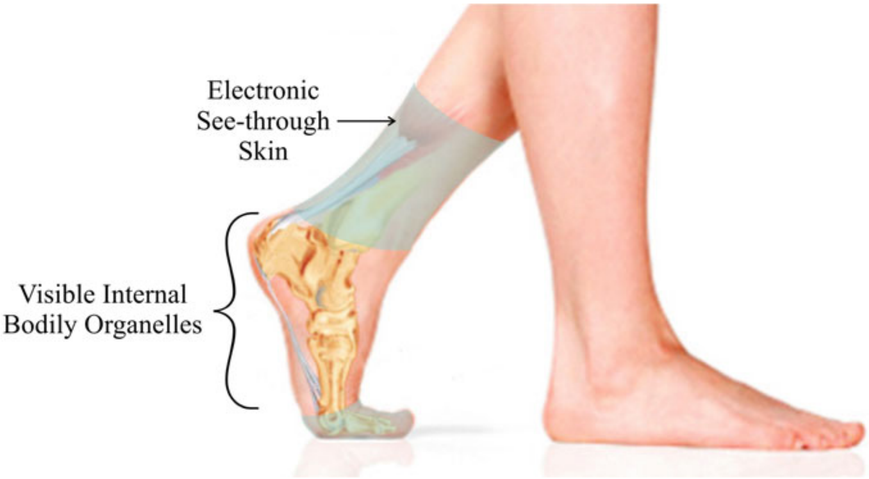


Fig. 4 "Electronic see-through skin"

refraction due to bodily organelles like blood vessels, cells, blood, WBC, ligaments, etc., the light rays reach the other end of the leg. On detection of light rays at the other end of SLM, the GA runs its course. As the SLM undergoes optimization, the light rays are eventually streamlined and the internal body components become visible on the upper transparent portion of the see-through skin.

Here, it is important to note that, as light reaches the other end, there are sensitive detectors in the SLM that detect even the slightest of incoming light rays from a distance and then optimize its optical path. The SLM starts its operation only when the light rays from the other end reach its surface.

5 Simulation Results

The simulation results of the evolutionary genetic algorithm for the optimization of SLM are disclosed in the below section.

The algorithm implemented in this paper follows the below steps:

1. Selection (Random pixel points)
2. Crossover (Cross point)
3. Mutation (Swap mutation)
4. Optimization (Iterative crossover and mutation).

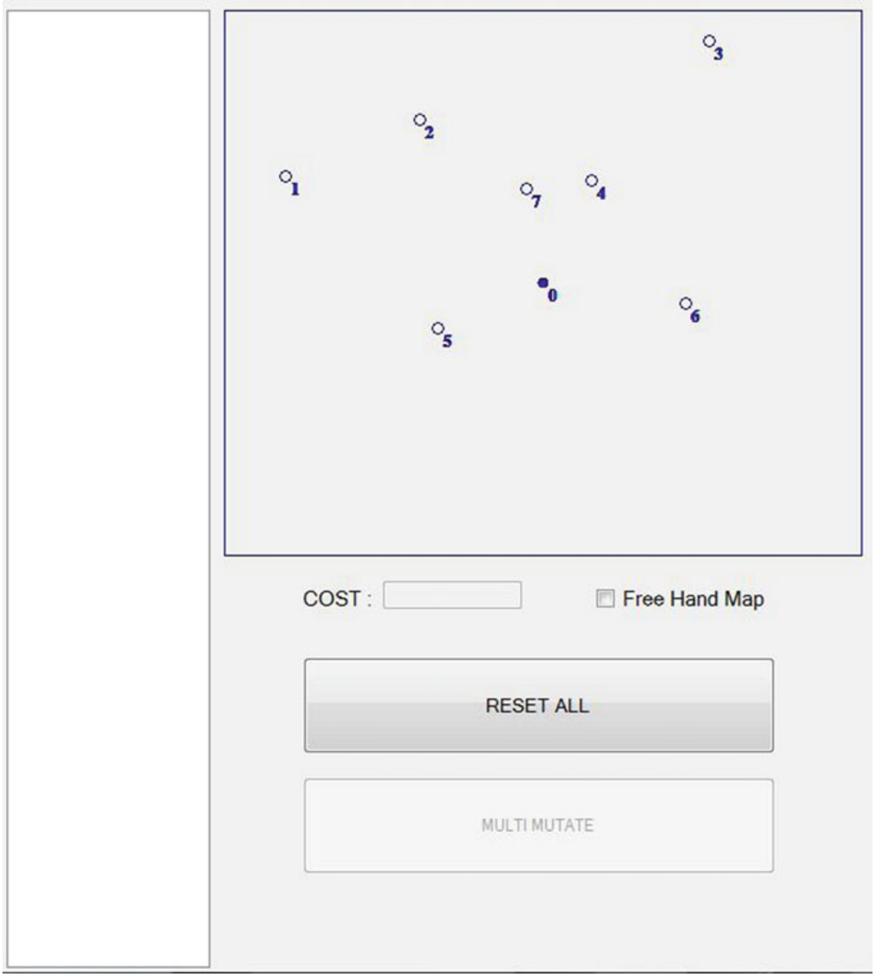
The algorithm begins by plotting the speckle on the GA generated map ([x, y] coordinates at [0, 0]) that needs maximum light intensity so as to make the speckle visible from the see-through skin. Next, the algorithm randomly selects a fixed number of pixel points and localizes them on the map.

Snapshot 1

Snapshot 1 discloses the plotted speckle (0, 0) at the center of the map along with the randomly selected and localized pixel points (1, 2, 3, 4, 5, 6, and 7) for producing maximum intensity at the target speckle.

Snapshot 2

Snapshot 2 discloses the "initial population" generation step, wherein the population set for the localized pixel points is generated and the cost parameter (i.e., sum of

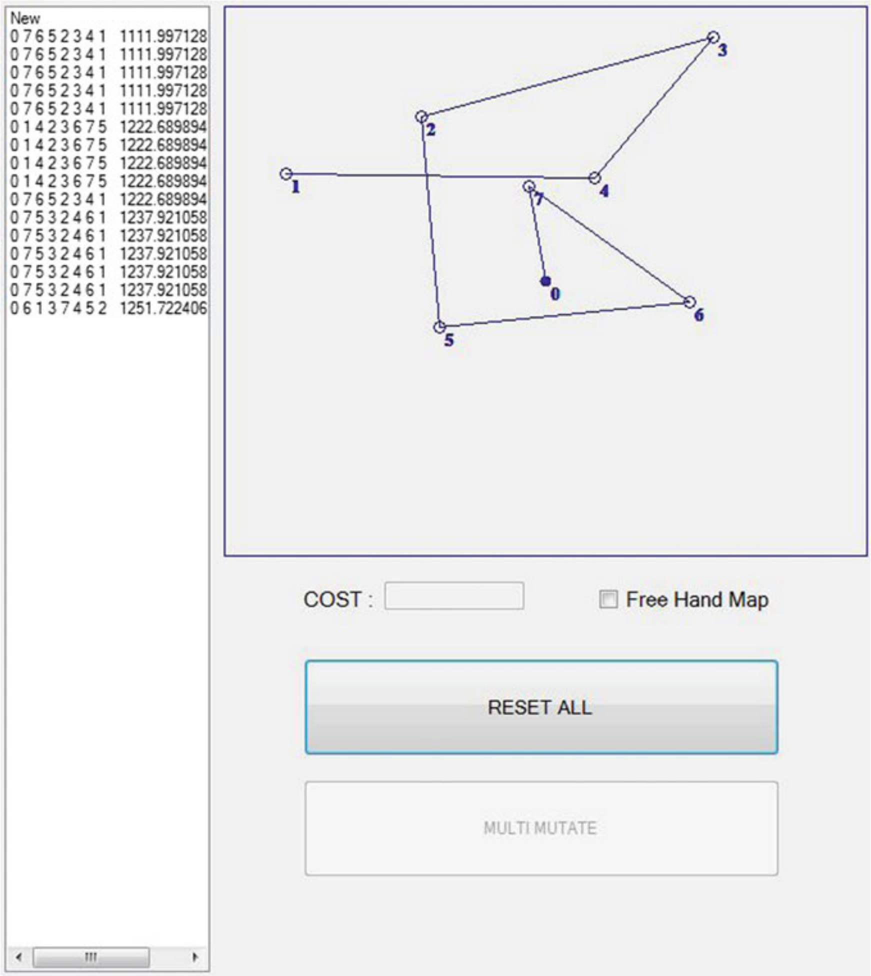


Snapshot 1 Step-I

distance of all pixel points from the speckle based on [x, y] coordinate position of the pixel points on the map) for each solution is calculated. The calculated cost parameter (i.e., in the above example the distance is 1111.997 units, 1222.689 units and so on) for each solution set along can be observed in the left section of the above snapshot (UI).

Snapshot 3

Snapshot 3 discloses the solution generated by the first iterative run of crossover and mutation. The cost parameter for the generated solution (i.e., in the above example the newly generated distance metric after first crossover and mutation is 785.682 units based on x, y coordinate position of the pixel points on the map) can be observed

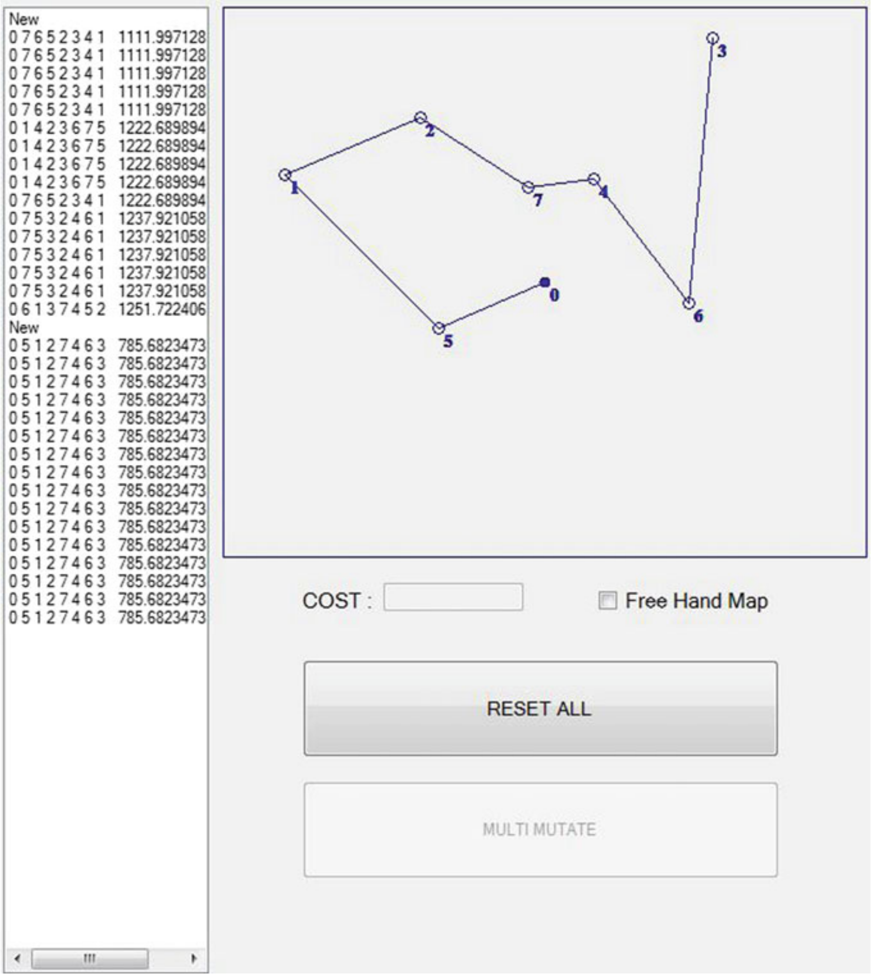


```
New
0 7 6 5 2 3 4 1  1111.997128
0 7 6 5 2 3 4 1  1111.997128
0 7 6 5 2 3 4 1  1111.997128
0 7 6 5 2 3 4 1  1111.997128
0 7 6 5 2 3 4 1  1111.997128
0 1 4 2 3 6 7 5  1222.689894
0 1 4 2 3 6 7 5  1222.689894
0 1 4 2 3 6 7 5  1222.689894
0 1 4 2 3 6 7 5  1222.689894
0 7 6 5 2 3 4 1  1222.689894
0 7 5 3 2 4 6 1  1237.921058
0 7 5 3 2 4 6 1  1237.921058
0 7 5 3 2 4 6 1  1237.921058
0 7 5 3 2 4 6 1  1237.921058
0 7 5 3 2 4 6 1  1237.921058
0 6 1 3 7 4 5 2  1251.722406
```

Snapshot 2 Step-II

in the left section of the snapshot (UI). However, the GA algorithm does not stop here, the various newly generated solutions undergo repeated crossover and mutation until the optimized result is obtained. Thus, all the older generations are killed and the newer offsprings are retained as offsprings yield better optimized result (i.e., minimum cost) or vice versa, i.e., the newly generated offsprings are killed in case the cost parameter of the parental generation is minimal in comparison to that of the offsprings.

Snapshot 4
Snapshot 4 discloses the final output of the GA algorithm, wherein optimized result for the SLM (i.e. pixel points) that produces maximum light intensity at the speckle

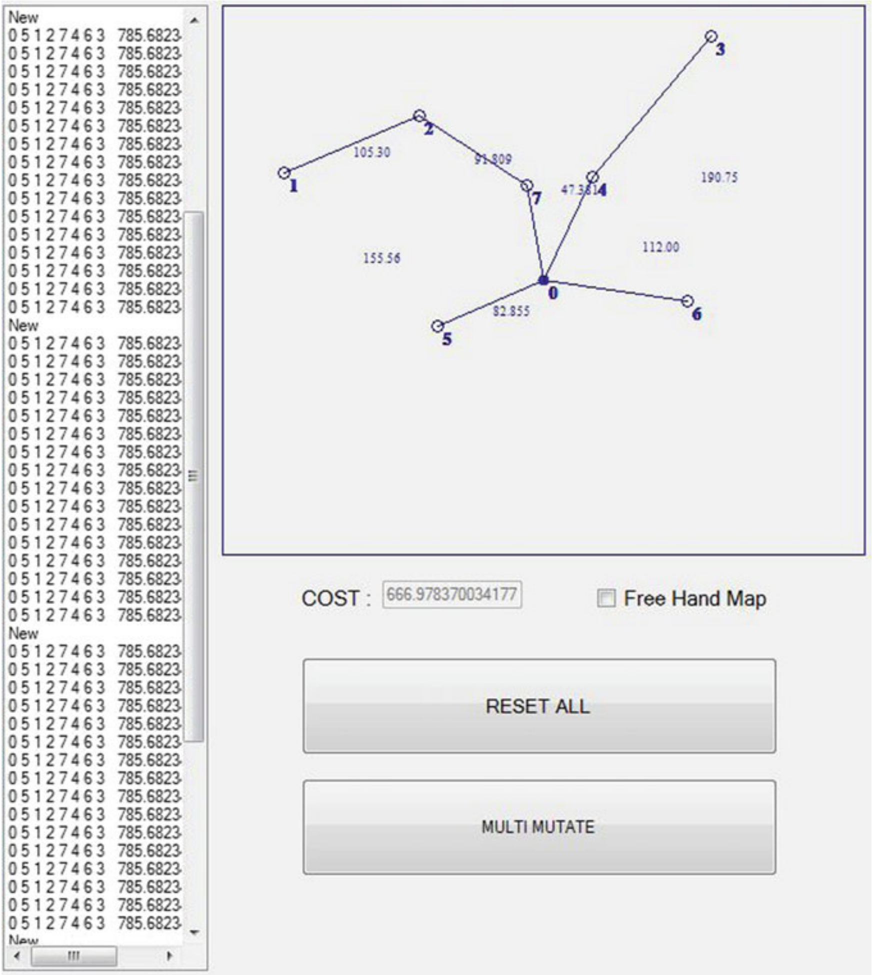


```
New
0 7 6 5 2 3 4 1   1111.997128
0 7 6 5 2 3 4 1   1111.997128
0 7 6 5 2 3 4 1   1111.997128
0 7 6 5 2 3 4 1   1111.997128
0 7 6 5 2 3 4 1   1111.997128
0 1 4 2 3 6 7 5   1222.689894
0 1 4 2 3 6 7 5   1222.689894
0 1 4 2 3 6 7 5   1222.689894
0 1 4 2 3 6 7 5   1222.689894
0 7 6 5 2 3 4 1   1222.689894
0 7 5 3 2 4 6 1   1237.921058
0 7 5 3 2 4 6 1   1237.921058
0 7 5 3 2 4 6 1   1237.921058
0 7 5 3 2 4 6 1   1237.921058
0 7 5 3 2 4 6 1   1237.921058
0 6 1 3 7 4 5 2   1251.722406
New
0 5 1 2 7 4 6 3   785.6823473
0 5 1 2 7 4 6 3   785.6823473
0 5 1 2 7 4 6 3   785.6823473
0 5 1 2 7 4 6 3   785.6823473
0 5 1 2 7 4 6 3   785.6823473
0 5 1 2 7 4 6 3   785.6823473
0 5 1 2 7 4 6 3   785.6823473
0 5 1 2 7 4 6 3   785.6823473
0 5 1 2 7 4 6 3   785.6823473
0 5 1 2 7 4 6 3   785.6823473
0 5 1 2 7 4 6 3   785.6823473
0 5 1 2 7 4 6 3   785.6823473
0 5 1 2 7 4 6 3   785.6823473
0 5 1 2 7 4 6 3   785.6823473
0 5 1 2 7 4 6 3   785.6823473
0 5 1 2 7 4 6 3   785.6823473
```

COST :

Snapshot 3 Step-III

is obtained. As per the implemented GA methodology, by performing multiple crossover and mutation operations, minimal cost parameter is obtained (i.e. as observed at the center of the UI). Here, the optimized result for the initially localized pixel points is 666.978 units. Along with the distance metric, a number of connection links originating from the speckle (0) and terminating at various pixel points (1–7) decide the pixel arrangement and their corresponding configuration. This implies, for achieving optimal light intensity at 0, the pixel configuration is altered in such a way that the pixel points can possibly change the phase of the light rays falling on them.



```
New
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
New
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
New
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
0 5 1 2 7 4 6 3  785.6823
New
```

COST : 666.978370034177

Snapshot 4 Step-IV

Table 1 Simulated optimization results

Randomly selected pixels	Optimized cost parameter	Phase change (Φ) possibilities
9	599.112 units	Φ (2)
12	785.261 units	Φ (4)
8	660.398 units	Φ (3)
14	1183.897 units	Φ (3)
18	1043.420 units	Φ (5)

Consider the above use case—the optimized cost parameter of 669.978 units is obtained. Now, in order to achieve bright spot at 0, consider the connection links from speckle 0 to each pixel point end traversed by the link. For each such connection link, the pixel points would be configured in such a way that they would be able to change the phase of incoming light rays at a predefined angle. Connection links in the above case along with the phase change configuration for those links are as below:

1. 0 ⟶ 5 ⟶ Φ = 0° (change phase by 0°)
2. 0 ⟶ 1 (0 – 7 – 2 – 1) ⟶ Φ = 90° (change phase by 90°)
3. 0 ⟶ 3 (0 – 4 – 3) ⟶ Φ = +180° (change phase by +180°)
4. 0 ⟶ 6 ⟶ Φ = - 90° (change phase by -90°)

Hence, as seen above, all the pixel points lying on the single connection link are configured in one manner, i.e. they are arranged in such a way that the light rays incident on those pixels would undergo a specific phase change. For example, as in the connection link [0 ⟶ 1 (0 – 7 – 2 – 1)],all the respective pixels lying on this link (i.e., 7, 2, and 1) are configured to bring about a phase change of about $\Phi = 90°$ on any incoming light ray. This arrangement is made in such a way that the speckle is easily visible from the see-through skin. The following table gives some optimization results for the implemented evolutionary GA obtained through simulation performed on "Visual Studio 2008" software platform, by using C# language (Table 1).

6 Applications

The proposed see-through skin has a relatively simple operative technique as it does not require the use of coherent light such as lasers. It could be used in a variety of applications that necessitate imaging or microscopy through turbid tissue or inhomogeneous media or living tissue, etc.

The research proposal enables numerable possibilities that could one day become a reality, such as what if we could see the activity unfolding within an egg and thus observe how a chick grows in high resolution and in real time, thanks to the ability disclosed in the research proposal to completely cancel out the effects of shell scattering? What if we could see-through the walls? What if we could see-through an atom, the activity of sub-atomic particles? This could possibly open up an avenue for unraveling the mysteries of quantum mechanics, and the universe. We are not quite there at that point yet, but the research proposal brings the possibility closer than ever before.

7 Conclusion

The research proposal discloses optimization of the SLM by using evolutionary genetic algorithm for advanced wavefront control as they pass through the "Electric See-Through Skin." The proposal validates the feasibility of the SLM that yields optimized results for a localized set of pixel points of the SLM. The optimized SLM helps to maximize the light intensity at the speckle (i.e., hidden object) that allows the see-through skin to recreate the image of the speckle that is not in the direct view. Hence, the research opens up a new paradigm that could significantly benefit medical imaging, quantum physics, etc.

Acknowledgments I would like to extend my sincere gratitude to Dr. A. S. Kanade for his relentless support during my research work.

References

1. Weller JA (2019) Beyond Scattering: identifying the boundaries of light control with a genetic wavefront optimization algorithm, Oregon State University, 27 May 2019
2. Spatial light modulators and modern optical systems, The University of Edinburg, Department of Physics, Applied Optics Group, Lecture notes
3. Owdy: Using walls as mirrors—spatial light modulators, Genetic algorithms and science, 4 September 2016
4. Anthony S (2012) The camera that can see through frosted glass and skin, and around corners, 16 July 2012
5. Hahn J et al (2008) Optimization of the spatial light modulation with twisted nematic liquid crystals by a genetic algorithm, 17 March 2008
6. Conkey DB et al (2012) Genetic algorithm optimization for focusing through turbid media in noisy environments, February 13, 2012/Vol 20, No 5/OPTICS EXPRESS
7. Vellekoop IM et al Focusing light through living tissue, Physics Institute, University of Zurich

Retrieval of Videos of Flowers Using Deep Features

V. K. Jyothi, D. S. Guru, N. Vinay Kumar, and V. N. Manjunath Aradhya

Abstract This paper presents an algorithmic model for the retrieval of natural flower videos using query by frame mechanism. To overcome the drawback of traditional algorithms, we propose an automated system using a deep convolutional neural network as a feature extractor for the retrieval of videos of flowers. Initially, each flower video is represented by a set of keyframes, then features are extracted from keyframes. For a given query frame, the system extracts deep features and retrieves similar videos from the database using k-nearest neighbor and multiclass support vector machine classifiers. Experiments have been conducted on our own dataset consisting of 1919 videos of flowers belonging to 20 different species of flowers. It can be observed that the proposed system outperforms the traditional flower video retrieval system.

Keywords Flower video retrieval system · Multiclass support vector machine · Deep convolutional neural network · KNN

V. K. Jyothi (✉) · D. S. Guru · N. V. Kumar
Department of Studies in Computer Science, University of Mysore, Manasagangotri, Mysore 570006, India
e-mail: jyothivk.mca@gmail.com

D. S. Guru
e-mail: dsg@compsci.unimysore.ac.in

N. V. Kumar
e-mail: vinaykumar.natraj@gmail.com

V. N. Manjunath Aradhya
Department of Computer Applications, JSS Science and Technology University, Mysore, India
e-mail: aradhya@sjce.ac.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_56

1 Introduction

Designing a system for the retrieval of videos is a significant research area. Research in video retrieval is growing rapidly due to the increase in Internet technology and storage capacity [1]. Most of the current video retrieval systems use traditional features such as edge [2], colors [2], bag-of-features [3], SURF descriptors [4], Texture [5], SIFT features [6], global and local geometric data information [7]. Retrieval of flower videos using traditional handcrafted features is tedious and requires experienced experts. To overcome these drawbacks, we propose an automated system using Deep Convolutional Neural Network (DCNN) as a feature extractor for the retrieval of videos of flowers.

DCNN has drastically improved the performance of computer vision and pattern recognition systems [8]. Initially, deep learning techniques have proposed for the classification of images [9, 10] and recognition of objects [11]. They are also used for the classification of videos [12]. Further, they are used for image retrieval [13, 14].

Flower video retrieval systems are applicable in the field of floriculture and useful in searching flower videos of users' interest for medicinal use, cosmetics, and decoration [15]. To automate the searching process designing and developing a flower video retrieval system is essential. Designing a flower video retrieval system is a challenging task when the videos of flowers are captured in different ecological conditions [16]. Flowers in the video include challenges like different viewpoints, scale variations, occlusions, and multiple instances [17].

2 Proposed Methodology

It includes three stages, namely, preprocessing, deep feature descriptors, and retrieval. Figure 1 shows the structure of the proposed flower video retrieval system.

2.1 *Preprocess*

In preprocessing, the system converts video to frames and resizes the frames into 256 × 256 for further processing.

Let $\mathrm{DB_F}$ be a database which consists of a collection of 'N' number of videos of flowers and can be written as

$$\mathrm{DB_F} = \{\mathrm{FV_1},\ \mathrm{FV_2},\ \mathrm{FV_3},\ \ldots,\ \mathrm{FV_i}, \ldots,\ \mathrm{FV_N}\} \tag{1}$$

Then any video $\mathrm{FV_i}$ in Eq. (1) consists of a set frames in 'n' number and can be written as follows,

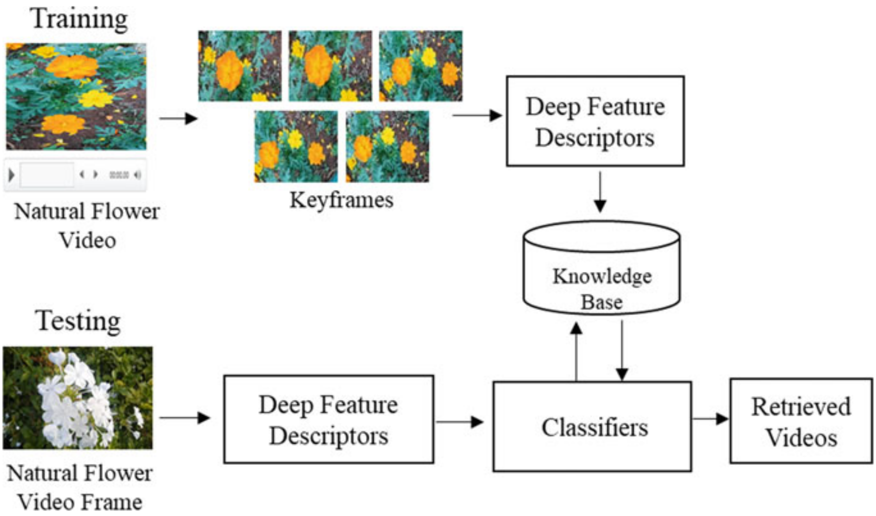


Fig. 1 Structure of the proposed flower video retrieval system [1]

$$FV_i = \{f_1, f_2, f_3, \ldots f_i, \ldots f_n\} \tag{2}$$

Then, keyframes are selected to represent the video and are explained in Sect. 2.1.1.

2.1.1 Selection of Keyframes

The selection of keyframes from a video is a vital field in video retrieval. The video contains redundant information, therefore it increases the computational burden. This can be avoided by a compact representation of a video. The most distinguishable and essential frames of each video of the database can be selected via a keyframe selection mechanism [18]. The selection of the most essential frames is an important process to design a flower video retrieval system.

In the proposed model, keyframes of the flower videos are selected using a Gaussian Mixture Model (GMM) clustering approach [17]. Here, the similar frames of the flower video are clustered using a block-wise entropy feature and GMM algorithm. The keyframes selected from a video FV_i can be defined as

$$FV_i = \left\{k_1, k_2, k_3, \ldots k_y\right\} \tag{3}$$

where KFV_i is a selected set of keyframes from the GMM clustering approach and ‘y’ says the number of selected keyframes among ‘n’ number of frames shown in Eq. (2).

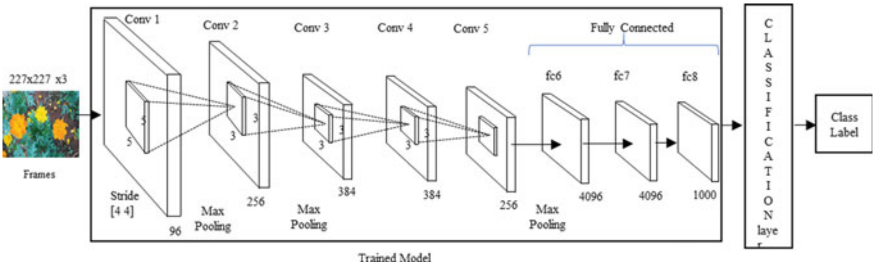


Fig. 2 The framework of a convolutional network used for flower video retrieval

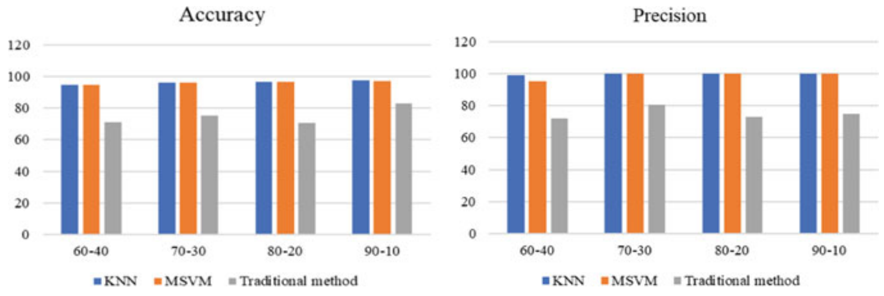
2.2 Deep Feature Descriptors

Features are extracted from deep learning techniques in layer-wise. It starts the extraction of low-level features from the lower layer. The output obtained from the previous layer is passed to the next layer. In the higher layer, it extracts parts and objects of the patterns. From an input image, DCNN architecture extracts features without depending on human designing feature extraction methods [19]. To design a flower video retrieval system, in the present work, we have utilized AlexNet [20] architecture for the extraction of deep features, the batch size is fixed to 32 and a fully connected layer is used. Figure 2 shows the framework of AlexNet architecture utilized in the proposed system. It is an eight-layered architecture. The functions in layers include Convolution, Pooling, and Rectified Linear Unit (ReLU) [21]. The convolution function passes a set of masks and it activates features. The function of the pooling is, it simplifies the output of the convolution function, by performing non-linear down-sampling. In the proposed system the max. The pooling mechanism of the AlexNet architecture is applied. ReLU maps all negative values to zero for faster processing. In the AlexNet architecture, the layers conv1 to conv5 are convolutional layers and the remaining fc6 to fc8 are fully connected layers. The number of kernels utilized in each layer is shown in Fig. 2.

2.3 Retrieval

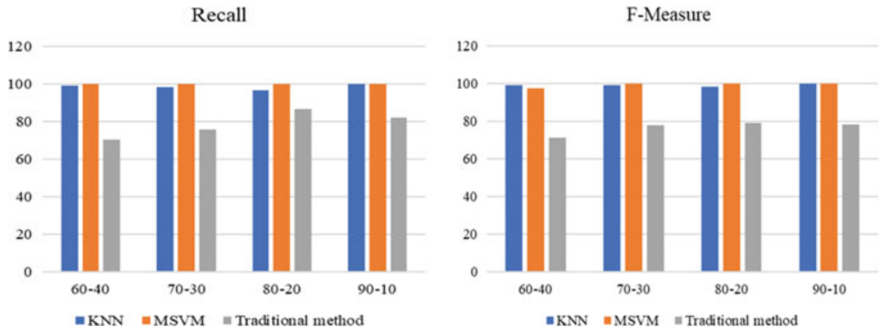
The videos in the database are represented as keyframes for the process of retrieval. For a given query frame of the video to recognize the similar videos, the system applies K-Nearest Neighbor (KNN) [22] and Multiclass Support Vector Machine (MSVM) [23] classifiers. Once the query finds the identity of the class, then the videos belonging to that class are retrieved.

Let the query video be Q_T, it contains ‘p’ number of frames say $Q_T = \{f_1, f_2, f_3, \ldots, f_p\}$, if we consider any frame as a query then the proposed system extracts DCNN features from query frame. And compares with the DCNN features of each



(a) (b)

Fig. 3 Comparison of proposed deep learning versus traditional approach: **a** Accuracy and **b** Precision



(a) (b)

Fig. 4 Comparison of proposed deep learning versus traditional approach: **c** Recall and **d** F-measure

keyframe of the video in the database using KNN and MSVM algorithms. The KNN and MSVM are explained in Sects. 2.3.1 and 2.3.2, respectively.

2.3.1 K-Nearest Neighbor

KNN is a supervised algorithm and is used for the classification of patterns. It applies the nearest neighbor approach to classify a sample. It essentially involves finding the similarity between the test pattern and every pattern in the training set.

For a given test pattern, the KNN algorithm finds the closest neighbor and assigns a class label. Let there be ‘m’ training flower videos and its corresponding class labels can be defined as (FV_1, C_1), (FV_2, C_2), (FV_3, C_3),…, (FV_m, C_m), where FV_i is represented with ‘d’ number of DCNN features and C_i is the class label of ith video. If a query frame be ‘p’ then $D(p, FV_k) = \min\{dist(p, FV_i)\}$, where $i = 1$ to n, ‘p’ is assigned to the class C_k associated with FV_k.

2.3.2 Multiclass Support Vector Machine

Support Vector Machine (SVM) is a supervised learning algorithm. Initially, SVM is designed to classify a problem consisting of two classes [24]. It works based on an optimal hyperplane, it generates maximum margin between two classes. In the proposed work, the dataset consists of multiclass of videos of flowers; therefore, support vectors are generated for all the classes to train the model. Each class separates from all other classes using these vectors to predict the class of the query.

To classify a pattern in multicategory classifier, it computes *'L'* linear discriminant functions and is defined as

$$g_i(x) = w^t x + T_i$$

where $i = 1,.., L$, w is the weight vector and T_i is the threshold. Assigns *'x'* to ω_i if $g_i(x) > g_j(x)$ for all $j \neq i$.

3 Dataset

Dataset is a basic requirement to design any pattern recognition and machine learning applications. It is essential to test the efficiency of the automatic machine learning system designed. To design an efficient flower video retrieval system, a flower video dataset is required for the conduction of experiments. Due to that there is an unavailability of the benchmark dataset in the literature; our own dataset is created. We have collected the videos of different species of flowers and each species include subspecies. Videos are captured using Canon camera of 16 megapixels. The dataset consists of 1919 videos of 20 different species, in which each class consists of 35–160 videos of flowers, captured in the duration of 4–60 s in the real environment during sunny, rainy, and winter seasons. Figure 5 shows the sample videos collected. The retrieved videos from the proposed deep learning system are shown in Fig. 6.

4 Experiments and Results

This section presents the retrieval performance of the proposed system. To test the efficacy of the retrieval system, our own dataset is used. In the training phase, keyframes are selected for the representation of a video. In the testing phase, the video frame is considered as a query. Deep features are extracted from both keyframes and a query frame. To identify the class of the query frame, the system matches the query with keyframes of the video using KNN and MSVM classifiers. Once the query identifies the class, then the system retrieves related videos to that class. The retrieved videos from the proposed deep learning system are shown in Fig. 6. To evaluate the proposed system, the measures, namely, accuracy (A), precision (P), recall (R), and

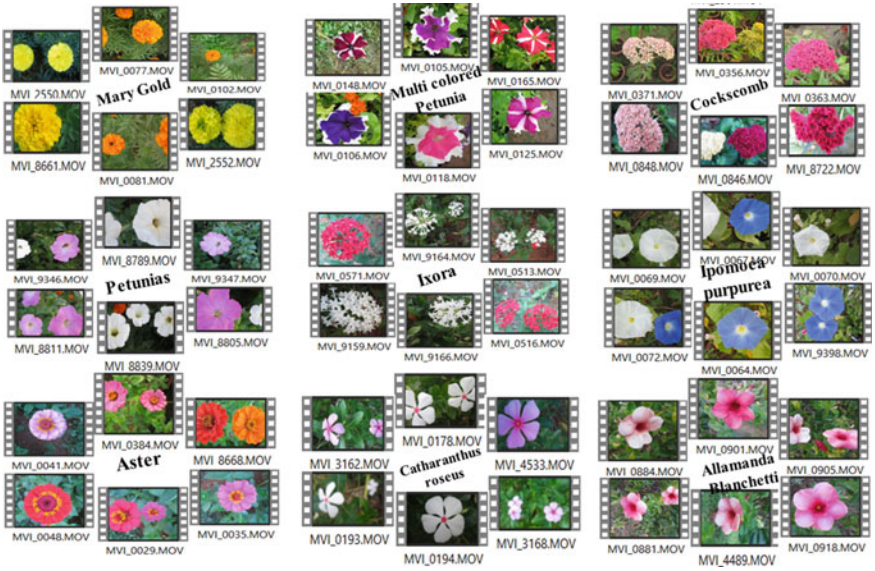


Fig. 5 Samples of videos of flowers with large variations in intraclass from 20 species

F-Measure (F-M) are used and are given below.

$$\text{Accuracy} = \frac{\text{Total number of videos retrieved correctly}}{\text{Total number of query videos}} \tag{4}$$

$$\text{Precision} = \frac{\text{Total number of correctly retrieved flower videos}}{\text{Total number of flower videos retrieved}} \tag{5}$$

$$\text{Recall} = \frac{\text{Total number of relevant flower videos retrieved}}{\text{Total number of similar flower videos in database}} \tag{6}$$

$$\text{F-Measure} = \frac{2 * \text{Precision} * \text{Recall}}{(\text{Precision} + \text{Recall})} \tag{7}$$

Tables 1 and 2 show the average retrieval results obtained using KNN and MSVM, respectively.

4.1 Comparative Study

The proposed deep learning system is compared with the traditional flower video retrieval system [5]. In [5], the videos are retrieved with the combination of features strategy, the features utilized are Scale Invariant Feature Transform (SIFT), Gray Level Co-occurrence Matrix (GLCM) and Local Binary Pattern (LBP). Experiments

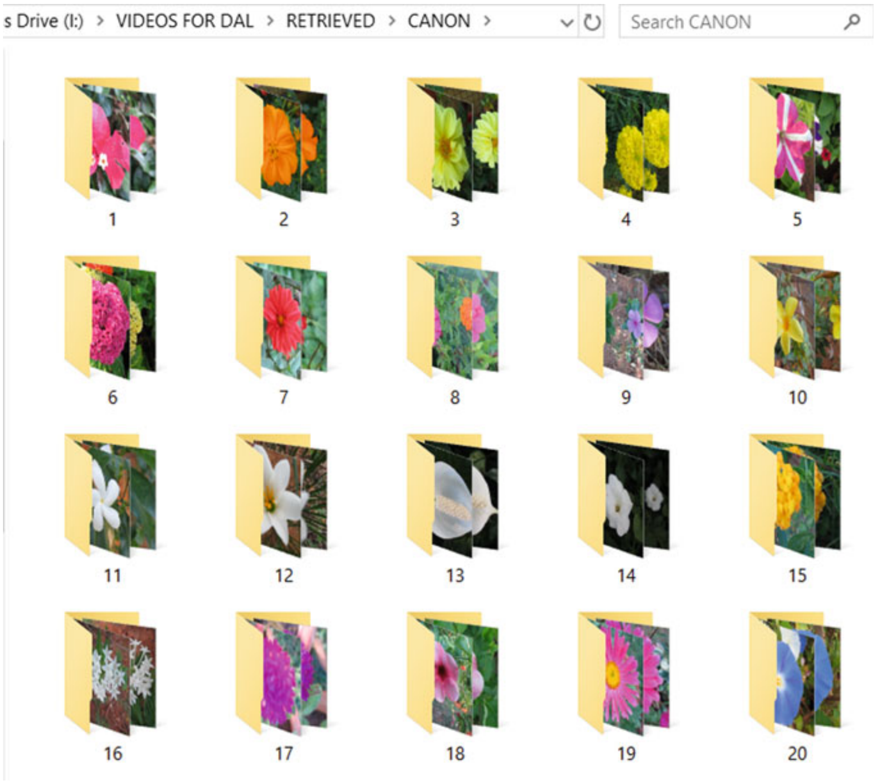


Fig. 6 Retrieved videos from the proposed deep learning retrieval system

Table 1 Result analysis using K-Nearest Neighbor algorithm

Train–Test in %	A	P	R	F-M
30–70	91.19	99.23	99.48	99.36
40–60	91.92	99.10	99.40	99.25
50–50	92.60	99.28	99.28	99.28
60–40	95.14	99.10	99.10	99.10
70–30	96.17	100	98.19	99.09
80–20	96.97	100	96.40	98.17
90–10	97.90	100	100	100

are conducted on individual features, the combination of two features, and the combination of all the three features. For the combination of features SIFT + GLCM + LBP the system achieved good results. Further, to improve the results of [5], we have proposed the present retrieval system. The comparison between the proposed system of KNN and MSVM approaches with the traditional approach [5] of retrieval

Table 2 Result analysis using Multiclass Support Vector Machine (MSVM)

Train - Test in %	A	P	R	F-M
30–70	90.58	96.76	100	98.35
40–60	92.85	95.14	100	97.51
50–50	92.29	96.18	100	98.05
60–40	94.70	95.28	100	97.58
70–30	96.51	100	100	100
80–20	96.87	100	100	100
90–10	97.27	100	100	100

of videos of flowers are shown in Figs. 3 and 4. The results show that the proposed system generates good results than the traditional system [5] (Figs. 5 and 6).

5 Conclusion

In this paper, we presented the retrieval of videos of flowers through a query by frame mechanism using DCNN feature descriptors. Videos are represented with keyframes in the training phase to avoid the computational burden, and keyframes are considered for the extraction of features. For a given query frame, the system retrieves similar videos using KNN and MSVM in the retrieval stage. We have made a comparative study to show the superiority of the proposed system.

References

1. Yu CC, Jou FD, Lee CC, Fan KC, Chuang TC (2008) Efficient multi-resolution histogram matching for fast image/video retrieval. Pattern Recogn Lett 29:1858–1867
2. Ling B, Xiao W, Liu X (2012) Design of video retrieval system using MPEG-7 descriptors. Proc Eng 29:2578–2582
3. Cui M, Cui J, Li H (2016) Dimensionality reduction for histogram features: a distance-adaptive approach. Neurocomputing 173:181–195
4. Asha S, Sreeraj M (2013) Content based video retrieval using SURF descriptor. In: Third international conf. on advances in computing and communications, pp 212–215. IEEE Xplore, ISSN:978-0-7695-5033-6
5. Guru DS, Jyothi VK, Kumar YHS (2019) Features fusion for retrieval of flower videos. Springer Nature. Lecture Notes in Networks and Systems, vol 43, pp 221-233
6. Zhu Y, Xuang X, Huang Q, Tian Q (2016) Large-scale video copy retrieval with temporal concentration SIFT. Neurocomputing 187:83–91
7. Mygdalis V, losifidis A, Tefas A, Pitas I (2018) Semi-supervised subclass support vector data description for image and video classification. Neurocomputing 278:51–61
8. Joe YHN, Matthew H, Sudheendra V (2015) Beyond short snippets: deep networks for video classification. IEEE Xplore, ISSN: 4694-4702

9. Xiaohong WG, Rui H, Zengmin T (2017) Classification of CT brain images based on deep learning networks. Comput Methods Programs Biomed 138:49–56
10. Carlos A, Rossi ALD, Vieira FHA, Andre C Deep learning for biological image classification. Expert systems with applications. 85:114–122
11. Xingcheng L, Ruihan S, Jian H, Jianhua D, Linji H, Qing G A deep convolutional neural network model for vehicle recognition and face recognition. Proc Comput Sci 107:715–720
12. Ming HC, Kao SH, Jyh HJ, Nai WL (2013) Classification based video super resolution using artificial neural networks. Sig Process 93:2612–2625
13. Adnan Q, Anwar SM, Muhammad A, Muhammad M (2017) Medical image retrieval using deep convolutional neural network. Neurocomputing 266:8–20
14. Xinggang W, Xiong D, Xiang B (2016) Deep sketch feature for cross-domain image retrieval. Neurocomputing 207:387–397
15. Das M, Manmatha R, Riseman EM (1999) Indexing flower patent images using domain knowledge. IEEE Intell Syst 14:24–33
16. Jyothi VK, Guru DS, Kumar YHS (2018) Deep learning for retrieval of natural flower videos. Proc Comput Sci 132:1533–1542
17. Guru DS, Jyothi VK, Kumar YHS (2017) Cluster based approaches for keyframe selection in natural flower videos. Intell Syst Des Appl 736:474–484
18. Ejaz N, Tariq TB, Baik SW (2012) Adaptive key frame extraction for video summarization using an aggregation mechanism. J Vis Com Image Rep 23(7):1031–1040
19. Yoshua B (2009) Learning deep architectures for AI. Found Trends Mach Learn 2(1):1–127. https://doi.org/10.1561/2200000006
20. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. International conference on neural information processing system, vol 1, pp 1097–1105
21. Guo Y, Liu Y, Oerlemans A, Lao S, We S, Lew MS (2016) Deep Learning for visual understanding: a review. Neurocomputing 187:27–48
22. Han J, Kamber M, Pei J (2012) Data mining. Concepts and techniques, Third ed. Morgan Kaufmann Publishers
23. Kumar AM, Gopal M A hybrid SVM based decision tree. Pat Rec 43:3977–3987
24. Alexandros I, Moncef G (2016) multi-class support vector machine classifiers using intrinsic and penalty graphs. Pat Rec 55:231–246

Analysis of Students Performance Using Learning Analytics—A Case Study

Manjula Sanjay Koti and Samyukta D. Kumta

Abstract Utilization of digital tools has been increased enormously in our daily learning activity by generating data on a huge scale. This huge amount of data generation provides exciting challenges to the researchers. Learning analytics effectively facilitates the evolution of pedagogies and instructional designs to improve and monitor the students' learning and predict students' performance, detects unusual learning behaviors, emotional states, identification of students who are at risk, and also provides guidance to the students. Data mining is considered a powerful tool in the education sector to enhance the understanding of the learning process. This study uses predictive analytics, which help teachers to identify student's at risk and monitor students progress over time, thereby providing the necessary support and intervention to students those are in need.

Keywords Learning analytics · Classification · Educational data mining

1 Introduction

Learning Management System uses various technological tools in teaching and learning thereby, enhancing the learning process. The students get benefits from Learning Management System as it provides the space to perform their academic activities and collaborate with their peers and teachers. While by doing so, a huge amount of information is gathered by LMS regarding students interaction with their peers, systems, teachers, contents of the course, etc., [1]. This information that is hidden can be extracted to obtain knowledge that helps in making decisions by those, who take care of the education program thereby improving the students' performance.

M. S. Koti (✉)
Sir M Visvesvaraya Institute of Technology, Bengaluru, India
e-mail: manjula.dsce@gmail.com

S. D. Kumta
VTU RRC, Bengaluru, India
e-mail: samyukta.citech@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_57

There is a drastic growth in digital learning information and educational data. Hence the greatest challenge now is to transform educational data into meaningful information and knowledge. This, of course, affects learning behaviors in students and the teaching modes of teachers along with the decision making of people who are involved in education and resource allocation. In order to extract meaningful information, it is necessary to represent the stored information in the most comprehensive manner [2]. The people who are involved in learning activity will able to identify the most important aspects of learning to make effective decisions [3].

Research needs to be done in students academic performance, and this performance is driven by factors viz., academic atmosphere, studying habits, skills with respect to education and personality traits, which helps to optimize students' academic performance and this can be achieved through learning Analytics, Academic Analytics, and Educational Data mining [4]. The data which has collected pertaining to student academic performance has to be analyzed through educational data mining and learning analytics, where the information extracted from the educational data can be used in decision making for the education sector [5].

In educational data mining, the emphasis is laid on techniques and methodologies while learning analytics deals with applications. The attributes of learning analytics can be shared with educational data mining, where the learning analytics helps in predicting the students who may quit the course or it can also be used to predict, which student needs special attention to improve the performances [6]. Currently, online teaching and learning have become a trend, many learning platforms that are available in education. The learners get benefits from the smart learning environment as it provides instant adaptive support by analyzing the needs of individual learners from various perspectives. Accordingly, most of the researchers are trying to explore the usage of education data in analyzing scientific and effective learning.

Learning analytics is a multi-disciplinary field involving machine learning, artificial intelligence, information retrieval, statistics, and visualization. Hence, it has emerged as a latest research field which focuses on computational techniques to measure the student practices. In general, Learning analytics is defined as the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environment in which it occurs [3, 5, 7]. The student learning process in a learning environment can be analyzed by taking into account the static and dynamic information of students and their learning patterns. Since it is more flexible and real-time availability of the data and personalized, many ongoing research works are observed in learning analytics [1]. Learning analytics uses techniques such as prediction, clustering, outlier detection, relationship mining, social network analysis, process mining, text mining, a discovery with models, Gamification, machine learning, and statistics, etc. [8].

The main advantage of Learning analytics is its flexibility, personalization and real time availability of data [9]. Learning Analytics can be categorized as descriptive, diagnostic, predictive, and prescriptive analytics. Fig. 1 depicts the different types of learning analytics.

Data mining is used in descriptive analytics to give complete details for the historical question such as "What has happened?". Diagnostic analytics analyzes the data

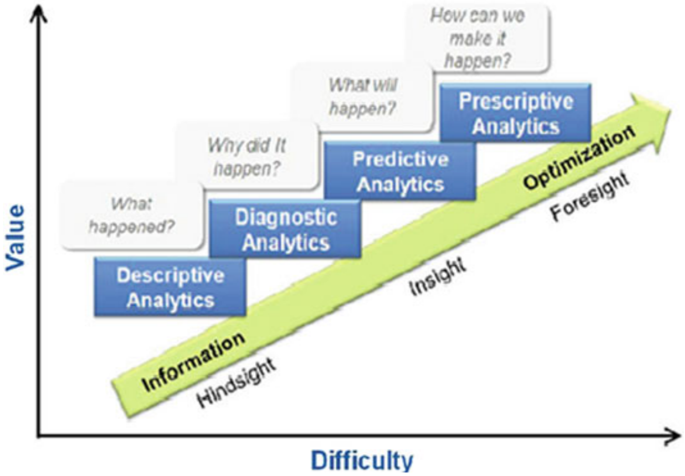


Fig. 1 Different types of learning analytics

or content, and answers to the question "Why did it happen?". Predictive analytics utilizes the statistical models to understand the future, to understand the question "What could happen?". Finally, prescriptive analytics uses algorithms related to optimization and simulation, tries to predict the possible outcome and also answers the question "What should we do?". Predictive analytics in education is used to analyze the students' performance. The stakeholders who are involved in learning analytics gets benefitted based on their vision and mission [10]. Certainly, they can improve the students' performance, and the teaching process of teachers. This leads to improvements in course models by the researchers. The newer methods of delivering educational information and the decision process are taken by educational institutions in order to achieve their goals [6].

The learning analytics has four components namely

i. Learning Environment- Where huge data is produced by the stakeholder.
ii. Big Data- Huge amount of dataset and repository of information.
iii. Analytics- Consists of various analytical techniques
iv. ACT- Optimization of Learning analytic environment by considering the objectives to be achieved.

Figure 2 presents the framework of learning analytics and its life cycle. A tremendous amount of data is collected in the learning environment. This big data can be analyzed using various methods which helps in discovering interesting and meaningful patterns in the educational dataset [11]. The analytical techniques make use of quantitative and statistical analysis. Interpretation of this analysis can be used to achieve the objectives. In this paper, one of the stakeholders is considered as a student where they produce an enormous amount of data in the learning environment. The student performance in the learning process is identified and analyzed using predictive analysis. The stakeholders take appropriate decisions based on the

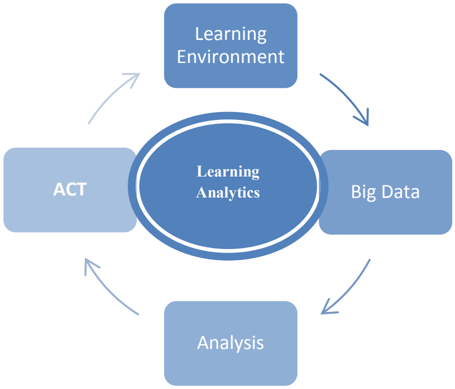


Fig. 2 Learning analytics framework and lifecycle

results obtained through analysis [3]. This paper is organized as Sect. 1 introduction to the topic of research and deals with importance of learning analytics, Sect. 2 deals with literature survey, Sect. 3 deals with the methodology, Sects. 4 and 5 presents the results and conclusion respectively.

2 Literature Survey

A lot of researches have been reported and have provided interesting results in predicting the students' academic performance. Several studies have utilized the data comprising of a non-academic parameter.

The authors in [12] explain the use of the classification algorithm to predict the students' academic performance in a big data environment and also the different classification methods are used to compare its accuracy. In [5, 13] the author Billy suggests that learning analytics can be used with other platforms such as the learning management systems where instructors can access various kinds of information online for providing feedback and support to students. The authors Isabela and Avanide explain the different dimensions in LA such as the types of the data to be collected to conduct the analysis, what is the relationship of these data, the stakeholders, objectives of the analysis, and techniques used to conduct the analysis. In [9] the authors considers that educational data mining is another analytic possibility to take into account in learning process. The authors in [3, 14] discuss different tools available to carry out the learning analysis such as general-purpose dashboards, ad hoc defined tools, learning analytics tools to analyze specific issue learning analytics framework and tools, etc. Based on this perspective the present study includes analyzing student's difficulty in scoring marks in respective sessions. In [15] authors explain how LA plays an important role across all educational sectors to find out the data which will tell us exactly what we need to know from learning methodologies in the education field. They also mention that educational technologies are not all easy to implement and educational practices are difficult to change. Wherein the push for LA in the education sector resolve most of the problems faced by an institution. The authors Austushi and Komoni [8] gave the suggestions to the lectures based on real-time learning system. This supports the teaching and learning process adaptively based on the situation in on-site classrooms and immediate improvements in the lecture plan. In [16] authors, compares the prediction algorithms used in classification techniques like Logistic Regression, Naïve Bayes, Random Forest, and KNN to predict their overall performance outcomes and these techniques in the study are able to make a prediction on the performance of the students. In the present study, we are using the KNN Classification algorithm to predict the student performance. In [7, 17] the authors discuss the different methods of Data Collection, Methods of Data Analysis such as Classification and regression, Evidence of Research in Learning Analytics, and authors also showed evidence where LA improves learning support and teaching has been dominating across all the year from 2012 to 2018.

3 Methodology

In this, we predict student performance by analyzing the student behavior patterns in the learning environment. We have considered online survey, log files, class attendance, and analyzing the examination grades which includes internal and external grades. R is a very powerful language that is widely used for data analysis and statistical computing. The student dataset for our study is considered from UCI machine repository. This dataset comprises 115 observations and 17 instances from log data file which contains information for each student-ID shows whether the student has logged in session or not, final grade and internal grade file contains the marks in the final and internal examination. In this study we have used R studio framework to classify the records.

3.1 Dataset Description

The dataset contains 115 observations of Student Result, Student Internal Exam, and Log Data which has been taken from the UCI machine learning repository. In Student Result Dataset we have 115 observations and a total of 17 instances. The variables are Student-ID, (Exam-Sessions) ES1.1, ES1.2, ES2.1, ES2.2, ES3.1-ES3.5, ES4.1, ES4.2, ES5.1-ES5.3, ES6.1, ES6.2, and TotalMarks. Adding each session marks we have reduced the variables to total 7. The new variable is renamed as “ES1”, “ES2”, “ES3”, “ES4”, “ES5”, “ES6”, and “FinalTotal”. The Exam Session Marks of ES1 and ES2 are with maximum marks of 5, where the marks for ES3, ES4, ES5, and ES6 are with 10,25,15, and 40, respectively. The marks range is corrected with the Normalization method in the implementation. The class labels are given based on the marks obtained by the students in both internal and external marks. From

Student Internal Exam Data we have 7 variables named Student-ID, (S-Session) S1, S2, S3, S4, S5, and S6. The Log data contains the log information for each students' individual sessions, which is used to analyze the time spent by the students in each session in the learning process.

3.2 Data Preparing

Data collected from the student dataset is the preprocessing technique which consists of cleaning the data, instance selection, normalization, transformation, and feature extraction and solution [4]. The erroneous values existed in the dataset have been converted into either upper bond or lower bond.

3.3 Data Selection

Data is selected from six various sessions of assessment where each session has a different allocation of marks. Based on the log data and the Internal marks obtained by the students the class labels considered are "FCD", "PASS", and "FAIL". The main objective is to analyze failures in the test sessions, analyzing the reason for students' failure in various courses and this certainly helps the teachers in improving their teaching-learning process which leads to improvising the learning ability of the student.

3.4 Classification Model

We have used the K-Nearest Neighbor classification to predict student performance. KNN is a non-parameterized method of classification and it is also known as lazy learners or instance-based learning. The steps involed in KNN are (i) Determine the parameter K (ii) Calculate the distance between the instances and all training samples (iii) Sort the distance and determine nearest neighbors based on the kth minimum distance (iv) Gather the category γ of the nearest neighbors (v) Use simple majority of the category of nearest neighbors as the prediction value of the query instance [18].

The classification is a two-step process. The first step is the learning phase of the classifier model. The dataset used in the learning phase is called as training data. The training data consists of a class label. The model will be built using the training data and KNN classifier algorithm, and it is applied to the test data to assign the class labels.

4 Results

The training and testing data sets are applied to the KNN classification model using R with the R Studio framework. The results have shown that the testing values and prediction values are accurate with 97 %. The accuracy of the KNN model is calculated using correlation analysis which is a method of statistical evaluation that is used to study the strength of a relationship between actual results and prediction results. The model successful prediction result is shown in Table 1 with the first six samples. The graph in Fig. 3 shows the accuracy of the classification model with actual and Prediction values.

The class labels FCD, PASS, and FAIL are assigned with numbers 1, 2, and 3 respectively.

The students' failure analysis in the subject is done by analyzing individual session marks. The following figures show the students' scores in all six sessions. Figures 4, 5, 6, 7, 8, and 9 give a complete picture of each session marks. Session-1 and session-3 are found to be scoring sessions for students. The maximum number of students have scored full marks. Where in session-2 and session-4, students have scored average marks. Therefore teachers need to concentrate on rectifying the students' inability to understand the concepts in this session. Maximum numbers of students are found to score very less score in session-5 and session-6. The failure student's data was analyzed with their log information and internal examination data with session-5 and

Table 1 Sample result of classification model

Obs	Actuals	Predictions
1	1	1
2	3	3
3	2	2
4	3	3
5	1	1
6	1	1

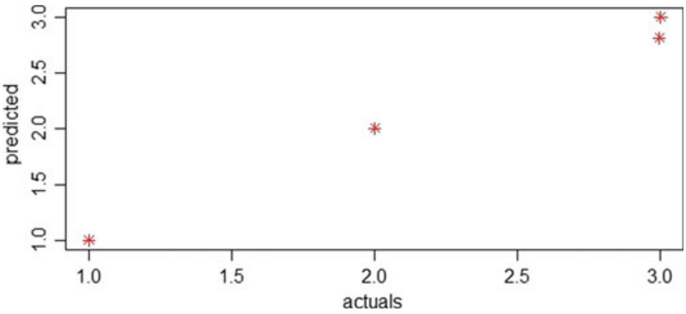


Fig. 3 Accuracy of actual and predicted values of classification model

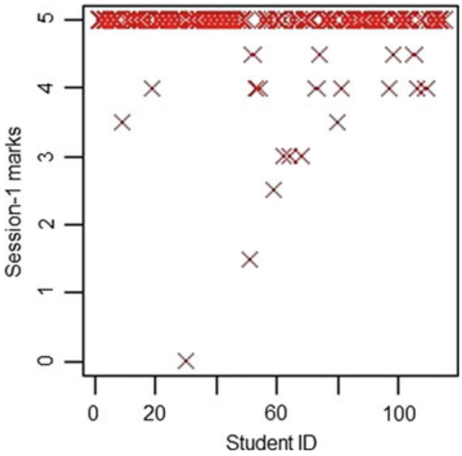


Fig. 4 Students' performance in section-1

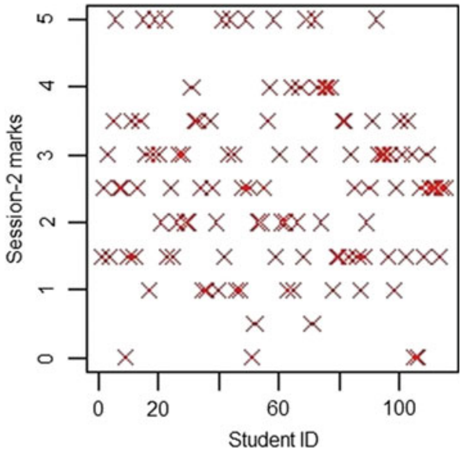


Fig. 5 Students' performance in section-2

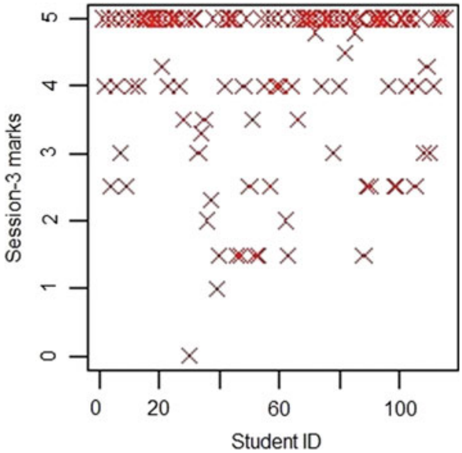


Fig. 6 Students' performance in section-3

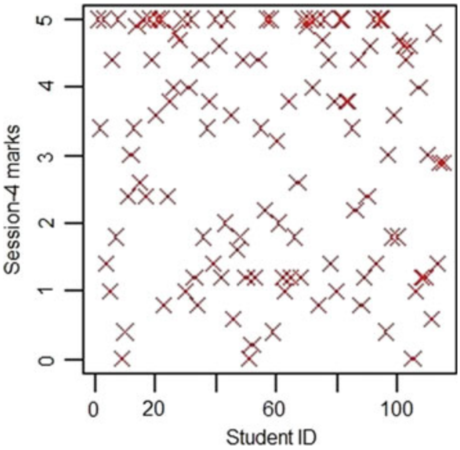


Fig. 7 Students' performance in section-4

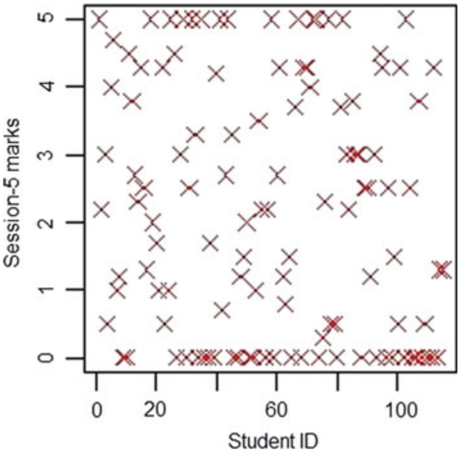


Fig. 8 Students' Performance in section-5

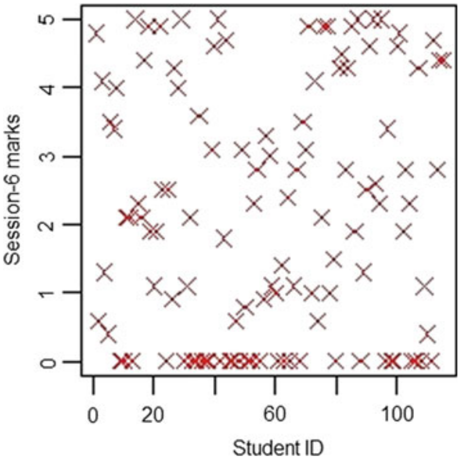


Fig. 9 Students' performance in section-6

session-6. It has been found that the failure percentage in internal exam for session-5 and session-6 is 29 % and 30 %, respectively, matches the fail percentage of final marks obtained by the students. This analysis from internal exam marks and log data for session-6 and session 5 shows that students did not utilize the available resources and scored less in internal exams. This results in a failure in the final examination. The analysis also helps the tutor to overcome the increasing numbers of failure students in respective sessions. Institutes can change the learning and teaching process to improve the students' performance.

5 Conclusion

Learning Management System generates huge data about learners and learning. The importance of learning analytics lies in analyzing and detecting various patterns in the data which leads to the development of methods in supporting the learners learning experience. This study is done to analyze students learning patterns by considering their logs, internal assessment, and external marks results. We have used the KNN classifier to predict the students' academic performance by using R tool. KNN predicts the class labels of the student taking into account various parameters from the student data set. The results indicate that the performance of students and also analyze the students who need mentoring and improve them in individual sessions. This improves the quality of education in the institutions.

References

1. Mason J, Chen W, Hoel T (2016) Questions as data: illuminationg the potential of learning analytics through questioning an emergent field. Res Pract Technol
2. Conde MÁ, García-Peñalvo FJ, Gómez-Aguilar DA, Theron R (2017) Visual learning analytics techniques applied in software engineering subjects. https://www.researchgate.net/publication/282792081(2017)
3. Armatas C, Spratt CF (2019) Applying learning analytics to program curriculum review
4. Liñán LC, Pérez ÁA (2015) Educational data mining and learning analytics: differences, similarities and time evolution: learning analytics: intelligent decision support system for learning environments
5. Wong BT (2019) Learning analytics in higher education: an analysis of case studies. University Research Centre, The Open University of Hong Kong
6. Fan C, Xiao F (2016) Assessment of building operational performance using data mining techniques: a case study
7. Viberg O, Hatakka M, Bälter O, Mavroudi A (2018) The current landscape of learning analytics in higher education: computers in human behaviour
8. Shimada A, Konomi SI, Ogata H (2019) Real-time learning analytics system for improvement of on-site lectures
9. Kuhnel M, Seiler L, Honal A, Ifenthaler D (2019) Mobile learning analytics in higher education: usability testing and evaluation of an app prototype
10. Clemens C, Kumar V, Mitchnick D (2013) Writing based learning analytics in education. In: IEEE ınternational conference on advanced learning technologies

11. Kumar K, Vivekanandan V (2019) Advancing learning through smart learning analytics:review of case studies
12. Seetharam A, Nagesh VS, Satyamurty KA (2017) Predicting student performance using KNN classification in bigdata environment. CVR J Sci Technol 13
13. Moissa B, Gasparini I, Kemczinski A (2015) Systematic mapping on the learning analytics field and its analysis in the massive open online courses context. Int J Distance Educ Technol 13(3):1–24
14. Kostagiolas P, Lavranos C, Korfiatis N (2019) Learninganalytics: survey data for measuring the impact of study satisfaction on students' academic self-efficacy and performance
15. Teasley SD (2019) Learning analytics: where information science and the learning sciences meet
16. Umer R, Susnjak T, Mathrani A, Suriadi S (2019) On predicting academic performance with process mining in learning analytics
17. Yamada M, Shimada A, Okubo F, Oi M, Kojima K, Ogata H et al (2017) Learning analytics of the relationships among self-regulated learning, learning behaviors, and learning performance. Res Pract Technol Enhance Learn
18. Conde MA, Garcia FJ, Gomez JA, Theron R (2015) Exploring software engineering subjects by usisng visial learning analytics techniques
19. Dalipi F, Imran AS, Kastrati Z (2018) MOOC dropout prediction using machine learning techniques: review and research challenges
20. Guenaga M, Garaizar P (2016) From analysis to improvement challenges and opportunities for learning analytics. IEEE Revista Technol 11(3)

A Case Study on Distributed Consensus Problem on Cloud-Based Systems

Ganeshayya Shidaganti, Ritu Pravakar, M. Shirisha, and H. R. Samyuktha

Abstract Fault tolerance is a vital element in cloud computing to achieve high performance. In cloud computing fault tolerance particularly cluster management, network discovery and consistent system master node replication, consensus and coordination play a major role. This paper provides a comprehensive overview of the topic of fault tolerance, i.e., the problem of consensus in cloud computing; highlighting the important concepts along with the explanation of Byzantine Agreement problem and consensus problems in multi-agent systems. There are multiple algorithms/protocols like RAFT and PAXOS available to approach this problem. We present generalized consensus implementation by solving consensus for dual failure nodes. We also describe Apache Zookeeper as our coordination service to obtain consensus in a distributed system.

Keywords Paxos · Raft · Byzantine agreement · Apache zookeeper · Zab

1 Introduction

Cloud computing has revolutionized the software Information Technology delivery model by allowing the obtainability of different softwares, technologies, and infrastructural resources as distributed services that can be used on request over the internet. A cloud network requires excellent collaboration by several clusters of nodes and servers that are located in different locations to effectively execute client requests

G. Shidaganti (✉) · R. Pravakar · M. Shirisha · H. R. Samyuktha
Ramaiah Institute of Technology, Bangalore, Karnataka, India
e-mail: ganeshayyashidaganti@msrit.edu

R. Pravakar
e-mail: ritupravarar22@gmail.com

M. Shirisha
e-mail: shirishasiri4199@gmail.com

H. R. Samyuktha
e-mail: samyureddyhr@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_58

and all of them are expected to exchange information with each other [1]. However, the service's performance is hampered by its vulnerability to failures due to the scale of operation. Cloud computing services can be used to their maximum potential when cloud service providers deal effectively with performance-related issues such as availability and throughput [2]. Paxos protocols and applications provide the problem of decentralized consensus with a fault-tolerant solution, attracting considerable attention along with creating a lot of confusion [3]. Distributed systems rely on dependencies such as Apache Zookeeper or Raft. While these systems differ with the features, they solve the basic fundamental problem: Agreement. Processes in a distributed system need to decide on a leader and also a lock holder. Having access to a consensus implementation, systems coordinate processes in a more effective manner, e.g., when managing the replica sets of Kafka.

This paper is organized as follows. Section 2 defines the related work describing consensus in distributed system and in multi-agent systems. Section 3 introduces Byzantine Agreement (BA) problem and an example of generalized consensus problem in cloud computing environments. In Sect. 4, we categorize the usage patterns of coordination in the cloud and examine the infrastructure of Google to look into how Paxos protocols and systems are used. The aim is to provide insights into the current consensus problem solutions as well as the problems that need to be addressed. Here, we also consider a few approaches that can be used for more productive solutions along with identifying important methodologies for research with respect to the topic considered.

2 Related Work

2.1 Consensus in Distributed Systems

Figure 1 depicts the protocols which are developed to deal with the consensus in distributed systems, which are: Paxos, Raft and Zab. The detailed explanation of each is explained in the following sections.

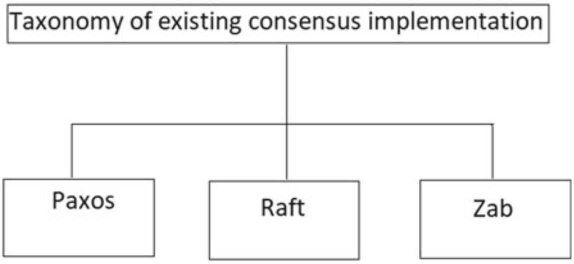


Fig. 1 Representational diagram

2.1.1 Paxos

Paxos protocols along with frameworks provide solution which is fault tolerant to the problem of distributed consensus along with gaining substantial attention distributed consensus causing major uncertainty. Paxos ensures safety (consistency) and is usually used where durability is required (for example, replicating a file or database) in which the amount of long-lasting state may be large. The protocol attempts to make progress even during periods when there is no response to some limited number of replicas [4].

Unfortunately, there are two major drawbacks to Paxos: One thing is Paxos is extremely hard to understand. Another issue with Paxos is that it is not suggestible for real-time implementation [5].

2.1.2 Raft

Raft is an algorithm of consensus to handle a replicated file. It generates a multi Paxos result which is similarly productive like Paxos, but its design's unlike Paxos [5]. Raft is more comprehensible which lays a proper base for developing real-time systems. It also segregates the key features of consensus like leader election, replication of log, safety, and so on. Here, it also foists impactful level of coherency which lowers the total states that must be examined. Here, a new methodology is employed which allows us to modify the cluster membership, using the concept of overlapping majorities thus ensuring protection. Raft, which proved more comprehensible than Paxos is providing a better basis for building the system [5].

2.1.3 Zab

Zab is an atomic broadcast protocol as it enables the nodes to carry out the functionalities in the same order for the same set of transactions (state updates). It implements a fundamental backup scheme where a primary process conducts server operations. Zab also propagates the proper step by step changes in the state to backup processes [5]. Zab allows multi-state changes by making sure that at most one primary will broadcast and at the same time will assimilate state changes into specific state along with using synchronization phase while a new primary is formed.

The assumption that each change in state is gradual in comparison to the previous state is crucial to Zab's layout, so there is an implicit reliance on the order of changes in the system. Finally, Zab uses a state change identification scheme that facilitates a process with finding out the left out changes easily which provides effective recovery [6].

2.2 Consensus Problem in Multi-agent Systems

User can use and download resources in cloud computing at any time by using any smart device with internet but should handle complex computing processes when accessing very large data storage facilities. Multi-Agents System (MAS) methodology is the perfect way for rapidly evolving flexible, scalable systems. In a multi-agent system, two different attack scenarios are identified: attacking agent dynamics (closed-loop dynamics) and attacking agent communications. Mathematics plays an important role in dealing with consensus problem of multi-agent systems like graph theory, matrix theory, and control theory.

A system of linear multi-agent is usually subjected to two types of attacks. Connected and disconnected are two types of topologies. The main issue is to achieve a secure consensus tracking for multi-agent systems with both the topology. To improve security performance, the attacks must be on the nodes in such systems [7].

At certain times consensus needs to be reached at different user preferences. In a framework, consumers are expected to buy energy to a platform and producers are expected to sell energy to the platform [8]. Minimizing a global network cost function guarantees a secure multi-agent system with an optimal control effort. It is found that Riccati equation isn't effective when it comes to the success of consensus. Linear Matrix Inequality (LMI) is preferred over this which also considers the constraint of consensus achievement.

This can be used to impose a structure which is controller specific upon the adjacent sets as an extra Linear Matrix Inequality (LMI) curtailment. Thus, each controller needs to know only the information it receives from the neighbors with which it is associated with the graph representation of the system. Finally, the optimal solution obtained involves all the constraints imposed thus suggesting an optimal global solution with global cost function.

2.3 ZooKeeper and Consensus

ZooKeeper is responsible for distributed synchronization across multiple nodes since it implements a protocol for distributed consensus [9]. A leader election protocol like Paxos is used by Zookeeper internally to select the master node. Clients can connect to any of the nodes in Zookeeper service and additional nodes forward facts which are agreed upon by majority back to clients. Master intervenes to update shared states and all updates are ordered by timestamps.

When a majority of nodes recognize an update, a quorum of nodes is said to hold it. Fact agreed upon by quorum is returned back to clients. Newest update is used when multiple states are updated on an individual node [10]. Leader election is held again if the elected leader by Zookeeper cluster fails. The cluster continues to function normally after the leader election.

3 Byzantine Problem

A Byzantine fault is a computer system condition, particularly distributed computer systems, where there are failed components and information about them is incomplete. Byzantine fault tolerance aims at defending against system component failures irrespective of symptoms that preclude other system parts from agreeing to an agreement among them which is necessary for the exact operation of the system [11]. In order to reach consensus, it requires separate right nodes where some of those nodes may be incorrect. Notably, if there's no impact or influence from components which have turned corrupted, a distributed system can achieve stability in terms of results [12].

3.1 Example: Executing Generalized Consensus Problem of Cloud Computing

In the example below nodes 3 and 6 are malicious and node 5 is dormant. Malicious nodes give unpredictable values when they participate in the intermediate stages while dormant nodes might crash or send faulty values (Fig. 2).

Figure 2 explains dormant nodes hold the value λ for any incoming values and also send λ to other nodes. Malicious nodes are not completely dead, instead, they manipulate the values. Correct nodes hold and send the values they received correspondingly.

Values:

1: Serving request

0: Not serving request

λ: Dead node

Fig. 2 Example cluster A

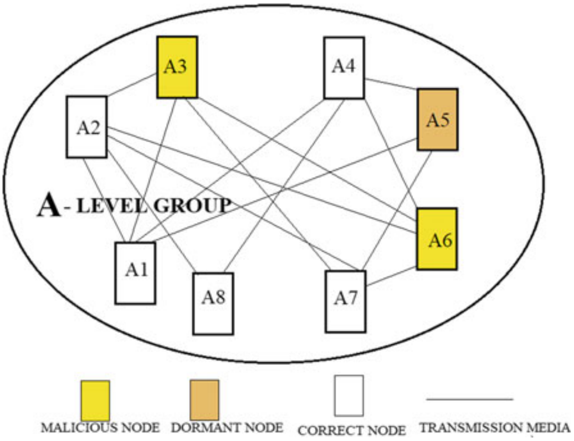
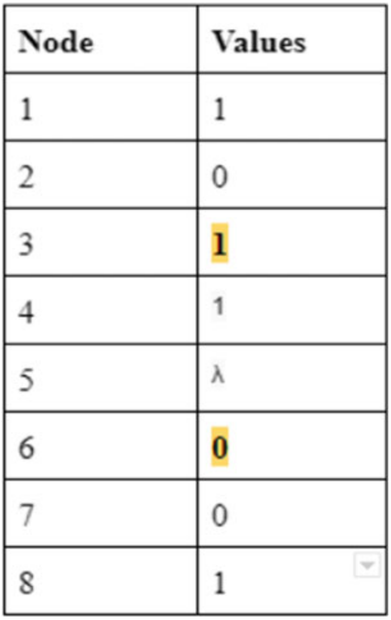


Table 1 Initial values



Node	Values
1	1
2	0
3	1
4	1
5	λ
6	0
7	0
8	1

Table 1 tells about the initial values at each node. The values of each node are not yet communicated between the other nodes.

Table 2 describes the first round, where each node sends its respective values to all other nodes including itself. Here, in this table, the values stored in a particular node are depicted.

In the second round, the values stored at each node are again communicated with all the other nodes. Table 3 shows how values at node are communicated with other nodes.

In Table 4, node 4 is selected and the corresponding values stored for each node are shown in the table.

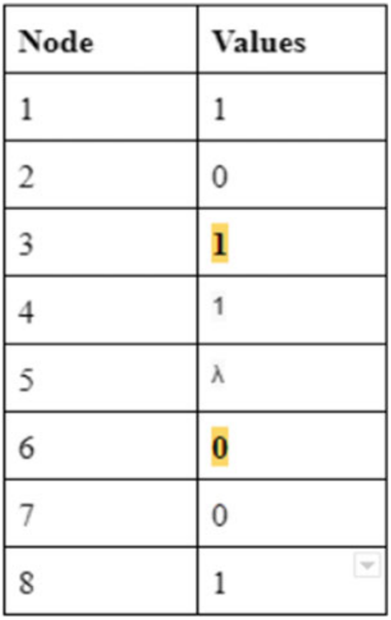
In level 2, the values stored at node 4 are again communicated with other nodes. Table 5 shows the level 2 representation for node 4.

The self-node values are omitted and the final mg-tree is obtained using majority voting method.

As values represented in level 2, now there is a set of values for each, from which a final value for the node to be determined. As mentioned in Table 6, through majority voting method the final Vote(4) = 1 considering the vector (1, 1, 1, 1, 1, 1, 1, 1) as obtained in the previous step. Similarly we calculate the majority vote for all the nodes stored in the node A1.

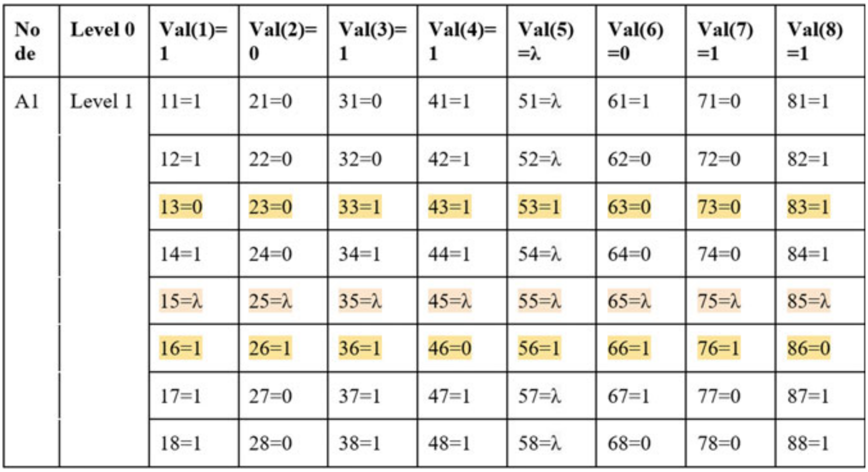
As determined for node 4 using majority voting method, the value for each node is determined in the same way. Table 7 shows the final value for each node correspondingly.

Table 2 First round



Node	Values
1	1
2	0
3	1
4	1
5	λ
6	0
7	0
8	1

Table 3 Second Round



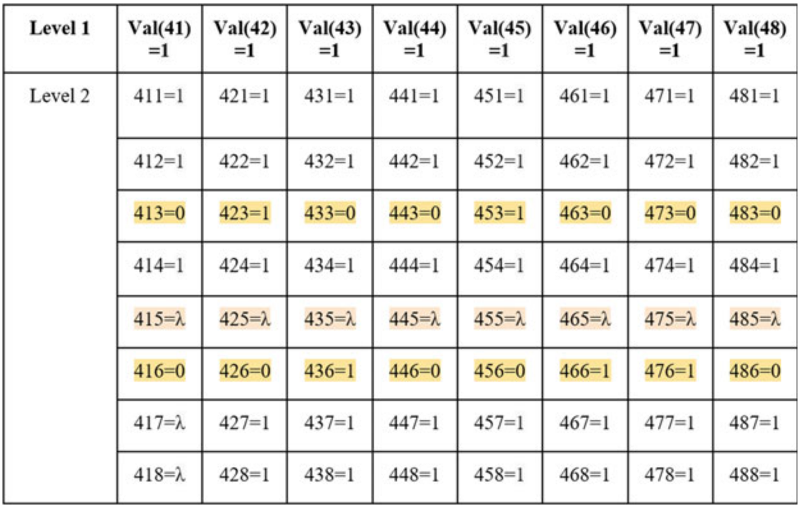
Node	Level 0	Val(1)= 1	Val(2)= 0	Val(3)= 1	Val(4)= 1	Val(5) =λ	Val(6) =0	Val(7) =1	Val(8) =1
A1	Level 1	11=1	21=0	31=0	41=1	51=λ	61=1	71=0	81=1
		12=1	22=0	32=0	42=1	52=λ	62=0	72=0	82=1
		13=0	23=0	33=1	43=1	53=1	63=0	73=0	83=1
		14=1	24=0	34=1	44=1	54=λ	64=0	74=0	84=1
		15=λ	25=λ	35=λ	45=λ	55=λ	65=λ	75=λ	85=λ
		16=1	26=1	36=1	46=0	56=1	66=1	76=1	86=0
		17=1	27=0	37=1	47=1	57=λ	67=1	77=0	87=1
		18=1	28=0	38=1	48=1	58=λ	68=0	78=0	88=1

All the nodes in Level A will have the same values when calculated as above. Hence the majority value is 1 for node A1. Similarly, all the nodes in level A will have their respective values what they have agreed upon initially. At the end of the majority voting method, the faulty node don’t have any effect on the majority value

Table 4 Node 4 values (node 4 selected)

Val (4) =1	41	42	43	44	45	46	47	48
	1	1	1	1	λ	0	1	1

Table 5 Level 2 for node 4



Level 1	Val(41) =1	Val(42) =1	Val(43) =1	Val(44) =1	Val(45) =1	Val(46) =1	Val(47) =1	Val(48) =1
Level 2	411=1	421=1	431=1	441=1	451=1	461=1	471=1	481=1
	412=1	422=1	432=1	442=1	452=1	462=1	472=1	482=1
	413=0	423=1	433=0	443=0	453=1	463=0	473=0	483=0
	414=1	424=1	434=1	444=1	454=1	464=1	474=1	484=1
	415=λ	425=λ	435=λ	445=λ	455=λ	465=λ	475=λ	485=λ
	416=0	426=0	436=1	446=0	456=0	466=1	476=1	486=0
	417=λ	427=1	437=1	447=1	457=1	467=1	477=1	487=1
	418=λ	428=1	438=1	448=1	458=1	468=1	478=1	488=1

agreed upon by all the nodes. Here malicious nodes 3, 6, and dormant node 5 affected intermediate stages but not the final value which is 1.

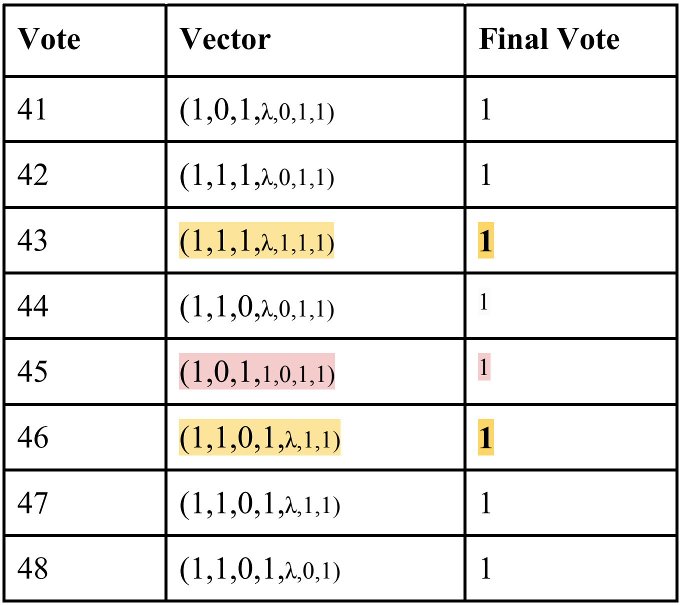
Similarly, the whole procedure is repeated in level B, and then the corresponding request is served accordingly.

4 Applications in Google

A Coordination service which was also faulted tolerant was required for Google File System. Hence Paxos was adopted by GFS to implement GFS lock service. It was named Google Chubby. It increased the interest of the industry in Paxos systems for fault-tolerant coordination.

Coordination and group management systems are at the bottom of the stack. The cluster manager for Google is Borg [13], which manages thousands of jobs in various clusters from different applications. For each job and running task submitted, Chubby is used by Borg as a Paxos store for holding metadata. Paxos is used to write the

Table 6 Majority voting



Vote	Vector	Final Vote
41	(1,0,1,λ,0,1,1)	1
42	(1,1,1,λ,0,1,1)	1
43	(1,1,1,λ,1,1,1)	**1**
44	(1,1,0,λ,0,1,1)	1
45	(1,0,1,1,0,1,1)	1
46	(1,1,0,1,λ,1,1)	**1**
47	(1,1,0,1,λ,1,1)	1
48	(1,1,0,1,λ,0,1)	1

Table 7 Final values

Node	1	2	3	4	5	6	7	8
Final vote	**1**	**0**	λ	**1**	λ	λ	**0**	**1**

hostname and port of all the tasks in Chubby and thus provides a naming service. It also implements Paxos for leader election, replication of master and as a queue of newly submitted jobs which helps scheduling.

For Linux containers Kubernetes [14] is Google's new open source project for cluster management. Kubernetes preserves state such as metadata of the cluster and membership of the resource pool. The stack's second layer includes many components for data storage. Google File System (GFS) is the most essential storage service. Bigtable [15] is a structured data distributed storage system developed by Google.

It relies heavily on Chubby to ensure that there is a total of one active manager, metadata management, group membership, and management of configuration. Megastore [16] is a database of the cross datacenter. It is the biggest system that has been deployed. It uses Paxos to duplicate primary user's data on each writes through data centers. This expands Paxos to synchronously replicate multiple write-ahead logs.

5 Conclusion

In this paper, the distributed consensus problem in the cloud-based systems is depicted through an example describing the generalized consensus issue. There are many fault-tolerant protocols like Paxos, Raft, Zab, etc., where each of them has certain limitations. A brief idea about these protocols is mentioned in this paper. Here, we are representing that common value can be achieved by all correct nodes in a cloud computing environment topology even though there are some malfunctioning nodes present. Implementations such as Paxos and Raft are efficient solutions but for cloud-based systems Zookeeper implements Zab. We have also done a case study on various applications by Google to provide fault-tolerant coordination service.

References

1. Chanchary FH, Islam S, Strong Consensus in Cloud Computing
2. Cheraghlou MN, Khadem-Zadeh A, Haghparast M (2016) A survey of fault tolerance architecture in cloud computing. J Netw Comput Appl 61:81–92
3. Ailijiang A, Charapko A, Demirbas M (2016) Consensus in the cloud: Paxos systems demystified. In: 2016 25th International conference on computer communication and networks (ICCCN). IEEE, pp 1–10
4. https://www.wikiwand.com/en/Paxos_(computer_science)#/overview
5. Ongaro D, Ousterhout J (2014) In search of an understandable consensus algorithm. In: 2014 {USENIX} annual technical conference ({USENIX}{ATC} 14), pp 305–319
6. Junqueira FP, Reed BC, Serafini M (2011) Zab: high-performance broadcast for primary-backup systems. In: 2011 IEEE/IFIP 41st international conference on dependable systems & networks (DSN). IEEE, pp 245–256
7. Feng Z, Hu G, Wen G (2016) Distributed consensus tracking for multi-agent systems under two types of attacks. Int J Robust Nonlinear Control 26(5):896–918
8. Bedo JS (2019) Consensus reaching with heterogeneous user preferences, private input and privacy-preservation output. Oper Res Perspect 100–138
9. Hunt P, Konar M, Junqueira FP, Reed B (2010) ZooKeeper: wait-free coordination for internet-scale systems. In USENIX annual technical conference, vol 8, no 9
10. http://zookeeper-tutorial.blogspot.com/2011/04/distributed-consensus.html
11. https://en.wikipedia.org/wiki/Byzantine_fault
12. Wang SS, Wang SC (2014) The consensus problem with dual failure nodes in a cloud computing environment. Inf Sci 279:213–228
13. Verma A, Pedrosa L, Korupolu M, Oppenheimer D, Tune E, Wilkes J (2015) Large- scale cluster management at google with borg. In: Proceedings of the tenth european conference on computer systems. ACM, p 18
14. Google kubernetes project. http://kubernetes.io/
15. Chang F, Dean J, Ghemawat S, Hsieh W, Wallach D, Burrows M, Chandra T, Fikes A, Gruber R (2008) Bigtable: A distributed storage system for structured data. ACM Trans Comput Syst (TOCS) 26(2):4
16. Baker J, Bond C, Corbett JC, Furman JJ, Khorlin A, Larson J, Leon JM, Li Y, Lloyd A, Yushprakh V (2011) Megastore: providing scalable, highly available storage for interactive services. In: CIDR, pp 223–234
17. https://www.confluent.io/blog/distributed-consensus-reloaded-apache-zookeeper-and-replication-in-kafka

An IoT Framework for Healthcare Monitoring and Machine Learning for Life Expectancy Prediction

Anna Merine George, Anudeep Nagaraja, L. Ananth Naik, and J. Naresh

Abstract The beginning of the IoT era, shrinking of devices and the concept of intelligent independently learning machines have led to improvements in the quality of human life. The application of machine learning to IoT data has led to the automation of the creation of analytical models. One key area of research has seen such a revolution in the health care sector. This work aims to design a wireless healthcare system that detects patients vitals using sensors, transfers data to cloud, and predicts the approximate life expectancy using machine learning techniques. The notion of the Internet of Things (IoT) interconnects devices and offers effective health care service to the patients. Here the IoT architecture gathers the sensor data and transfers it to the cloud where processing and analyses take place. Based on the analyzed data, feedback inputs are sent back to the doctor and using the present pulse rate of the patient, nominal or approximate value of life expectancy is predicted using machine learning algorithms.

Keywords Wireless sensor · Healthcare · Thinkspeak · Arduino · Machine learning

1 Introduction

Internet of Things (IoT) is the association of physical devices such as sensors and actuators to enable remote access to objects. It is anticipated that by 2020 more than 50 billion objects will be linked to the internet [1, 2].

IoT is the unification and integration of communication devices to enable a new generation of assistance services. The hype surrounding IoT is complemented by the application of Machine Learning making meaningful correlations, better decisions,

A. M. George (✉) · A. Nagaraja · L. A. Naik · J. Naresh
Department of ECE, Dayananda Sagar University, Bangalore, India
e-mail: anna_thanku@yahoo.com

A. Nagaraja
e-mail: anudeepnagaraja1998@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_59

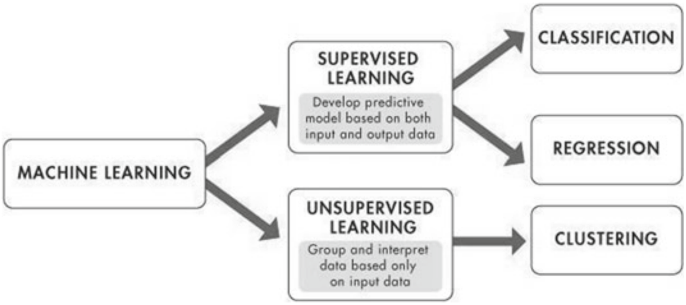

MACHINE LEARNING
SUPERVISED LEARNING
Develop predictive model based on both input and output data
UNSUPERVISED LEARNING
Group and interpret data based only on input data
CLASSIFICATION
REGRESSION
CLUSTERING

Fig. 1 Block diagram for machine learning [8]

and predictions based on learned user patterns and analysis of data collected. Machine Learning algorithms are grouped as supervised, unsupervised, and reinforcement learning depending on the learning style.

Machine learning relies on computational models to find natural patterns in data to learn and make decisions and predictions. The algorithm depends on the number of samples available for learning. They can be used to predict the energy load or battery life of IoT devices.

Supervised learning trains the model using a known set of input and output data to make reasonable predictions and decisions. Machine learning algorithm makes a trade-off in terms of training speed, memory usage, accuracy, and transparency on new data. Figure 1 shows the block diagram for machine learning.

Machine learning can be used when:

- Mathematical equations and rules are complex.
- The equations and rules related to a task are changing.
- The kind of data keeps changing as in the case of energy demand prediction.

Regression models make predictions about variables based on the behavior of variables and their trends. Pattern classification is to assign discrete class labels to particular observations as consequences of a prediction. Figure 2 shows the detailed workflow for machine learning model development and real-time implementation.

1.1 Common Classification Algorithms

k-Nearest Neighbor: Here the predictions assume that objects adjacent to each other are alike. The algorithm works well in those applications where the memory usage of the trained model and prediction speed is of less concern.

Support Vector Machine: The algorithm classifies data by finding the best hyperplane between two or more classes. It is best used for data that has exactly two classes and high dimensional non-separable data.

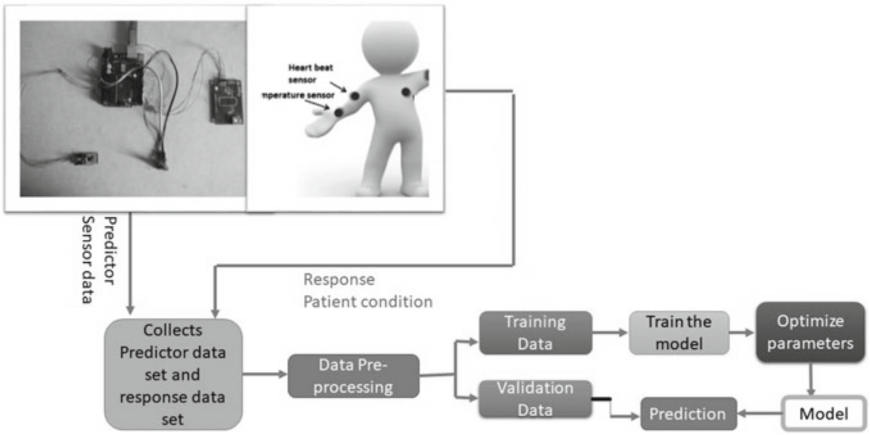


Fig. 2 Detailed flow diagram for machine learning model development and real-time implementation [9]

Naïve Bayes: The algorithm categorizes new data based on the highest probability of it fitting to a particular class. It is best suited for small dataset having many parameters.

2 Related Work

Sarfraz Fayaz Khan, in the year 2017 [3], proposed that the staffs and doctors were given a chance to improve and adopt new services using Wi-Fi. This is done by using the Internet of Things hardware which is fused with the Wi-Fi module of the RFID, NFC tags, and some small sensor nodes. The different ways in which IoT can be implemented in healthcare institutions and the combination of microcontrollers with the sensors added to improve efficiency is discussed in this paper. The result includes robust output against medical emergencies. The work, however, focused on RFID-based health monitoring that restricts the coverage area.

Nilanjan Dey [4] proposed that the internet of things can be converged with the healthcare sector to transform into more advanced and efficient services. The conjunction of the Internet of Things technology and medical field makes a great impact in the healthcare sector. IoT has physical devices network, embedded system, sensor, servers, software, and network connectivity to communicate remotely and collect data from the system components. IoT integrates the automation, sensor networks, embedded system, these facilities make IoT a great convenience. He also mentioned the healthcare application supported with IoT system can be connected and used anywhere, anytime, and at any place of our convenience which leads to the smart healthcare usage.

Prasanth [5] introduced a user-friendly desktop application for monitoring the health status of the patient themselves.

Luca Catarinucci [6] proposed the designing of health care system which collects the patient's health condition and environmental conditions at real time and sends to the control center where it is analyzed and sends an alert based on the emergency condition. The RFID-based system reduces manpower and allows online monitoring of medical reports.

Moeen Hassanalieragh et al. in 2015 [7] proposed an IoT-based smart health care system where the physical and mental health status is collected by various embedded or environmental sensors. These data are processed and analyzed and made available always. It aids in reducing the cost of healthcare by simultaneously improving the quality of services.

3 Methodology

The project is implemented using Arduino Uno, LM35 temperature sensor, an analog heartbeat sensor, esp8266 Wi-Fi module and Thing Speak as the cloud platform as shown in Fig. 3. The temperature is measured in degrees and the heart rate in bpm. Emergency conditions are defined for pulse rate greater than 100 and lesser than 40 and based on the pulse rate values a machine learning-based predictive analysis is done where data is fed to a controller in real time so that the controller takes decision based on the previous data values and gives the average life expectancy of the patient.

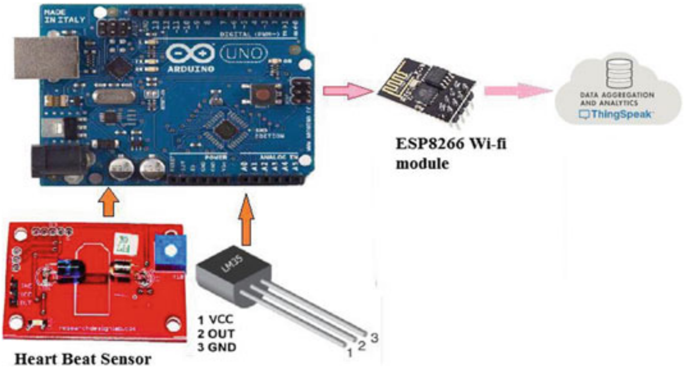


Fig. 3 General block diagram

3.1 Steps Involved in Transferring Data to Cloud

Temperature and heartbeat data are acquired by biosensors, preprocessed, and transmitted through suitable communication platform. Sensors are connected to the network through immediate data aggregator or concentrator. The components of data transmission system are responsible for transmitting the patient records from home to the data center of healthcare organization. Often a storage/processing devices in vicinity of client, is referred to as Cloudlet. Cloudlet is local processing unit which is used to transfer to cloud in case of any limitations such as lack of connectivity. Distinct components of cloud processing are Storage, Analysis, and Visualization.

Steps involved in transferring data to cloud include

Step 1: Setup the ThingSpeak, i.e., create an account on ThingSpeak website.
Step 2: Prepare and build ESP8266 hardware for appropriate communication.
Step 3: Write the code to send data from Arduino to ThingSpeak.

- Connect the ESP8266 Wi-Fi module using AT commands.
- The data will be sent over HTTP connection. Define SSIO, Password, API Key to ThingSpeak.

Step 4: Run the code.

3.2 Steps Involved in Predictive Analysis Using Machine Learning Algorithm

Predictive analytics uses new statistical patterns and machine learning algorithm to analyze past data and predict future ones. Predictive models are developed in MATLAB with classification algorithms and tested using K-fold validation. Figure 4 shows the block diagram for predictive analysis using Machine Learning.

4 Results

Figure 5 shows the results of the temperature sensor in temperature versus time which is uploaded in the thinkspeak cloud.

Figure 6 shows the results of the pulse sensor in pulse rate versus time which is uploaded in the thinkspeak cloud.

Figure 7 is used to indicate an optimum patient condition (green light) when the pulse rate of the patient is between 40 and 100.

Figure 8 is used to display an emergency condition using red light when the pulse rate is below 40 or greater than 100.

Figure 9 is used to indicate the approximate life expectancy of the patient obtained using classification learner. When the pulse rate is below 40 or greater than 100, the

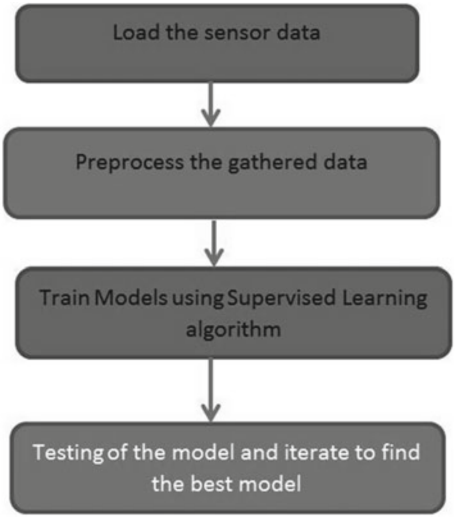

Load the sensor data
Preprocess the gathered data
Train Models using Supervised Learning algorithm
Testing of the model and iterate to find the best model

Fig. 4 Flow chart for Predictive analysis using machine learning

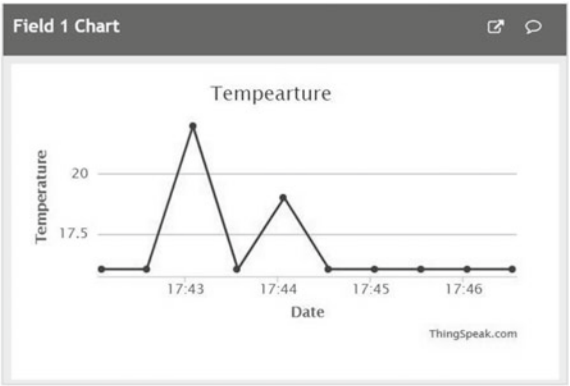


Fig. 5 Temperature sensor data uploaded to cloud

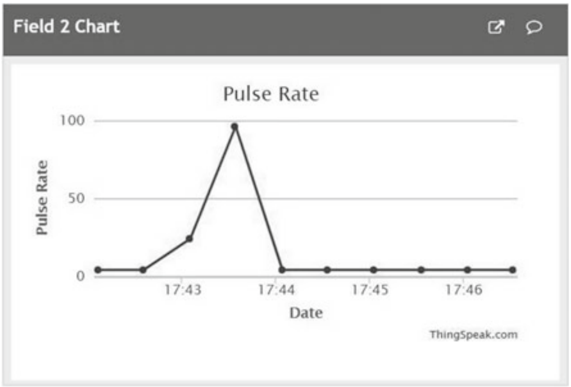


Fig. 6 Pulse sensor data uploaded to cloud

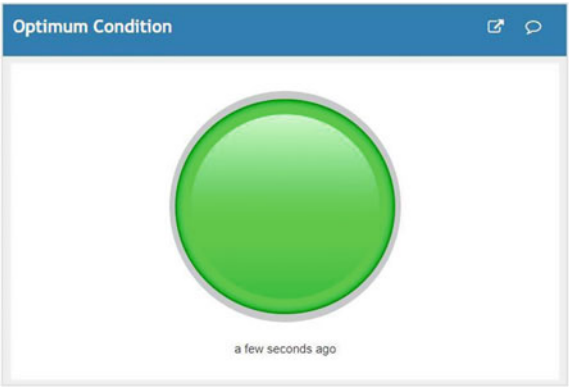


Fig. 7 Optimum condition indication

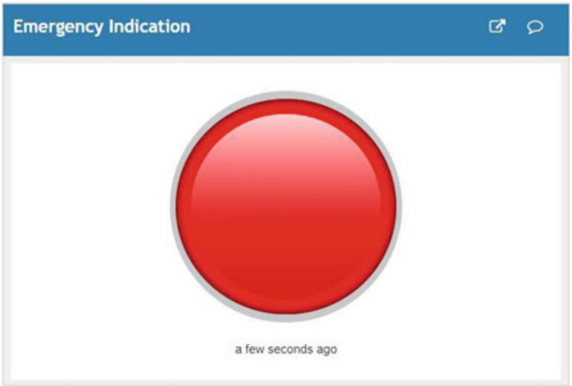


Fig. 8 Emergency condition indication

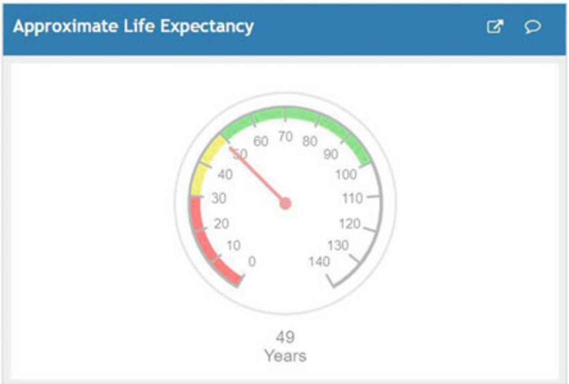


Fig. 9 Approximate life expectancy based on machine learning

life expectancy of the patient will be lower and the doctor will be intimated about an emergency condition.

5 Conclusion and Future Scope

The patient's temperature and heartbeat rate are measured and monitored in the cloud (ThingSpeak). For pulse and temperature variations the optimum (GREEN) and emergency conditions (RED) are obtained. Using prediction-based regression model approximate life expectancy of the patient is obtained. The developed system can be used to detect tumors and several abnormalities in the brain. More research needs to be carried on security algorithms and data privacy as IoT is managed and run by multiple technologies and multiple vendors. The present system requires the sensors to be connected to the body, whereas in future contactless devices can be built. Moreover, self-powered and low power health monitoring systems need to be designed.

References

1. George AM, George VI, George MA (2018) IoT based smart traffic light control system. In: International conference on control, power, communication and computing technologies (ICCPCCT). IEEE, pp 148–151
2. Mahdavinejad MS et al (2018) Machine learning for internet of things data analysis: a survey. Digit Commun Netw 4(3):161–175
3. Khan SF (2017) Health care monitoring system in internet of things (IoT) by using RFID. In: VIth International conference on industrial technology and management (ICITM). IEEE, pp 198–204
4. Bhatt C, Dey N, Ashour AS (2017) Internet of things and big data technologies for next generation healthcare
5. Natarajan K, Prasath B, Kokila P (2016) Smart health care system using internet of things. J Netw Commun Emerg Technol (JNCET) 6(3)
6. Catarinucci L, De Donno D, Mainetti L, Palano L, Patrono L et al (2015) An IoT-aware architecture for smart healthcare systems. IEEE Internet Things J 2(6):515–526
7. Hassanalieragh M, Page A, Soyata T, Sharma G, Aktas M et al (2015) Health monitoring and management using internet of things (IoT) sensing with cloud-based processing: opportunities and challenges. In: International conference on services computing. IEEE, pp 285–292
8. www.mathworks.com
9. Kurian CP, Shetty S (2019) A data driven approach for the control of a daylight–artificial light integrated scheme. Light Res Technol 1477153519841104
10. Baker Stephanie B, Xiang Wei, Atkinson Ian (2017) Internet of things for smart healthcare: technologies. IEEE Access Chall Oppor 5:26521–26544
11. Sunehra D, Ramakrishna P (2016) Web based patient health monitoring system using Arduino. In: 2nd International conference on contemporary computing and informatics. IEEE
12. Krishnan DS, Gupta SC, Choudhury T (2018) An IoT based patient health monitoring system. In: International conference on advances in computing and communication engineering (ICACCE). IEEE

A Study on Discernment of Fake News Using Machine Learning Algorithms

Utkarsh, Sujit, Syed Nabeel Azeez, B. C. Darshan, and H. A. Chaya Kumari

Abstract Due to recent events in world politics, fake news, or malevolently established media has taken a major role in world politics discouraging the opinion of the people. There is a great impact of fake news on our modern world as it enhances a sense of discretion among people. Various sectors like security, education and social media are intensely researching in order to find improvised methods to label and recognize fake news to protect the public from disingenuous information. In the following paper, we have conducted a survey on the existing machine learning algorithm which is deployed to sense the fake news. The three algorithms used are Naïve Bayes, Neural Network and Support Vector Machine (SVM). Normalization is used to cleanse the information before implementing the algorithm.

Keywords News · Fabricated media · Influence · Misleading ınformation · Naïve Bayes · Neural network · Support vector machine · Normalization method

Utkarsh (✉) · Sujit · S. N. Azeez · B. C. Darshan · H. A. Chaya Kumari
Department of Computer Science and Engineering, Vidyavardhaka College of Engineering,
Mysuru, Karnataka, India
e-mail: utkarshsteve7@gmail.com

Sujit
e-mail: sujithashapur123@gmail.com

S. N. Azeez
e-mail: syednabeelazeez@gmail.com

B. C. Darshan
e-mail: darshangowda101@gmail.com

H. A. Chaya Kumari
e-mail: chayakumari44@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_60

1 Introduction

With the advancement of technology, unclassified facts are available to everyone for free. This is a leap in the advancement of mankind but at the expense of blurring the lines between true media and maliciously fabricated media. News, stories or hoaxes created intentionally to misguide readers can be defined as fake news. Fake news can be deliberately created in order to influence one's views, promote a political agenda or to cause a misperception which can be profitable to certain business communities. It can deceive a person by personifying a trusted website or by using analogous names and addresses of reputable organizations. Conventionally, we always got broadcast information from trusted sources, journalists and media which essentially need to trail certain codes of practice. With the invent of online news broadcasting, there are very little editorial standards leading ungenuine news movement.

Due to the circulation of news and facts through social media and other networking sites, it is often difficult to differentiate among credible and fake information. Due to the generation of an excess of information and little knowledge among the general public about the working of the internet, hoax news generation has found immense growth. The increasing outreach of these stories is majorly due to social media sites where tremendous data is generated. It acts as a platform for the public to discuss events that have occurred leading to the formation of various conspiracy theories.

The associations monitoring the online frameworks have made it mandatory to contact beast onlookers before streaming any content on a website, blog or profiles. Business communities, diverse substance markers/traders have channelized the fake news generation into their favour. Counterfeit news can be emphasized as a profitable business, publicizing pays for distributors who generate and distribute stories that travel throughout the web. The more breakthroughs a story gets, the more money online news creators make for driving the web traffic in the light of this news. So, it becomes an essential criterion to categorize a news article as true or false, so as to improvise the quality and accuracy of news public receives each day. In this work, we mainly put forth the approaches to discernment fake news. The rudiment method being Naïve Bayes and the sophisticated methods are Neural Network and Support Machine Network (SVM).

2 Related Work

Trustworthiness, believability, reliability, accuracy, fairness, objectivity can be used to define the credibility of the information.

Contents of fake news acknowledge people to believe falsified information and sometimes it can be a sensitive message. These messages upon being received, disperse rapidly among communities. The dissemination of hoax stories adversely affects various people beyond specific clusters. The main confusion is due to the incapability to separate believable and unbelievable data being circulated via social media

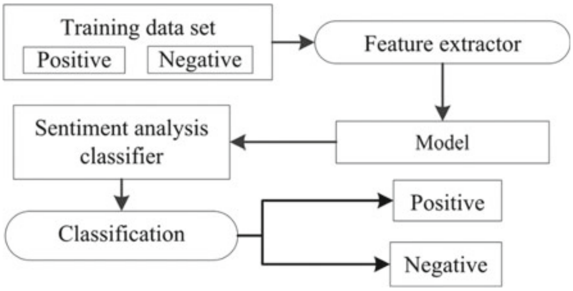


Fig. 1 Sentiment analysis flow diagram

outlets. Presence of fake news imposes a greater threat to one's life and property. Fake new proliferation can take place due to misinformation, that is the distributor believes the news is true or due to disinformation, which occurs when the distributer intentionally circulates a hoax [1, 2].

For example, fake news generated states that, Donald Trump donates his entire $400,00 salary to re-establish cemeteries. This could possibly not be true as he cannot donate his entire annual salary for this cause, as he has already assigned first quarter's worth of his salary for a different initiative under Department of Veterans Affairs.

As shown in Fig. 1 researches use mostly sentimental study [2] and segregation to detect the hoax stories, nonetheless, it always depends on the dialect's context [3].

2.1 Machine Learning Methods

2.1.1 Naïve Bayes

It is a straightforward machine learning algorithm. It is a very prominent calculation that can be deployed to determine the accuracy of news is credible or not, by utilizing multinomial NB and pipelining ideas. Normal standards are persuaded by the number of algorithms, so it cannot be the main calculation for formulating such classifiers.

Naïve Bayes is a simple method to detect whether a news is fake or not.

It is a type of calculation which is applied in content characterization. The utilization of token is mainly concerned with detecting whether the news is reliable or unreliable in Naïve Bayes classifier. After that, the exactness of the information is obtained by applying Bayes postulate.

Naïve Bayes Formula and Details

The next concept is the equation for Naïve Bayes order that exploits the likelihood of the past occasion and distinguishes it with the current occasion. Each and every likelihood of the event or occasion is determined. Finally, the general likelihood of the news that is contrasted with dataset is determined.

Along these lines on computing the general probability, we can get the estimated esteem and can recognize whether the news is genuine or fake

$$\mathrm{P(X|Y)} = \mathrm{P(Y|X)} \cdot \mathrm{P(X)}\big/\mathrm{P(Y)} \tag{1}$$

To find the probability of a circumstance, given X when event Y is assumed to be TRUE.

2.1.2 Support Vector Machine

In the mid-60's scientists proposed the primary Support Vector Machine (SVM). The model can simply coordinate classification and it misses out the most practical issues that are being faced while classifying. In the early 90's researchers, established SVM that promotes discontinuous classification. This made SVM more efficient for users. Radial Basis Function is used in our implementation. The clarification we implement in this part is two Doc2Vec feature vectors which are adjacent to each other. If their addressing chronicles are relative, the detachment is figured by the bit limit, regardless of addressing the principal partition. The formula is as follows:

$$K\left(x, x'\right) = \exp\left(-\frac{x - x'^2}{2\sigma^2}\right)$$

It precisely addresses the required dependency and it is a regular kernel for SVM.

We use the speculation familiar in [4] with executing the SVM. The guideline thought of the SVM is to disconnect multivariate classes of data by the vastest 'street'. This target can be addressed as the improvement issue.

$$\arg\max_{w,b}\left\{\frac{1}{W}\min_{n}[t_n\left(w^T\emptyset\left(x_{n)} + b\right)\right]\right\}$$

$$s.t.t_n(w^T\emptyset(x_n) + b \leq 1n = 1, 2, \ldots N$$

At that point, we utilize the Lagrangian function to dispose of limitations.

$$L(w, b, a) = \frac{1}{2}w^2 - \sum_{n=1}^{N} a_n\left\{t_n\left(w^T\emptyset(x_n) + b\right) - 1\right\}$$

where an $\geq$0, n = 1, N.

At long last we take care of this improvement issue utilizing the raised advancement devices gave by Python bundle CVXOPT.

2.1.3 Neural Network

Huge Feedforward networks or rather multilayer perceptron's can be regarded as the most critical learning models establishments. CNNs and RNNs are only certain remarkable examples of Feedforward networks. Controlled AI assignments are implemented using these networks. This is where we undeniably understand the possible outcome, we need our network to perform. These can be noted as fundamentals for rehearsing AIs. It is to mainly structure the description of different business applications, regions, for example, PC vision and NLP that was basically influenced by the nearness of these networks.

The rule focus of a feedforward network is to incorrect some utmost f*. For instance, an apostatize work y = f*(x) maps an information x to a worth y. It depicts a mapping y = f (x; θ) and learns the estimation of the parameters θ that outcome in the unsurpassed work deduce.

We mainly comprehended two feedforward neural networks. One of them by using TensorFlow and the other one using Keras. Present-day NLP applications deploy neural frameworks on the huge scale [5], instead of using straight models like SVM's and logistic regression that revolved around progressive techniques. Three hidden layers are used in our neural framework. For introductory analysis, we deployed Rectified Linear Unit (ReLU), which is regarded to be appropriately suitable for NLP applications [5].

It has a constant magnitude feed of x R 1 × 300

$$h1 = ReLU(W_1 x + b_1)$$

$$h2 = ReLU(W_1 h_2 + b_2)$$

$$y = Logits(W_3 h_2 + b_3)$$

3 Discussion

3.1 Limitations of Existing System

There are some limitations in the studied systems and they are mentioned below:

1. Naïve Bayes
   - In Naïve Bayes it makes a strong assumption on the distribution of data.
   - If a variable has a category which was not observed on training dataset, then the model will give (assign) the result as zero (0).
   - Naïve Bayes is also known as a bad estimator, so the output is not taken too seriously.
2. Support Vector Machine (SVM)
   - The main drawback of SVM algorithm is that there are several important parameters that need to be set correctly to get the best classification.
   - It is not suitable for large datasets.
   - It does not perform very well when target classes are overlapping.
   - The algorithm will over-fit, if the number of features is much greater than the number of samples also it does not provide probability estimates.
3. Neural Network
   - The artificial neural network requires processors with parallel processing power according to their structure. For this reason realization of hardware is essential.
   - No clue about how results are acquired (obtained), so you cannot know what causes the output and how it is obtained.
   - ANN works with numerical data, so problems have to translate into numerical values.
   - It works on large datasets so the training time is very high and the duration of the network is unknown.

4 Proposed System

In Fig. 4 the system architecture is shown which uses the state of the art algorithm Long Short Term Memory(LSTM) to overcome the drawbacks proposed by the above stated algorithms.

4.1 Data

Kaggle is the source for extracting dataset for our implementation model [2]. The dataset consists of about sixteen thousand six hundred lines of data extracted from numerous reports available online.

A lot of pre-processing is done on the dataset in order to get it ready for the implementation. This can be clearly notable in the source code [3] that will be performed to set up training models. The attributes that our dataset have, are as follows:

- id: This attribute refers to the exclusive identification
- heading: This is the label of the news report
- editor: editor of the news columns
- script: This is the data of the report which can be partially written.
- marker: To mark whether the source is credible or not
- F: un-credible
- T: credible.

4.2 Data Cleansing and Attribute Retrieval

Data Cleansing refers to the task of transformations applied to our data before it can be considered apt to feed it to the algorithm. The technique used to convert the raw data into clean, usable data set is often referred to as data pre-processing. Usually, when we tend to collect data from various sources, it is in the raw form. It is not feasible for analysis, hence it must be pre-processed to match our needs. In Fig. 2, the seuqential steps for data preprocessing is shown which involves collection of data, structuring the data into a proper format, performing preprocessing and then performing graphical analysis of the results.

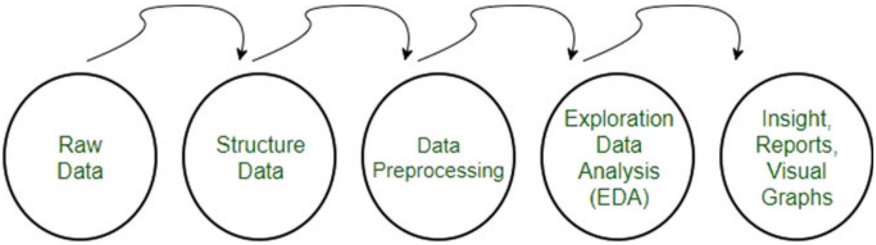

Raw Data
Structure Data
Data Preprocessing
Exploration Data Analysis (EDA)
Insight, Reports, Visual Graphs

Fig. 2 Data pre-processing

4.2.1 Need for Data Pre-processing

- Information arrangement must be regarded as highly specific whenever a Machine Learning venture is taking place. This leads to better outcomes. Few of the machine learning models prefer data to be organized in a predetermined format for processing.
- Information collection through various sources and streams is highly necessary as it accounts for running more than one Machine Learning and Deep Learning calculations that can be executed in one informational index pertaining to the selection of the best algorithm for deployment.

The pre-processing of data involves

- Removing of unrelated texts
- Removing empty cells
- Removing stop words
- Truncating data without labels.
- Converting all text to lowercase.

Upon performing these steps, we obtain a CSV file, which is fed to Doc2Vec algorithm as an input.

4.3 Doc2Vec

The main agenda of Doc2Vec is to generate a numeric representation of a document, irrespective of its length. Words are in logical structure, but documents are not. Hence an alternative method must be devised for numeric representation creation. A Doc2Vec as shown in Fig. 3 can be utilized in order to perform this task. As an

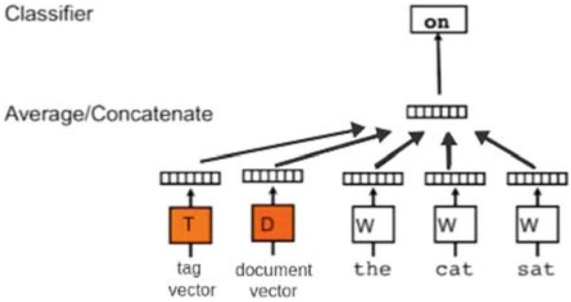


Fig. 3 Doc2Vec

initial step of preparation, a lot of records must be collected. For each word, a word vector W is produced. For each archive, record vector D is assigned. The model also formulates loads of data for SoftMax concealed layer. In the case of derivation organize, another manner can be emphasized. This leads to fix loads of ascertaining document vector.

Word2Vec communicates with documents by connecting vectors of individual words. This, however, leads to loss of all word request data. A Word2Vec is generated by the Doc2Vec by the involvement of a ‘document vector’, that yields a portrayal containing some information about the document overall. This enables the familiarity of the data about the word request. We are expecting an output that differentiates the unpretentious contrasts between content documents. Hence, conservation of word request data makes Doc2Vec very useful for our application.

4.4 Text Encoding and Word Embeddings

It is necessary to convert text data into vector representation in order to feed words into a machine learning algorithm. One of the methods is to use word embeddings.

In order to understand Word embeddings in simpler terms, it can be expressed as writings that are changed over into numbers and there might be diverse numerical representations of similar book. It is prominently stated that for unknown reasons, many machine learning calculations and practically all Deep Learning infrastructures are not capable of formulating sentences or plain context in a rudimentary manner. They need figures so as to carry out a particular given task, be it order relapse and so on in expansive standings. After the extensive degree of knowledge being extracted from content organization, it is a most basic task to remove data out of it and assemble applications. Some certifiable utilization of content applications are—slant scrutiny of audits by Amazon and so on, collection or broadcast characterization or clustering by Google, and henceforth.

Word Embeddings cluster and describe words by making attempts to utilize a lexicon to a vector. The following example separates the sequence of words into more probable subtleties to have a reasonable view. Investigating the model—sentence = 'Word Embeddings are Word changed over into numbers'. A word in a sentence can be considered to be 'Embeddings' or 'numbers' and so on. A lexicon can review every single one of kind words in the sentence. Along these lines, a lexicon may resemble—['Facts', 'and', 'figures', 'will, 'represent', 'numbers']. Vector description of a word might be a one-hot encoded vector where 1 represents the position where the word exists and 0 wherever else. The vector portrayal of 'numbers' in this organization as per the above lexicon is [0, 0, 0, 0, 0, 1] and of altered into [0, 0, 0, 1, 0, 0].

4.5 The Long Short-Term Memory (LSTM)

Hoch Reiter ad Schmid Huber proposed the Long Short-Term Memory (LSTM) unit [6]. It can be regarded as an extensively useful tool in describing serialized objects. This is because it makes a gauge by explicitly taking the past data and utilizes that to put together the present commitment. The content of the news we are concerned about is usually serialized. The adjuration pf sentences are critically based on the words. So, the LSTM model is best preferred for our implementation idea.

It is a general idea to schedule our events of the day, based on appointments based on work. Whenever we encounter an important task, we adjust it with fewer priority works, that can be performed later. Using LSTMs, the information drifts through a mechanism that is referred to as cell states. This enables LSTMs to selectively remember or forget things. There are mainly three dependencies for a particular cell state information. We can instantiate this for predicting stock prices for a particular stock.

For a particular day, the stock price will be detected based on the following factors:

- The previous day trend of stock which can be a downtrend or an uptrend.
- The traders compare previous day's stock price before buying them, so it is necessary to address the value of stock on the previous day.
- It is necessary to consider the factors that mainly affect the price of stock in the present day. The influencing factors can be a policy that is implemented by a company which is widely unaccepted, drop in the profit of a company or a change in the high position of a company unexpectedly.

The dependencies can be generalized as follows:

- The state of the previous cell, that is the information that was present in the memory previously.
- The hidden state's previous cell information which is regarded as the output of the previous state.
- The current time step taking in the new information

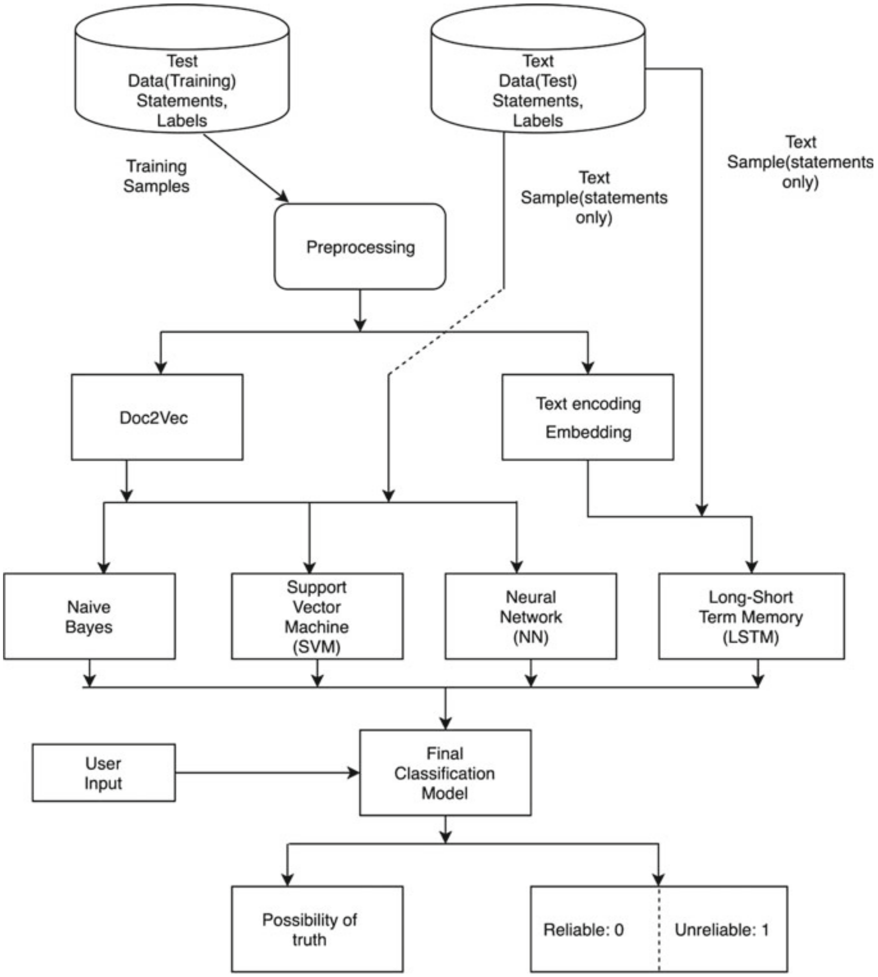


Fig. 4 System architecture

Since the request for the words is significant for the LSTM unit, we can't utilize the Doc2Vec for pre-processing on the grounds that it will move the whole archive into one vector and lose the request data. To forestall that, we utilize the word embedding (Fig. 4).

5 Conclusion

Fake news makes it difficult for the general public to believe in what is right and what is wrong because the rumours make it hard to identify the truthiness of a fact

[6]. Due to the failure of the capability of furnishing a legible content, a corpus used in IBM's Watson led to the let-down of the initial archetype examination in late 2016 [7]. A tremendous idea needs to be formulated to detect the proliferation of truth and fake news through various streams [8]. A model built on this purpose will prove to be definitely useful in this modern era [9, 10].

Fake news can be identified using machine learning methods. In this experiment machine learning methods used are Naïve Bayes, Neural Network and Support Vector Machine (SVM) which detects the fake news with high confidence. We can for future enhancement use Long Short-Term Memory (LSTM) to improve the results as LSTM works like the human brain. It keeps the information which is useful and discards the unnecessary information which is false or is not required.

References

1. Campan A, Cuzzocrea A, Truta TM (2017) Fighting fake news spread in online social networks: actual trends and future research directions. In: IEEE International conference on big data (BIGDATA), pp 4453–4457
2. Castillo C, Mendoza M, Poblete B (2011) Information credibility on twitter. In: Proceedings of the 20th international conference on World wide web (WWW'11). ACM, New York, NY, USA, pp 675–684. http://dx.doi.org/10.1145/1963405.1963500
3. Lorek K, Suehiro-Wiciński J, Jankowski-Lorek M (2015) Automated credibility assessment on twitter. Comput Sci 16(2):157–168. http://doi.org/10.7494/csci.2015.16.2.157
4. AlRubaian M, Al-Qurishi M, Al-Rakhami M, Rahman SM, Alamri A (2015) A multistage credibility analysis model for Microblogs. In: Pei J, Silvestri F, Tang J (eds) Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015 (ASONAM'15). ACM, New York, NY, USA, 1434–1440. http://dx.doi.org/10.1145/2808797.2810065J. Maxwell C (1892) A treatise on electricity and magnetism, 3rd ed, vol 2. Oxford, Clarendon, pp 68–73
5. Goldberg Y (2015) A primer on neural network models for natural language processing. https://arxiv.org/pdf/1510.00726.pdf
6. Aker A, Bontcheva K, Liakata M, Procter R, Zubiaga A (2017) Detection and resolution of rumours in social media: a survey. CoRR, http://arxiv.org/abs/1704.00656
7. Vorhies W (2017) Using algorithms to detect fake news—the state of the art. http://www.datasciencecentral.com/profiles/blogs/using-algorithms-to-detect-fake-news-the-state-of-the-art
8. Ehsanfar A, Mansouri M (2017) Incentivizing the dissemination of truth versus fake news in social networks." 2017 12th System of systems engineering conference (SoSE), 1–6
9. Berghel H (2017) Alt-news and post-truths in the "fake news" era. Computer 50(4): 10–114. https://doi.org/10.1109/MC.2017.104
10. Buntain C, Golbeck J (2017) Automatically Identifying fake news in popular Twitter threads. In: 2017 IEEE ınternational conference on smart cloud (Smart Cloud), pp 208–215
11. El Ballouli R, El-Hajj W, Ghandour A, Elbassuoni S, Hajj H, Shaban K (2017) CAT: credibility analysis of Arabic content on Twitter. WANLP@EACL
12. Hochreiter S, Jrgen S (1997) Long short-term memory. http://www.bioinf.jku.at/publications/older/2604.pdf
13. Granik M, Mesyura V (2017) Fake news detection using naive Bayes classifier. In: 2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON), pp 900–903
14. Alrubaian M, Al-Qurishi M, Hassan MM, Alamri A A credibility analysis system for assessing information on Twitter. IEEE Trans Depend Secure Comput 1–14. https://doi.org/10.1109/tdsc.2016.2602338k.Elissa, "Title of paper if known," unpublished

15. Hertz J, Palmer RG, Krogh AS (1990) Introduction to the theory of neural computation, Perseus Books. ISBN 0-201-51560-1
16. Thandar M., Usanavasin S. 2015 Measuring opinion credibility in Twitter. In: Unger H, Meesad P, Boonkrong S (eds) Recent advances in information and communication technology 2015. Advances in intelligent systems and computing, vol 361. Springer, Cham
17. Gupta M, Zhao P, Han J (2012) Evaluating event credibility on Twitter. In: Proceedings of the 2012 SIAM international conference on data mining, pp 153–164. http://epubs.siam.org/doi/abs/10.1137/1.9781611972825.14
18. Conroy NJ, Rubin VL, Chen Y (2015) Automatic deception detection: methods for finding fake news. In: Proceedings of the 78th ASIS&T annual meeting: information science with impact: research in and for the community (ASIST'15). American Society for Information Science, Silver Springs, MD, USA, Article 82, 4 p

Detection of Diseased Plants by Using Convolutional Neural Network

M. Maheswari, P. Daniel, R. Srinivash, and N. Radha

Abstract Agricultural takes a major percentage in a country's economic growth. Crop production plays an essential role in agriculture. Countries' economical growth rate is reduced due to less crop production. Foods are essential for every living being, since we need proper food for survival. Hence, it is essential for every farmer to cultivate a healthy plant to increase the crop production. However, in nature, every plant can get attacked by some sort of disease but the level of damage occurred to the crops are different for every plant. If a fully matured plant get affected by a simple disease, it will not affect the full plant but if a small plant gets affected by the same disease, it causes severe damage to the plant, as we cannot manually monitor the plants and cannot detect the disease occurring in the plants everyday. Huge manpower is needed to monitor every plant in the farm so it needs time for monitoring every crop in the field. In this paper, image recognition using conventional neural network (CNN) has been proposed to reduce the time complexity and manpower requirement. The proposed algorithm accurately detects the type of diseases that occurs in the plants.

Keywords Plant disease detection · Image recognition · CNN

M. Maheswari (✉) · P. Daniel · N. Radha
Department of Electronics and Communication Engineering, K. Ramakrishnan College of Engineering, Tiruchirappalli, India
e-mail: kousi.rhithi@gmail.com

P. Daniel
e-mail: danielraja783@gmail.com

N. Radha
e-mail: radha1290@gmail.com

R. Srinivash
Department of Computer Science and Engineering, K. Ramakrishnan College of Engineering, Tiruchirappalli, India
e-mail: srinivashramasamy1706@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_61

1 Introduction

Western Asia is considered to be the birthplace of agricultural revolution where wild ancestors of wheat and barley and domesticated animals like goat, sheep, pig, and cattle are found. The period from 7500 to 6500 B.C. was the period of discovery of agriculture. Agriculture in India is considered as a primary activity because a large chunk of the population in India depends on farming for their livelihood. In India, more than 70% of the population is depending on agriculture in one form or the other. The present populace is around 1000 million which is relied upon to balance out at around 1500 million by the center of the present century. This pattern of populace development made disturbing circumstances as the extent of expanding region under development is limited. The importance of agribusiness can be estimated by the offer of farming in national salary and work design and so forth. Knowing the significance of agribusiness, the farming area can be connected with the modern division [1]. According to data published by the Ministry of Food Processing Industries recently, Harvest and postharvest loss of India's major agricultural produce is estimated at Rs 92,600 crores approximately. The new spending plan for the farming segment expanded to 44% from Rs 24,909 crore in 2015–'16 to Rs 35,984 crore in 2016–'17 [1].

2 Proposed Algorithm

Due to heavy loss in the crop productivity, everyday farmers are facing severe problems in the crop production. These are happening due to the increase in the disease on the plants and so it causes the loss in the productivity. Hence in the large farming lands, the plants with the disease cannot be identified easily by the human-races. So we decided to improve the crop productivity by introducing a PLANT DISEASE DETECTOR which will be useful for every farmer. It helps in identifying the plants with the disease and gives the severity level of the attacked disease. We also include a special method to give an alert to the farmers through mailing them the notification message that includes the image of the plants with the disease and the affected area and gives the tips to cure the disease or to remove the crop from that area. We use Tensor flow library for object detection and SMTP library for the mailing process. Tensor flow library used to detect the disease occurred in the plants through training the datasets by CNN. We mainly use CNN for accuracy and clear classification of the image sets. SMTP files include mailing process if the disease is detected, then automatically the mailing process is called and then they are delivered to the receiver.

Detection of disease by visualizing through our own eye is difficult. Therefore in the field of agriculture, detection of disease plays an important role in visualization of plants through computer-based automation. These automatic disease detection will be beneficial in the field of crop cultivation. An automatic detection of those diseases are identified just by analyzing with the datasets provide to the machine. In plants, some general diseases are brown, black, and yellow spots, and others are fungal, viral, and bacterial diseases. Image recognition is a technique to identify the area of affected plant disease and to determine the difference in their color of the affected area [2].

Digital Image Processing is a growing technology which is helpful to enhance the quality of the image [3]. Image recognition is the process of identifying people, objects, places, and letters by the process of machine learning. These can be achieved using Neural networks for the machine to learn the images. In this paper, it is proposed to identify the disease that occurs in the plants by using Convolutional Neural Networks [4, 5]. There are several types of Neural networks Convolutional neural network outperforms the other neural networks in terms of memory occupancies and performances. The neural network works in a similar way like neurons of the human brain in the way how it stores information and passes the information to us fast. The same neural network can be used for the prediction of glaucoma in medical images [6] and in wireless sensor networks for optimizing the network performance [7]. The information that we are feeding to it are the inputs they get split into numerous tiles. The tiles are get matched with the hidden layers that we have trained from the datasets that we have. The hidden layer consists of the convolutional layer and pooling layer [5]. The convolutional layer is the main block of the neural network, when we give an image as an input. The pooling layer has the work to form a cluster of neuron to carry the outputs into a single neuron. In this, max pooling is common to do these works. Then a single neuron carries the information and gives the output from the output layer [8, 9].

When we use CNN in the detection of disease that occurs in the plants, it produces us a minimum error rate hence the object can be classified perfectly. The network is trained with 800 database that includes all diseases that occur in the plants. They used 120 images for every 10,000 iterations [4, 10]. The conventional neural network structure is given in (Fig. 1).

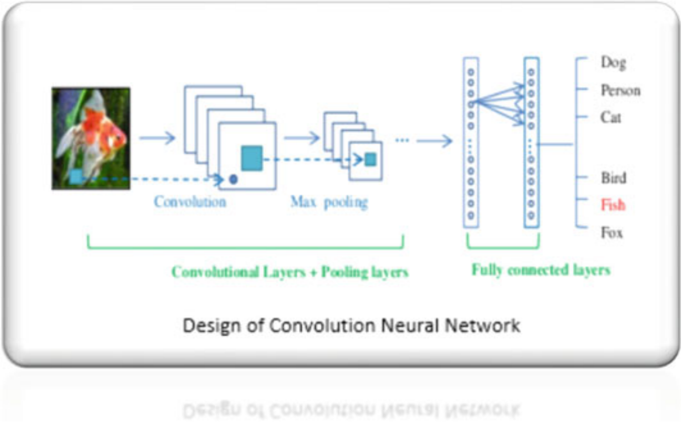

Convolution
Max pooling
Dog
Person
Cat
Bird
Fish
Fox
Convolutional Layers + Pooling layers
Fully connected layers
Design of Convolution Neural Network

Fig. 1 Structure of conventional neural network

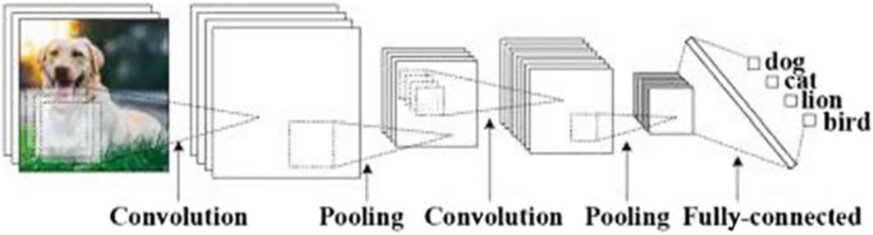


Fig. 2 Training of CNN

In this paper, we address another important aspect of Convolutional neural network architecture design, which is feasible due to the use of very small convolution filters in all layers. During the training of the images, they are computed to the trained set as RGB images from each pixel [4, 11]. The training method of CNN is shown in (Fig. 2).

The convolutional neural network is trained on the newly trained set. We can train the dataset from the training set and also they are tested from the test sets. The neural structures have five hidden layers, there is no overlap between any trained and test datasets [12, 13]. The connection between the layers of CNN is shown in (Fig. 3).

The completely associated layer contains neurons of which are specifically associated with the neurons in the two contiguous layers, without being associated with any layers inside them. Here, the images from the camera are the input to the neural networks [14]. The output layer gives the information if the images provided matches while training the datasets [8]. The output layer can also be able to identify the severity level of the disease occurred to the crops. The information we provided can give an error if all the information are represented at a time, so the output gives the specified information of the input we fed. For example, if the hidden layer contains two units one that occurs in leaves and other that occurs in fruits or vegetables or crops and we are feeding an image of some plants which have the disease in leaves,

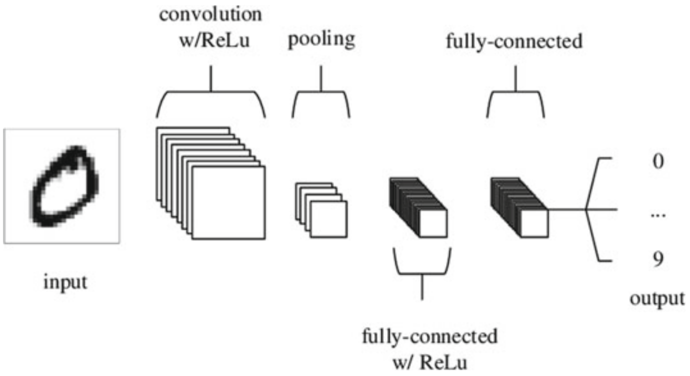


Fig. 3 Connection between the layers of CNN

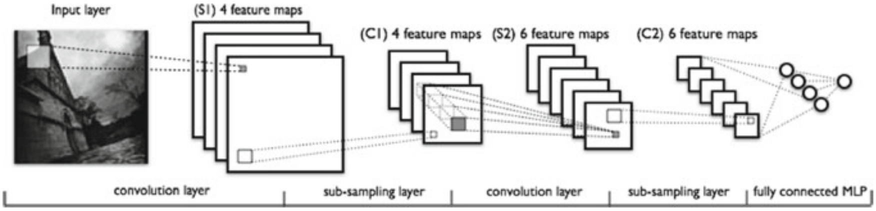


Fig. 4 Layer structure of CNN

then the first hidden layered units value get increased and the other units values are decreased, if input is given as the plant having a disease that occurs in whole plants, then their values are summed and gives the output from the values [8]. The complete layer structure of CNN is shown in (Fig. 4).

In the working process of Neural network, the images are split into small tiles of images and they are transmitted to the neurons as the information. In CNN, not every neuron is going to carry all information; any one neuron accepts the inputs in the form of small layers of the subsection. In the subsampling area, the tiled images are stored in the form of arrays and these featured arrays are then mapped in the pooling layer and they are carried as one single information to the output and provide the accurate classification of disease. The featured selection process helps to give fast information and execution and higher classification. This technique produces more accuracy than other detection techniques. In first, the trained image of the neural networks is taken that have some bacterial, viral, and fungal diseases in plants. The experimental detection gives the classification of the images with the automatic detection of the disease occurred in the plants. Based on the best performance results, it is well capable of detecting or classifying the disease accurately around 99% [15].

3 Tensor Flow Object Detection API

Tensor Flow's Object Detection API [16] is an incredible asset that makes it simple to develop, train, and convey object identification models. In the majority of the cases, preparing a whole convolutional organize without any preparation is tedious and requires extensive datasets. This issue can be fathomed by utilizing the benefit of exchange learning with a pre-prepared model utilizing the Tensor Flow API. Before diving into the specialized subtleties of executing the API, how about we examine the idea of exchange learning. Exchange learning is an exploration issue in machine discovering that centers around putting away the information picked up from taking care of one issue and applying it to an alternate yet related issue. Exchange learning can be connected in three noteworthy ways; Convolutional neural system (CNN) as a settled component extractor: In this strategy, the last completely associated layer of a CNN is evacuated, and whatever remains of the CNN is treated as a settled element extractor for the new dataset. Adjusting the CNN: This technique is like the past strategy; however, the thing that matters is that the loads of the pertained arrange are tweaked by proceeding with backpropagation. Pre-prepared models: Since present day CNN takes a long time to prepare without any preparation, usually to see individuals discharge their last CNN checkpoints to help other people who can utilize the systems for calibrating. For instance, [17] Tensor Flow Zoo5 is one such place where individuals share their prepared models/checkpoints. In this trial, we utilized a pre-prepared model for the exchange learning. The benefit of utilizing a pre-prepared model is that as opposed to building the model without any preparation, a model prepared for a comparative issue can be utilized as a beginning stage for preparing the system. Numerous pre-prepared models are accessible. This analysis utilized the COCO pre-prepared model/checkpoints SSD [18]. Mobile Net from the Tensor Flow Zoo [17, 19]. This model was utilized as an introduction checkpoint for preparing. The model was additionally prepared with pictures of traffic lights from Image Net. This calibrated model was utilized for induction. Steps involved in [16] Tensor flow object detection API is shown in (Flow chart 5).

771 images have been used over of diseased plants as dataset. For training, 650 images have been used and for testing, 121 images. The dataset, training, and testing images have been shown in (Figs. 6 and 7).

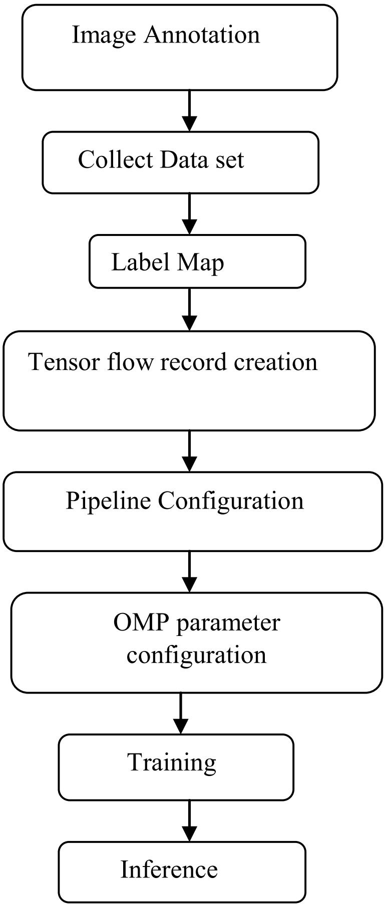
4 Image Annotation

For labeling image, Label Image has been used [20]. Label Image is a graphical image annotation tool. All the images have been manually labeled for test and training dataset. The severity level of the disease is also classified that may be useful to feed the correct amount of fertilizer to the crop. The Label Image is shown in (Fig. 8).

The images are saved in the XML format for example

```
<?xml version="1.0"?>

<annotation>

<folder>train</folder>

<filename>image010.jpg</filename>

<path>C:\tensorflow1\models\research\object_detection\images\train\im
age010.jpg</path>

<source><database>Unknown</database></source>

<size><width>448</width>

<height>448</height>

<depth>3</depth></size>

<segmented>0</segmented>

 </annotation>
```

Flow Chart 5 Design flow



XML FILE FORMAT

Training images play a major role in image recognition. The images have been trained until the loss becomes below 0.05. This is quite accurate. The loss graph is shown in the below image. This graph was plotted automatically with the help of the tensor board (Fig. 9).

In the proposed model, it is designed such that the notification alert will be sent to the user if the accuracy of the disease is above 80%. The above statement will be suitable for both low severity and high severity condition. The sample run of the model is shown in Figs. 10 and 11.

The inferencing video was first changed over into edges utilizing MoviePy, a Python* module for video altering. These arrangements of casings are given to our model prepared utilizing exchange learning. After the edges go through the Object Detection pipeline, the jumping boxes will be drawn on the distinguished casings. These edges are at long last converged to frame the surmised video. For notifications,

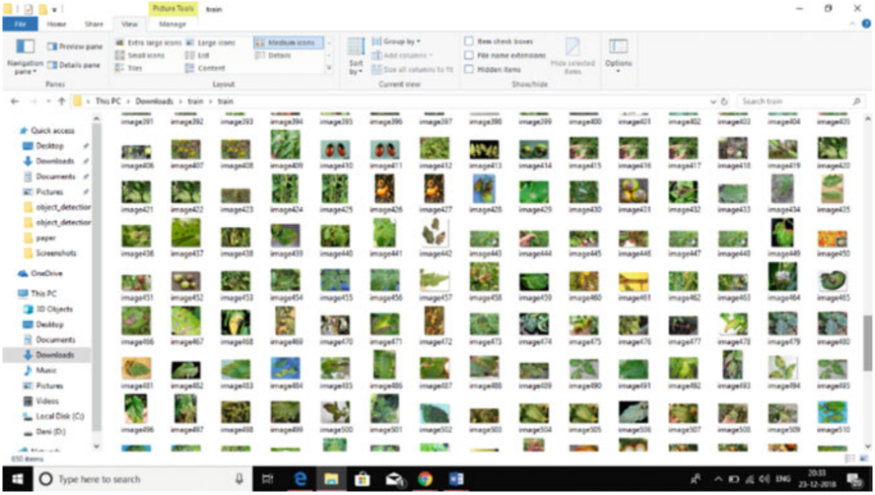


Fig. 6 Images of training DATASET

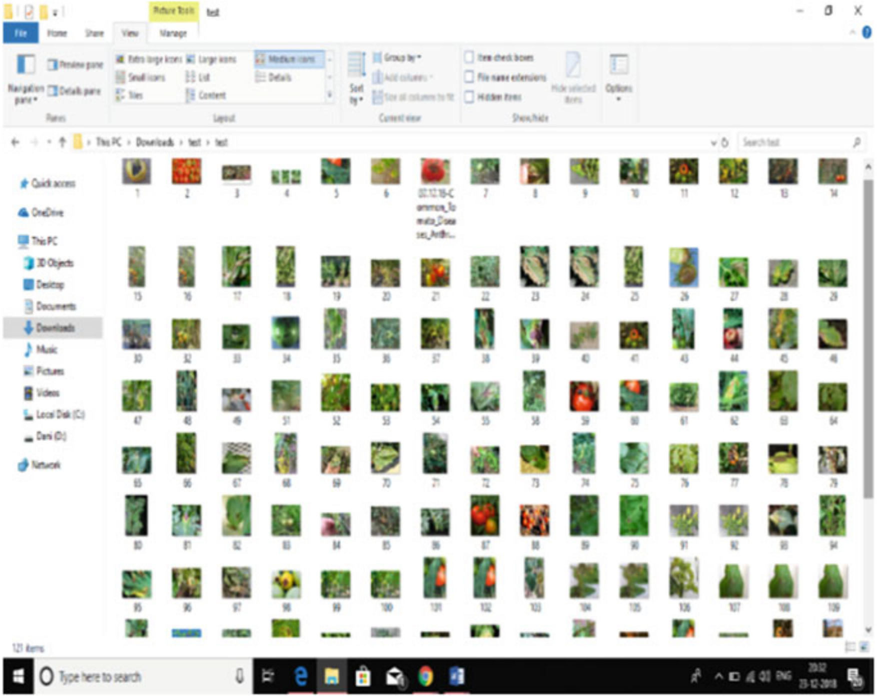
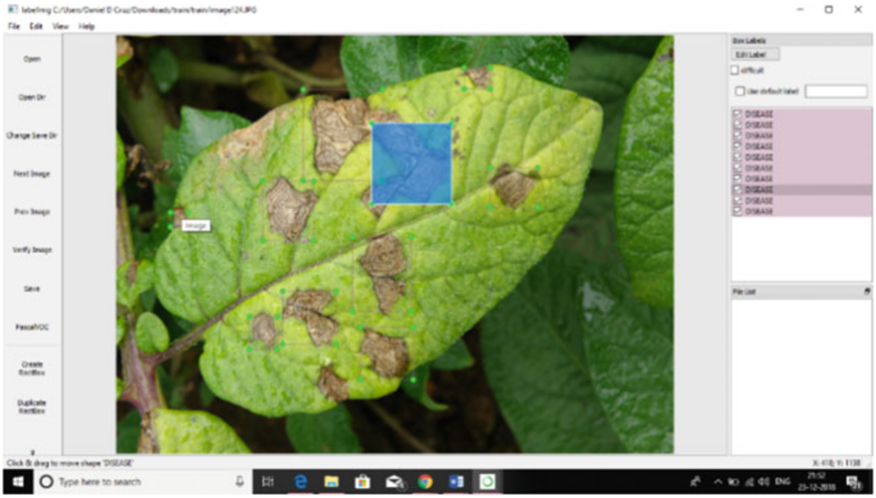


Fig. 7 More images of training DATA SHEET

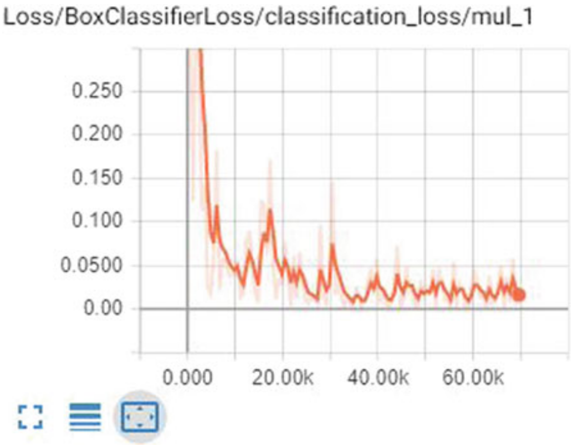


Click & drag to move shape 'DISEASE'

X: 418; Y: 1138

Fig. 8 Labelling image

Fig. 9 Loss graph



we have used SMTP (Simple Mail Transfer Protocol) [21, 22]. In that notification, we attached the defected plant image and some suggestions for that.

5 SMTP

SMTP is a protocol used for mailing purpose which is sent by the sender and received by the receiver through the host line. It is connected with our code for disease identification to find if any disease that occur in the plant, and if the disease severity level

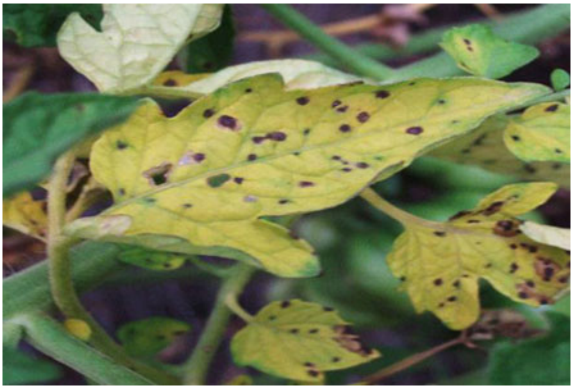


Fig. 10 Sample run image before recognizing

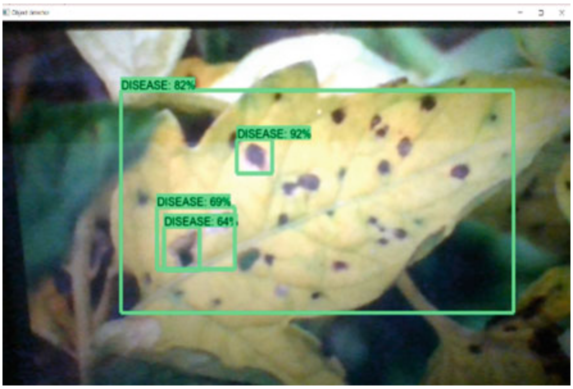


Fig. 11 Sample run image after recognizing

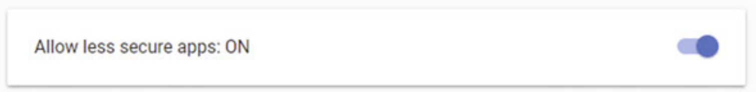
is above and equal to 80%, then it will send the message to the receiver mail provided in the code [23]. There are two modes of SMTP models; they are

- End to end method,
- Store and forward method.

In these two methods, the store and forward message through the sender to the receiver is used. The mail can be sent with the header, body title, and images to the receiver. To do this process, we have to change some of the settings in the sender's mail settings. If we have done two-step verification on Gmail, we have to OFF the Two-step verification. We have to ON the less secure app allowance as shown in Fig. 12.

Now by using the SMTP library in the code by installing the SMTP by "pip install smtplib" command [23, 22]. Next, we can create the mailing code and we need to provide the sender and receiver's mail address. Next, we need to provide the common server for login to our mail.

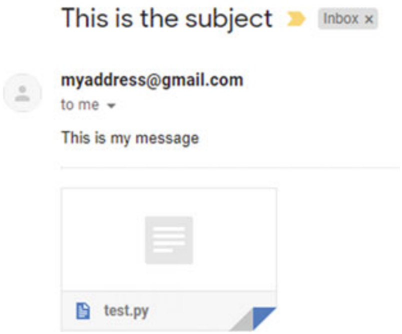
```
server=smtplib.SMTP('smtp.gmail.com',587)
```



Allow less secure apps: ON

Fig. 12 Image for less secure app ON condition

We need to provide the sender's mail with the password then only we can send mail to the receiver what we need to send as a message. If we need to attach images, we can insert attachment in the code [22].



This is the subject

myaddress@gmail.com

to me

This is my message

test.py

If any disease occurs that is identified by the system, then the image is get captured by the system and they get ready to be mailed to the receiver we have provided. The mail contains the header, body content, and attachments of image. When the image gets captured they are attached with the mail automatically and the header provides which disease has occurred and the body content provides the ideas to prevent the plants from disease or to remove the disease. By those suggestions, we can protect the other plants from getting affected [23].

6 Conclusion

From this concept, we have concluded that by using this detection method, we can identify the diseased plants soon and this reduces our time of identifying the diseased plants by our own eye. The system has been tested with the different set of images and it easily detects the disease just by keeping camera pointing on the plants or using auto bots with circulating cameras moved across the plants, it captures the images of the diseased plants and sends to the receiver immediately. It acts as an efficient system by reducing our clustering time on finding the disease and the area of infection. It serves as a good tool for identifying the disease in the plants. The usage of the object detection for the identification of disease in plants helps to improve the accuracy of identification. Everyday, the technology provides new techniques and methods to easy our work. However, automation is the best in technologies but not used for agricultural purposes and it will be very helpful in the field of agriculture. We use the automation here for the automatic identification of the plant diseases and deliver the image of those plants with which area the plant is present and give the suggestions to avoid the disease in the future to the user by mailing them.

References

1. Ghosh P Article published in shareyouressays.com named as essay on Indian agriculture. Mayukh Bardhan, Digital Marketer, explains the present situation of Indian agriculture in Quora.com for the question “What is the present scenario of Indian agriculture”
2. Ghaiwat SN, Arora P (2014) Detection and classification of plant leaf diseases using image processing techniques: a review. Int J Recent Adv Eng Technol 2(3): 2347–2812. ISSN (Online)
3. Priyadarshini K (2019) Navigating visually impaired and sightless people using auditory guidelines. Int J Adv Res Comput Commun Eng 8(5) May
4. Karpathy A et al (2014) Large-scale video classification with convolutional neural networks. In: IEEE conference on computer vision and pattern recognition (CVPR)
5. Tsoi AC, Back AD, Giles CL (1997) Member face recognition: a convolutional neural-network approach Steve Lawrence. IEEE Trans Neural Netw 8(1) January
6. Preethi Rajaiah R, John Britto R (2014) Optic disc boundary detection and cup segmentation for prediction of glaucoma, Int J Sci Eng Technol Res (IJSETR) 3(10), 2665–2672
7. Shabina S (2014) Smart Helmet using RF and WSN technology for underground mines safety. In: Proceedings of international conference on intelligent computing applications, pp 305–309
8. H. Martin Hunke (1994) Locating and tracking of human faces with neural networks. CMU{CS{94{155 School of Computer Science Carnegie Mellon University Pittsburgh, PA 15213 c H. Martin Hunke, August
9. Sudhakar S Towards Data Science. Deep-Learning Deep Learning projects implemented by Shreenidhi Sudhakar.github.com. https://towardsdatascience.com/convolution-neural-network-e9b864ac1e6cshree6791/Deep-Learning
10. Baluja S, Rowley HA, Kanade T Neural network-based face detection. School of Computer Science, Carnegie Mellon University 5000 Forbes Avenue, Pittsburgh, PA 15213, USA. har@cs.cmu.edu http://www.cs.cmu.edu/˜har, baluja@cs.cmu.edu, http://www.cs.cmu.edu/˜baluja, tk@cs.cmu.edu, http://www.cs.cmu.edu/˜tk
11. Convolutional Neural Networks for Document Image Classification Le Kang, Jayant Kumar, Peng Ye, YiLi†, DavidDoermann, University of Maryland, College Park, MD, USA {lekang, jayant, pengye, doermann}@umiacs.umd.edu †NICTA and ANU yi.li@cecs.anu.edu.au
12. Lawrence S, Lee Giles C, Tsoi AC, Back AD (1997) Face recognition: a convolutional neural network approach. IEEE Trans Neural Netw 8(1):98–113
13. Hinton J (2012) Coursera lectures on neural networks. https://www.coursera.org/learn/neural-networks
14. Xu L, Ren JSJ, Liu C, Jia J Deep convolutional neural network for image deconvolution. Lenovo Research & Technology, xulihk@lenovo.com. Lenovo Research & Technology, jimmy.sj.ren@gmail.com. Microsoft Research, celiu@microsoft.com. The Chinese University of Hong Kong, leojia@cse.cuhk.edu.hk
15. An introduction to convolutional neural networks. https://www.researchgate.net/publication/285164623_An_Introduction_to_Convolutional_Neural_Networks. Accessed 23 Dec 2018
16. TensorFlow Object Detection API. https://github.com/tensorflow/models/tree/master/research/object_detection
17. TensorFlow detection model zoo. https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/detection_model_zoo.md
18. Ren S, He K, Girshick R, Sun J, Faster R-CNN: towards real-time object detection with region proposal networks

19. Kagaya H, Aizawa K, Food Detection and Recognition Using Convolutional Neural Network, Graduate School of Interdisciplinary Information Studies The University of Tokyo, kagaya@hal.t.utokyo.ac.jp. Dept. Information and Communication Eng. The University of Tokyo, aizawa@hal.t.utokyo.ac.jp, Makoto Ogawa foo.log Inc. ogawa@foo-log.co.jp
20. LabelImg. https://github.com/tzutalin/labelImg
21. Sureswaran R, Bazar HA, Abouabdalla O, Manasrah AM, El-Taj H (2009) Active e-mail system SMTP protocol monitoring algorithm. In: 2009 2nd IEEE international conference on broadband network & multimedia technology. https://doi.org/10.1109/icbnmt.2009.5348490
22. Klensin J, Editor, Network Working Group, Request forComments:2821 AT&T Laboratories,Obsoletes:821,974, 1869 April 2001, Updates: 1123Category:Standards Track
23. Klensin J (2001) SMTP Simple mail transfer protocol, RFC 2821, April

Emoticon: Toward the Simulation of Emotion Using Android Music Application

Aditya Sahu, Anuj Kumar, and Akash Parekh

Abstract Music unquestionably affects our emotions. We tend to listen to music that reflects our mood. Music can affect our current emotional state drastically. Earlier, the user used to manually browse songs through the playlist. Over the period, recommendation systems have used collaborative and content-based filtering for creating playlist but not the current emotional state of the user. This paper proposes an idea of an android music player application which recommends songs after determining the user's emotion by facial recognition at that particular moment using deep learning techniques. And create a playlist by considering the emotion of the user and recommending songs according to the current emotion of the user.

Keywords Convolutional neural network · Music application · Emotion recognition

1 Introduction

We know from psychology, that music induces emotional responses in human beings and it has tremendous application in the current research [1]. Emotion recognition research helps to analyse the process of listening and human interaction with the help of human robots and it will be a more efficient technique [2, 3]. Here, several techniques of emotion recognition have been discussed such as smile, angry, sad, neutral, disgust and fear.

Happy emotion helps to hear happy music with high beats and in sad emotion, can able to listen to slow motion songs. The main focus of this research work is to

A. Sahu (✉) · A. Kumar · A. Parekh
Siddaganga Institute of Technology, Tumkur, India
e-mail: aditya.1si16cs004@gmail.com

A. Kumar
e-mail: anuj.1si16cs017@gmail.com

A. Parekh
e-mail: akash.1si16cs009@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_62

design an efficient yet simple android music player which recommends and plays songs based on the current emotional state of the user. This paper consists of basically two modules, Emotion module and Android module. Emotion module detects the emotional state of the user. Android module integrates the Emotion module in android platform and recommends a playlist according to the output of emotion module and song database using firebase.

2 Related Work

Many works have been done in the field of emotion detection and how music can affect the user's emotion. There are several articles and papers which describe how music can affect our emotion. According to Annemiek Vink's paper, music can be used for therapy as well as psychological treatment [4]. It described how music can drastically change one's emotional state. Tanner Gilligan and Baris Akis published a paper on Real-time emotion detection using CNN [5]. They used the data from Cohn-Kanade dataset and their own custom images to train their model. Their model achieved an outstanding accuracy of 97%. Another paper uses the same dataset as well as MMI dataset using FeatEx block as the main component in their Neural Network Architecture [6].

Another paper focused on facial muscles [7] moments using neural networks to recognise fine-grained changes in facial expression based on the Facial Action Coding System (FACS) action units (AUs) [8]. Another paper by Dachapally, Prudhvi Raj uses Autoencoder Units concept on JAFFE dataset with an accuracy of 86.38% [9].

As we can see, several past works have been done on emotion recognition through various datasets like Cohn-Kanade, MMI, JAFFE, etc but, this paper proposes an application of emotion recognition process to be integrated to the android application. This will allow a totally exceptional user experience with on-spot emotion detection and recommendation of songs.

3 Proposed Work

The proposed model takes input, i.e. the facial expression of the user from the front facing camera of the mobile phone. The image is passed to the Emotion module which is stored in the form of .pb file. The preprocessing of the image takes place such that the image will be converted into Grey-scale 48 × 48-pixel image. According to the trained Emotion module, it determines the percentage of all emotions and takes the maximum emotion. Then, the feed function is called which outputs the predicted result. The output will be used to create a playlist of the songs from the database and stores it in the temporary buffer. Later, the temporary buffer will be used to

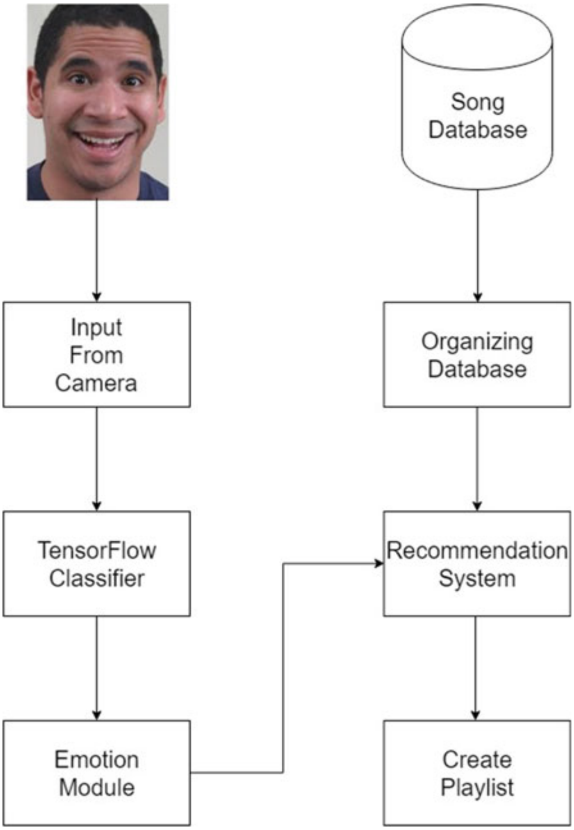


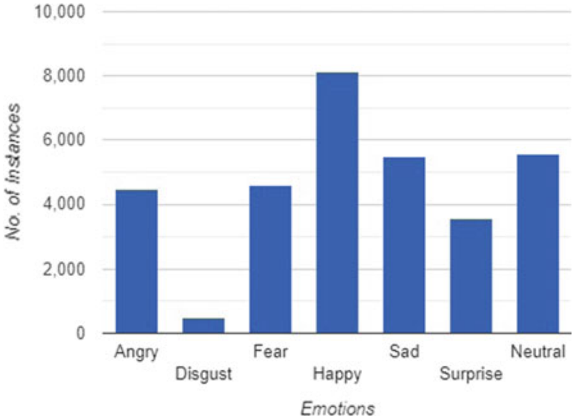
Fig. 1 Work flow of the proposed model [4]

create a playlist object in the android application. The graphical representation of the workflow is given in Fig. 1. In order to implement this, we consider the following modules.

3.1 Emotion Module

1. **Dataset**: FER refers to facial emotion recognition as this study deals with the general aspects of recognition of facial emotion expression. Fer2013 dataset has been taken from Kaggle Competition held in 2013 [10]. The dataset is already defined in such a way that label for each image is given either training or testing. So here there is no need for splitting of data manually. It has 28,709 training and 3,589 testing image information. Each image has a dimension of 48 × 48 pixel; thus, the whole dataset contains $48 \times 48 = 2304$ pixel values for each individual image. Target attribute gives the numerical value from 0 to 6 according to the emotion(0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

Fig. 2 Number of emotion instances



Dataset is not normalised as the ratio of number of each instance is different which can be observed in Fig. 2. For example, the number of Happy images is much more compared to other emotions as well as the number of Disgust images is very less compared to other emotions.

2. **Convolutional Neural Network**: A multi-layered convolutional neural network is programmed to evaluate the features of the user image [11, 12]. CNN is the deep learning algorithm that takes input, i.e. an image as an array of pixels which is in a range of 0–255. The matrix is passed through many layers of CNN to get a class score.

3. **Activation function ReLu**: It is widely used activation function in deep learning, what this function does is that it takes the maximum of input x and converts all negative values to zero, i.e. threshold at zero [13].

The Emotion module is implemented in Python programming language. This uses a sequential model of keras, and after building it, three 2D convolutional neural networks and a Fully connected neural network are added to process the FER dataset. In the first Conv2D() layer, the first argument passed is the number of output channels, i.e. 64 output channels with 5 × 5 as kernel_size and finally, it is supplied with the size of input layer (48, 48, 1). Next, a 2D Max Pooling layer is added with pool size (5, 5) and stride as (2, 2) in x and y direction, respectively. For the next convolutional layer again, a Conv2D() layer is added with 64 output channels with 5 × 5 as kernel_size with a 2D Max Pooling layer is added with pool size (3, 3) and stride size as (2, 2). Again in the third convolutional layer, a Conv2D() layer is added with 128 output channels with 3 × 3 as kernel_size with a 2D Max Pooling layer is added with pool size (3, 3) and stride size as (2, 2). Also, in all the above layers, Rectified Linear Unit (ReLu) is taken as the activation function. Now since convolutional layers are built, the output is flattened in the Fully connected layer by using Dense() layer and

20% of the data is dropped out. Figure 3 graphically represents the architecture of CNN. Here Softmax is used as the Activation function. At last, while compiling the model, Adam Optimizer is used with Categorical_crossentropy as loss functions. For training purpose, 20 epochs and 256 batch_size is used which gives an accuracy of 80.35% for the training set. Some of the computational examples of emotion detection using our Emotion module in python language are given in Fig. 4.

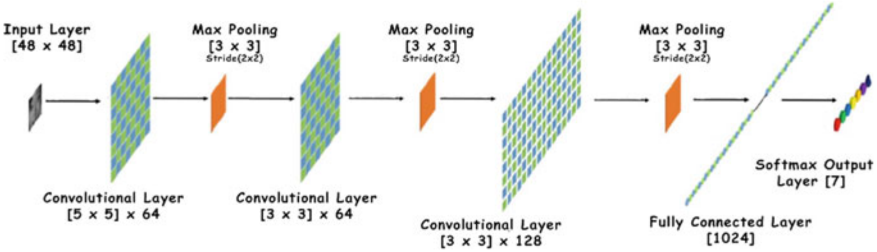


Fig. 3 Architecture of proposed CNN

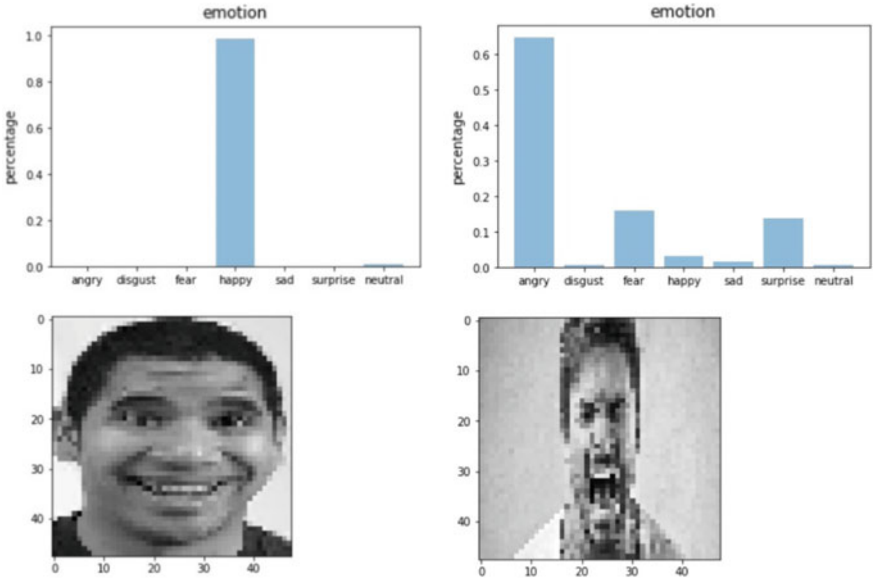


Fig. 4 Computational results for happy and angry

3.2 Android Module

A .pb (protobuff) file is created of the trained Emotion module, which can be copied to assets folder in the android app. Another file with the labels of the target values, i.e. 7 emotions are created. This file tells the android application about the possible outputs of the model. The android NDK is used to inference with the TensorFlow. The Java code then calls functions in our native library through the Java Native Interface (JNI) interface. The model is fed with an image either from the camera of the device or from the gallery. The TensorFlow Classifier then takes the image and converts it into a 48 × 48-pixel grayscale image and sends the images to the classifier. The classifier determines the output percentage for each emotion on the image. Then it selects the highest percentage and shows that as the emotion for the image.

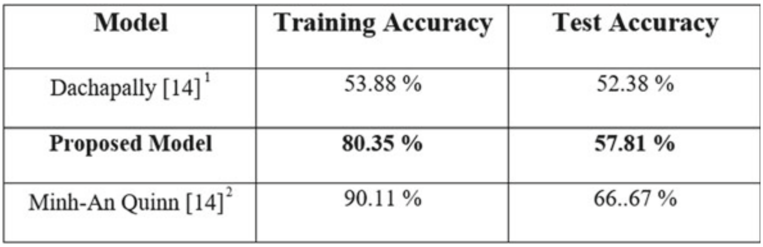
Once the emotion has been fetched and returned, the music player is active. In android, a Media Player class is used to handle all the operations regarding any media operations. Since the already trained module is used which is stored in .pb file. The time taken to recognise emotion is just one second. The songs used in the applications are all copyright free songs and are stored in the firebase cloud storage. The songs selected are in such a way that it would counter any negative emotions and would promote a positive emotion. Here for emotions 'Fear' and 'Disgust', we try to recommend spiritual and funky music, respectively, to counter these emotions. For other emotions, we recommend songs that suite their genre. Jolly for 'Happy', Slow beat for 'Sad', Heavy metal for 'Angry', Jazz/Country for 'Surprise' and Ambient for 'Neutral'. Every song stored has a unique URI (uniform resource identifier). The songs are already sorted into different emotions and stored in the cloud storage. Once the emotion of the image is determined, then the folder for that emotion is selected and a song is fetched at random from all the available songs. The song is fetched and stored in the buffer for the Media Player class for android. The buffer is used to create the playlist in the music player. Now talking about the creation of playlist, the current research on emotion-based recommendation system focuses on collaborative and content-based filtering. But here we use a combination of number of hits per song for each user and random generation, i.e. selecting songs randomly from the database to recommend and add it to the playlist of songs.

4 Result Analysis

An emotional facial recognition-based music player has been proposed and implemented using the android platform. The result of the proposed idea is highly promising. The quick response time of the application makes it suitable for real-time application. Although the dataset is not normalised and there are huge differences in the number of instances in emotions, we were able to achieve a remarkable accuracy of 80.35% while training the neural networks. Talking about testing, an accuracy of 57.81% was achieved while taking all 7 emotions into consideration.

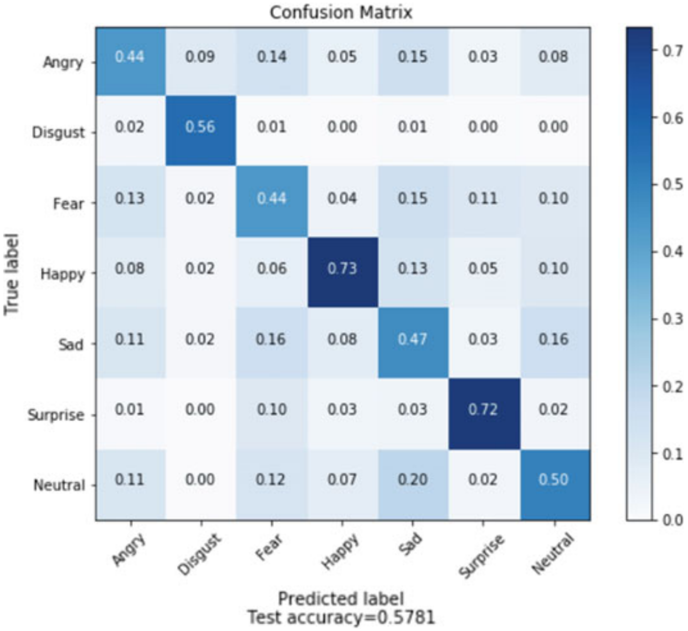
Figure 5 describes the comparison between three models, the proposed model and two of them taken from Minh-An's paper [14] on the same dataset. Dachapally [14][1] model is originally not trained on FER dataset, but this paper implements it on FER dataset, and results are given. Whereas the Minh-An Quinn [14][2] model has been implemented on FER dataset with some preprocessing and using Subtracting Mean concept to get higher accuracy. So our model stands in between, implemented with the original FER-2013 dataset without any preprocessing.

Figure 6 plots confusion matrix, which describes that our precision is high for Happy, Surprise and Disgust for test data. Figure 7 is the snapshot of Application developed, which predicts the current state of the user with the use of front facing camera.



Model	Training Accuracy	Test Accuracy
Dachapally [14][1]	53.88 %	52.38 %
Proposed Model	**80.35 %**	**57.81 %**
Minh-An Quinn [14][2]	90.11 %	66..67 %

Fig. 5 Comparison table



Confusion Matrix

True label \ Predicted label	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	0.44	0.09	0.14	0.05	0.15	0.03	0.08
Disgust	0.02	0.56	0.01	0.00	0.01	0.00	0.00
Fear	0.13	0.02	0.44	0.04	0.15	0.11	0.10
Happy	0.08	0.02	0.06	0.73	0.13	0.05	0.10
Sad	0.11	0.02	0.16	0.08	0.47	0.03	0.16
Surprise	0.01	0.00	0.10	0.03	0.03	0.72	0.02
Neutral	0.11	0.00	0.12	0.07	0.20	0.02	0.50

Fig. 6 Confusion matrix generated by the proposed model

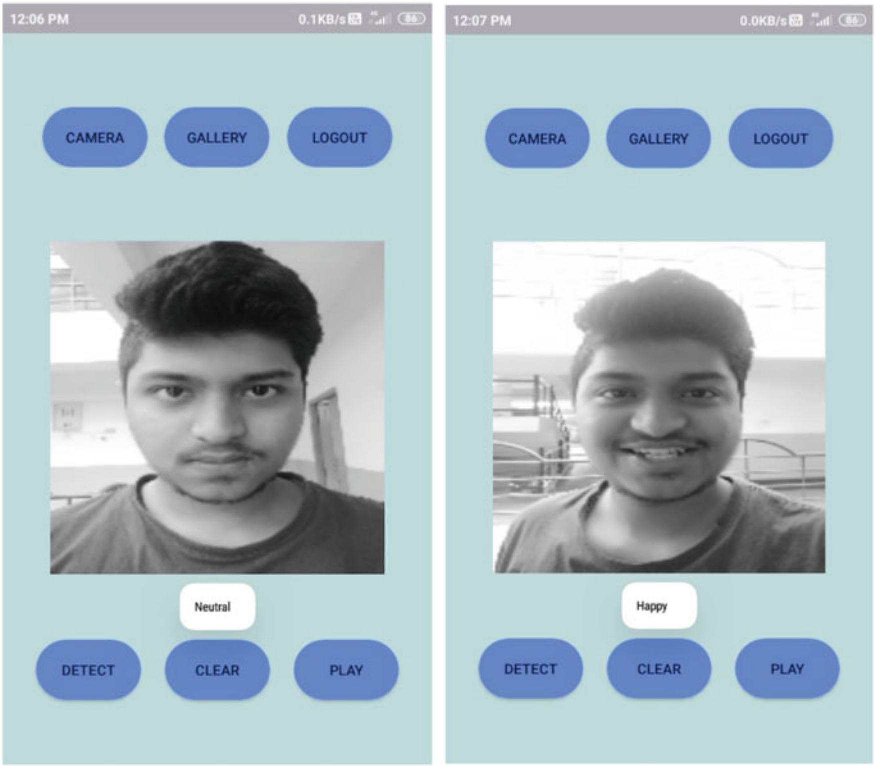


Fig. 7 Snapshots for neutral and happy emotion from the android application

5 Conclusion

We can conclude that our proposed module achieved a good accuracy of 57.81% while testing. Also, this ideology of real-time emotion recognition for prediction of songs seems to be the upcoming state of the art. We also recognise room for improvement. It would be interesting to see how the android module works for a dynamic song list instead of a fixed database. Also, we would like to build a recommendation system using content-based filtering to make it more robust. In the future work, we would like to extend our model to a dataset which provides colour images that will allow us to investigate the efficiency of pre-trained models such as AlexNet [15] or VGGNet [16] for facial emotion recognition such that it is compatible for mobile applications. We also plan to increase the accuracy of our emotion module.

Informed Consent Informed consent was obtained from all individual participants included in the study. We have used the author's photo in the result section.

References

1. Swaminathan S, Schellenberg EG (2015) Current emotion research in music psychology. Emotion Rev 7(2), Spp. 189–197, Apr
2. Thibeault CM, Sessions O, Goodman PH, Harris FC Jr. (2010) RealTime emotional speech processing for neurobotics applications. In: Proceedings of international conference computer applications in industry and engineering, Las Vegas, NV, USA, pp 239–244
3. Sebe N, Cohen I, Huang TS (2005) Multimodal emotion recognition. In: Handbook of pattern recognition and computer vision, World Scientific, pp. 1–23
4. Vink A Living apart together: a relationship between music psychology and music therapy. Nordic J Music Ther 10(2): 144–158
5. Emotion AI, Real-time emotion detection using CNN by Tanner Gilligan and Baris Akis
6. Burkert P et al Dexpressıon: deep convolutıonal neural network for expressıon recognıtıon. Arxıv:1509.05371v2, arxıv.org/pdf/1509.05371.pdf
7. Kim JH, Lee S, Yoo WY (2013) Implementation and analysis of mood-based music recommendation system. In: 2013 15th international conference on advanced communications technology (ICACT), PyeongChang, IEEE, pp 740–743
8. Tian Y-L, Kanade T, Cohn J (2000) Recognizing lower. Face action units for facial expression analysis. In: Proceedings of the 4th IEEE international conference on automatic face and gesture recognition (FG'00), Mar 2000, pp 484–490
9. Dachapally PR Facıal emotıon detectıon usıng convolutıonal neural networks and representatıonal autoencoder unıts. http://arxiv.org/abs/1706.01509
10. Goodfellow IJ et al Challenges in representation learning: a report on three machine learning contests
11. Lawrence S, Giles CL, Tsoi AC, Back AD (1997) Face recognition: a convolutional neural-network approach. IEEE Trans Neural Netw 8(1): 98–113, Jan
12. Kołakowska A, Landowska A, Szwoch M, Szwoch W, Wriobel MR (2014) Human-Computer systems interaction: back-grounds and applications. In: Emotion recognition and its applications, ch 3. Springer, Cham, pp 51–62
13. Scherer D, Muller A, Behnke S (2010) Evaluation of pooling operations in convolutional architectures for object recognition. In: 20th international conference on artificial neural networks (ICANN), Thessaloniki, Greece, September
14. Quinn M-A, Sivesind G, Reis G Real-time emotion recognition from facial expressions. CS 229 - Stanford University {minhan, gsivesin, greis}@stanford.edu
15. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems
16. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. http://arxiv.org/abs/1409.1556. arXiv preprint

Multi-document Text Summarization Tool

Richeeka Bathija, Pranav Agarwal, Rakshith Somanna, and G. B. Pallavi

Abstract In today's world, there is a massive amount of data being continuously generated every minute. This data can be utilised to gain a large amount of information that can have numerous uses. However, it is difficult to obtain this information because of the speed and volume of data being generated. One of the tools that can be useful in extracting useful information from textual data is a text summarization and analysis tool. Many text summarization tools are being developed but largely focus on summarising a single document effectively. This project aims to create a text summarization tool using abstractive and extractive text summarization techniques that can extract the relevant and important information from multiple documents and present it as a concise summary. The tool also performs multiple analyses on the data to obtain more useful information and make inferences based on the contents of the input textual data. This tool has various use cases as it can greatly reduce the time spent in gathering information from a large number of different documents such as surveys and feedback forms from various sources by providing an effective summary and analysis of the relevant data in these text documents.

Keywords Text summarizer · Abstractive summarization · Extractive summarization · ROUGE · BERT

R. Bathija (✉) · P. Agarwal · R. Somanna · G. B. Pallavi
B.M.S College of Engineering, Bengaluru, India
e-mail: richeeka@gmail.com

P. Agarwal
e-mail: singhal.pranav12@gmail.com

R. Somanna
e-mail: rakshithsomanna@gmail.com

G. B. Pallavi
e-mail: pallavi.cse@bmsce.ac.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_63

1 Introduction

With the advent of new fields like data science and big data, the world is clearly becoming more data-centric. It has become apparent that data about various business processes in organisations across almost all industries contains useful information that can be used to improve the organisation's performance and give it an edge over other competitors in that field. Therefore, it is becoming increasingly common that organisations look into data to obtain useful information like skill sets required by the employees to succeed in particular roles, trends in customer purchases and interactions, the risk associated with the projects being planned, etc. Organisations also obtain integral data through surveys and feedback forms presented to their own employees as well as to customers or the general public. The main challenge associated with using this data is that data is being generated at an incredible rate and volume constantly. A large amount of this data is not useful and it is very time consuming and requires a large amount of effort to extract the right information. This can be greatly reduced through a summarization tool built to generate a summary of only the useful information present in this large store of data.

A large amount of research is being done in various text summarization techniques. Text summarization can be widely divided majorly into two categories—Extractive and Abstractive Summarization. Summarization can also be used to make the process of conducting surveys better. Surveys tend to collect a lot of data and it takes time to find the results of the survey. People try to make surveys as objective as possible so that they are easier to analyse. This results in restricting the user's answers to the options that are provided by the surveyor. If we present more subjective type questions, we can extract more information from the user's answer. This project aims to help surveyors reduce the time taken to analyse subjective answers by using techniques of text summarization (Fig. 1).

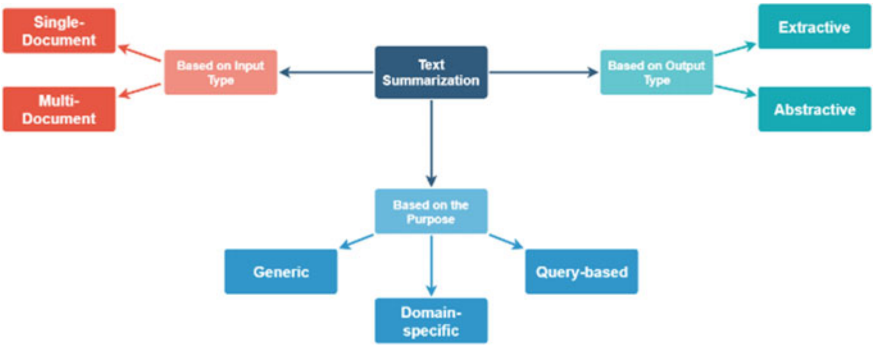


Fig. 1 Types of text summarization [1]

2 Related Work

There are two major types of text summarization [2]:

Extractive summarization—In this method, the most important sentences from a document are selected and included in the summary verbatim. The sentences are included from their parent document without making any changes. There are many different techniques that are used to perform extractive summarization. There are three main tasks involved. First, we construct an intermediate representation of the text. Next, we score the sentences based on the representation. And then we finally have to select the relevant sentences [3].

Abstractive summarization [4]—In this method, the document is studied and interpreted as a whole to understand its contents. The summary created consists of newly generated sentences that are not from the parent document but are still grammatically correct and convey the tone and meaning of the most important parts of the parent document. Abstractive summarization is generally modelled using Attentional Encoder–Decoder Recurrent Neural Networks [5].

Extractive summarization is a simpler process than abstractive summarization but it is more grammatically limited as the contents of the summary are restricted to sentences from the parent documents. For a summarization tool that takes multiple and varied documents as input, it is important to give weight to all of the documents in the final summary. Therefore, using an extractive approach to select sentences from all the documents and an abstractive approach to combine these sentences into a concise and coherent summary is the most appropriate approach.

Text summarization as a field is becoming more and more popular in terms of research as well as practical applications. Therefore, there are many significant developments in the tools and techniques of summarization.

Text summarization tools are being used in various fields including:

Media monitoring—to help deal with problems of information overload and content shock, summarization tools have been developed [6].

Inshorts—This is a news delivery app that selects the trending news stories domestically and internationally. The app obtains information about each story from the Internet and other news sources and summarises this information into a short 60-word article to cover each of the stories.

Financial research—Investment banks go through large amounts of market information to drive their decision making. This has led to the development of text summarises tailored to financial documents like earning reports and financial news. This can help analysts quickly determine market trends [7].

The summarization tools being developed have different architectures and features. Each of these has its own advantages and can be used in particular use cases. Some of the most common and successful text summarization models being used are

Sequence to sequence models—The most successful approach to abstractive text summarization has been the seq 2seq model. In this method, there is an encoder–decoder architecture. The encoder and decoder can be developed using Recurrent

Neural Networks (RNNs) and Long Short Term Memory (LSTM). The input document is first sent to the encoder, word by word, which processes the document and stores the contextual information in its hidden and cell states of its neural network. These states are passed to the decoder which generates the summarised text by predicting the next word in the summary using the previously generated words and the state information from the encoder network.

Structure-Based Approach—In this approach, there is a fixed structure for the final summary. The important information from the input text is selected and fit into the specified structure without altering its meaning to obtain the final summary. The structure can be (1) a template for the final summary, (2) a tree-based structure where similar sentences are grouped together in a single tree, (3) an ontology-based structure where different entities in the domain of input documents are modelled and the relationships between them are determined or other structures. The final summary can easily be obtained after the correct construction of any of these structures.

Semantic-Based Approach—In this approach, the input document is first converted into a semantic representation of itself. This representation is then fed into a Natural Language Generator that is used to create coherent sentences that represent the semantic information which creates the required summary. The semantic representation can be done as (1) multimodal semantic model where the representation is done as important concepts in the text and their relationships using ontology, (2) information item model in which the information items are obtained by performing a syntactical analysis of the text and their importance is noted, (3) semantic graph model where the document is initially represented as a rich semantic graph which is then reduced in complexity to a simpler semantic graph using certain heuristic rules. From any of these semantic representations, we can use natural language generators to obtain the final summary.

3 Proposed System

The proposed system is a generic tool that can be used to summarise and analyse multiple paragraphs. With small variations, it can be fit into various use cases and be applied effectively for different kinds of users.

The system can be divided into 3 major subsystems:

1. React app for the front end.
2. The backend services—RESTful exposed via an API gateway.
3. Database layer.

The front end makes API calls to the API gateway using HTTP. The backend services are independent, each with its own database. This separation of concerns aligns with the microservices architecture that is proposed for this tool (Fig. 2).

The main functionalities provided by this tool are

1. User authentication—This service provides the necessary user authentication feature. It is necessary to distinguish the types of users—Surveyor, Respondent.

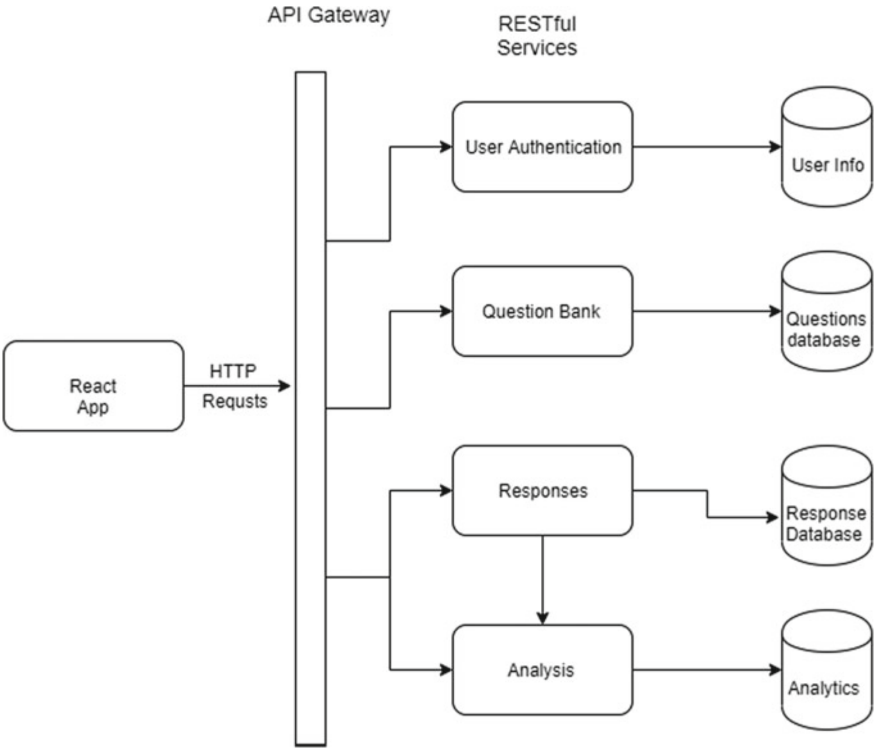


Fig. 2 Microservices architecture

2. Survey Creation—This service enables surveyors to create new surveys with the required questions. The surveyor can also select any of the suggested questions to add to his survey.
3. User responses—Each of the users are sent a link through which he can answer the questions selected for him by the surveyor. The responses are collected and stored in the database.
4. Summarization—The responses recorded for the surveys are summarised and displayed. The surveyor can also view any from a list of contextualised summaries to gain more information. This will be done using the techniques of multi-document text summarization [8]. In our application, we will first have to perform multi-document text summarization. There could be information overlap between the documents since some of the user's answers to a particular question could be similar. We aim to use the technique of Multi-Sentence Compression (MSC) to solve the above problem [9]. Another technique that we can use to combine the sentences is ILP-Based Multi-Sentence Compression. This approach is better than MSC because it prevents redundant information from being included in the summary. It uses an inter-sentence redundancy constraint and hence generates better summaries that are more informative [10].

This tool can then be modified to meet user's requirements in various ways including

1. Surveys and Feedback forms for end-users—Surveys are very commonly used to obtain the opinions of various people through the same questions and useful information can be gleaned by going through all the answers. For example, organisations can learn a lot of useful information about the good features and flaws in their products by obtaining feedback from customers through surveys. The various responses to a survey can be provided to the summarization tool as an input. The tool can then provide an overall summary of what the opinion is based on the received answers as well as useful analytics on the answers provided such as the trends in customer behaviour. Thus, the useful information contained in a large number of answers can be quickly obtained and used to make decisions. There is no need to spend time analysing various individual responses.
2. Training needs analysis—Organisations often need to determine the skills possessed by their employees and what they need to be trained to improve the productivity of the organisation. This is done by making questionnaires and conducting interviews with the employees to obtain data which is then analysed to determine the training needs of employees in various roles throughout the organisation. This process requires a large amount of effort and time which can greatly be reduced by using the summarization tool. The answers given by the employees to the questionnaires and interviews can be given to the tool as input. The tool can then analyse and summarise all the answers to provide the final training requirements—the employees that need to be trained, what they should be trained on and how long the training process should take. We can also perform some sort of sentiment analysis of the user responses. There are many techniques that can be used to perform sentiment analysis as mentioned in [11]. Therefore, the tool can be used to greatly simplify and quicken the training needs analysis process that is performed at various organisations.
3. BERT: Bidirectional Encoder Representations from transformers. It is an open-source tool by Google which specialises in NLP tasks like question answers and natural language inferences. The unique feature of this tool is that it views a sentence in two directions from left to right and from right to left in contrast to all previous approaches. This new approach has shown significant improvement in accuracy [12].

BERT is bidirectional in the sense that it processes an entire sentence at once rather than the traditional sequential approach. Hence it can also be called as non-directional. BERT makes use of attention mechanism. The attention mechanism takes two sentences and relates parts of one sentence with another and it finds out which parts of a sentence correlate with which parts of another sentence. This mechanism is useful in machine translations. However, if a sentence is related to itself, i.e. if the correlation between a part of a sentence is found with respect to every other part of the same sentence, then it is a bidirectional model. The BERT model has the biggest set of parameters and is trained on the biggest corpus amongst its competitors. This model shows a significant increase in accuracy with an increase in training steps.

The converging time for this model is slower compared to its competing models but it outperforms them with some basic steps of pre-training.

4 Result Evaluation Methods

With the increase in popularity of text summarization techniques, the importance of the evaluation methods of summarization has also increased. The evaluation methods aim to determine the quality of summarization. The different types of quality metrics are [13]

1. Extrinsic Evaluation: It is a metric to identify how useful the summarization output is for the execution of other tasks like answering questions, relevance assessment and sentiment analysis.
2. Intrinsic Evaluation: The quality is determined internally during the process of summarization. Intrinsic evaluation can be used to determine how informative summarization is and the quality of summarization.

The most recent and popular tool to evaluate how informative the summarization is Recall Oriented Understudy of Gisting Evaluation (ROUGE) [14].

ROUGE-N: computes the length of the common unigrams, bigrams or n-grams between the output and a reference output or the original text.

ROUGE-W: is a weighted longest common subsequence that is based on the longest common subsequence approach but gives more weightage to sequences with close proximity.

F-score: is a comparative analysis between an ideal summary and the summarization output.

ParaEval: evaluation method for paraphrases in the summarization output.

5 Conclusion

Our project aims to help users perform and analyse surveys in an easy and effective manner. The project uses NLP techniques to summarise multiple survey responses to provide one single response that includes all the key points from all the surveys. This makes it easier for the surveyor to view the results of the survey. The surveyor does not have to spend a lot of time reading each and every answer. We also provide many different types of analysis that can be useful to the surveyor. This project makes use of the concept of multi-document text summarization to make the whole process of conducting and analysing surveys very easy.

Some of the future enhancements involve

Automatic generation of surveys: Based on the user's previous surveys, and the type of survey the user requires, templates for surveys can be suggested automatically. This makes the process of creating new surveys much easier.

Automated Training Needs Analysis for companies: TNA is a very important process in many companies. This process involves conducting multiple surveys and interviews to gauge the training needs of the employees. This type of analysis generally has one specific end goal in mind that needs to be accomplished. The TNA consultant then has to draft surveys and questionnaires while keeping the end goals in mind. The responses of surveys and questionnaires need to be analysed and a Needs Analysis document must be generated. This document is then used to create a training plan. The whole process of TNA can be automated by using this tool. We can even try to automate the generation of questions if some documents like the Business Process Documents are available that have the information of the process flow.

References

1. Unsupervised text summarization using sentence embeddings. https://medium.com/jatana/unsupervised-text-summarization-using-sentence-embeddings-adb15ce83db1
2. Andhale N, Bewoor LA (2016) An overview of text summarization techniques. In: 2016 international conference on computing communication control and automation (ICCUBEA), Pune, 2016. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7860024&isnumber=7859963
3. Towards automatic text summarization: extractive methods. https://medium.com/sciforce/towards-automatic-text-summarization-extractive-methods-e8439cd54715
4. Modi S, Oza R (2018) Review on abstractive text summarization techniques (ATST) for single and multi documents. In: 2018 international conference on computing, power and communication technologies (GUCON), Greater Noida, Uttar Pradesh, India, 2018, pp 1173–1176. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8674894&isnumber=8674884
5. Yeasmin S, Tumpa PB, Nitu AM, Uddin MP, Ali E, Afjal MI (2017) Study of abstractive text summarization techniques. Am J Eng Res 6(8): 253–260
6. Sethi P, Sonawane S, Khanwalker S, Keskar R (2017). Automatic text summarization of news articles, pp 23–29. https://doi.org/10.1109/bid.2017.8336568
7. Filippova K, Surdeanu M, Ciaramita, M, Zaragoza H (2009) Company-oriented extractive summarization of Financial News, pp 246–254. https://doi.org/10.3115/1609067.1609094
8. Mohod R, Kamble V (2018) A literature study on different multi-document summarization techniques. In: 2018 2nd international conference on trends in electronics and informatics (ICOEI), Tirunelveli. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8553936&isnumber=8553678
9. Nayeem MT, Fuad TA, Chali Y (2018) Abstractive unsupervised multi-document summarization using paraphrastic sentence fusion. In: Proceedings of the 27th international conference on computational linguistics, August. https://www.aclweb.org/anthology/C18-1102
10. Banerjee S, Mitra P, Sugiyama K (2015) Multi-document summarization using ILP based multi-sentence compression. In: Twenty-Fourth international joint conference on artificial intelligence (IJCAI)
11. Chakraborty K, Bhattacharyya S, Bag R (2020) A survey of sentiment analysis from social media data. In: IEEE transactions on computational social systems. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8951256&isnumber=6780646
12. Devlin J, Chang M-W, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. http://arxiv.org/abs/1810.04805v2 [cs.CL] 24 May 2019

13. Steinberger J, Jeˇzek K (2009) Evaluation methods for text summarization. Comput Inf 28: 1001–1026, 2 Mar
14. ROUGE—Tool to evaluate summarization. http://kavita-ganesan.com/what-is-rouge-and-how-it-works-for-evaluation-of-summaries/#.Xcz0IlczbIU

Cognitive Computing Technologies, Products, and Applications

N. Divyashree and Prasad K. S. Nandini

Abstract Cognitive computing has made industries and business organizations to operate in a different paradigm with respect to the use of technology right from carrying business operations to high-level decision-making strategy. The ability of human experts in any field to think and make right decisions varies from person to person which creates the demand and necessary requirement of a high skilled person in an industry, but it becomes difficult for any human when it comes to obtaining useful insights to carry out business operations and to take right decisions from a huge amount of data that gets generated every day. Different technologies and platforms are necessary to process almost petabytes of data and make proper use of it to obtain patterns and insights.

Keywords Cognitive science · Cognitive computing · Artificial intelligence · Machine learning

1 Introduction

The word Cognition is obtained from the Latin word Cognosco ("con" means with "gnosco" means know) which means "conceptualize" or "recognize". Cognitive science is a study of the interdisciplinary field [1] (cognitive psychology, cognitive linguistics, biology, psychological science, computer science, and neuroscience) on the human brain (mind and intelligence). Human cognition involves memory, perception, concept formation, concept association, concept recognition (involves pattern recognition), consciousness, and mental ability to understand and solve any problem. For instance, the way a person tries to remember a learned concept in different ways and tries to link what is learnt to any previously encountered situations or examples.

N. Divyashree (✉) · P. K. S. Nandini
Dr. Ambedkar Institute of Technology, Bengaluru, India
e-mail: ndivya.shree17@gmail.com

P. K. S. Nandini
e-mail: nandiniks1@dr-ait.og

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_64

Cognitive computing [2] is a discipline under science which does all the effort with a theme of human psychology and mental ability behind it. Cognitive computing is yet another subfield under Artificial Intelligence (AI) that makes use of Machine Learning (ML), Natural Language Processing (NLP), Human–Computer Interaction (HCI), sentiment analysis, vision, Artificial Neural Networks (ANN), and big data technologies to build a successful, self-learning information processing models that simulate human cognition and intelligence.

The goal of cognitive computing systems is not to substitute humans but to mimic the human thought process along with an extraordinary capability of processing complex and the large amount of data that no human brain can process or retain so as to assist human experts in the better decision-making process. Cognitive computing tries to explore and implement human cognition concepts to solve vast problems through penetration into the complexities of big data. Cognitive Computing Consortium (CCC) has mentioned the following features that describe/identify a cognitive system [3].

1. **Adaptive**: Cognitive systems should have the ability to learn and adapt in real time be it the data gathering or understanding goals as surrounding environment change in order to get desired results.

2. **Interactive**: Cognitive systems must be able to understand the requirements of humans, take input from humans, and interact with its own system components like processors, interconnected devices, and services to provide appropriate results.

3. **Iterative and Stateful**: Cognitive systems must be able to retain and use previously encountered knowledge/situation and use it on present problems statements if it occurs similar or incomplete.

4. Contextual: Cognitive systems must be able to identify the requirements and extract only necessary information from multiple sources of any data types (structured, unstructured, semi-structured data) that suits the problem context such as syntax, meaning, time, place, or domain.

This paper tries to explain the idea behind Cognitive computing areas where Cognitive computing can be and are successfully implemented, about different cognitive computing companies, their products and solutions for business that requires sloution through cognitive computing.

2 Technologies That Aid Cognitive Computing

2.1 Artificial Intelligence

AI is a study on how to mimic human intelligence through machines. AI can be classified into three different levels:

- Narrow AI,

- General AI, and
- Strong AI

A system is said to exhibit AI if it has the capabilities of problem-solving and learning abilities from examples.

2.2 Machine Learning

ML is a branch under AI, where an algorithm/program learns from historical and existing data without being explicitly programmed called Machine Learning Models. ML algorithms.

Individual or combinations of these algorithms are used to build models to perform complex tasks like prediction forecasting, analytics, estimation, etc.

2.3 Natural Language Processing

NLP deals with deciphering unstructured data to extract, understand, analyze, and process meaningful information to drive intelligent solutions. Different ML algorithms are employed to build NLP systems.

NLP is used in several areas like speech recognition, sentiment analysis, customer service, advertisement, text summarization, text analytics, social media monitoring, etc. Trends that impact NLP are machine learning algorithms, Deep learning, semantic search, and Cognitive communication.

2.4 Computer Vision

Computer vision deals with the automation of high-level understanding and information gathering from digital videos and images. According to British Machine Vision Association and Society for Pattern Recognition (BMVA) [4], "Computer Vision is concerned with the automatic extraction, analysis and understanding of useful information from a single image or sequence of images". Computer vision is used in many fields like

• Biometrics	• Agriculture
• Image restoration	• Forensics
• Robotics	• Transport and many more
• Augmented reality	• Facial recognition
• Security and surveillance	• Autonomous vehicles

2.5 Human–Computer Interaction

Human–Computer Interaction (HCI) deals with designing of technologies and interfaces that assist interaction between human and computer in a novel way and also tries to mimic human–human interaction as well. According to Association for Computing Machinery (ACM), "Human–Computer Interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them". Examples of HCI include

• Graphical User Interface(GUI)	• Natural language question and answering
• Voice User Interface (VUI)	• Google voice search
• Facial recognition systems	

3 Cognitive Computing Companies and their Products

3.1 Sparkcognition

Sparkcognition is a leading AI solution company [5]. It has provided many software solutions for customers namely, Analysis of complex data stores, providing actionable insights, Automation and identification of optimal responses, and Infrastructure protection against cyber threats.

i. *Areas where Sparkcognition has provided solutions*

• Aviation	• Financial services
• Maritime	• Manufacturing
• Defense	• Energy and utilities
• Cybersecurity	• Federal industry
• Oil and gas	• Telecommunication

ii. *Products/solutions of Sparkcognition*

• SparkPredict	• DeepNLP
• DeepArmor	• Darwin

3.2 SparkSecure

SparkSecure is a Cloud-based platform developed by Sparkcognition [6] which augments and optimizes cybersecurity team performance. SparkSecure is an innovative approach that combines cognitive analysis with big data and large security corpus through IBM Watson. SparkSecure automatically collects large amounts of security data of both structured and unstructured data in the form of natural language or in the form of server logs through IoT (Internet of Things)and applies its patented Machine Learning algorithms for Cognitive analysis to understand about the data and performs detection of insider threats, malicious software and identifying bot traffics in weblogs.

3.3 Mindfabric

Mindfabric is yet another Cognitive security analytics platform [7] developed by Sparkcognition for protection against cyber threats. Industries that are connected to mobile devices and servers in the cloud are very vulnerable to cyber threats that no humans can possibly contend due to its diverse nature. Mindfabric collects digital information (be it the infrastructure security data or real-time data through IoT devices), and automatically builds cognitive models to learn from data and project/predict outcome (be it for constantly updating security policies, prevention of disasters before occurrence, or prediction of system failure) into the future.

3.4 Microsoft Cognitive Services

Microsoft Cognitive Services provides APIs (Application Programming Interface) with inbuilt AI algorithms for apps, bots, and websites [8]. Microsoft Cognitive Services has a collection of around 25 tools (collection of AI algorithms and APIs) that allows developers to include various features like sentiment detection, emojis, vision, etc., to the applications with no requirement of prior knowledge on machine learning. E.g., Skype, Crotona, and Bing provide features like conversation translation, understanding of spoken words, and intent.

Azure [9] is a cloud-based platform developed by Microsoft for managing applications and services by providing Infrastructure as a Service (IaaS), Software as a Service (SaaS), and Platform as a Service (PaaS).

i. *Cognitive Services provided by Azure*

• Vision	• Language

(continued)

(continued)

• Speech	• Search
• Knowledge	• Anomaly detection

3.5 IBM Watson

IBM Watson is a supercomputer named after its founder Thomas J. Watson. Watson is a cloud-based suite of applications, enterprise-ready AI services and tools [10]. Watson has several applications that apply Cognitive computing as the underlying technology. IBM has provided AI-based solutions in several fields like advertising, IoT, health, customer engagement, financial services, education, Media, collaborated solutions for teamwork, and talent management solutions.

i. *Some APIs developed by IBM Watson*

• Watson assistant	• Natural language understanding
• Watson openScale	• Personality insights
• Visual recognition	• Watson studio and many more
• Tone analyzer	

3.6 Numenta

Numenta [11] is a company that works on machine intelligence technology and applications based on neocortex principles. Numenta has developed an open-source project called NuPIC (Numenta Platform for Intelligent Computing) using an underlying technology called HTM (Hierarchical Temporal Memory).

i. *Commercial products created by Numenta*

- Grok (for anomaly detection, and for reduction of downtime in business)
- Cortical.io (natural language processor)
- Hierarchical Temporal Memory (HTM) an open-source technology inspired by the neocortex.

ii. *Some of the example applications available under NuPIC*

- Stock monitoring
- Rogue behavior [12]
- Geospatial tracking [13]

3.7 Expert System

Expert System [14] is a company that develops Text analytics software and Cognitive computing software using AI algorithms to explore and understand the intended meaning of written language and provides solutions like insights gaining, better decision-making, customer engagement, automation of processes that are information intense, and mitigation of operational risks. Solutions provided by Expert System are Knowledge management, Corporate intelligence, Entity extraction software, Cognitive automation, and Automatic classification.

i. *Products developed by Expert Systems*

• Cogito cognitive technology	• Cogito studio
• Cogito intelligence platform	• Cogito for underwriting
• Cogito discover	• Biopharma navigator
• Cogito claims	• API and Integrations
• Cogito answers	

3.8 Cisco Cognitive Threat Analytics

Cisco Cognitive Threat Analytics is a cloud-based service [15] to detect malicious attacks and suspicious web-based traffic for the websites. Cognitive threat analytics is able to analyze nearly ten billion web requests per day and zeroes in malicious activities using statistical models and machine learning models for analyzing web traffic and device behavior over networks.

3.9 HPE Haven OnDemand

Haven OnDemand [16] is introduced as Machine learning as a service by HPE (Hewlett Packard Enterprise) on Microsoft Azure as a Cognitive computing service. Haven OnDemand provides services and collection of nearly more than sixty APIs that are developed using machine learning algorithms for data scientists and enterprise application developers to extract and analyze multiple data formats from big data in a faster and easier way for building data-rich applications.

3.10 CognitiveScale

CognitiveScale [17] is an AI-powered cognitive foundation company and is first in the world to release a cognitive platform called Cortex. Cortex software augments machine intelligence to human intelligence to enhance business operations and customer experience by simplifying business processes including development, deployment, and management by provision cognitive insights for business actions.

i. *Products by CognitiveScale*

- Engage AI
- Amplify AI

3.11 Deepmind

Deepmind [18] is a leader company in the world for its research in Artificial intelligence and its applications. Deepmind has built many relationships-time applications using AI technology.

i. *Projects of Deepmind*

• Deepmind health	• Games
• Deepmind for Google	• Differential neural computer
• Deepmind ethics and society	• Deep Q-Network

4 Application Areas of Cognitive Computing

• Law firms	• Personal shopping bots
• Industrial sector	• Customer support bots
• Financial sector	• Travel agents
• Education	• HealthCare
• Customer behavior analysis	• Security

These are some of the application areas where cognitive computing is currently being applied, but not limited to.

5 Conclusion

SparkCognition, SparkSecure, Mindfabric, Microsoft Cognitive Services, IBM Watson, Numenta, Expert System, Cisco Cognitive Threat Analytics, HPE Haven OnDemand, CognitiveScale, and Deepmind are the main companies that are currently contributing to cognitive computing, but not limited to. Cognitive computing technologies can be used to only those areas which generate a large amount of data and cognitive decisions to be made out of it.

References

1. Chen M, Herrera F, Hwang K (2018) Cognitive computing: architecture, technologies and intelligent applications. IEEE Access 6:19774–19783. https://doi.org/10.1109/ACCESS.2018.2791469
2. https://www.predictiveanalyticstoday.com/what-is-cognitive-computing/
3. https://cognitivecomputingconsortium.com/
4. http://www.bmva.org/visionoverview
5. https://www.sparkcognition.com
6. https://spark.apache.org
7. https://mind-fabric.com
8. https://azure.microsoft.com/en-in/services/cognitive-services
9. https://stackify.com/azure-cognitive-services-2
10. https://www.jenunderwood.com/2017/03/28/ibm-watson-cognitive-computing/
11. https://numenta.com
12. https://numenta.com/assets/pdf/whitepapers/Rogue%20Behavior%20Detection%20White%20Paper.pdf
13. https://numenta.com/assets/pdf/whitepapers/Geospatial%20Tracking%20White%20Paper.pdf
14. https://www.expertsystem.com/company
15. www.cisco.com/go/cognitive
16. https://www.havenondemand.com
17. https://www.cognitivescale.com
18. https://deepmind.com/

Breast Cancer Prognosis Using Machine Learning Techniques and Genetic Algorithm: Experiment on Six Different Datasets

S. Jijitha and Thangavel Amudha

Abstract The strategy used in this research is to select the best features using the genetic algorithm from various breast cancer dataset for getting better prediction results using machine learning algorithms. This research involves two main phases. One is feature selection using Genetic Algorithm (GA) and second is breast cancer prediction using Logistic Regression (LR) and k-Nearest Neighbor techniques (k-NN).

Keywords Genetic algorithm · Feature selection · Machine learning · Breast cancer prediction · Logistic regression · k-NN

1 Introduction

Breast cancer is one of the diseases which cause a number of deaths every year across the world. Early detection of such type of disease is a challenging task in order to reduce several deaths. Various techniques of machine learning and data mining are used for medical diagnosis, supported by prediction in the fields of chronic diseases like cancer. For prediction; most possible datasets should be considered because of the unique features each dataset holds to predict cancer, with good accuracy. This is only possible with critical features selection from intricate Breast cancer datasets. Machine learning (ML) with Genetic Algorithm (GA) is the methodology of prediction and selection used in this research work. The objectives of this research work are,

- To develop a Genetic Algorithm based selection framework to select the best features from various benchmark datasets to predict breast cancer.

S. Jijitha (✉) · T. Amudha
Department of Computer Applications, Bharathiar University, Coimbatore 641046, Tamil Nadu, India
e-mail: jijithasivan@gmail.com

T. Amudha
e-mail: amudhaswamynathan@buc.edu.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_65

Table 1 Breast cancer datasets

Datasets	No. of attributes	No. of instances	No. of class/diagnosis
Breast Cancer Wisconsin–Diagnosis dataset(BCWD)	32	569	2
Breast Cancer Wisconsin–Original dataset (BCWO)	11	699	2
Breast Cancer Wisconsin– Prognosis dataset(BCWP)	34	198	2
ISPY1 clinical trial dataset (ISPY1)	17	169	2
Breast cancer dataset (BCD)	5	569	2
Breast Cancer Coimbra Dataset (BCCD)	10	116	2

- To apply the machine learning algorithms, Logistic Regression (LR) and k-Nearest Neighbor (k-NN) for breast cancer prediction from the features selected by GA and to evaluate their performance.

The multiplicity of features in a dataset is one of the ultimatums in the diagnostic system. Irrelevant and redundant features can increase the confusion in classification algorithms and leads to inaccuracy [1, 2]. A method to deal with this challenge is feature selection [3, 15]. Table 1 display the datasets used for this research work, which includes breast cancer survival datasets, breast cancer tumor identification datasets, follow-up (recurrence or non-recurrence) datasets and blood result analysis dataset for breast cancer. This model evaluated on Wisconsin breast cancer databases, Breast Cancer Coimbra dataset, Breast cancer dataset and ISPY1 clinical trial dataset.

2 Review of Literature

Aalaei et al. [3] addressed a comparison for feature selection especially for breast cancer diagnosis by using a wrapper method with GA-based on feature selection and PS-classifier. The dataset used for this work is Wisconsin breast cancer datasets. PS-classifier, artificial neural network (ANN) and genetic algorithm based classifier (GA-classifier) are the 3 classifiers used for evaluating the usefulness of the anticipated feature selection method for 3 different datasets. Wisconsin diagnosis breast cancer (WDBC), Wisconsin breast cancer dataset (WBC) and Wisconsin prognosis breast cancer (WPBC) are the 3 datasets. The comparison results for all the 3 datasets were produced. The result concluded that feature selection can outdo the specificity and sensitivity, the accuracy of the 3 classifiers.

Pawlovsky et al. [4] draw up a paper on the prognosis of breast cancer using k-Nearest Neighbors (kNN). Dataset from UCI repository is used for deriving 73% accuracy on the reappearance of cancer.

Agarwal and Saxena [5] published an article on machine learning and effective models for finding out cancer tumors. Basic steps are followed to prepare a very strong machine learning program to spot malignant or benign tumor using Python and its open source libraries. Logistic Regression and KNN classifier are used for this purpose.

3 Materials and Methods

3.1 Dataset Description

- Breast Cancer Wisconsin-Diagnostic dataset (BCWD)

This benchmarked dataset is available in UCI ML Repository [6, 7]. It is a quantitative dataset with features of breast masses. The dataset includes 30 features and 569 instances and one binary classification variable to diagnose Malignant or Benign tumor from the breast mass. Dr. Wolberg, Street and Olvi are the creators of this dataset. The attributes which are distributive were obtained from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe the properties of the cell nuclei present in the image. The case diagnosis includes 357 cases of benign breast tumor and 212 cases of malignant breast cancer. This is a tumor identification dataset.

- Breast Cancer Wisconsin—Original dataset (BCWO)

Benchmarked dataset available in UCI ML Repository [6–9]. Original Wisconsin Breast Cancer Database has the number of instances of 699 and features is 10. Dr. William H. Wolberg is the creator of this dataset. This dataset reflects the chronological grouping of data from 1981 to 1991. It represents the numerical details of clump thickness, cell size and shape, nuclei etc. and one diagnosis column to find Malignant or Benign tumor. This is a tumor identification dataset.

- Breast Cancer Wisconsin—Prognostic dataset (BCWP)

Prediction models based on these predictors can potentially be used as a biomarker of breast cancer if they are accurate. This dataset predicts the presence and absence of cancerous cells.

- Breast cancer data set (BCD)

This a subset of breast cancer Wisconsin (Diagnostic) dataset with 5 attributes for diagnosing Malignant or Benign [10, 11]. This dataset also identifies the tumors.

- ISPY1 clinical trial dataset (ISPY1)

The I-SPY TRIAL (Investigation of Serial Studies to Predict Your Therapeutic Response With Imaging and Molecular Analysis) are changing the way new treatments are developed for breast cancer, making available new, better and more personalized treatments, faster. All data for the 222 patients treated for breast cancer in the ISPY-1 clinical trial was obtained from the cancer imaging archive and the Breast Imaging Research Program at the University of California at San Francisco (UCSF). To facilitate the dissemination and reproducibility of this analysis, the raw data and all code were posted at the websites Data.World and Github which are available under an MIT license [12]. It has the survival status as the outcome and 17 other attributes. This dataset has the features of the survival data from breast cancer patients.

3.2 Feature Selection Using a Genetic Algorithm

The working of the Genetic Algorithm reflects in the process of natural selection for finding the fittest individuals selected for reproduction in order to generate offspring for the next generation. GA is search-based algorithms since they search for the best gene for the next generation from the population.

Genetic Algorithm is used to select the best subset of features in the preprocessing step to improve the quality of the final result. A criterion to get the candidate solutions and fitness values with larger values will be considered. These solutions are considered as *population* in GA. An *individual* in the population is each solution and these individuals will have the best fitness values. Then these individuals are grouped together to generate off-springs at random which in-turn makes the next population. Behind the process of making next generation, there are two processes called *cross-over* and *mutation*. For every iteration, a better solution will be created, maybe less good than the previous one in certain cases.

In feature selection prediction, to identify whether a feature is included or not in the feature subset is represented binomial that is by binary values. The values for fitness could be the appraisal of the performance of a model or the accuracy of classification [13]. Table 2 shows the operators and parameters used for the feature selection using genetic algorithm.

3.3 Breast Cancer Prediction Models

The expansion of computer programs which can access data and use it to learn by themselves is where ML mostly focuses on. The first step in ML process is

Table 2 Genetic algorithm parameters

Sl. no	GA parameters	Values
1	Population	Based on the data size
2	Number of generation	10–50 generations
3	Selection	Tournament selection
4	Crossover	One point crossover
5	Mutation	Bit flip mutation
6	Termination criteria	End of the iteration

data collection. It could be by observations or direct or indirect experiences or by instructions given by the programmer. After collecting the data, the next step is to study the data to find patterns and based on these patterns; learn to make better decisions in the future. This is where the computers learn by themselves, without any external interventions or support or assistance. The computers will learn to take and adjust their own actions according to the situations. Two types of ML algorithms are often catalogued as supervised or unsupervised [14]. Supervised ML can be applied when the system has to learn from the past and apply it to the presently available data with labeled samples for predicting future events [16, 17]. The two machine learning algorithms, used form comparison, are Logistic Regression and k-Nearest Neighbor.

Logistic Regression

Even though Logistic Regression (LR) is a sub-part of linear regression, in Python it is considered as a classification algorithm in machine learning. LR is a statistical technique to analyze a dataset which has one or more independent objective variables that could decide the end result. The outcome of the prediction probability of LR could only be binomial distribution. The result is based on the dichotomous dependent variable. Binomial predictions can only have two values- 1 or 0 / True or false. So it is used in the area of categorizing alive/dead, win/lose, pass/fail, spam/not spam, healthy/sick and so on.

k-Nearest Neighbor

The k-Nearest Neighbor algorithm is very easy to carry but can perform extremely complex tasks at hand. For a dataset, k-NN can forecast the assessment of a variable of interest for each element of a target. The plurality vote of its neighbors is considered by k-NN while classifying an object. On gaining efficient votes then the object will be assigned simply to the class of that single nearest neighbor.

For evaluating classification accuracy, metrics of accuracy, sensitivity and specificity have been calculated by the use of confusion matrix [18]. They are calculated from

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{1}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100 \tag{2}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100 \tag{3}$$

4 Results

The strategy discussed above was applied to six different types of datasets. Table 3 displays the relevant features selected using genetic algorithm. Table 4 displays the Accuracy of LR and k-NN (in %), with and without feature selection. Table 5 shows the sensitivity and specificity of LR and k-NN, with and without feature selection.

Table 3 Selected features by genetic algorithm

Datasets	No. of selected features
Breast Cancer Wisconsin—Diagnosis dataset (BCWD)	12
Breast Cancer Wisconsin—Original dataset (BCWO)	7
Breast Cancer Wisconsin—Prognosis dataset (BCWP)	19
ISPY1 clinical trial dataset (ISPY1)	7
Breast cancer data set (BCD)	3
Breast Cancer Coimbra Data Set (BCCD)	3

Table 4 Accuracy of LR and k-NN

Datasets	Accuracy			
	Without feature selection		With feature selection using GA	
	LR	k-NN	LR	k-NN
BCWD	96.5	79.02	98.24	95.1
BCWO	97.83	95.6	99.27	98.55
BCWP	71.42	76.27	86.2	78.33
ISPY1	92.85	95.23	97.05	97.61
BCD	91.98	92.39	93.85	94.73
BCCD	75.86	54.16	79.16	58.33

Table 5 Sensitivity and specificity of LR and k-NN

Datasets	Sensitivity				Specificity			
	Without feature selection		With feature selection using GA		Without feature selection		With feature selection using GA	
	LR	k-NN	LR	k-NN	LR	k-NN	LR	k-NN
BCWD	98.64	96.59	99.73	97.68	89.09	50.9	95.23	92.72
BCWO	98.72	98.42	99.46	100	96.66	91.02	97.87	96.66
BCWP	83.33	93.61	99.63	97.82	38.46	0	27.27	23.07
ISPY1	99.65	99.64	99.87	100	92.1	97.36	96.42	94.73
BCD	80.95	88.88	83.72	94.11	99.72	94.44	100	95
BCCD	88.88	33.33	75	55.55	70	61.11	83.33	66.66

5 Discussion

The objective was to implement Genetic Algorithm for selecting the best features subset from the existing attributes available in the different types of datasets and apply the selected features in the prediction models of Logistic Regression and k-NN. Six various types of secondary datasets are used for the comparison. For each dataset, the features are selected using GA to produce better accuracy in machine learning prediction. Both LR and k-NN shows improvement inaccuracy. Out of 6 datasets, k-NN scored better accuracy in almost 2 datasets. They are ISPY1 clinical trial dataset and Breast cancer dataset. All other datasets show the improved accuracy in Logistic regression model which is higher than the accuracy rates of k-NN. In case of sensitivity, logistic regression with feature selection is showing almost higher sensitivity rates other than the two datasets. In the case of specificity, the results are the same. So this work has arrived at the conclusion that the Logistic Regression model with GA selected features is capable of accurate breast cancer prediction. Table 6 displays the comparison of accuracy with the considered existing paper with

Table 6 Comparison of results with existing and proposed

Dataset	ANN [3]	PS-Classifier [3]	GA-Classifier [3]	This work	
				With feature selection	
				LR	k-NN
BCWO	96.7	96.9	96.6	99.27	98.55
BCWD	97.3	97.2	96.6	98.24	55.1
BCWP	79.2	78.2	78.1	86.2	78.33
ISPY1	–	–	–	97.05	97.61
BCD	–	–	–	93.85	94.73
BCCD	–	–	–	79.16	58.33

the proposed work which shows feature selection with prediction models proposed shows better accuracy than the others.

6 Conclusion

In this work, we compared the breast cancer prediction accuracy of two machine learning techniques, logistic regression and k-NN by using the best features selected by genetic algorithm. Six various types of secondary datasets are used for the comparison. For each dataset, the features are selected using GA to produce better accuracy in machine learning prediction. Both LR and k-NN shows improvement in accuracy. Logistic Regression with feature selection using GA obtained better accuracy than k-NN. Out of 6 datasets k-NN scored better accuracy in 2 datasets. All other datasets show the improved accuracy in LR model which is higher than the accuracy rates of k-NN. So this work has arrived at the conclusion that LR model with GA selected features is capable of accurate breast cancer prediction.

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

1. Abe N, Kudo M, Toyama J, Shimbo M (2000). A divergence criterion for classifier-independent feature selection. Adv Pattern Recognit: Springer: 668–676
2. Kohavi R, John GH (1997) Wrappers for feature subset selection. Artif Intell 97:273–324
3. Aalaei S, Shahraki H, Rowhanimanesh AR, Eslami S (2016) Feature selection using genetic algorithm for breast cancer diagnosis: experiment on three different datasets. Iran J Basic Med Sci. 19:476–482
4. Pawlovsky AP, Nagahashi M (2014). A method to select a good setting for the kNN algorithm when using it for breast cancer prognosis. In: IEEE-EMBS international conference on biomedical and health informatics (BHI). IEEE, pp 189–192
5. Agarwal A, Saxena A (2018) Malignant tumor detection using machine learning through scikit-learn. Int J Pure Appl Math 119(15):2863–2874
6. Wolberg WH et al Breast Cancer Wisconsin (Diagnostic) data set. archive.ics.uci.edu. https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic). Accessed 4 Nov 2019
7. UCI Machine Learning Repository. Breast Cancer Wisconsin data set. archive.ics.uci.edu. http://archive.ics.uci.edu/ml/index.php. Accessed 4 Nov 2019
8. Wolberg WH et al Breast Cancer Wisconsin (Prognostic) data set. archive.ics.uci.edu. https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Prognostic). Accessed 4 Nov 2019
9. Wolberg WH et al Breast Cancer Wisconsin (Original) data set. archive.ics.uci.edu. https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original). Accessed 4 Nov 2019
10. Singh Suwal M Dataset created for: 'AI for social good: women coders' Bootcamp. kaggle.com. https://www.kaggle.com/merishnasuwal/breast-cancer-prediction-dataset#Breast_cancer_data.csv. Accessed 4 Nov 2019
11. Patrício M, Pereira J, Crisóstomo J et al Using Resistin, glucose, age and BMI to predict the presence of breast cancer. BMC Cancer. https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra. Accessed 4 Nov 2019

12. Cardenas-Rodriguez J julio/ispy-1-trial. data.world. https://data.world/julio/ispy-1-trial. Accessed 4 Nov 2019
13. Shetty K FeatureSelectionGA. github.com. https://github.com/kaushalshetty/FeatureSelectionGA. Accessed 4 Nov 2019
14. Varian. Artificial intelligence smartens up cancer care. physicsworld.com. https://physicsworld.com/a/artificial-intelligence-smartens-up-cancer-care/. Accessed 4 Nov 2019
15. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3:1157–1182
16. Dubey D, Kharya S, Soni S (2013) Predictive machine learning techniques for breast cancer detection. Int J Comput Sci Inf Technol 4(6):1023–1028
17. Kumari M, Singh V (2018) Breast cancer prediction system. Procedia Comput Sci 132:371–376
18. Narkhede S Understanding confusion matrix. towardsdatascience.com. https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62. Accessed 4 Nov 2019

Detection Classification and Cutting of Fruits and Vegetables Using Tensorflow Algorithm

Rupal Mayo Diline D'Souza, S. R. Deepthi, and K. Aarya Shri

Abstract This paper presents automatic fruit and vegetable recognition and chopping mechanism. This work aims to reduce the time taken for cutting fruits and vegetables in a large kitchen. This mechanism has two parts. The first part of the system is based on image processing, which consists of capturing an image of the object, comparing it with the images stored in the database, and identification of objects. For doing this, Tensor flow algorithm is used and the accuracy obtained is more than 90%. The second part of the project is the segregation of the fruits/vegetables and cutting them depending on the requirement.

Keywords Identification of fruits and vegetables · TensorFlow · Nvidia graphics · Cuda · Cutting mechanism based on type

1 Introduction

Human beings are taught from their childhood to identify fruits and vegetables. But when it comes to the large kitchen where dozens of fruits and vegetables are brought in, the cutting process takes lots of time. Especially while cutting these fruits/vegetables in different shapes. In order to reduce the time taken to cut these fruits and vegetables, an automated system is designed which can identify different types of fruits and vegetables and cut them according to the requirement. There are different methods proposed for identifying fruits and vegetables [1]. The most commonly used techniques are color and shape based analysis methods [2] for image

R. M. D. D'Souza (✉) · S. R. Deepthi · K. A. Shri
Electronics and Communication Engineering Department, St Joseph Engineering College, Mangaluru, India
e-mail: rupaldilinedsouza@gmail.com

S. R. Deepthi
e-mail: deepthi.srg27@gmail.com

K. A. Shri
e-mail: aaryashri05@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_66

detection, TensorFlow algorithm is used. Tensorflow is an open-source platform developed by Google. A hardware system is designed which has a conveyor belt where these fruits/vegetables to be cut are placed and the camera mounted on the conveyor belt captures the image and sends it for preprocessing and identification. Once the fruit/vegetable is identified depending on requirements, it can be cut. For cutting these fruits/vegetables in different shapes, four types of blade are designed.

There were some other systems proposed that uses Artificial Neural Network (ANN) for automatic identification and sorting of fruits. In this paper, the TensorFlow algorithm is used for fruit classification and a cutting mechanism is also implemented.

2 Related Works

In the paper titled "Faster R-CNN Implementation Method for Multi-Fruit Detection Using Tensorflow Platform" [3], the author proposed a Deep learning concept using Faster Region-based Convolutional Neural Network (Faster R-CNN) to detect and classify different fruits.

Kanade et al. [4] proposed a method to classify the ripeness level of fruit as green, ripe, overripe, and spoiled using Computer Vision System. This developed system analyzes RGB color distribution and classifies fruits using Principal Component Analysis (PCA).

The paper titled "A Modified Canny Edge Detection Algorithm for Fruit Detection & Classification" [5] uses a method to recognize fruits by extracting features like color and shape. Canny Edge Detection (CED) algorithm is used to identify and classify the fruit.

In paper [6], the author reviewed some image processing approaches used in the classification of fruits.

Seng [7] proposed a method that classifies and recognizes fruit images with the help of nearest neighbor classification. The author makes a comment that the results obtained are greatly affected by the scalar values, fruit size, which is selected by the consumer. The system tester plays a vital role in determining the accuracy of the recognition results.

The book titled "Deep Learning Classifiers with Memristive Networks" provides an explanation of the deep learning concepts. Deep Learning can be used for a wide variety of applications like object classification and detection. This book also explains how deep learning concepts can be built using memristive systems [8].

The paper titled "Automatic Fruit Recognition from Natural Images using Color and Texture Features" discusses a method for classifying different types of fruits efficiently and precisely with the help of a SVM [9].

The book titled "Learning TensorFlow" by Hope, Resheff, and Lieder gives guidance on image processing, and comparison of the object is done using Tensorflow [10].

In the paper [11], the author developed a cost-effective hand-operated elephant apple cutter by considering the machine parameters like capacity, machine efficiency, and loss percentage. The developed cutter can be used for commercial applications.

3 Proposed Method

The proposed method of fruit recognition consists of various stages like image acquisition, preprocessing of the images, feature extraction, and classification using TensorFlow Algorithm. The cutting mechanism is controlled by using Arduino UNO based on the data obtained from the image processing and classification technique. The block diagram for the fruit cutting and recognition system is shown in Fig. 1.

The detection and cutting system consists of a USB webcam for image acquisition, PC inbuilt with NVIDIA GPU for feature extraction and classification, and an Arduino Uno Controller to interface the object detector model with the cutting and segregation mechanism.

In the detection phase, the first step is image acquisition which is done with the help of a webcam. The acquired image is sent for feature extraction phase where the characteristics or attributes of an image are extracted.

For the image classification, the following steps are performed. A database is created by capturing images of different types of fruits or vegetables for the training process. The image of a fruit or a vegetable to be cut is captured using a webcam for recognition. The TensorFlow algorithm is used for fruit identification. This algorithm performs classification by comparing it with the database. To use TensorFlow algorithm, several environment variables have to be set up and several libraries and files need to be installed. To NVIDIA, GPU system was used for implementing the TensorFlow. After fruit recognition, its name will be displayed on the monitor of the PC and then the fruit is sent to cutting. A mechanical system is designed to cut fruits or vegetables based on their type. After a fruit or vegetable is cut, it is segregated according to its type (Fig. 2).

The cutting and segregation system consists of a conveyor belt along with Wiper Motor, Shafts, Bearings, Clamps, Sheet metal, pneumatic cylinders, Blades, and tray.

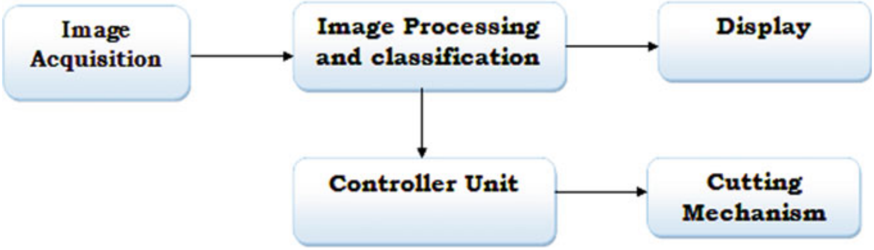
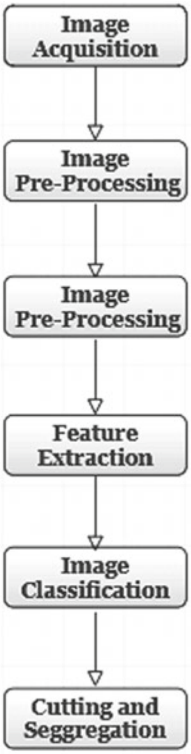


Fig. 1 Block diagram for fruit recognition and cutting system

Fig. 2 Fruit detection and classification



The conveyor belt is of size 1 m length and 0.15 m breadth. The fruit or vegetable to be cut is placed at one end of the conveyor left side where the webcam is attached to capture the image of fruit. The image thus captured goes through a series of steps for recognition of the type of fruit. Separate blades are used in the cutting process, i.e., each blade of particular shape for each fruit. Once the cutting process is complete, segregation is done. Four different fruits are considered for the experiment namely apple, potato, guava, and ivy gourd (Fig. 3).

4 Experimental Results

The model with conveyor belt, cutting tool, and segregation mechanism has been designed and implemented. The Tensor Flow algorithm is considered for classification which provides the high computational speed and used for real-time applications. Interfacing of object detector model with the mechanical system is done through serial communication with Arduino controller (Fig. 4).

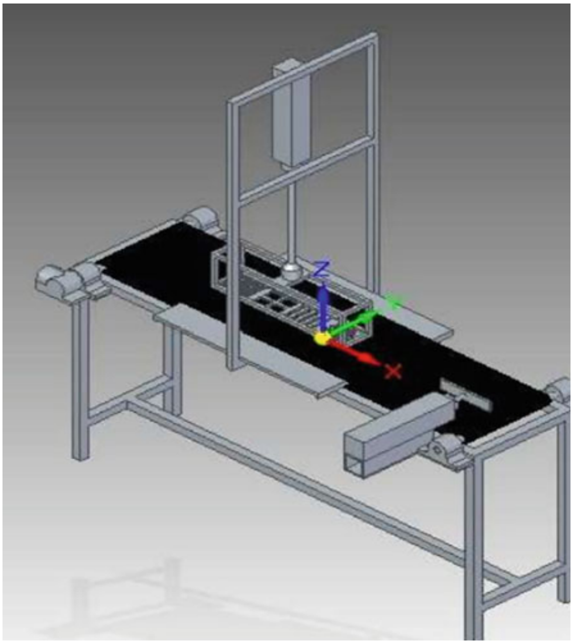


Fig. 3 Fruit cutting and segregation mechanism model

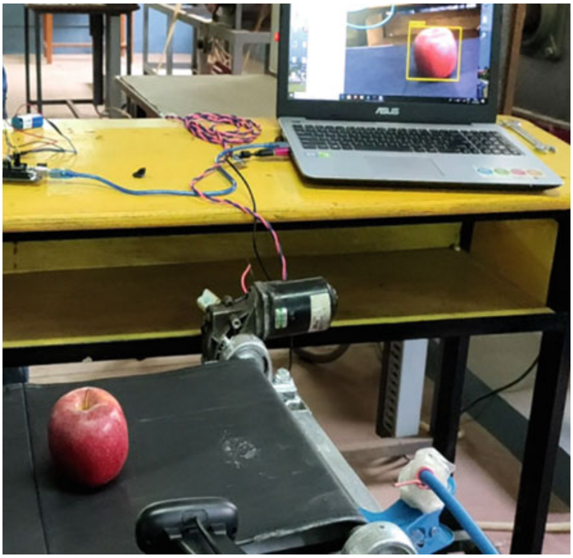


Fig. 4 Detection of fruit placed on the conveyor belt

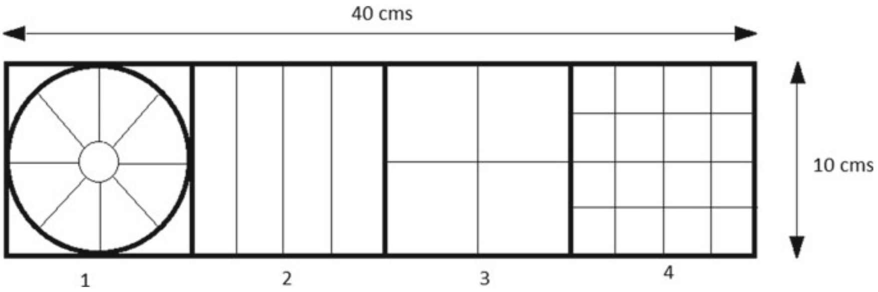
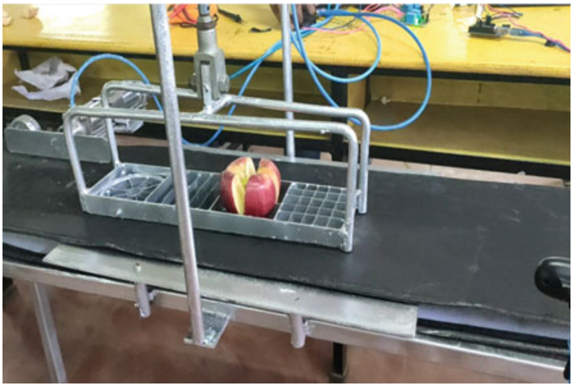


Fig. 5 Four types of blades used

Fig. 6 Automated fruit/vegetable cutting process



Once the fruit is detected, depending on the requirement, the type of blade can be selected to cut it. Four different types of blades are designed for this purpose (Figs. 5 and 6).

Once the cutting process is done, then the fruit is collected in the tray placed at the end of conveyor belt. İf there are more than one fruit or combination of fruits and vegetables, it can be detected and cut accordingly. In this paper, Apple, Guava, potato, and ivy gourd are detected (Figs. 7 and 8).

5 Conclusion and Future Work

In order to minimize the labor, cost, and time for cutting of fruits and vegetables in the large kitchens, a system is designed. This system identifies apple, guava, potato, and Ivy gourd and cuts it into different shapes. Here, the fruit/vegetable is first placed on the conveyor belt. A webcam is attached along the conveyor belt which captures the image of the object; here fruit or vegetable and sends it for processing. TensorFlow software is used to identify the object. Once the fruit/vegetable is identified, then it will enter the cutting mechanism. A cutting mechanism is designed with four blades,

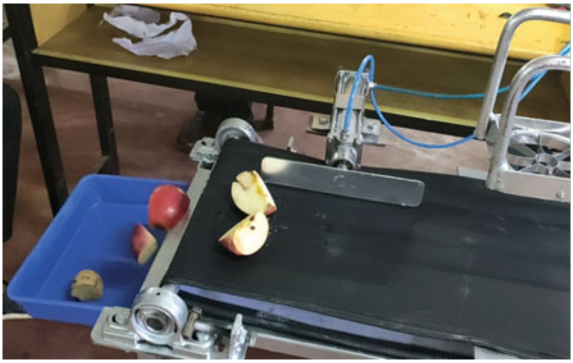


Fig. 7 Collecting of fruit which is cut

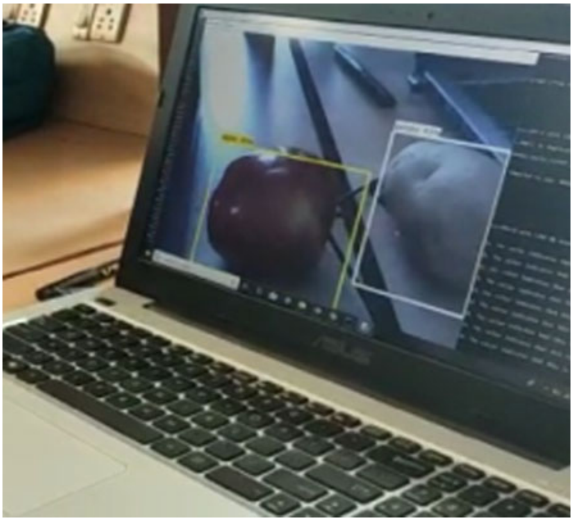


Fig. 8 Detection of Apple

each used for giving different shapes for fruits or vegetables. Once the cutting is done, the fruits and vegetables are segregated according to the type. Arduino board is used to interface the object detector model with the cutting mechanism. Air compressor is used to pass the compressed air to the pneumatic cylinder for cutting and segregation mechanism, and the wiper motor is used to control the conveyor belt.

The work can be improved further by including the peeling mechanism along with the existing system. This system is used for the cutting of fruit/vegetable in large kitchens; in order to use this in household kitchens, a compact device has to be designed. This system can be further extended to identify more fruits and vegetables.

References

1. Zawbaa HM, Abbass M, Hazman M, Hassenian AE (2014) Automatic fruit image recognition system based on shape and color features. In: AMLTA 2014, CCIS 488. Springer International Publishing Switzerland 2014, pp 278–290
2. Anuj JM, Revankar SG, Shettigar S, Mendonsa RS, D'souza RMD, Rashmi H (2019) Identification of fruit and subsequent automated peeling and cutting based on the type. Jnanasangama. ISBN 97881-934215-29
3. Basri H, Syarif I, Sukaridhoto S. Faster R-CNN implementation method for multi-fruit detection using tensorflow platform. In: 2018 international electronics symposium on knowledge creation and intelligent computing (IES-KCIC)
4. Kanade A, Shaligram A (2015) Development of machine vision based system for classification of Guava fruits on the basis of CIE1931 chromaticity coordinates. In: 2015 2nd international symposium on physics and technology of sensors (ISPTS)
5. Monir Rabby MK, Chowdhury B, Kim JH (2018) A modified canny edge detection algorithm for fruit detection & classification. In: 2018 10th international conference on electrical and computer engineering (ICECE)
6. Meruliya T, Dhameliya P, Patel J, Panchal D, Kadam P, Naik S (2015) Article: image processing for fruit shape and texture feature extraction—Review. Int J Comput Appl 129(8):30–33, November 2015. Published by Foundation of Computer Science (FCS), NY, USA
7. Seng WC, Mirisaee SH (2009) A new method for fruits recognition system. In: 2009 international conference on electrical engineering and informatics
8. Toleubay Y, James AP (2020) Getting started with tensorflow deep learning. In: James A. (eds) Deep learning classifiers with memristive networks. Modeling and optimization in science and technologies, vol 14. Springer, Cham
9. Jana S, Basak S, Parekh R (2017) Automatic fruit recognition from natural images using color and texture features. 2017 Devices for Integrated Circuit (DevIC)
10. Hope T, Resheff YS, Lieder I (2017) Learning TensorFlow. Printed in the United States of America, Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, August 2017
11. Nayak PK, Rayaguru K (2017) Design, development and performance evaluation of elephant apple core cutter. J Food Sci Technol 54:4060–4066

Detection of Counterfeit Cosmetic Products Using Image Processing and Text Extraction

Siddharth Mehta, Prajakta Divekar, Aditi Kolambekar, and Amol Deshpande

Abstract The growing popularity of the cosmetics industry has made it vulnerable to duplication and counterfeiting. One in every five individuals purchases counterfeits cosmetics. Consumption of such cosmetics can be harmful to health. They might sometimes contain chemicals such as arsenic or other carcinogenic chemicals. The packagings of the counterfeit products are designed in such a manner that someone can barely differentiate it from the original one. The packaging of the duplicate products differs from that of the original in terms of certain features such as dimension, font style, font size, color, and ingredients. This paper reports the development of a system that uses the application of image processing and text extraction techniques that will enable a user to determine the authenticity of the test product. The system uses the features of the authentic product and compares them with the features of the test product to determine its authenticity. The system has been tested with some of the counterfeit and original cosmetic products.

Keywords Authenticate · Cosmetics · Image processing · Text extraction · Preprocessing techniques · Web portal

S. Mehta (✉) · P. Divekar · A. Kolambekar · A. Deshpande
Bharatiya Vidya Bhavan's Sardar Patel Institute of Technology, Mumbai, India
e-mail: mehtasiddharth2512@gmail.com

P. Divekar
e-mail: prajaktadivekar70@gmail.com

A. Kolambekar
e-mail: kolambekaradati23@gmail.com

A. Deshpande
e-mail: amol_deshpande@spit.ac.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_67

1 Introduction

Today, cosmetics are widely used, by individuals of all genders and demographics.

A significant number of cosmetics seen in the market are counterfeit. With more and more duplication, counterfeit cosmetics have become increasingly difficult to detect with the human eye. Compounding this problem is the fact that these cosmetics are sold at a fraction of the price of the original cosmetics. Thus, consumers are frequently attracted to these counterfeit products. This motivates the authors to study the aspects related to original and counterfeit products and implement a system to distinguish between the two using the image processing technique.

On careful observation, the counterfeit product differs from the original product in terms of

- Dimensions of the packaging
- Overall packaging and its color
- Font style and font size of the text
- Ingredients
- Hallmark sticker

In the case of the example depicted in Fig. 1, with the left one being original and the right one being counterfeit, careful observation shows that the counterfeit product has broader and different colored packaging, no hallmark sticker, slightly different font style, and absolutely no text written on the product itself.

It is extremely difficult to detect the differences between these two products with a naked eye. The differences between the original and the duplicate products are far too subtle. However, when image processing techniques are used to compare these two products, these differences can be spotted efficiently as discussed in this paper.

Currently, there are no user-friendly systems in place that can detect these counterfeit products. At the same time, mobile phones are extremely common and are used by almost everyone today.

This project aims to build a user-friendly and efficient system to detect counterfeit cosmetics, by capturing the image of the product using a mobile phone. The user is expected to take pictures of the test product with the help of their mobile phone, and the system uses image processing and text extraction techniques to detect a counterfeit product.

Images of the test product are captured using a phone camera and uploaded on the web portal, with each image being used for a particular test. The physical features of the packaging of the product are compared to those of the authentic product. The system then calculates the degree of authenticity, and if that is above a certain threshold, the product is classified as original, else is classified as counterfeit. Also, dimensions of the packaged product are detected and compared with that of the original product. Similarly, salient features of the text are recognized, such as the set of ingredients or the type and size of the fonts, and are compared with the same features of the original product. Keeping all of this in check, the counterfeit products are detected.

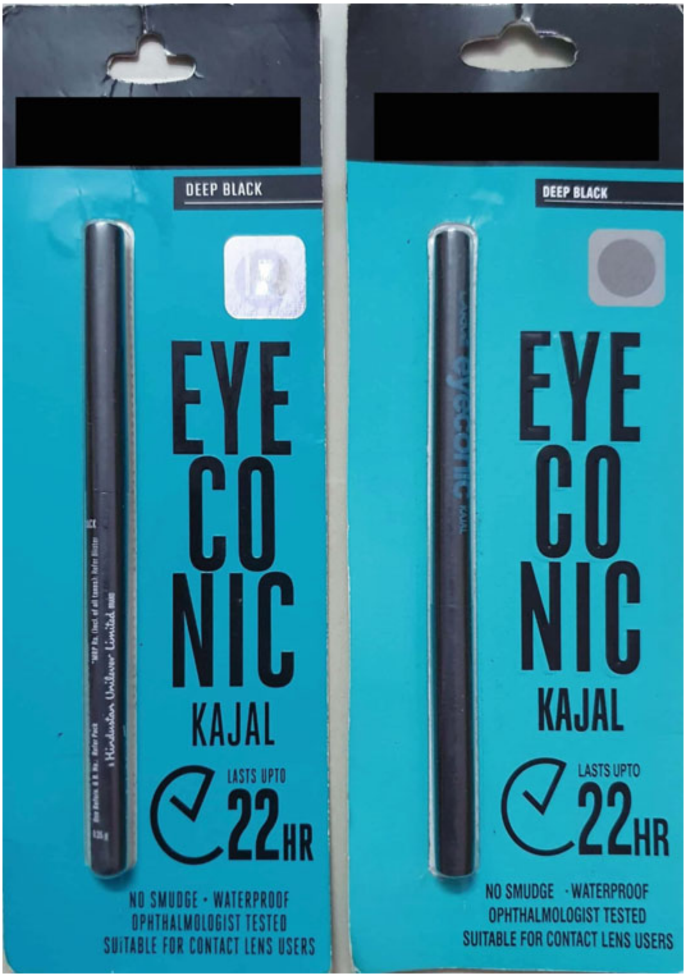


Fig. 1 Diagram depicting original and counterfeit versions of the same product [8]

The work is motivated by considering the need to build a system in place that can protect a consumer from firstly unintentionally harming themselves by purchasing a cosmetic product. Secondly, this system is being built to protect the consumer from fraud, and also to raise awareness among the consumer against counterfeit products. The system aims to

- Design a simple and efficient system that can detect a counterfeit product.
- Apply image preprocessing techniques efficiently in order to obtain accurate results irrespective of the quality of the user uploaded image.
- Detect the authenticity of the product directly through its packaging.
- Identify the ingredients of the products through its packaging.

- Raise customer awareness against fraud.
- Caution the consumer about any harmful ingredients that the product that the cosmetic may contain.

2 Related Work

The proposed solution uses the image of the product to determine its authenticity. It is based on image processing techniques which involve image preprocessing as well as image analysis with respect to text recognition and text extraction, image classification based on the extracted text. The input image is also further used to detect dimensions of the actual product.

2.1 Image Preprocessing Techniques

The Gaussian filter has advantages over a Bilateral filter in the presence of increasing noise. Bilateral filter, though known to preserve edges better while denoising, fails to perform well in the presence of noise which can be concluded from the results and observations made in [1].

This paper [2] aims to find a good trade-off between high image smoothing quality and fast processing which finds many applications in real-time image processing. These requirements are met by using Local Binary Patterns and Graphical Processing Unit. LBP is used to analysze local texture around the central pixel, and the result is used to decide the size of Gaussian kernel. The overall smoothing performance of Gaussian filters is indicated by measuring the Peak Signal-to-Noise Ratio (PSNR) before and after the processing. The paper discusses results after applying 3×3 and 9×9 Gaussian kernels and concludes on the use of large and small kernels.

The bilateral filter is implemented using MATLAB, provides a easy and simple way to blur the image. The results are analyzed by using two performance merits—Mean Square Error (MSE) and PSNR which conclude the accurate and reliable denoising technique using bilateral filter [3].

A novel texture-based design of the multilateral filter is described, and its advantages over a conventional bilateral filter are discussed [4]. The limitations of conventional filtering technique, failing to preserve edges with similar color, can be successfully overcome using the proposed multilateral filter. This technique makes use of the additional spatial texture to preserve clear region boundary. The paper concludes improved performance using multilateral filter based on the comparisons made [4].

2.2 Text Detection and Text Extraction

A business card reader application using Optical Character Recognition (OCR) engine Tesseract is described, which plays a significant role in extracting text from various image formats such as Joint Photographic Experts Group (JPEG) and Portable Network Graphics (PNG) [5]. It uses a high resolution smartphone camera to capture the image and to further extract the text on it. The proposed method provides an accuracy of up to 74% in terms of both data detection and text recognition. This system implements Scrum methodology and after each sprint completion performance is tested and improved further. It uses User Interface (UI) kit framework for implementing interfaces, Audio Visual (AV) foundation framework to configure i-phone cameras, and Core image framework for image processing along with Tesseract OCR for text detection and classification.

The described system uses Open-source Computer Vision (OpenCV) and Tesseract OCR for real-time license plate detection [6]. It describes the effects of using image preprocessing techniques such as thresholding along with image filtering methods such as median blur. The software performance is further improved by using Neuro Fuzzy networks to minimize errors and to improve the ability of the system to predict results based on previous predictions.

This system [7] proposes new methods to extract text from shadowed images to further improve the performance of Tesseract OCR. Starting with contour detection of text from the binary image, it then deletes the salt-pepper noise from shadowed areas using a double filtering algorithm. Projection method removes the noise between the texts, and the median filter removes the noise between characters to produce optimum results.

Kikuchi et al. [8] Propose a stereo vision system to detect the dimensions of the object. The system detects objects from the stereo images and consists of blob extraction and size calculation. It uses image preprocessing techniques followed by object detection and image segmentation. It provides a considerably faster approach for object detection using cameras aligned at a specific position and with a particular orientation. The error between the results obtained and actual data is negligible.

2.3 String Matching for Product Classification

This paper [9] describes a novel and fast string matching algorithm which is essential in network security. The proposed algorithm is based on evaluating the weight of the search pattern by adding the weights of individual characters. A sliding window equal to the length of the search pattern covers the entire text and compares the weight of the characters encountered. The calculated weight is then compared with the weight of the search pattern to produce results. The paper compares the results of the proposed algorithm with BM and Brute Force algorithm and concludes that the proposed algorithm has constant search time independent of pattern length.

2.4 Contour Detection Techniques

As contour detection holds significant importance for 2D image processing, a novel algorithm for contour transversal based on morphological operations is described here [10]. The technique detects a pixel wide contour around objects in binary images. This method provides faster contour detection with slight compensations in accuracy. The method is evaluated on manually created ground truth contours. For accurate contour detection, the width of the object must be greater than 3 pixels.

This paper [11] describes an automated method for road detection using a region-based active contour model and distance transform. It requires image preprocessing to increase the contrast between different contours and the use of median blur to enhance the quality of the original image. The proposed system is effective in detecting all the contours along with some cases false contour detection. The method has limitations while processing complex or high resolution images.

2.5 Structural Similarity and Color-Tone Similarity

The use of Structural Similarity (SSIM) and Color Similarity (CSIM) is done to compare two images with respect to hue, brightness, contrast, and color space. While the structural similarity index provides similarity measurement of grayscale images, it cannot be used to understand the role of color information. PSNR is used as a performance indicator before and after image smoothing. The system uses fixed Gaussian kernel size and concludes on the kernel size required for processing different images.

3 Proposed Work

3.1 Process Flow Model of the System

The system is based on image processing algorithms that compare two images to give a final verdict stating the authenticity of the product under test. The proposed system facilitates the user to conduct three tests viz. Similarity check, Dimension detection, and Ingredient matching. These tests are made available on a web portal which can be accessed by any user with a mobile phone. Each test needs distinct images to be captured considering some specific guidelines. Each of the three tests has a threshold value of result depending on which authenticity of the product is predicted. The users that want to determine the authenticity of the product must upload the images of the same on the web portal which is accessible on the browser. The user is expected to capture the images of the test product from different orientations such as front view and back view for similarity test, product image with a reference object for dimension detection, and image of the ingredients for ingredient matching. The image/s for each test are to be uploaded on a web portal using a mobile phone camera and in a particular format. Instructions to be followed while capturing the image as per the system's requirement are made available for the user.

As delineated in Fig. 2, the uploaded image will be compared with the database of the original product. The proposed system consists of a database containing information about various parameters of the original products of different cosmetic brands. The database mainly includes

- Images of the packaged product from different orientations—front view and back view.
- Dimensions of the packaged product.
- Text extracted (ingredients) from the packaged product.

At the end of each test, appropriate results will be displayed.

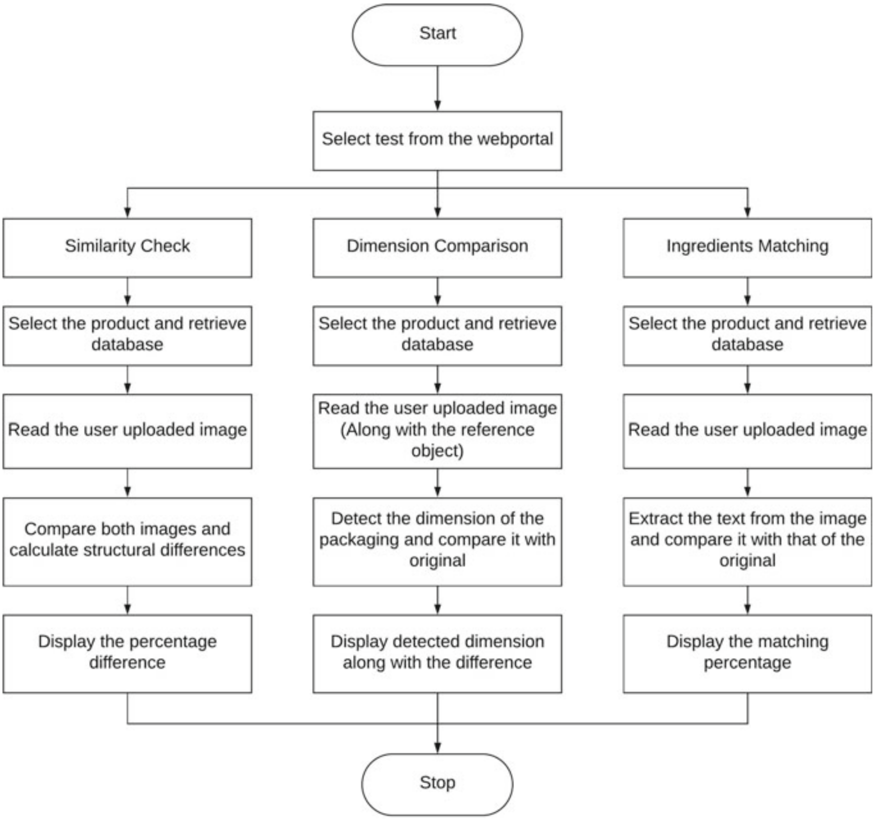

Start
Select test from the webportal
Similarity Check
Dimension Comparison
Ingredients Matching
Select the product and retrieve database
Select the product and retrieve database
Select the product and retrieve database
Read the user uploaded image
Read the user uploaded image (Along with the reference object)
Read the user uploaded image
Compare both images and calculate structural differences
Detect the dimension of the packaging and compare it with original
Extract the text from the image and compare it with that of the original
Display the percentage difference
Display detected dimension along with the difference
Display the matching percentage
Stop

Fig. 2 Process flow diagram of the proposed work

4 Algorithm and Preprocessing Techniques

4.1 Image Preprocessing Techniques

Gray scaling and filtering are the preprocessing techniques applied to the input image before conducting each test viz. Dimension detection, Similarity check, and Text extraction. These techniques can be explained in detail as follows:

- **Gray scaling**: It is a technique by which a colored image is represented using multiple shades of gray which can be manipulated by the computer. The result of the operations performed on the gray scaled image of the product is shown in Fig. 3, where all the images depicted have been gray scaled. Conventional software and hardware systems can represent the image only in a limited number of shades of gray usually 16 or 256. Gray scaling typically uses large amounts of memory as each dot is represented using either 4 or 8 bits. This is a preprocessing technique which involves the removal of all the color information and representation of each pixel in terms of intensity and illumination.
- **Filtering**: Filtering process is an integral part of image preprocessing as it tends to eliminate as much noise as possible from the image while preserving its quality. The presence of noise in the image can lead to misleading results in contour

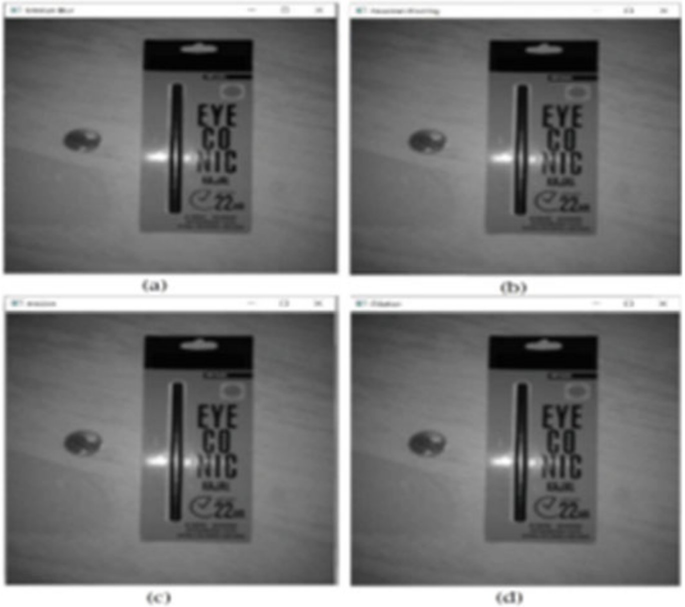


Fig. 3 Result of different preprocessing techniques (Parenthesis) **a** Result of median blur, **b** Result of Gaussion blur, **c** Result of erosion, and **d** Result of Dilation

detection and edge detection which can affect the process of dimension detection. Filtering process plays an important role in obtaining maximum accuracy in the results of different tests irrespective of the quality of the uploaded image. Following filtering techniques are used to preprocess the uploaded image before each test is conducted. The choice of the filtering technique depends on the noise level, intensity distribution, and the resolution of the input image, as discussed in [1].

1. Median Blur Filter: It is an averaging technique in which each pixel is replaced by the median of the pixels in a kernel area of that pixel. This technique makes the overall image smoother, with an aperture size of kernel X kernel. The smoothing operation results in noise removal from the input image. For the system under consideration, median blur filter is used as a preprocessing technique for mainly two tests—Dimension detection and Text extraction. This filtering technique provides accurate results in 90% of the cases. The result of median blur is shown in Fig. 3a.
2. Gaussian Blur Filter: The Gaussian Blur filter is used to reduce the image detail noise. In this technique, an image is multiplied by a Gaussian Blur function matrix, which results in producing an overall effect of viewing the image through a translucent screen. The Gaussian function with small variance is generally used, whereas a Gaussian function with a larger variance is used for image segmentation [2]. The Gaussian function in two dimensions is

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{\frac{-\left(x^2+y^2\right)}{2\sigma^2}} \tag{1}$$

where x and y are the coordinates of the pixel in the image, and σ is the standard deviation of the distribution. Gaussian filter response is a function of Gaussian Variance and it is independent of the variance of signal-to-noise ratio of the image under consideration\cite{singhal}. Gaussian filter proves to be effective in the presence of significant noise levels or images with low Signal-To-Noise Ratio. Figure 3b shows the result of Gaussian filtering.

4.2 Dimension Detection

Other than the above-mentioned techniques, dimension detection requires the following preprocessing techniques:

- **Erosion**: Erosion is a technique which removes pixels from the image boundary and makes the boundary clear. The result of erosion can be observed from Fig. 3c.

- **Dilation**: Dilation is a technique which tends to make the boundaries of the image sharper by comparing the values of neighboring pixels. The dilated image of the product is shown in Fig. 3d.
- **Contour Detection**: Contour detection plays an important role in image preprocessing since most of the image information is contained in the edges. Contour detection in the proposed system makes use of filtering and blurring techniques which separates the object in the image from its background [4]. Different filtering techniques produce different results based on the presence of noise, with some filters resulting in blurring the edges while others resulting in preserving the edges. Contour detection has a significant effect while determining the dimensions of the product in the image under consideration [5]. Among all the detected contours, only one is of utmost importance which depicts the contour of the test product. That contour is selected according to the following algorithm.
- **Algorithm for Contour Selection**: For determining the dimensions of the object in the image, contour detection plays a significant role [6]. Every object in the image has its contour associated with it which is detected in the preprocessing itself. But for the system under consideration, the test image consists of two objects of interest—one being the product itself while the other being a reference object. The algorithm is designed to discard all the detected contours excluding the contours of two objects which in this case are the coin and the product. After selecting two contours of interest, the height and width of each contour are calculated [7]. For the reference object, i.e., the coin, actual dimensions are already known to the system and hence the algorithm calculates a ratio of the calculated dimensions of the coin with respect to the original dimensions. This ratio is further used to detect the dimensions of the test product in the user uploaded image.

4.3 Similarity Check

After converting the uploaded image into grayscale and applying appropriate filtering techniques, the following algorithm is used to detect the structural similarity between the packaging of the original product and the test product.

This is one of the significant tests for authenticity check. This step is used to check for structural differences between the captured images of the test product and those of the authentic product. The structural difference can be extracted using the two functions SSIM (Structural Similarity Index) and MSE (Mean Square Error) available in Python. The SSIM compares the two images with respect to hue, brightness, contrast, and color space. Another function CSIM is the same as SSIM but can be used for the colored images also. While the MSE gives the mean-square difference between both the images.

While the structural similarity index provides similarity measurement of grayscale images, it cannot be used to understand the role of color information. Peak signal-to-noise ratio is used as a performance indicator before and after image smoothing. The system uses a fixed Gaussian kernel size and concludes on the kernel size required for processing different images [8].

4.4 *Text Extraction*

- **Thresholding**: Thresholding being the most important image segmentation technique is used to distinguish the desired object from its background [9]. After converting the test image into grayscale format, different thresholding techniques can be applied to the image in order to separate only the desired portion of the image from the background. The thresholding technique can be chosen depending on the quality of the image, color of the background.
- **Algorithm**: The algorithm is divided into two parts—Text extraction and text matching which are implemented in Python. The algorithm detects the text embedded in the image and can read all the image types. Also, a similarity score out of 100 is obtained, which is a metric to gauge how similar the extracted text is in comparison to the text extracted from the original product in the database.

5 Web Portal

The web portal has been developed using flask framework provided by Python. The modules such as WTF-forms, flask-login, werkzeug, and flask-bcrypt are used to provide the various facilities such as form fields, registration for user, debugging, and hashing, respectively.

The application under consideration facilitates to carry out various tests in order to determine product authenticity. A user interactive environment is developed where the user can upload images of the packaged product under test such as front view and back view of the product, product image with a reference object for dimension detection, and image displaying product ingredients.

Other facilities such as new user registration, log in for existing users, and account information are also incorporated in the web portal. The verification of credentials of a user is also performed on the web portal.

Each of the tests which determines the product authenticity is associated which different image processing techniques. User can go for any of the mentioned tests and view the results on the web portal itself. The process flow and logic behind determining the authenticity is written in Python.

Table 1 Similarity check between images of the same products

Test no.	Images of the same product	MSE	SSIM
Test 1	Original product with distortion (flash ON)	826.05	0.77
Test 2	Original product on a different phone	670.52	0.73

Table 2 Similarity check between images of different products

Test no.	Images of the different products	MSE	SSIM
Test 1	Fake product	2016.92	0.59
Test 2	Fake product with distortion (flash ON)	3378.01	0.50

6 Results and Analysis

6.1 *Similarity Check*

As discussed earlier, MSE and SSIM are the available methods for similarity check. Various experiments are performed with the various categories of images for conducting the Similarity check test. For checking similarity, the product package of one of the brands (say brand1) is used. From Table 1, it is clear that the similarity cannot be 100% for the image of the same product. If the image is captured by the same camera but under the different light conditions, then the similarity index will become 77%. The similarity index will become 73% if the same image is captured from another camera, that is, with different pixel resolution. While the results of Table 2 show the result similarity of the counterfeit product with the original, more than 50% which is due to the exact duplicate design of the product package.

The user is prompted to upload the captured image of the front view of the product. Once the image is uploaded, the system authenticates the image with the front view of the original product, against which the test product is supposed to be tested. The result of the authentication is then displayed to the user, as shown in Fig. 4.

6.2 *Dimension Detection*

The actual dimension of the brand1 are

L = 2.6 inch and H = 7.8 inch (from front view), where

D = Distance of the product from the reference object,

L = Length obtained in inches,

H = Height obtained in inches.

From Table 3, it is clear that the result is more accurate if the distance between the reference object and the product is 5 cm.

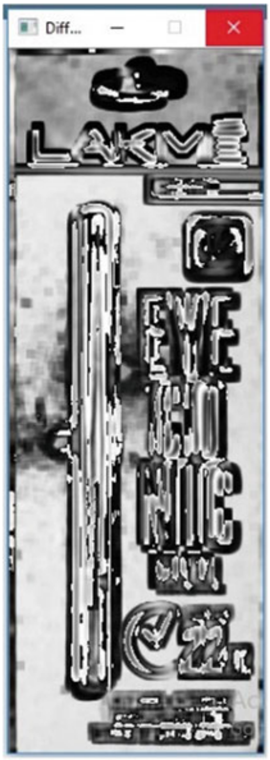


Fig. 4 Similarity check on the web portal

Table 3 Detected Dimensions of the Product with varying distance from the camera

D (cm)	L (inches)	H (inches)
5	2.7	7.8
7	2.7	7.6
9	2.6	7.5
11	2.6	7.3
13	2.7	7.6
15	2.6	7.5
17	2.9	8.2
21	2.6	7.5
23	2.9	8.1

Now, the front and the back views of the test product are already available on the portal which was uploaded for earlier test. The system now automatically detects the dimensions of the product using the uploaded images. The length and the breadth of the product are returned as markers on the uploaded image in a separate window, as shown in Fig. 5.

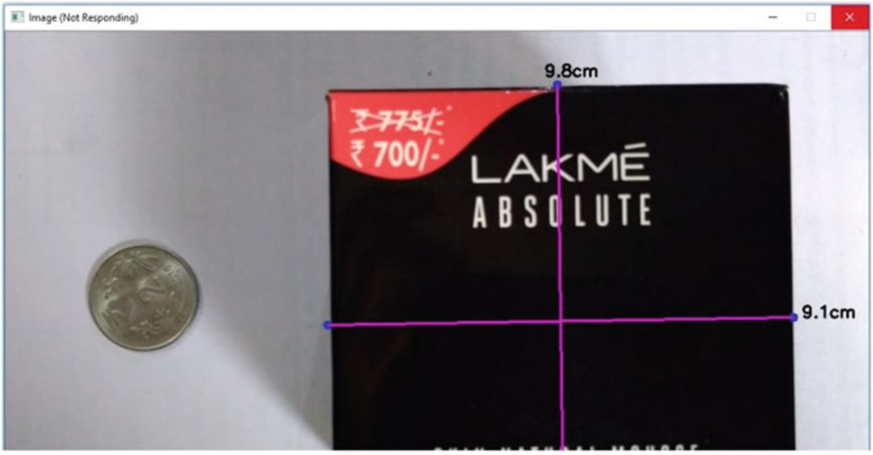
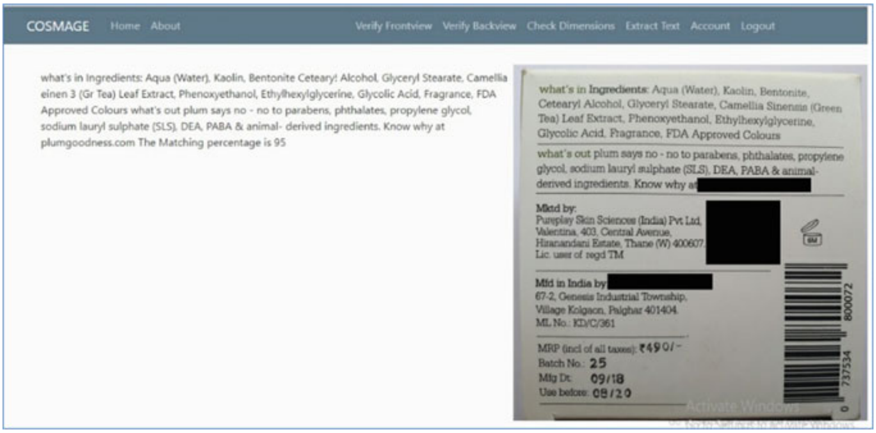


Fig. 5 Dimensions of the product as displayed on the web portal

6.3 *Text Extraction and Detection*

For efficient text extraction, the different preprocessing techniques are applied to the different images, based on parameters such as text color, background color, and ambient light. In the preprocessing, the type of image smoothing filters such as Gaussian blur, Median blur, and bilateral filter varies according to the image. Also, the size of kernel and the thresholding techniques vary according to the image properties.



what's in Ingredients: Aqua (Water), Kaolin, Bentonite Cetearyl Alcohol, Glyceryl Stearate, Camellia einen 3 (Gr Tea) Leaf Extract, Phenoxyethanol, Ethylhexylglycerine, Glycolic Acid, Fragrance, FDA Approved Colours what's out plum says no - no to parabens, phthalates, propylene glycol, sodium lauryl sulphate (SLS), DEA, PABA & animal- derived ingredients. Know why at plumgoodness.com The Matching percentage is 95

what's in Ingredients: Aqua (Water), Kaolin, Bentonite, Cetearyl Alcohol, Glyceryl Stearate, Camellia Sinensis (Green Tea) Leaf Extract, Phenoxyethanol, Ethylhexylglycerine, Glycolic Acid, Fragrance, FDA Approved Colours

what's out plum says no - no to parabens, phthalates, propylene glycol, sodium lauryl sulphate (SLS), DEA, PABA & animal-derived ingredients. Know why at [redacted]

Mktd by:
Pureplay Skin Sciences (India) Pvt Ltd,
Valentina, 403, Central Avenue,
Hiranandani Estate, Thane (W) 400607,
Lic. user of regd TM
[redacted]

Mfd in India by [redacted]
67-2, Genesis Industrial Township,
Village Kolgaon, Palghar 401404
ML No.: KD/C/361

MRP (incl of all taxes): ₹490/-
Batch No.: 25
Mfg Dt: 09/18
Use before: 08/20

Fig. 6 Result of text extraction displayed on a web portal

Figure 6 shows one such instance when white-colored text appears against a black colored background. The right half of Fig. 6 shows one such instance whereas the left half of Fig. 6 shows the output of text extraction.

7 Conclusion and Future Scope

The proposed system is capable of authenticating an original product from a counterfeit product. The authenticity of the product under test is not determined by a single test but by using a series of tests which provide reliable results with justification for each test. Dimension detection, ingredient matching, and similarity check provide significant data to predict the authenticity of the given product. In the case of dimension detection, the accuracy of the system depends upon the distance between the packaging and the reference object in the image and also the angle of capturing the image. While for similarity check, the results depend upon the resolution of the camera through which image is captured and the brightness level of the image. Similarly, for ingredients matching, the results depend upon the image quality, font style, and font size of the text written on the packaging. The overall system performance can be improved further by using better string matching algorithms and image smoothing techniques.

The result of each test prominently depends on preprocessing techniques used. The preprocessing techniques vary from product to product according to the image quality, product color and size, and ambient lighting. While applying filtering as a preprocessing technique, the kernel size is determined based on the quality of the uploaded image.

After considering a number of products for each of the three tests, the threshold for concluding a product to be authentic is as follows. The difference obtained from the Similarity Check should be less than 20%. The difference between the detected dimension of the test product and that of the original product should be less than 0.3 inches. The ingredient matching percentage should be greater than 90%.

The system provides an adequate solution to the problem and will go a long way in protecting the end consumers as well as vendors from fraud and counterfeiting.

In order to make the system more efficient, the restrictions on the user should be minimized. Thus overall system performance can be improved further by incorporating machine learning and artificial intelligence.

This system can be modified to recommend appropriate cosmetics based on the consumer's previous uses and preferences using Machine Learning. With the extensive data set the system has at its disposal, the system can recognize the product that the consumer uploads and recommend a more suitable product from its database. This improves the consumer experience as well as provides them with adequate market intelligence on the product that they wish to purchase.

Acknowledgments The authors are thankful to the Sardar Patel Institute of Technology, India for providing the necessary facilities for carrying out this work.

References

1. Teutsch M, Trantelle P, Beyerer J (2013) Adaptive real-time image smoothing using local binary patterns and Gaussian filters. In: 2013 IEEE international conference on image processing, Melbourne, VIC, 2013, pp 1120–1124. https://doi.org/10.1109/icip.2013.6738231
2. Gedraite ES, Hadad M (2011) Investigation on the effect of a Gaussian Blur in image filtering and segmentation. In: Proceedings ELMAR-2011, Zadar, 2011, pp 393–396
3. Singhal P, Verma A, Garg A (2017) A study in finding effectiveness of Gaussian blur filter over bilateral filter in natural scenes for graph based image segmentation. In: 2017 4th international conference on advanced computing and communication systems (ICACCS), Coimbatore, 2017, pp 1–6. https://doi.org/10.1109/icaccs.2017.8014612
4. Leventić H., Keser T, Vdovjak K (2018) A fast one-pixel wide contour detection method for shapes contour traversal in binary images. In: 2018 international conference on smart systems and technologies (SST), Osijek, 2018, pp 11–14. https://doi.org/10.1109/sst.2018.8564595
5. Zhang Y, Wang X, Peng L (2013) Text based bilateral filter for color image edge preserving. In: 2013 5th IEEE international conference on broadband network & multimedia technology, Guilin, 2013, pp 49–52. https://doi.org/10.1109/icbnmt.2013.6823913
6. Yasir MM, Noor R, Hasbi H, Azman A (2012) Stereo vision images processing for real-time object distance and size measurements. 659–663. https://doi.org/10.1109/iccce.2012.6271270
7. Dangiwa BA, Kumar SS (2018) A Business card reader application for iOS devices based on Tesseract. In: 2018 international conference on signal processing and information security (ICSPIS), DUBAI, United Arab Emirates, 2018, pp 1–4. https://doi.org/10.1109/cspis.2018.8642727
8. Kikuchi H, Huttunen H, Hwang J, Yukawa M, Muramatsu S, Shin J (2012) Color-tone similarity on digital images. 1–4. https://doi.org/10.13140/rg.2.1.5055.6565
9. Lu H, Guo B, Liu J, Yan X (2017) A shadow removal method for tesseract text recognition. In: 2017 10th international congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI), Shanghai, 2017, pp 1–5. https://doi.org/10.1109/cisp-bmei.2017.8301946
10. Palekar RR, Parab SU, Parikh DP, Kamble DP (2017) Real time license plate detection using openCV and tesseract. In: 2017 international conference on communication and signal processing (ICCSP), Chennai, 2017, pp 2111–2115. https://doi.org/10.1109/iccsp.2017.8286778
11. Singla N, Garg D (2012) String matching algorithms and their applicability in various applications. International J Soft Comput Eng (IJSCE) 1

Multiclass Weighted Associative Classifier with Application-Based Rule Selection for Data Gathered Using Wireless Sensor Networks

Disha J. Shah and Neetu Agarwal

Abstract The twenty-first century has seen an explosion in the amount of data that is available to decision-makers. Wireless Sensor Networks are everywhere continuously collecting and feeding data to various systems. Translating these enormous amounts of data into information and making decisions quickly and effectively is a constant challenge that decision-makers face. Decision Support Systems that analyze data using different approaches are constantly evolving. There are many techniques that analyze these large datasets collected from Wireless Sensor Networks. Associative Classifier is one such technique which combines Associative Rule Mining along with Classification to find different patterns generated from large datasets. A common challenge associated with sensor data is that it is unstructured and heterogeneous. In this research, a Multiclass Weighted Associative Classifier with Application-based Rule Selection is proposed which translates unstructured and heterogeneous data into a set of rules that enable decision-makers in any domain to make decisions.

Keywords Wireless sensor network · Data mining · Associative classifier

1 Introduction

Wireless Sensor Network is used in several domains like military [1, 2], environment monitoring [3–5], healthcare monitoring, habitat monitoring [6, 7], object tracking [8, 9], disaster management [10], etc., and it generates diverse and voluminous data in different formats which may be structured or unstructured. This raw data needs to be efficiently analyzed to generate information which will help in decision making.

D. J. Shah (✉)
Pacific University, Udaipur, Rajasthan, India
e-mail: disha.shah@glsuniveristy.ac.in

FCAIT, GLS University, Ahmedabad, Gujarat, India

N. Agarwal
Department of Computer Science, Pacific University, Udaipur, Rajasthan, India
e-mail: neetu.agarwal1508@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_68

Knowledge Discovery Databases (KDD) is a process which takes raw data, processes it, and produces valuable information. The process consists of data cleaning, data integration, data transformation, data mining, and data evaluation and implementation [11]. Data mining helps to discover and extract hidden and novel patterns from gathered data. Different approaches of data mining are clustering [12–15], association mining [16, 17], frequent patterns [18–21], sequential patterns [22–24], and classification [25, 26].

Association is a technique used to identify the relationship between two or more items. Association Rule Mining are if-then statements that show the probability of relationships between data items within large datasets. The Apriori Algorithm, AIS Algorithm, and SETM Algorithms are some of the techniques used to search frequently used itemsets [27–29]. Classification is a technique that helps to construct a classifier for large datasets. Some techniques for classification are Naive Bayesian Classifier, Support Vector Machines, Decision Trees, Neural Networks, and C4.5 [30]. An approach of combining Association Rule Mining and Classifier is known as Association Classification. The main aim of Associative classifiers is to use association rules to link attribute values with its class labels. In recent years, many associative classifier techniques like CBA, CMAR, CPAR, etc,. are being recommended [31–34].

The main issues for integration of association and classification are efficiency, accuracy, and scalability. In associative classifier, all items are given the same importance. It also avoids the difference between the number of transactions and its importance. To overcome this problem, weights can be given to attributes to tell its importance to the user [35–37].

The paper is divided into four sections. Section 2 focuses on the related work. Section 3 indicates the proposed Associative classifier for data mining for data gathered using wireless sensor network and Sect. 4 emphasize on conclusion.

2 Related Work

Different Associative Classifiers have been used in many domains having large datasets. This section shows how associative classifiers have been used in different scenarios like MCAR [38], MAC [39], MMAC [40], CMR [41], and CWAC [42].

Thabtah et al. [38] proposed Multiclass classification based on Association Rules (MCAR) which discovers items and rules in one phase. It has also introduced a rule ranking technique that minimizes the use of randomization when a choice point would be between two or more rules.

Abdelhamid et al. [39] showed that how Multiclass Associative Classification (MAC) Algorithm generates less number of rules compared to MCAR in order to enable the user to understand it.

Thabtah et al. [40] provided an approach for Multiclass Multilabel Associative Classification (MMAC) which produces classifiers that contain rules with multiple labels and multiple evaluation measures for accuracy rate. It also discovers the rules in one scan a ranking technique which ensures effective rules.

Zhou [41] presented Classification based on Multiple Classification Rules (MCR) that generates many classification rules and provides more accuracy and efficiency in classification.

Ibrahim and Chandran [42] proposed Compact Weighted Class Association Rule Mining (CWAC) a classifier that uses HITS algorithm which does not require preassigned weights. It prefers information gain attribute and generates all the rules based on an attribute.

3 Proposed Multiclass Weighted Associative Classifier with Application-Based Rule Selection of Data Gathered Using Wireless Sensor Networks

As proposed in [43, 44], there are numerous applications of wireless sensor network where each application generates a huge amount of data. Each application needs to collect the data, analyze it, and do mining for decision making. The system aims to construct a multiclass classifier model which effectively uses the data gathered by different locations of an application of WSN. The data are distributed among various locations in WSN. This proposed system implements distributed data mining where the data are distributed over various locations.

The proposed algorithm is basically divided into three steps: Data Cleaning, Rule generation, and Classifier Builder. In the first step, a training data file is taken as an input and data preprocessing is implemented. Step 2 generates rules based on appropriate weights and importance. In the final step, it selects the generated rules and based on that, a classifier is generated effectively. The general diagram of the proposed Associative Classifier is depicted in Fig. 1.

The dataset will be taken as input from different applications like Weather monitoring, Health care monitoring, Agriculture monitoring, etc. The dataset would be a location application-based dataset. The data shown in Table 1 is a sample training dataset of weather monitoring application. Here, the dataset D has n distinct attributes namely A1, A2, A3, …, An.

The first process in this system is data cleaning. Data cleaning is the process to ensure that the data is consistent, complete, correct, and usable. Cleaned data can be categorical or continuous data. For continuous attributes, a discretization technique is to be implemented.

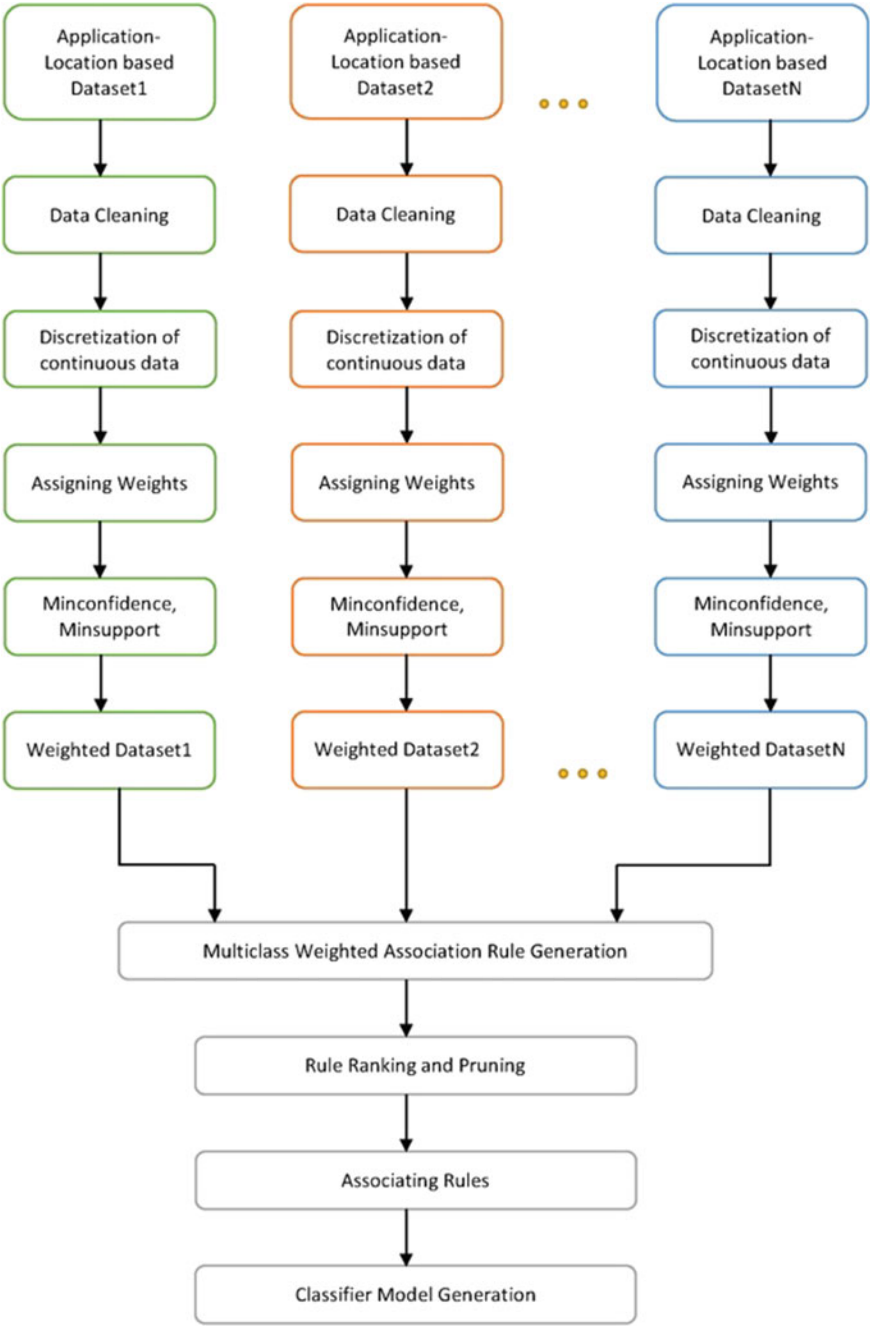

Application-Location based Dataset1
Data Cleaning
Discretization of continuous data
Assigning Weights
Minconfidence, Minsupport
Weighted Dataset1
Application-Location based Dataset2
Data Cleaning
Discretization of continuous data
Assigning Weights
Minconfidence, Minsupport
Weighted Dataset2
Application-Location based DatasetN
Data Cleaning
Discretization of continuous data
Assigning Weights
Minconfidence, Minsupport
Weighted DatasetN
Multiclass Weighted Association Rule Generation
Rule Ranking and Pruning
Associating Rules
Classifier Model Generation

Fig. 1 Diagram of the proposed multiclass weighted associative classifier

Table 1 Training dataset of weather monitoring

timestamp	boardid	temp_max	temp_min	temp_avg	light_max	light_min	light_avg	humidity_min	humidity_max	humidity_avg	model	latitude	longitude	elevation
12/15/2014 01:40:00 AM	508	21.6	21.6	21.6	96.4	96.4	96.4	41.2	41.2	41.2	ENV	−37.813408	144.979492	30.1
12/15/2014 01:40:00 AM	505	23.2	23.2	23.2	93.5	93.5	93.5	48.3	48.3	48.3	ENV	−37.813073	144.980406	29.91
12/15/2014 01:45:00 AM	507	21.6	21.6	21.6	97.2	97.2	97.2	44.8	44.8	44.8	ENV	−37.814922	144.982258	38.79
12/15/2014 01:45:00 AM	502	21.3	21.3	21.3	97.4	97.4	97.4	45.3	45.3	45.3	ENV	−37.81461	144.979018	22.57
12/15/2014 01:45:00 AM	501	22.9	22.9	22.9	97.4	97.4	97.4	43.1	43.1	43.1	ENV	−37.814808	144.980999	29.96
12/15/2014 01:50:00 AM	508	21.3	21.3	21.3	96.4	96.4	96.4	41.2	41.2	41.2	ENV	−37.813408	144.979492	30.1
12/15/2014 01:55:00 AM	505	23.2	23.2	23.2	93.5	93.5	93.5	47	47	47	ENV	−31.813073	144.980406	29.91
12/15/2014 01:55:00 AM	507	21.6	21.6	21.6	97.2	97.2	97.2	44.5	44.5	44.5	ENV	−37.814922	144.982258	38.79
12/15/2014 01:55:00 AM	502	21.6	21.6	21.6	97.4	97.4	97.4	45.3	45.3	45.3	ENV	−37.81461	144.979018	22.57
12/15/2014 02:00:00 AM	501	23.5	23.5	23.5	97.4	97.4	97.4	43.7	43.7	43.7	ENV	−37.814808	144.980999	29.96

```
Input: N applications, Training Dataset (D[n]) a set of distributed datasets of N applications, S minSupport, C minConfidence, W weights

Output: Set of classification rules

Scan D[n] for data preprocessing
Do
    For each transaction, check inaccurate data
    If found
        Perform data cleansing
    Else
        Read Next transaction

    For each attribute, check datatype
    If continuous
        Convert using Discretization technique
    else
        Read Next attribute
Until no items are found

Scan D[n] to apply weights
Do
    Apply Weights to each transaction
Until no items are found

Scan D[n] for set S of frequent items
Do
    For each item I in S
        If (sup(I) > minSupport)
            $S \leftarrow S \cup I$
    Next
Until no items greater than minSupport found

Scan D[n] for set C of classes
Do
    For each item I in S
        If (con(I) > minConfidence)
            Generate rules
            $C \leftarrow C \cup I$
    Next
Until no items greater than minConfidence found

Rank all rules that are generated.

Remove all rules $I' \rightarrow C'$ from S where there is higher rank rule and $I \subseteq I'$.
```

The proposed technique gives weights to attributes to give importance to transactions. Then minimum support and confidence are given by the user. This will help in generating rules and building classifiers. Each item is compared with minimum support and if the item is greater than the minimum support, then the item is added in frequent item set.

Once an item has been identified as a frequent item, the algorithm finds all the rules. The multiple set of classes are added and compared with minimum confidence. Here, when two rules with identical confidence are found, both the rules are taken into consideration. Rules are generated and added in the classifier. Rule ranking and rule pruning play an important role in any associative classifier. The rules are added in the class which has high confidence. In the end, once the rules are generated and ranked, each rule which does not fulfill the condition is removed from the class. In the end, a set of R rules are generated from the classifier.

4 Conclusion

In this paper, we have proposed how Application-based Association Classifier can be used for data gathered using Wireless Sensor Network. In this system, any WSN application location-based dataset can be used which can provide decision-makers with rules that are more meaningful in decision making and can compare with previous results. This technique can also be used in any domain like Finance, Defense, Agriculture, Non-Profit Organizations, etc.

References

1. Yick J, Mukherjee B, Ghosal D (2008) Wireless sensor network survey. Comput Netw 52(12):2292–2330
2. Arampatzis T, Lygeros J, Manesis S (2005) A survey of applications of wireless sensors and wireless sensor networks. In: Proceedings of the 20th IEEE international symposium on intelligent control (ISIC '05), June 2005, pp 719–724
3. Lee LT, Chen CW (2008) Synchronizing sensor networks with pulse coupled and cluster based approaches. Inf Technol J 7(5):737–745
4. Sabri N, Aljunid SA, Ahmad B, Yahya A, Kamaruddin R, Salim MS (2011) Wireless sensor actor network based on fuzzy inference system for greenhouse climate control. J Appl Sci 11(17):3104–3116
5. Kumar D (2011) Monitoring forest cover changes using remote sensing and GIS: a global prospective. Res J Environ Sci 5:105–123
6. Rozyyev A, Hasbullah H, Subhan F (2011) Indoor child tracking in wireless sensor network using fuzzy logic technique. Res J Inf Technol 3(2):81–92
7. Szewczyk R, Osterweil E, Polastre J, Hamilton M, Mainwaring A, Estrin D (2004) Habitat monitoring with sensor networks. Commun ACM 47(6):34–40
8. Chauhdary SH, Bashir AK, Shah SC, Park MS (2009) EOATR: energy efficient object tracking by auto adjusting transmission range in wireless sensor network. J Appl Sci 9(24):4247–4252
9. Biswas PK, Phoha S (2006) Self-organizing sensor networks for integrated target surveillance. IEEE Trans Comput 55(8):1033–1047

10. Tseng YC, Pan MS, Tsai YY (2006) Wireless sensor networks for emergency navigation. Computer 39(7):55–62
11. Nwagu CK, Omankwu OC, Inyiama H (2017) Knowledge Discovery in Databases (KDD): an overview. Int J Comput Sci Inf Secur (IJCSIS) 15(12):13–16 Dec
12. Chau M, Cheng R, Kao B, Ng J. Uncertain data mining: an example in clustering location data, advances in knowledge discovery and data mining. In: PAKDD 2006. Lecture Notes in Computer Science, vol 3918. Springer, Berlin, Heidelberg
13. Zakir Hossain M, Nasim Akhtar M, Ahmad RB, Rahman M (2019) A dynamic K-means clustering for data mining. Indones J Electric Eng Comput Sci 13(2):521–526. Feb. ISSN: 2502-4752
14. Gama J, Rodrigues PP, Lopes L (2011) Clustering distributed sensor data streams using local processing and reduced communication. Intell Data Anal 15(1):3–28
15. Aghbari ZA, Kamel I, Awad T (2012) On clustering large number of data streams. Intell Data Anal 16(1):69–91
16. Boukerche A, Samarah S (2007) An efficient data extraction mechanism for mining association rules from wireless sensor networks. In Proceedings of the IEEE international conference on communications (ICC ’07), June 2007, pp 3936–3941
17. Kumar V, Chadha A (2012) Mining association rules in students assessment data. Int J Comput Sci Issues 9(5) No 3:221–219, Sep
18. Deypir M, Sadreddini MH (2011) EclatDS: an efficient sliding window based frequent pattern mining method for data streams. Intell Data Anal 15(4):571–587
19. Hassani M, Tows D, Cuzzocrea A, Seidl T (2019) BFSPMiner: an effective and efficient batch-free algorithm for mining sequential patterns over data streams. Int J Data Sci Anal 8:223–239
20. Chuang P-J, Tu Y-S (2019) Efficient frequent pattern mining in data streams. 2019 IOP Conf Ser: Earth Environ Sci 234: 012066
21. Mahmood A, Shi K, Khatoon S (2012) Mining data generated by sensor networks: a survey. Inf Technol J 11:1534–1543
22. Cook DJ, Youngblood M, Heierman III EO et al (2003) MavHome: an agent-based smart home. In: Proceedings of the 1st IEEE international conference on pervasive computing and communications (PerCom ’03), Mar 2003, pp 521–524
23. Adamo J-M Data mining for association rules and sequential patterns: sequential and parallel algorithms. Springer. ISBN: 146130086X 9781461300861
24. Aggarwal CC, Zhai CX (2012) Mining text data. Kluwer Academic Publishers, Springer. ISBN: 978-1-4614-3222-7
25. Aggarwal CC (2015) Data classification: algorithms and applications. CRC Press. ISBN: 978-1-4665-8675-8
26. Phyu TN (2009) Survey of classification techniques in data mining. In: Proceedings of the international multiconference of engineers and computer scientists, vol I IMECS 2009, 18–20 March 2009
27. Kumbhare TA, Chobe SV (2014) An overview of association rule mining algorithms. Int J Comput Sci Inf Technol 5(1):927–930
28. Zhang C, Zhang S (2002) Association rule mining- models and algorithms. Lecture notes in computer science, vol. 2307. Lecture notes in artificial intelligence, Springer
29. Singh G, Jassi S (2017) A comparative analysis on association rule mining algorithms. Int J Recent Technol Eng (IJRTE) 6(2). ISSN: 2277-3878, May 2017
30. Phyu TN (2009) Survey of classification techniques in data mining. In: Proceedings of the international multi conference of engineers and computer scientists 2009, vol I, IMECS 2009, 18–20 March 2009, Hong Kong
31. Thabtah F, Mahmood Q, McCluskey L, Abdel-Jaber H (2010) A new classification based on association algorithm. J Inf Knowl Manag 9(1):55–64
32. Xiaoxin Y, Han J (2003) CPAR: classification based on predictive association rules. In: Proceedings of the 2003 SIAM international conference on data mining. ISBN: 978-0-89871-545-3

33. Li W, Han J, Pei J (2001) CMAR: accurate and efficient classification based on multiple class-association rules. In: Proceedings 2001 IEEE international conference on data mining, Dec 2001. ISBN: 0-7695-1119-8
34. Ramani D, Kanani H, Pandya C (2013) Ensemble of classifiers based on association rule mining. Int J Adv Res Comput Eng Technol 2(11):2963–2967, Nov 2013. ISSN: 2278-1323
35. Chhikara S, Sharma P (2014) Weighted association rule mining: a survey. Int J Res Appl Sci Eng Technol 2(IV):85–88, April 2014. ISSN: 2321-9653
36. Madhuri R, Pushpa Latha P, Prasad Rao K (2013) Weighted association rule without pre-determined weights. Int J Eng Res Technol (IJERT) 2(2), Feb. ISSN: 2278-0181
37. Ibrahim S, Sivabalakrishnan (2019) An enhanced weighted associative classification algorithm without preassigned weight based on ranking hubs. Int J Adv Comput Sci Appl 10(10): 290–297
38. Thabtah F, Cowling P, Peng Y (2005) MCAR: multi-class classification based on association rule. In: The 3rd ACS/IEEE international conference on computer systems and applications. ISSN: 2161-5322
39. Abdelhamid N, Ayesh A, Thabtah F (2012) MAC: a multiclass associative classification algorithm. J Inf Knowl Manag 11(2)
40. Thabtah FA, Cowling P, Peng Y. MMAC: a new multi-class, multi-label associative classification approach. In: Proceedings of the fourth IEEE international conference on data mining (ICDM'04). 0-7695-2142-8/04
41. Zhou Z (2014) A new classification approach based on multiple classification rules. Hindawi Publishing Corporation Mathematical Problems in Engineering 2014, 7 pp. Article ID 818253
42. Syed Ibrahim SP, Chandran KR (2011) Compact weighted class association rule mining using information gain. Int J Data Min Knowl Manag Process (IJDKP) 1(6), November
43. Shah DJ, Arolkar HA (2012) Single point interface for data analysis in wireless sensor networks. Int J Comput Appl 47(9):0975–888, June
44. Shah DJ, Arolkar HA (2013) Overview of data mining technique for data gathered using wireless sensor network. Anveshanam—J Comput Sci Appl II(1), August

Void-Aware Routing Protocols for Underwater Communication Networks: A Survey

Pradeep Nazareth and B. R. Chandavarkar

Abstract Underwater Acoustic Sensor Networks (UASNs) is a technology used in several marine applications like environment prediction, defense applications, and discovering mineral resources. UASNs has several challenges like high bit error rate, high latency, low bandwidth, and void-node problem during routing. In the context of routing protocols for underwater communication networks, the void-node problem is one of the major challenging issues. The void-node problem arises in the underwater communication during the greedy-forwarding technique, due to which packet will not be forwarded further toward the sink. In this review paper, we analyze the void-node problem in underwater networks and issues related to the void-node. Also, we elaborate on the significant classification of void-handling routing protocols. We analyze both location-based and pressure-based void-handling routing protocols.

Keywords Underwater communication · Void-node · Routing protocols

1 Introduction

About 70% of the earth's surface is covered by water, and oceans are the primary source of water. Even though oceans are playing a significant role in many aspects like military and commercial activities, we know little about it. Therefore, it is critical to explore the ocean for different applications like coastal surveillance, environmental monitoring, underwater resource exploration, and disaster detection [1–4]. All these applications find relevance to the underwater communication network. Sensor nodes present in UASN collect data from the underwater environment. Further, relay nodes forward the received data from other sensor or relay nodes. For a long time,

We would like to thank the Science and Engineering Research Board (SERB), Govt. of India for providing financial support (Ref. No. EEQ/2018/001036).

P. Nazareth (✉) · B. R. Chandavarkar
Wireless Information Networking Group (WiNG), Department of Computer Science and Engineering, National Institute of Technology Karnataka, Surathkal, Mangalore 575025, India
e-mail: pradeep.nitk2017@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_69

ocean sensing and data collection are limited to the surface and nearby water due to limitations of wire-line (cabled) instruments. Due to advancements in wireless technologies over the past few decades, underwater wireless communication is recommended to collect sensed data from the ocean. In underwater, radio waves require low-frequency results in higher energy consumption. Optical waves are suffering from scattering. Thus, Radio and optical waves are not suitable for long-range communication [1, 5]. Acoustic communication is ideal for underwater communication for long-range communication. Some of the significant challenges of designing UASNs are the limited bandwidth of only a few kbps, limited energy, high bit error rate, and void-node during routing.

Void-node problem is a severe problem in underwater routing. It occurs in the greedy routing strategy. Void-nodes in underwater routing results in packet drop if suitable void-node handling strategy is not defined. This survey paper reviews the problems of void-node in underwater and issues related to void-node. Also, we elaborate on the important categorization of routing protocols that are capable of handling void-node. We analyze location-based and pressure-based void-handling routing protocols.

The following sections are organized as follows: In Sect. 2, we provide more insight into void-node. Section 3 describes and analyzes various state-of-the-art routing protocols capable of handling void-node problem. Section 4 discusses some of the research gaps in existing protocols in the literature, and Sect. 5 contains the conclusion.

2 Void-Node in UASNs

A typical void-node scenario is shown in Fig. 1. UASNs consist of anchored nodes that collect information and propagate it to relay nodes. Both anchored nodes and relay nodes are equipped with an acoustic modem and deployed in the 3D underwater environment. UASNs includes one or more sink nodes, deployed at the water surface. UASN contains one or many sink nodes. Sink nodes are deployed at the water surface. Sink node has an acoustic modem that enables communication with an underwater relay or sensor nodes. Radio modem is used to communicate with the terrestrial network.

In a greedy-forwarding technique, every node transmits packets to a neighboring node with distance to the sink closer than itself (highest positive advancement) [4, 6, 7]. The non-availability of positive advancement node to any neighbor results in itself a void-node or local maxima node or stuck node. The void-node can emerge in underwater in convex or concave shapes. With reference to Fig. 1, node f is a concave void. The advancement of the neighboring node is computed as described as follows:

Let X_i be the node that has a packet to forward. X_j be a neighboring node, X_j is in the transmission range of X_i, and S is the sink node. The node X_j its advancement toward the sink S with respect to X_i can be calculated as follows [8]:

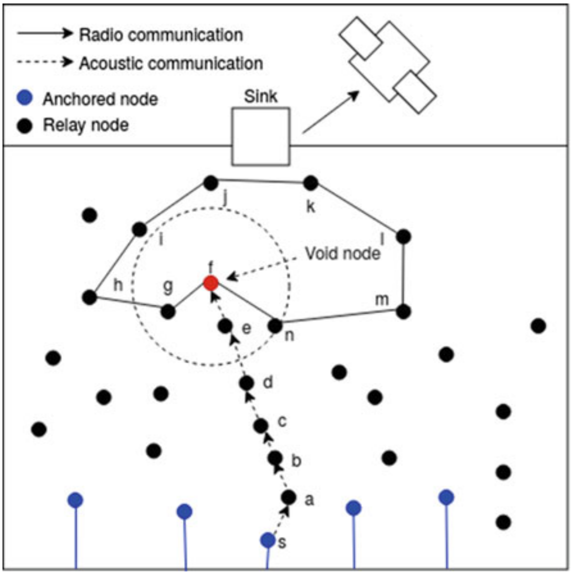


Fig. 1 Void-node scenario in underwater [4]

$$ADV\left(X_j\right) = D(X_i, S) - D\left(X_j, S\right) \quad (1)$$

where *D(X,S)* is the distance from node *X* to the sink *S*. *ADV*(X_j) is the advancement of node X_j toward the sink. In location-based routing, *D(X,S)* is the Euclidean distance between node *X* and *S*. Suppose $X(X_x, X_y, X_z)$ are the 3D coordinates of node *X* and $S(S_x, S_y, S_z)$ are the 3D coordinates of node *S*. Then Euclidean distance between node *X* and *S* can be calculated as per Eq. (2):

$$D(X, S) = \sqrt{(S_x - X_x)^2 + \left(S_y - X_y\right)^2 + (S_z - X_z)^2} \quad (2)$$

In pressure-based routing, *D(X,S)* is the difference in depths of nodes *X* and *S*. Then one of the nodes which are having the highest advancement in candidate forwarding set will be selected as next hop. Candidate forwarding set C_{X_i} of node X_i is calculated as follows [9]:

$$C_{X_i} = \{X_j \in N_{X_i} : ADV\left(X_j\right) > 0\} \quad (3)$$

where N_{X_i} is the set of neighboring nodes of X_i. An empty C_{X_i} indicates node X_i is void-node.

The distance between nodes is not only the reason for the void-node situation. The number of other factors, alone or in combination, are responsible for it. Some of the reasons for the void-node are ship movement between/over the nodes, sparse deployment of nodes, and noise. Packets received by a void-node will not be forwarded further toward the sink and dropped by the void-node if void-handling/recovery/processing mechanisms are not defined. If the void-node scenario is not handled properly, the performance of UASNs will be affected. Some repercussion of the mishandling of void problems are as follows:

- Packet loss
- Increased end-to-end delay
- Increased hop-count
- looping during routing

Hence, it is essential to define the proper application dependent void-handling strategy. A node that forward the packet to a void-node is referred to as a trapped node [9]. In Fig. 1, node *f* is void-node because it is not having any neighbors with positive advancement. In Fig. 1, nodes *s, a, b, c, d,* and *e* are trap nodes because packet forwarded by all these nodes are stuck at void-node *f*. The area covered by *f-g-h-i-j-k-l-m-n-f* (as shown in Fig. 1) is called a communication void region.

3 Void-Handling Routing Protocols

In literature, many underwater void-handling (void-aware) routing protocols are proposed and classified based on various criteria. Though, one of the main classifications is based on whether nodes in UASNs require 3D location information or not. Based on it, routing protocols are classified into two types: Location- and Pressure-based routing protocols, as shown in Fig. 2. In location-based routing protocols, underwater nodes must be aware of their 3D location by using localization mechanisms [10]. Next hop (or relay) is selected based on 3D locations of the node, its neighbors, and the sink node. In the pressure-based routing protocols, decisions on the next-hop selection are taken based on the only depth information of nodes. Nodes themself can easily obtain the depth of the node by pressure gauge embedded on it. Further, both location-based and pressure-based routing protocols are classified into opportunistic routing and non-opportunistic routing. Opportunistic routing is a scheme in which a subset of neighboring nodes is involved in packet forwarding [4, 11]. The following sub-sections present a significant classification of routing in detail.

3.1 Location-Based Routing Protocols

This sub-section discusses location-based void-handling (void-aware) routing protocols. Some of the location-based void-handling routing protocols are Vector-Based Void Avoidance (VBVA) [12], Adaptive Hop-by-Hop Vector-Based Forwarding (AHH-VBF) [13], Routing protocol, Focused Beam Routing (FBR) [14], and Directional Flooding-based Routing (DFR) [15].

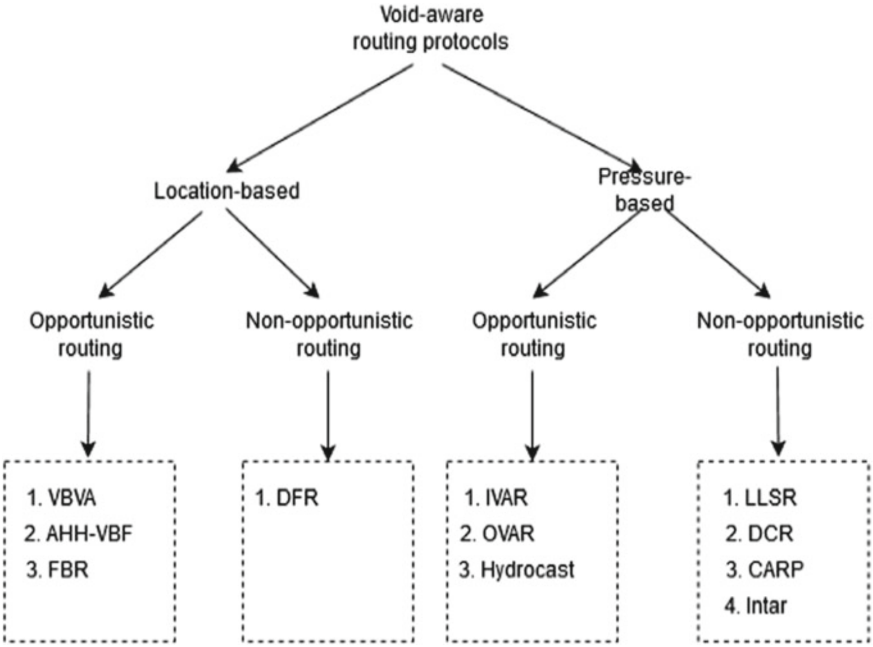


Fig. 2 The classification of void-aware routing protocols

A. Vector-Based Void Avoidance (VBVA):

In VBVA [12], packet forwarding is based on a vector-based approach. Initially, the routing path of packets is defined by a vector between source and the sink node, similar to the routing approach used in Vector-Based Forwarding (VBF) protocol [16]. In VBVA routing, every node in UASN is aware of the 3D location information of the source, the sink, and itself. During routing in VBVA, it may face void-node. Based on whether it is convex void or concave void, it selects the appropriate recovery mechanism to recover from void-node. When a void is a convex void, it tries to recover by a vector-shift mechanism. In Fig. 3a, node s and d are the sender and the sink nodes, respectively. Node s forwards the packet through the $\overrightarrow{sd}$. Node s will wait for overhearing the forwarding of the same packet by the nodes present within its forwarding vector until predefined time expires. If it does not overhear the transmission of the packet within a predefined time, it will conclude that it is a void-node. To recover from void, it broadcasts vector-shift control packet, asking neighbors s and b to change current forwarding vector to $\overrightarrow{ad}$ and $\overrightarrow{bd}$. Neighbors a and b repeat the same procedure to forward the packets. If the sender does not find any neighbor during vector-shift, this scenario indicates that the sender is a concave node. The back-pressure mechanism is used to recover from a concave node until it finds a suitable node capable of performing vector-shift, as shown in Fig. 3b.

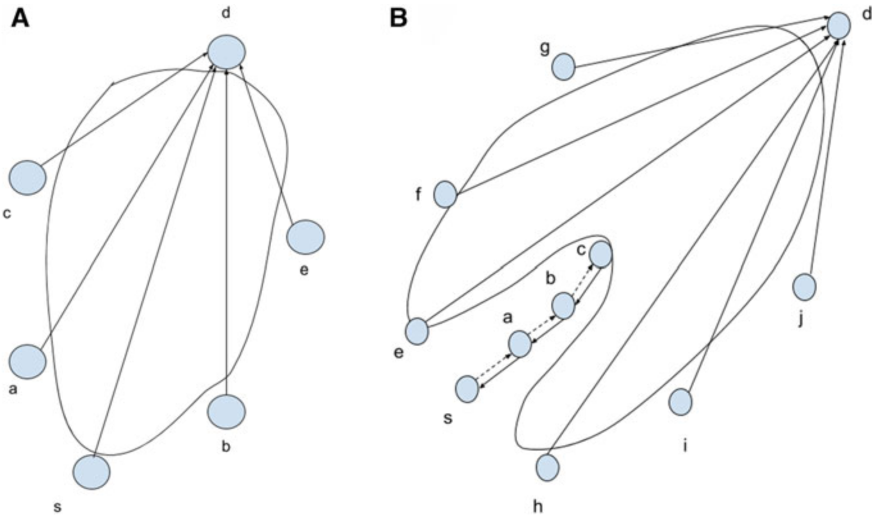


Fig. 3 Void-handling in VBVA **A** Vector-shift mechanism **B** Back-pressure mechanism [12]

Conclusion: VBVA is a reactive void-handling underwater routing protocol without the need to store additional information about neighbors and the void-node status of neighboring nodes. Even in the large underwater network, it is capable of handling the void-node problem. The major drawback of VBVA is, void-recovery mechanism activated only after the data packet trapped at void-node. It results in a higher delay in delivering the packet in the void-node scenario.

B. Adaptive Hop-by-Hop Vector-Based Forwarding (AHH-VBF)}:

AHH-VBF [13] is a location-based routing protocol. It requires every node in UASN to be aware of the 3D location of itself, the sink, one-hop neighbors, and the sender of the packet. It is based on Hop-by-Hop VBF (HH-VBF) [17]. In HH-VBF, the direction of the vector is changed from hop to hop. In AHH-VBF, in addition to changing the direction of vector hop-by-hop, the radius of the vector can be changed based on the density of network topology. In the case of sparse topology, the radius of the vector is increased to cover many candidate nodes. AHH-VBF is also capable of adjusting transmission power according to the density of topology.

Conclusion: Adapting the radius of vector and transmission power, small void-nodes can be efficiently handled. However, AHH-VBF is not able to handle larger void-holes in the network. Adaptive transmission power control efficiently manages the energy of the node.

C. Focused Beam Routing (FBR):

FBR [14] exploits the power control mechanism to conserve the energy of the sending/forwarding node. Every node knows its 3D location. FBR uses Request to Send (RTS)/Clear to Send (CTS) control packets to select the next hop. RTS packet includes the location of the sender/forwarder and the sink node. CTS packet contains locations of the node which receive the RTS packet and address of the node from which it receives RTS. A receiver of the RTS packet sends the CTS packet only if it is having positive progress toward the sink. Initially, a sender broadcasts an RTS packet at the lowest power. If the sender failed to receive any CTS packet in response to transmitted RTS, it indicates that in the transmitted region, there are no suitable nodes to forward the packet as per the greedy-forwarding strategy. Thus, the sender re-transmits the same RTS packet at the next higher power level so that it can find a suitable forwarding node in the new transmission region. If the receiver of RTS is suitable for forwarding the packet according to greedy-forwarding, it sends the CTS packet. In the case of many CTS packets received by the sender, it will find a suitable candidate node which is having the highest positive progress toward the sink as the next hop and forwards the data packet to selected next hop. Upon receiving the data packet by the selected next hop, it will use the same procedure to select its next hop.

Conclusion: The FBR is an energy-efficient protocol due to the usage of the power control mechanism. However, in sparse network topology, selection of next hop is time-consuming.

D. Directional Flooding-based Routing (DFR):

DFR [15] is the location-based routing protocol. It follows a receiver-based routing approach. DFR requires each node to know the 3D location of the neighbor, the sink, and its own. One of the main features of DFR is it uses the quality of the link to decide the flooding zone. In the case of good link quality, a smaller flooding zone to be used, resulting in few candidate nodes involved in becoming forwarder. If link quality is poor, larger flooding zone is used so that more nodes participate in packet forwarding.

Conclusion: DFR is capable of handle void-node problems. DFR is scalable to a large network. However, in the case of poor link quality or void-node scenario, the flooding zone becomes huge and results in many nodes involved in forwarding the same packets due to the hidden-node problem.

3.2 Pressure-Based Routing Protocols

This sub-section elaborates on pressure-based routing protocols. These protocols rely on the depth of the node; it can be obtained through the pressure-sensors embedded on the node.

A. Location-free Link State Routing (LLSR):

LLSR [18] is a beacon-based routing protocol. It uses hop-count as the primary parameter to ensure the progress during routing while selecting next hop by avoiding void-node. During the beacon dissemination phase, every node prepares beacon and broadcasts a beacon with hop-count, path quality, and the pressure. The path quality is the number of redundant paths available toward the sink. The pressure depends on the depth of the node deployment, obtained through pressure-sensors. The beaconing process is initiated by the sink node, with hop-count and pressure value equal to 0, path quality as 1. The receiver of the beacon updates its next hop to the neighbor node, which is having minimum hop-count value. If more than one nodes have same hop-count value, then it considers neighbor with higher path quality among them as its next hop. If there is a tie between path quality of those nodes, it considers node with lowest pressure value among them as the next hop. Further, these nodes propagate the beacon, so that all nodes in the network update their next hop. To reflect the changes in topology, beacons are exchanged at regular intervals. If a node detects itself as a void-node, it sends a beacon with hop-count value as infinity and path quality as 0. Upon receiving the beacon, neighbors make appropriate changes in their routing table.

Conclusion: The LLSR is effective in detecting and avoiding void-node by using depth information and positive progress. It is a loop-free routing algorithm with the capability of handling node movement. LLSR is not using opportunistic routing. Instead, it forwards data packets only to a selected next hop. Thus, it will not ensure reliability. LLSR is not considering link quality during the process of selecting the next hop. In LLSR, path quality has given more priority over the pressure (or depth) of the node; this may result in a higher end-to-end delay.

B. Inherently Void-Aware Routing (IVAR):

IVAR [19] is a receiver-based routing protocol. It will overcome some of the limitations of LLSR by using opportunistic routing. The hop-count and depth of nodes decide the suitable forwarding node. The beaconing process obtains the hop-count of each node. The sink node initiates the beaconing process. Initially, the sink node prepared a beacon with hop-count value 0 and broadcasts the beacon. Upon receiving the beacon, neighboring nodes are updating their hop-count information. Further, nodes are propagating the beacon so that all nodes in the network obtain their hop-count information. To maintain up-to-date information about the topology, beacons are sent at regular intervals. When a node broadcasts packets, all neighbors with lower hop-count value are potential candidate nodes to forward the packets, as shown in Fig. 4. Node *s* is the sender, hop-count (hc) of corresponding nodes are shown in Fig. 4 Even though it has two neighboring nodes, *b* and *c,* there is the only node *b* that will be the potential candidate to forward the packet. Since the only node *b* is having a lower hop-count (hc) value than sender *s*. If there is more than one potential candidate to forward the packets, there is a need to avoid the transmission of the same packet by more than one node. This can be achieved by using their depth as the second parameter to decide the forwarding timer of candidate nodes. A candidate node with the least forwarding timer value, whose timer expires first, will forward the packet first. Other nodes in its close vicinity are suppressing the transmission of the same packet after overhearing, to avoid the duplicate transmission.

Conclusion: It is a receiver-based routing protocol. It requires only hop-count and depth information of the node to decide the forwarding node. When a void-node or trap node receives the packet, it simply drops the packet since their hop-count value is equal or more than the sender. However, if candidate nodes are not present in each

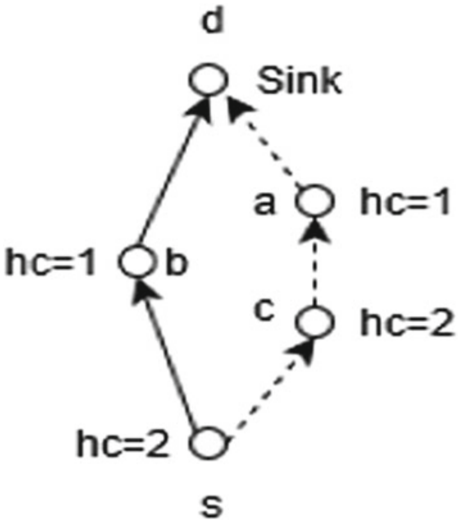


Fig. 4 Next-hop selection in IVAR

other's vicinity, there is a chance of transmission of duplicate packets which also results in a hidden-node problem. Duplicate packet transmission and hidden-node problem waste a considerable amount of energy of nodes.

C. Opportunistic Void Avoidance Routing (OVAR):

OVAR [20] is similar to IVAR except, OVAR follows a sender-based routing approach, and nodes in networks hold one-hop information. It avoids the limitations of IVAR by avoiding the duplicate packet transmission and hidden-node problem. In OVAR, the sender node selects the set of candidate forwarding nodes within the vicinity of each other. The number of nodes in the forwarding set is adjusted based on dense or sparse topology.

Conclusion: Although it overcomes the problems of hidden-node and duplicate packet transmission problems, the beacon period is critical in the correctness of the routing path. If the beacon period is too small, frequently beacon to be transmitted results in higher energy dissipation during beaconing. If the beacon period is too large, then hop-count information may be invalid.

D. Depth-Controlled Routing (DCR):

DCR [21] is based on the topology control mechanism. DCR requires location information of all nodes. X-Y coordinates are detected through Autonomous Underwater Vehicle (AUV) and Z coordinate obtained through on\-board sensors present in the individual nodes. DCR adjusts the depth of the disconnected (isolated) or void-node so that those nodes can communicate with suitable neighbors as per the greedy-forwarding mechanism. DCR works in two phases. First, it detects void-nodes and then, in the second phase, finds the candidate nodes and adjusts the depth of the void-node so that it can communicate with a suitable neighbor. To detect all void-nodes, a centralized Depth-first search algorithm is executed by all sink nodes as a root node. A node that is not reachable from any of the sink nodes or the sum of its distance to all sink nodes is zero are non-reachable nodes to sink. All such nodes are sorted from shallow water to deep water. In the second phase, starting from shallow water, a void-node at the center of the cylinder of a specified radius considers nodes reachable to sink as candidate nodes. Distance between void-node and all candidate node is calculated. The void-node will be moved toward a candidate node that requires minimum displacement.

Conclusion: In DCR, there is no void-recovery technique. Void-node is handled through node movement to the proper depth. Although there are some limitations like localization, detection of the void and depth adjustment of detected void-node require a considerable amount of time. Further, localization is costly in terms of energy consumption.

E. Hydrocast:

Hydrocast [22] operates in two phases greedy-forwarding based on pressure and void-recovery. In greedy-forwarding based on pressure, all nodes receive the packet to determine their progress toward the sink. A node nearer to the sink have more progress and high priority. Initially, the node with higher priority forwards the packet, and other lower priority nodes hear the packet transmission and suppress their transmission. If all higher priority nodes fail, then lower priority nodes will forward the packet. It can be done using the setting of a backoff timer. A higher priority node has a lower backoff time. To recover from void-node, every node can find its void status by comparing its and neighboring node depth. If a node is void, then it proactively finds the recovery path with a node having least-negative progress toward the sink.

As shown in Fig. 5, nodes *b* and *e* are void-nodes. Node *b* has two neighbors, nodes *c* and *d*. Node *b* selects *d* as its next hop because it is in the least-negative progress

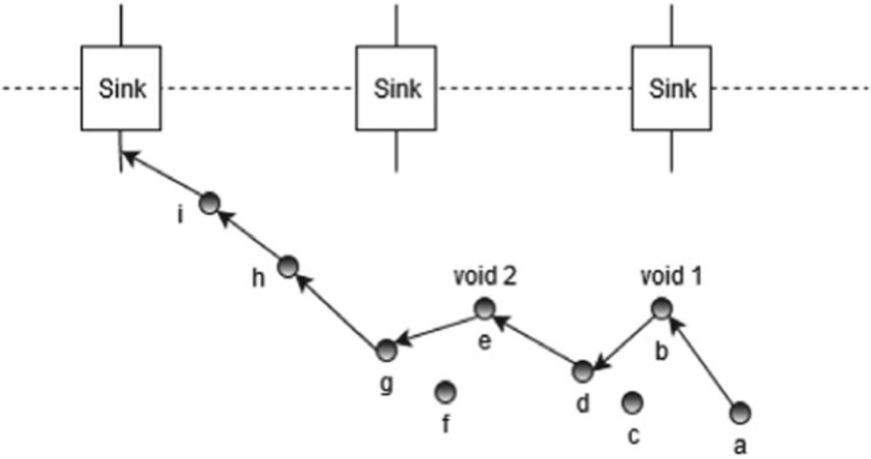


Fig. 5 Void-node recovery in Hydrocast [22]

than node *c*. Accordingly node *e* selects *g* as its next hop. Thus, every void-node is ready with their recovery path.

Conclusion: In Hydrocast, void-nodes proactively maintain recovery paths. However, void-handling in deep water was not addressed in protocol.

F. Channel-Aware Routing Protocol (CARP):

CARP [23] is an underwater routing protocol, it selects the next hop based on link quality. The successful packet forwarding ratio to the neighbor is maintained, and that information is used in the selection of next hop. Apart from link quality, CARP also uses hop-count to route the packets around the void region, residual energy of nodes, and buffer size in selecting the next hop.

Conclusion: CARP is an efficient routing protocol for dynamically changing underwater conditions by considering link quality in the selection of the next hop.

G. Interference-aware routing (Intar):

Interference-aware routing (Intar) [24] consists of two phases, the network setup phase and the data forwarding phase. During the network setup phase, the beacon is generated by the sink node and propagates throughout the network. Thereby every node updates its neighbors, hop-count of their neighbor, and Euclidean distance to the neighbor in the neighbor table.

In the data forwarding phase, every node identifies a Potential Forwarding Node (PFN). PDFs of a node are neighboring nodes present at lower depth. Further, sender/source node *i* finds the cost-function (*CFn*) of its every PFN node *j*. The CFn is computed as follows:

$$CFn(i, j) = \frac{D(i, j)}{\text{Hop}(j) * \text{Neighbor}(j)} \tag{4}$$

where $D(i, j)$ is the Euclidean distance between source/sender i to PFN node j, $\text{Hop}(j)$ is the hop-count distance between PFN node j and the sink, and $\text{Neighbor}(j)$ is the total number of neighbors to PFN node j. Finally, the source/sender node selects a PFN node with maximum CFn value as its next hop and forwards the data packets through the selected node.

Conclusion: Intar selects next hop from PFN having the least hop-count value which results in the selection of non-void-node as next hop.

4 Research Gaps

The features of the void-handling routing protocols are shown in Table 1. The location-based void-handling protocols are not efficient in terms of energy utilization due to the exchange of many control signals for localization compared with pressure-based routing. The void-handling strategy may be proactive or reactive or preventative. Proactive or reactive approaches have higher overhead. Routing can

Table 1 Features of the void-handling routing protocols

Protocol	Void-handling method	Beacon oriented	Void-recovery approach	Opportunistic routing	Architecture
VBVA	Back-pressure and vector-shift	No	Reactive	Yes	Single sink
AHH-VBF	Change in the width of the pipe and transmission power adjustment	No	Preventative	Yes	Single sink
FBR	Transmission power adjustment	Yes	Preventative	No	Single sink
DFR	flooding zone adjustment based on link quality	No	Hybrid	No	Single sink
LLSR	Hop-count	Yes	Preventative	No	Single sink
IVAR	Hop-count	Yes	Preventative	Yes	Single sink
OVAR	Hop-count	Yes	Preventative	Yes	Single sink
DCR	Depth adjustment	Yes	Preventative	No	Multiple sink
Hydrocast	Recovery path	No	Proactive	Yes	Multiple sink
CARP	Hop-count	Yes	Preventative	No	Single sink
Intar	Hop-count	Yes	Preventative	No	Multiple sink

be opportunistic or non-opportunistic. In opportunistic routing, a subset of neighboring nodes receives the data packets to forward. Among them, single/few nodes will forward the packet. If the transmission fails, other nodes will forward packets. The major pros of opportunistic results in improved reliability. On the other end, opportunistic routing causes duplicate packet transmissions if nodes are not in each other's communication range and results in the hidden-node problem, waste of node energy. In non-opportunistic routing, a node selects its next hop and then forwards packets to it. It may not ensure high reliability. The routing protocol architecture may be a single sink or multiple sinks. If protocol architecture supports multiple sink nodes, it is enough to deliver the packet to anyone sink out of many available sink nodes.

5 Conclusion

In this paper, we discussed the challenges and negative impacts of void-node in underwater communication. The major classification of Void-aware protocols is discussed. Further, next-hop selection criteria, void-handling strategy, and their pros and cons of each void-handling routing protocols are theoretically analyzed. Each void-handling routing protocols have pros and cons. It is important to select appropriate routing protocols depending on the application. In the end, some research gaps are discussed.

References

1. Akyildiz IF, Pompili D, Melodia T (2004) Challenges for efficient communication in underwater acoustic sensor networks. ACM Sigbed Rev 1(2):3–8
2. Heidemann J, Ye W, Wills J, A Syed, Li Y (2006) Research challenges and applications for underwater sensor networking. In: IEEE wireless communications and networking conference, 2006. WCNC 2006, vol 1. IEEE, pp 228–235
3. Akyildiz IF, Pompili D, Melodia T (2005) Underwater acoustic sensor networks: research challenges. Ad hoc Netw 3(3):257–279
4. Ghoreyshi SM, Shahrabi A, Boutaleb T (2017) Void-handling techniques for routing protocols in underwater sensor networks: survey and challenges. IEEE Commun Surv Tutor 19(2):800–827
5. Heidemann J, Stojanovic M, Zorzi M (2012) Underwater sensor networks: applications, advances and challenges. Philos Trans Royal Soc A Math Phys Eng Sci 370(1958):158–175
6. Kheirabadi TM, Mohamad MM (2013) Greedy routing in underwater acoustic sensor networks: a survey. Int J Distrib Sens Netw 9:701834
7. Tzu-Chiang C, Jia-Lin C, Yue-Fu T, Sha-Pai L (2013) Greedy geo-graphical void routing for wireless sensor networks. Int J Comput Electric Autom Control Inf Eng 7:769–777
8. Coutinho RWL, Boukerche A, Vieira LFM, Loureiro AAF (2016) Geographic and opportunistic routing for underwater sensor networks. IEEE Trans Comput 65:548–561
9. Noh Y, Lee U, Wang P, Choi BSC, Gerla M (2013) Vapr: Void-aware pressure routing for underwater sensor networks. IEEE Trans Mob Comput 12:895–908
10. Erol-Kantarci M, Mouftah HT, Oktug S (2011) A survey of architectures and localization techniques for underwater acoustic sensor networks. IEEE Commun Surv Tutor 13(3):487–502
11. Ang Eu Z, Tan H-P, Seah WKG (2010) Opportunistic routing in wireless sensor networks powered by ambient energy harvesting. Comput Netw 54(17):2943–2966
12. Xie P, Zhou Z, Peng Z, Cui J-H, Shi Z (2009) Void avoidance in three-dimensional mobile underwater sensor networks. In: International conference on wireless algorithms, systems, and applications. Springer, pp 305–314
13. Haitao Y, Yao N, Liu J (2015) An adaptive routing protocol in underwater sparse acoustic sensor networks. Ad Hoc Netw 34:121–143
14. Jornet JM, Stojanovic M, Zorzi M (2008) Focused beam routing protocol for underwater acoustic networks. In: Proceedings of the third ACM international workshop on Underwater Networks. ACM, pp 75–82
15. Hwang D, Kim D (2008) DFR: directional flooding-based routing protocol for underwater sensor networks. In: OCEANS. IEEE, pp 1–7
16. Xie P, Cui J-H, Lao L (2006) Vbf: vector-based forwarding protocol for underwater sensor networks. In: International conference on research in networking. Springer, pp 1216–1221
17. Nicolaou N, See A, Xie P, Cui J-H, Maggiorini D (2007) Improving the robustness of location-based routing for underwater sensor networks. In: Oceans 2007-Europe, p 18

18. Barbeau M, Blouin S, Cervera G, Garcia-Alfaro J, Kranakis E (2015) Location-free link state routing for underwater acoustic sensor networks. In: 2015 IEEE 28th Canadian conference on electrical and computer engineering (CCECE). IEEE, pp 1544–1549
19. Ghoreyshi SM, Shahrabi A, Boutaleb T (2015) An inherently void avoidance routing protocol for underwater sensor networks. In: 2015 international symposium on wireless communication systems (ISWCS). IEEE, pp 361–365
20. Ghoreyshi SM, Shahrabi A, Boutaleb T (2016) An opportunistic void avoidance routing protocol for underwater sensor networks. In: 2016 IEEE 30th international conference on advanced information networking and applications (AINA), pp 316–323
21. Rodolfo WL Coutinho, Luiz FM Vieira, Antonio AF Loureiro (2013) DCR: Depth-controlled routing protocol for underwater sensor networks. In: 2013 IEEE symposium on computers and communications (ISCC). IEEE, pp 000453–000458
22. Noh Y, Lee U, Lee S, Wang P, Vieira LFM, Cui J-H, Gerla M, Kim K (2015) Hydrocast: pressure routing for underwater sensor networks. IEEE Trans Vehi Technol 65(1):333–347
23. Basagni S, Petrioli C, Petroccia R, Spaccini D (2015) Carp: a channel-aware routing protocol for underwater acoustic wireless networks. AdHoc Netw 34:92–104
24. Javaid N, Majid A, Sher A, Khan W, Aal-salem M (2018) Avoiding void holes and collisions with reliable and interference-aware routing in underwater wsns. Sensors 18(9):3038

A Comprehensive Survey on Content Analysis and Its Challenges

Ankitha A. Nayak and L. Dharmanna

Abstract In recent years the explosive development and growth of video technology and multimedia have created a new challenge in the field of computer vision. The tremendous increase in multimedia exchange like video, audio, image, etc., through the internet and other social media has led the way to content analysis. Content analysis is a technique to interpret textual data, multimedia data, and communication artifacts. In this paper we have provided an overview of content analysis and the more profound interpretation of video content analysis in a different area is shown. Specifically, in this paper, we have focused more on affective content analysis, its methodology, trends, and challenges.

Keywords Affective content analysis · Content analysis · Direct approach · Emotion descriptor · Image · Video

1 Introduction

The advancement of network-based technologies and multimedia service has steadily promoted multimedia content in the network. This advancement has led to more difficulty in organizing, analyzing, and searching for required media content. And it is observed from the survey that content analysis can be used to overcome these issues.

Content analysis is a research approach to describe textual, multimedia data, and communication artifact. It is an efficient and replicated method to study different patterns in communication. The view of content analysis differs from domain to domain. It uses various analytical strategies to identify leading trends in data. Basically, Content Analysis consists of four steps: (1) Identification of data source: In this

A. A. Nayak (✉)
Department of CSE, NMAMIT, Nitte, India
e-mail: ankitu.1234@gmail.com

L. Dharmanna
Department of ISE, SDMIT, Ujire, India

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_70

phase, an appropriate data source is identified and analyzed. And the identified data should be able to answer all the raised queries. (2) Data categorization: In this phase, the refining of data is carried out. Here ambiguous data is removed. And the refined data are further categorized as a group. (3) Code data: This phase mainly works on decision rule. Based on the preliminary review, it decides the actual group the data belongs. (4) Access Reliability: In this phase, inter rate reliability is decided. (5) Result analysis: The outcome is reviewed and analyzed in this phase. The overview of content analysis methodology is shown in Fig. 1.

In general, the content analysis is carried out in two approaches, quantitative and qualitative approach, as shown in Fig. 2. The Quantitative approach explains data in terms of numeric and statistics. And Qualitative approach is exploratory research to analyze the descriptive data.

Since content analysis can be achieved on multimodal data like text, video, audio, image, etc., in this paper, we have primarily focused on Image and Video Content Analysis. The paper is organized as follows: in Sect. 2, we have discussed existing work on Image content analysis with a comparative study. In Sect. 3, the detailed survey on the Video content analysis approach and its methodology in a different area is outlined. Section 4 is about gaps, challenges in affective content analysis, and in Sect. 5 conclusion is outlined.

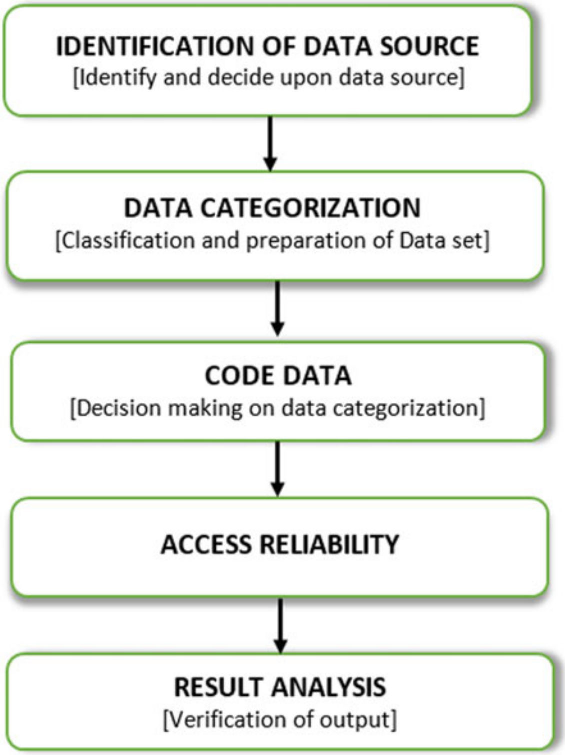

IDENTIFICATION OF DATA SOURCE
[Identify and decide upon data source]
DATA CATEGORIZATION
[Classification and preparation of Data set]
CODE DATA
[Decision making on data categorization]
ACCESS RELIABILITY
RESULT ANALYSIS
[Verification of output]

Fig. 1 Overview of content analysis methodology

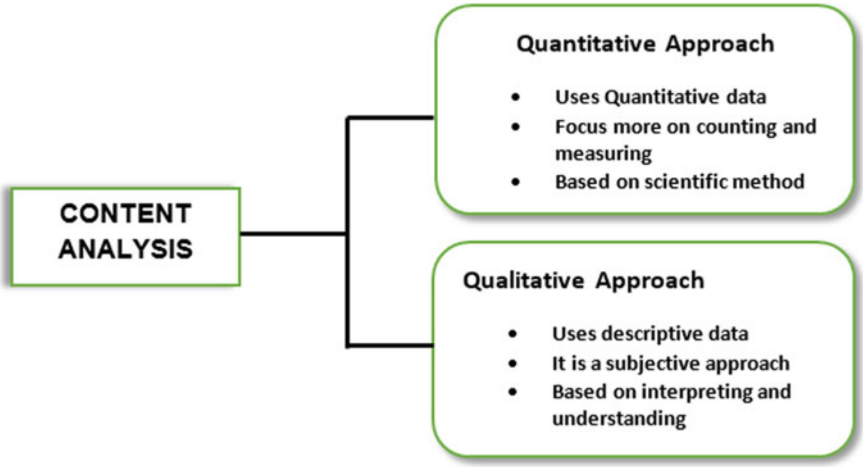

CONTENT ANALYSIS
Quantitative Approach
Uses Quantitative data
Focus more on counting and measuring
Based on scientific method
Qualitative Approach
Uses descriptive data
It is a subjective approach
Based on interpreting and understanding

Fig. 2 Overview of content analysis two different approach

2 Affective Image Content Analysis

A digital image is an array of a square aligned pixel in the form of rows and columns. It is a binary representation of visual information. Since there is rapid growth in photography technology, the need for image content analysis is expanded. In this section, we presented a brief review of existing work in image content analysis.

Sicheng Zhao et al. surveyed the affective image content analysis in [1]. They outlined research trends and methods in affective content analysis using different emotional representation model. The proposed Analysis of Image content module consists of three steps, (1) Human annotation (2) Visual feature extraction (3) Mapping between recognized emotion and visual features. In this paper, they summarized the emotion model, emotion feature extraction, and emotion distribution learning.

Zang et al. [2] designed a system to extract emotion from images in emerging news topics. The extracted emotion from the image is classified as positive, negative, and neutral. The work is carried out using two facial image data set. It is observed that high consistency is shown in facial expression for positive emotion and low correlation for neutral emotion. Mittal et al. conducted a survey on Image sentimental analysis using deep learning technique in [3]. In this paper, they have given an outline of existing work and analyzed suitability and limitation of the model used for image content analysis.

Garcia [4] explored the deep neural network for affective content analysis in the image. He used the semi-supervised model to overcome the challenges like lack of available data, unknown stimulus feature, and individuality of annotation and uncertainty of data. The classification accuracy achieved is 50%. He concluded that the deep learning model not only gives better classified output but also bring proper insight for visual information and emotion in the image. Bao et al. [5] proposed an Image system called ThuPIS for clinical mental health diagnosis. The system is used to diagnose different emotional reactions of people with different psychological states. The analysis is carried out using images captured in a specific time slot.

Li et al. [6] proposed a new approach to classify emotion in color image using a semi-supervised hierarchical classification algorithm. In this approach, two level classification is carried out using SVM and Adboost technique. The work concentrated on extracting three groups of features. Rao et al. [7] designed an emotion classification method based on Multi Scale block. This approach uses multi Instance learning technique. In this technique, the first extraction of the image block is achieved using pyramid segmentation and linear iterative clustering technique. Later each extracted block is represented using a bag of visual words method. To achieve the affective gap, probabilistic latent semantic analysis is used. Comparative Study on Affective Image Content Analysis is shown in Table 1.

3 Video Content Analysis

Video is a set of frames with attributes like size, speed, and clarity. Video Content Analysis is the methodology of identifying and analyzing temporal and spatial content. The video has different features to analyze like theme, emotion or particular user required data. Video Content Analysis is the trending and exciting area of research in recent years. Since the video is entertaining and informative multimedia, analyzing the content of the video is beneficiary for the consumer. Video Content Analysis has a vast spectrum of applications like safety enhancement, facial recognition, health care, home automation, smoke detection, intelligent recognition, and so forth. In many existing research work several machine learning, Data Mining, and Deep Learning models are combined with video processing to analyze the content of the video.

The structured video content analysis can be done in two ways. (1) Spatio-temporal structural analysis: Analyzing the content of video using pixel structure and objects of the same class is known as Spatio-temporal Structure Analysis. (2) Multimodal structure analysis: The chronological order event like running, jumping or speaking are combined with audio. Analyzing these kinds of features is known as Multimodal Structure Analysis. In most of the current work, it is seen that the video content analysis consists of the main four processes: Structure Analysis, Feature Extraction, Abstraction, and Indexing [8].

- **Feature Extraction**: The technique of extracting ROI from videos like color, shape or texture is known as Feature Extraction.
- **Structure Analysis**: In this process, temporal structure information of video is extracted and organized according to their relation and order.
- **Video Abstraction**: The process of presenting a piece of brief information about video structure is known as Video Abstraction.

Table 1 Comparative study on affective image content analysis

Reference	Objective	Methodology	Emotion Descriptor	Remarks
Zhao et al. [1]	Survey on affective image content analysis	Discussed the emotion model, emotion feature extraction	Not mentioned	Not mentioned
Zang et al. [2]	Affective classification of facial images	mRmR-feature extraction SVM-classifier	Positive, Negative, Neutral	Positive emotion accuracy—85% Negative emotion accuracy—33.3% Neutral emotion accuracy—39.4%
Mittal et al. [3]	Survey on image sentimental analysis using deep learning	Deep learning	Not mentioned	Not mentioned
Garcia [4]	Neural Network model for affective content analysis	Neural network semi supervised model	Not mentioned	Classification accuracy—50%
Bao et al. [5]	Image system for mental health diagnoses	MMPI	Positive Negative	Positive emotion has more accuracy
Li et al. [6]	Classify emotion in color image	Semi supervised hierarchical classification algorithm SVM, Adaboost	Static image, dynamic image, heavy and light image, warm and cool image	Better accuracy than the existing system
Rao et al. [7]	Emotion classification based on multi scale block	Linear iterative clustering technique	Positive emotions-amusement, awe, contentment, excitement Negative emotions - Anger, Disgust, Fear, Sad	Classification efficiency increased by 5.1% compared to the existing system

- **Indexing**: The method to generate video indices from extracted features, parsed video for clustering process or other process is known as Indexing.

Further, in this paper, we are presenting a detailed survey on Existing Sports video content analysis and affective video content analysis with a comparative study.

3.1 Sports Video Content Analysis

In this section, we are discussing existing work on sports video content analysis, the algorithm used and efficiency achieved by each work.

Shih et al. [9] surveyed on video content analysis for sports. They have provided a different view on themes and techniques available for sports video content analysis. The work is carried out by dividing a video content into three groups namely, event, context-oriented, and object. Finally, they have summarized challenges and future work in the discussed area. Guan et al. [10] developed a novel method to analyze and synthesize sports video based on visual content. In this technique, the analysis of video is achieved by automatic and manual interception of video. The analyzed video is stored in a separate folder. The group of pictures stored in the same folder is further used for content analysis. It is observed that the proposed method is an easy and efficient way to analyze sports video content.

Huayong et al. proposed a new approach for the sports video retrieval technique in [11]. The approach is designed for multimodal data. In this technique, the analysis is carried out for both video and audio feature. Video stream analysis is achieved by Shot by shot indexing. Audio analysis is achieved using audio energy called Interesting level. The designed approach has 91% precession value and 97% recall value which shows the designed approach is efficient and robust.

Russo et al. [12] contributed to work on the classification of sports video based on action. In this work, they have combined convolution and recurrent neural network with the deep neural network to achieve high accuracy. The accuracy achieved by the proposed system is 96.61%. Russo et al. [13] proposed an approach to classify five distinct classes of sports. In this work features extracted from the CNN model are combined with RNN temporal analysis. The accuracy achieved by this is 96.66% for different frame sequences. Comparative Study on Sports Video Content Analysis is shown in Table 2.

3.2 Affective Video Content Analysis

In general, the video content is analyzed and retrieved in two different levels. They are

- Cognitive Level
- Affective Level

The cognitive video content analysis intent to extract the information from the video, which defines the fact. Affective Video content analysis points at the extraction of feeling and emotion type in the video. In this section, we reviewed the affective content analysis. The survey is categorized based on the direct and implicit approach. The direct approach focus on analyzing affective content directly from visual and

Table 2 Comparative Study on Sports Video Content Analysis

Reference	Objective	Methodology	No. of dataset	Remarks
Shih et al. [9]	Survey on sports video Content analysis	Survey	Not mentioned	Not mentioned
Guan et al. [10]	Novel method to analyze and synthesize sports video	Automatic and manual interpretation of video	Not mentioned	Better accuracy than the existing system
Huayong et al. [11]	Sports video retrieval	Shot by shot indexing	3 data set	91% precession value
Russo et al. [12]	Classification of sports video	Convolution and recurrent neural network	2 data set	96.611%
Russo et al. [13]	Classification of 5 different sports	CNN and RNN temporal analysis	50 data set	96.66% for different frames

audio features of the video. The implicit approach focus on extracting the affective content from user response during video consumption

Affective Video Content Analysis- Direct Approach

In this section, we have discussed various work carried out on the direct affective content analysis.

Ashwin et al. [14] proposed a novel hybrid classifier named SVM-RBM classifier. Using audio-visual features this classifier finds the emotion for both stored video data and live streaming video. Compared to RBM and SVM, it is observed that the SVM-RBM classifier gives a better output for the annotated data set. Hanjalic et al. [15], proposed a new method “dimensional approach to affect” in multimedia content analysis. In this method, the affective content is feeling and emotion in the video towards a viewer, and the detected affective feature points are plotted in 2D emotion Plane. As an output, two curves are obtained representing arousal and valence video characters.

Hanjalic [16] designed an approach to analyze the excitement level of the user while watching the video. The approach is designed to enhance user comfort. In this methodology, the excitement level is calculated using motion activity and feature selection density cut. The excitement is calculated based on two levels comparability and smoothness. Kang [17] proposed a new technique to find the emotional events like sadness, fear, and joy from video data using the relevance feedback scheme. In this approach, the user is asked to give feedback on the video and the emotional content of the video. Further, this feedback is taken as data set. The system is trained to detect the emotion from the video and the data stored. In this technique, the low level features like color and motion is used to analyze the emotions.

Zheng et al. [18] proposed a model named visual-aural attention based on video content analysis approach. This method is used for automatic detection of highlights in the video. The visual and aural affective features are extracted from the highlighted section. A fusion strategy called ordinal decision is used in this methodology to form the attention curve for a video by identifying a change of human attention while watching the video. Furini [19] designed a novel mechanism ViMood to enhance and improve the indexing of video. In this approach, the indexing of the video is improved by integrating objective and subjective emotion. This approach indexes every emotion in a video. The objective is achieved by combining two methodologies on-the-fly viewers' emotion annotation and low level feature analysis. Comparative Study on Affective Video Content Analysis is shown in Table 3.

Affective Video Content Analysis- Implicit Approach

In this section, we have discussed various work carried out on Implicit affective content analysis.

Zhu et al. [21] designed a new approach to identify the user emotion from EEG signals and stimulus video together. In this methodology from each channel of EEG signal, both visual and audio features are extracted. Statistical analysis and canonical correlation analysis is used to select the extracted features. SVM is used as a classifier for further construction of EEG feature. Shahnaz et al. [22] proposed a new method to recognize the emotion of music videos based on wavelet analysis of empirical mode decomposed EEG signals. In this approach, DWT is applied to selected intrinsic mode function which is the outcome of EMD operation. The variance, kurtosis, skewness of a suitable DWT coefficient is used to form the feature vector. The feature vector is reduced and fed to SVM to perform emotion classification.

Wang et al. conducted a research to connect human brain waves with mind prediction [23]. In this approach, the brain waves are measured using the low pass, high pass, and notch filter. The captured signals are converted to DC and sent to the computer. The signals are processed and passed to support vector machine for classification. The classified signatures are processed by the Genetic Algorithms method. Later processed signals are categorized. Through this approach, it is possible to measure emotion and brain wave signal for a specific disease.

4 Challenges and Proposed Methodology

In this survey, we observed some of the gaps and challenges in the current work. The observed difficulties are discussed in brief in this section.

4.1 Challenges and Gaps Analyzed

In image content analysis, it is observed that understanding image content and context is difficult. Most of the current work is discussed either about the metadata of images like tags, description or contexts like a pattern, pixel relationship. Approaches on

Table 3 Comparative study on affective video content analysis

Reference	Objective	Methodology	Emotion descriptor	Remarks	Platform
Ashwin et al. [14]	Video affective content analysis	SVM-RBM classifier, Feature extraction	Not specified	78% accuracy	Machine learning, python
Hanjalic et al. [15]	Affective content analysis like emotion/feeling	Arousal modelling, Valence modelling, Motion activity	Arousal and valence	Arousal, valence and affective curve is obtained	MATLAB
Hanjalic et al. [16]	To find excitement of user while watching video	Density of cuts, motion activity and block based motion estimation	Arousal and valence	Excitement level extracted and labelled for different sequences	MATLAB
Kang [17]	Emotional event detection using relevance feedback	Using low level features, histogram difference, short term memory model	Fear, sadness, joy	Detection rate with 76% accuracy	MATLAB
Zheng et al. [18]	Visual Aural attention modelling from video highlight extraction	Visual Arousal and valence arousal attention model, support vector machine	Not specified	Encouraging result with 86% accuracy.	Not specified
Furini [19]	Computation of objective and subjective emotion	Pultchick model, video segmentation, emotion visualization, 5 point likert scale	Joy, sad, surprise	Viewers' evaluation is done on 3 factors: satisfaction, ease of use and perceptiveness and obtained result with mean 4, 3, 3 respectively	Not specified
Hanjalic et al. [20]	Affective video content analysis	2D emotion space, mapping low level feature with valence and arousal curve	Arousal and valence	Separate arousal and valence curve is achieved with affect curve	MATLAB

the context and content analysis of the image are presented in few papers and the accuracy achieved is 85–86%. Interpretation of both the content and meaning of the image can increase efficiency.

Videos and Images consist of a group of object, emotion recognition for a group of objects is another challenging task. In video content analysis, a more accurate and efficient result is obtained by using machine learning algorithms and CNN. And maintaining GPU Memory is a challenging task in the CNN platform. No efficiency or execution time analysis is achieved for current literature work, and the approach discussed can be considered as better if adaptability and execution time is better.

It is observed that an increase in the dataset is decreasing the efficiency of the current approaches. Most of the procedure is using the SVM classifier for the content classification. In the machine learning algorithm, it is essential to pick a proper plan, algorithm, and data set. A very less novel algorithm and approach are seen in the existing system.

Since we are focusing on affective content analysis of video, the current approach is working on a single modality. Very few techniques work on the multi-modality of video and the efficiency achieved is not satisfying. The classification and analysis of the content is performed separately, which increases the execution time.

5 Conclusion

In this paper, we have illustrated a thorough review of the content analysis approach. Different field of content analysis is discussed. The existing algorithm and methodology in image content analysis and video content analysis is analyzed and discussed briefly with a comparative study. The more focus is given on affective content analysis in image and video. The main contribution of this paper is the gaps, limitations, and challenges of the current system which are illustrated.

References

1. Zhao S, Ding G, Chua T-S, Schuller BW, Keutzer K (2018) Affective image content analysis: a comprehensive survey. In: Proceedings of the twenty-seventh international joint conference on artificial intelligence (IJCAI-18)-2018
2. Zhang L, Han Lau AC, Tjondronegoro D, Chandran V (2014) A pilot study on affective classification of facial images for emerging news topics. In: IEEE 16th international workshop on multimedia signal processing (MMSP), Jakarta, Indonesia, 22–24 Sep 2014
3. Mittal N, Sharma D, Joshi ML (2018) Image sentiment analysis using deep learning. In: IEEE/WIC/ACM international conference on web intelligence (WI)
4. Garcia AH (2017) Perceived emotion from images through deep interaction neural networks. In: 2017 seventh international conference on affective computing and intelligent (ACII)
5. Bao S, Ma H, Li W (2014) ThuPIS: a new affective image system for psychological analysis. In: 2014 IEEE international symposium on bioelectronics and bioinformatics (IEEE ISBB 2014)

6. Li N, Xia Y, Xia Y (2015) Semi-Supervised emotional classification of color images by learning from cloud. In: 2015 international conference on affective computing and intelligent interaction (ACII)
7. Rao T, Xu M, Liu H, Wang J, Burnett I (2016) Multi-Scale blocks based image emotion classification using multiple instance learning IEEE
8. Dimitrova N, Zhang H-J, Shahraray B, Sezan I, Huang T, Zakhor A (2002) Applications of video-content analysis and retrieval. IEEE
9. Shih H-C (2017) A survey on content-aware video analysis for sports IEEE Trans Circuits Syst Video Technol 99(9), Jan
10. Guan Y-P, Li JJ, Ye Y, Si J, Zhang H (2011) Content based sports video sequence analysis and synthesis-IEEE
11. Liu H-Y, Zhang H (2005) A sports video browsing and retrieval system based on multimodal analysis: Sportsbr. In: Proceedings of the fourth international conference on machine learning and cybernetics, Guangzhou, 18–21 Aug 2005
12. Russo MA, Kurniang goro L, Jo K-H (2019) Classification of sports videos with combination of deep learning models and transfer learning. In: 2019 international conference on electrical, computer and communication engineering (ECCE), 7–9 Feb 2019
13. Russo MA, Filonenko A, Jo K-H (2017) Sports classification in sequential frames using CNN and RNN, IEEE
14. Ashwin TS, Saran S, Mohana Reddy GR (2016) Video affective content analysis based on multimodalfeatures using a novel hybrid SVM-RBM classifier. In: 2016 IEEE Uttar Pradesh section international conference on electrical, computer and electronics engineering (UPCON), Indian Institute of Technology (Banaras Hindu University) Varanasi, India, 9–11 Dec 2016
15. Hanjalic A, Xu L-Q (2001) User-oriented affective video content analysis. IEEE
16. Hanjalic A (2003) Multimodal approach to measuring excitement in video. In: ICME, IEEE
17. Kang H-B (2003) Emotional event detection using relevance feedback. IEEE
18. Zheng Y, Zhu G, Jiang S, Huang Q, Gao W (2008) Visual-aural attention modeling for talk show video highlight detection. ICASSP, IEEE
19. Furini M (2015) ViMood: using social emotions to improve video indexing. In: 2015 12th annual ieee consumer communications and networking conference (CCNC)
20. Hanjalic A, Xu L-Q (2005) Affective video content representation and modeling. IEEE Trans Multimed 7(1), Feb
21. Zhu Y, Wang S, Ji Q (2015) Emotion recognition from users' eeg signals with the help of stimulus videos. IEEE
22. Shahnaz C, Shoaib-Bin-Masud, Shafiul Hasan SM (2016) Emotion recognition based on wavelet analysis of empirical mode decomposed EEG signals responsive to music videos. In: 2016 IEEE region 10 conference (TENCON)—Proceedings of the international conference
23. Wang M-L, Lin C-W, Mayer NM, Hu M-H, Lee P-Y. An brain-computer interface for video content analysis system for perceive emotions by using EEG. In: 2016 international conference on consumer electronics-Taiwan

Applications of Blockchain and Smart Contract for Sustainable Tourism Ecosystems

Jaehun Joo, Joungkoo Park, and Yuming Han

Abstract Blockchain is one of the most promising technologies for innovating business ecosystems in tourism. Blockchain enables secure, reliable, and efficient decentralized management systems without trusted third parties which are a core part of centralized management systems. Many tourism business activities have been doing under globalized and decentralized environments. Thus, It is necessary to do research regarding applications of blockchain technology to the tourism industry for building sustainable tourism business ecosystems. The purpose of the present paper is to examine application cases of blockchain and smart contract to the tourism industry and to identify the opportunity to innovate existing tourism business ecosystems. Stakeholders of tourism business ecosystems are able to innovate their ecosystems being conducive to their business using blockchain technology. Application cases of blockchain to tourism such as TripEcosys and TravelChain can be helpful to tourism business managers who seek new opportunities for innovative businesses.

Keywords Blockchain · Ethereum · Smart contract · Bitcoin · Tourism · Business ecosystem · Tourism business ecosystem

J. Joo (✉)
Department of Information Management, Dongguk University, Gyeongju-si, South Korea
e-mail: givej@dongguk.ac.kr

J. Park
Department of Hotel Tourism & Food Management, Dongguk University, Gyeongju-si, South Korea
e-mail: jkpark85@dongguk.ac.kr

Y. Han
Department of International Business Management, Dongguk University, Gyeongju-si, South Korea

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_71

1 Introduction

Blockchain is one of the most promising technologies for innovating business ecosystems in tourism. Blockchain enables secure, reliable, and efficient decentralized management systems without trusted third parties which are a core part of centralized management systems. A variety of industrial sectors, as well as finance and cryptocurrency, have been introducing and applying blockchain technology with the smart contract because of uncertainty releasing, transparency improvement, trust-building, and transaction cost reduction. Many tourism business activities have been doing under globalized and decentralized environments. Thus, It is necessary to do research regarding applications of blockchain technology to the tourism industry for building sustainable tourism business ecosystems.

The purpose of the present paper is to examine application cases of blockchain and smart contract to the tourism industry and to identify the opportunity to innovate existing tourism business ecosystems. After this introductory section, Sect. 2 introduces the concept and current trends of blockchain-related technologies. Section 3 identifies the opportunity to innovate business ecosystems in tourism through blockchain technology. Section 4 examines application cases of blockchain and smart contract in tourism and its business ecosystems. Section 5 concludes and suggests implications.

2 Blockchain and Smart Contract

2.1 Bitcoin and Blockchain

Bitcoin is the first cryptocurrency based on blockchain [1]. Components of blockchain technology are as follows:

- Distributed peer-to-peer network: Unstructured P2P network using flooding algorithm and TCP/IP.
- Public key cryptography and hash algorithm: Elliptic-curve cryptography (ECC), SHA-256 hash algorithm, and Merkle tree for verifying data integrity.
- Consensus algorithm: A method of consensus decision making among participants which are nodes of the blockchain network, where a new block which is a set of valid transactions for a given time is added to the existing blockchain. There are various consensus algorithms such as proof-of-work (PoW), Proof-of-stake (PoS), Delegated proof-of-stake (DPoS), and Tendermint.
- Smart contract: a self-executing contract with the agreement of contractural parties in the case of satisfying common contractural conditions written in computer codes containing a set of rules.
- Distributed ledger: storage of transaction records that are consensually shared, replicated, and synchronized among participants in a distributed network.

Each block is composed of a header and a list of transactions represented as a Merkle tree. Each block contains a hash value of the previous block, a hash value of the block, a nonce, a timestamp, and a list of transaction data. Blocks are chained chronologically in the blockchain using a hash. The hash function has two characteristics: one-way function and unique value. If any data of a block are changed, the hash value should be changed. In order for an actor of a node in the blockchain network to change the content of a block, he or she must change all block of the blockchain. Thus, in reality, it is computationally hard to tamper the distributed ledge, although each node of the blockchain network stores a copy of the ledger. Cryptography and digital signature assure security services such as authenticity, confidentiality, data integrity, and non-repudiation.

Blockchain is classified into three categories as follows:

- Public blockchain: Known as permissionless blockchains such as Bitcoin and Ethereum.
- Private blockchain: Permissioned blockchain such as Hyperledger Fabric.
- Consortium blockchain: A shared ledger can be operated by multiple banks.

2.2 Ehereum and Smart Contract

Ethereum is a programmable blockchain and an open platform based on blockchain technology with a cryptocurrency (Ether, ETH). Ethereum enables a community of developers who are protocol developers, blockchain researchers, miners, and applications, developers, to build and deploy decentralized applications for various areas of finance, game, and other markets. Ethereum ultimately seeks to create "a World Computer: a huge network of many private computers that run various internet applications without any third parties [2]."

Since Szabo [3] introduced the concept of the smart contract which is defined as "a computerized transaction protocol that executes the terms of a contract", the smart contract has been diffused and extended a blockchain functionality. Smart Contracts. Smart contracts are stored on blocks of the blockchain as scripts. There are many computer programming languages for smart contracts such as Java, C++, Python, and JavaScript. Solidity which is turning complete scripting language is used to write smart contracts in Ethereum. Ethereum virtual machine is used to execute smart contracts [4].

3 Tourism Business Ecosystems

Business ecosystems are defined as mutually dependent systems which are composed of various actors such as partners, suppliers, customers, financial organizations, trade associations, standards bodies, labor unions, government, and quasi-governmental organizations as stakeholders [5, 6]. Moore [7] viewed the business ecosystem as an

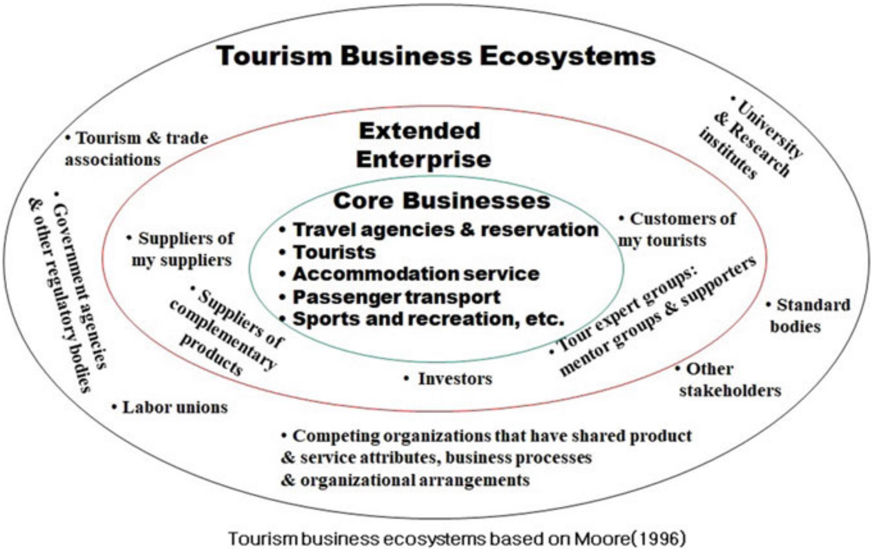

Tourism Business Ecosystems
Extended Enterprise
Core Businesses
• Travel agencies & reservation
• Tourists
• Accommodation service
• Passenger transport
• Sports and recreation, etc.
• Tourism & trade associations
• Government agencies & other regulatory bodies
• Suppliers of my suppliers
• Suppliers of complementary products
• Investors
• Tour expert groups: mentor groups & supporters
• Customers of my tourists
• University & Research institutes
• Standard bodies
• Other stakeholders
• Labor unions
• Competing organizations that have shared product & service attributes, business processes & organizational arrangements
Tourism business ecosystems based on Moore(1996)

Fig. 1 Tourism business ecosystems

economic community which is in the same boat and has a shared fate in the business world. The business ecosystem includes competitors as its actor [5]. According to [5], productivity, robustness, and niche creation are three measures of the business ecosystem health. They suggested three players including keystones, physical dominators, and niche operators [5].

Figure 1 shows tourism business ecosystems based on Moore's [7] study regarding business ecosystems. Core businesses of tourism business ecosystems include accommodation services, food & beverage services, travel transportation services, transportation supporting services, rental services, recreation & sports services, and souvenir retailers, as well as online & offline travel agencies and reservation services.

4 Application Cases of Blockchain to Tourism Industry

There are various application areas of blockchain to the tourism industry such as improving trust among partners, facilitating better disintermediation, securing travel transactions, and providing trustworthy loyalty programs, traceable tourism products and activities, and reliable online travel reviews [8].

Blockchain technology enables tourism consumers to achieve the optimization of their experiences through their personalized services from travel service providers. Customers as travelers increasingly become co-creators of travel businesses by forging more cooperative relationships in tourism ecosystems [8]. Travel customers

Table 1 Application areas of blockchain technology

Areas	Overviews
The whole trip and travel reservations	There are some limitations for applying public blockchains to whole trip processes because of scalability and speed of transaction validity
Travel identity	This area of tourism is suitable to apply blockchain
Payments of travel transactions	This is a suitable area under the collaboration of various financial organizations
Loyalty services	This service is a suitable area under the collaboration of various travel service organizations and financial institutions
Guarantee of reservation and payments	This area is partially suitable to apply blockchain
Ticketing	This area is also partially suitable to apply blockchain but has some risks for processing legacy systems
Business to business payment & settlement	This area is suitable to apply private or consortium blockchain
Inventory management	Blockchain and smart contract can be used to control inventory in the tourism industry
Tour exchange	Currency exchange for travelers across the partner's ledgers

Source [9, p. 9]

can play a role of proactive actors having controllability or influence over tour-related information and content. According to [8], Decentralized systems based on the blockchain technology improve following functionalities rather than existing centralized systems in the tourism industry: transparency because of providing full auditability and traceability; securing travel-related transactions; disintermediation in travel activities; new opportunities for building creative loyalty programs to travel service providers; enhancing food tourism traceability and baggage tracking through a smart contract paying automatically for any overweighting luggage; authentic and reliable online reviews.

Table 1 shows suitable application areas of blockchain to the tourism industry [9].

4.1 TripEcoSys

TripEcosys is a project building a travel platform enabling the following functions: Travel blog, social travel review, adverting service for travel service providers, and cryptocurrency. "TripEcoSys is a decentralized travel ecosystem that supports community building and social interaction with cryptocurrency rewards" [10].

TRIP-TALK is a platform based on blockchain providing a tour blog allowing travelers to log their tour stories and travel experiences, and to have the cryptocurrency called Tripcash as an incentive. One of the Triptalk features is to evaluate users' reputation through a Triptalk voting mechanism [11].

TripPay is a cryptocurrency based on Ethereum aiming to secure the interoperability enabling to merge, divide, and exchange of the travel assets. Furthermore, the project is developing an open APIs for providing interconnection to partners who may employ to programmatically get interactions to their wallet.

4.2 Travelchain

TravelChain is a blockchain-based travel platform in which travel-related participants can share their data and experiences. All players in the tourism marketplace can develop their travel service like a hotel, reservation, and trip schedule management by using APIs of the TravelChain and can access to data exchange for providing personalized services. Travelers receive TravelTokens as rewards for sharing their own data. Data in the Travelchain are stored with ontology standards and rules. Travelchain uses DPOS (delegated proof-of-stake) as the consensus algorithm which appends a block to the blockchain for validating secure transactions. TravelTokens is a cryptocurrency and employed as a payment token for all transactions in the Travelchain ecosystem.

About 20% of travel data is opened to players of the tourism market, whereas, 80% of the data are stored in servers of a few enterprises such as Google, Apple, and Amazon. TavelChain aims to allow users or players of the tourism market who are registering to the TravelChain system to use travel data for providing their customized services.

4.3 DeskBell Chain

DeskBell Chain is a blockchain-based platform project for marketing in areas of hotels and travel businesses based on the existing DeskBell service which is a business service for hotels with mobile and service applications. DeskBell Chain aims to help hotel businesses offer information to travelers and interact with them via encrypted chat [12].

Ethereum based DeskBell token is used as not only the currency for monetization of all operations on the platform but also used to reward participants and players on the DeskBell system.

4.4 Winding Tree

Winding Tree is a decentralized travel distribution platform based on the blockchain and smart contract. It connects travel service providers and consumers by allowing participants to use a set of smart contracts, open APIs, and Lif token as a cryptocurrency, which is operating on the Ethereum platform [13].

4.5 TravelFlex

TravelFlex is a cryptocurrency based on blockchain supporting a decentralized social travel network. It aims to reduce the high fees of currency conversions for realizing an efficient travel network and also to serve as an escrow service for secure transactions between travel service providers and travelers [14, 15].

5 Conclusions

Blockchain technology contributes to improving transparency, security, trust, convenience, and transaction cost reduction through decentralized management. Stakeholders of tourism business ecosystems are able to innovate their ecosystems being conducive to their business using blockchain technology. Application cases of blockchain to tourism such as TripEcosys and TravelChain can be helpful to tourism business managers who seek new opportunities for innovative businesses.

Acknowledgments This work was supported by the Dongguk University-Gyeongju Research Fund of 2019.

References

1. Nakamoto S (2008) Bitcoin: a peer-to-peer electronic cash system
2. district0x, What is Ethereum. https://education.district0x.io/general-topics/understanding-ethereum/what-is-ethereum
3. Szabo N (1994) Smart contact
4. Christidis K, Devetsikiotis M (2016) Blockchains and smart contracts for the internet of things. IEEE Access 42:2922–2303
5. Iansiti M, Levien R (2004) Strategy as ecology. Harvard Bus Rev 82(3):68–81
6. Baig Z, Zeadally S (2018) Cyber-security risk assessment framework for critical infrastructures. Intell Autom Soft Comput. 1–1
7. Joo J, Eom T, Shin M (2016) Executive practices for corporate sustainability: a business ecosystems perspective. Int J Bus Res 16(1):133–146
8. Moore JF (1996) Death of competition: leadership and strategy in the age of business ecosystems. HarperBusiness, New York, NY

9. Rejeb A, Rejeb K (2019) Blockchain technology in tourism: applications and possibilities. World Sci News: Int Sci J 137:119–144
10. Travelport (2018) Blockchain and distributed ledger technology at Travelport. https://www.travelport.com/sites/default/files/travelport-blockchain-whitepaper.pdf
11. Tripecosys (2020) Whitepaper. https://www.tripecosys.com/web/default/doc/whitepaper.pdf?v=120
12. TravelChain (2018) TravelChain whitepaper. https://icorating.com/upload/whitepaper/17GgBCxNd1PEP2kPDGUrfP8UPKTqUzgdpCbP5Wju.pdf
13. DeskBell (2018) Blockchain platform of the hotel and tourism industry DeskBell chain v.1.0. https://icorating.com/upload/whitepaper/fbK49ILm9wD2X1VEwsVcaMNzC9lPnSIfmAybmJ77.pdf
14. Izmaylov M, Anderson P, Lemble A, Vysoky J (2018) A practical application of blockchain for the travel industry. https://windingtree.com/White_Paper_EN.pdf
15. TravelFlex (2017) TravelFlex whitepaper. https://icosbull.com/whitepapers/10361/TravelFlex_Coin_whitepaper.pdf

Friendship and Location-Based Routing in Delay Tolerant Networks

Nidhi Sonkar, Sudhakar Pandey, and Sanjay Kumar

Abstract The network with the delay tolerant capabilities enables communication in the environment where intermittent connectivity or peer-to-peer connection is available. Routing for these networks is a challenging problem, therefore, the concept of Social Delay Tolerant Networks has been proposed by researchers. Friendship is a property of Social DTN where the friendship between nodes is used for transmission of a message between nodes. The problem arises in friendship-based routing is when any node doesn't have any friend it cannot send the data to the destination. Therefore, we are proposing a novel algorithm for routing based on friendship and location to overcome the problem in friendship based routing by adding the location-based similarity which is another social property of DTN with friendship based routing algorithm. Therefore, Delivery probability is heightened and the average latencies caused to find a friend can be decreased.

Keywords Delay tolerant networks · Social DTN · Friendship · Similarity · Delivery probability · Delay · ONE sımulator

1 Introduction

Network paradigm with the tolerable delays [1] is a type of paradigm with the capability to communicate in sparse networks where the peer-to-peer connectivity is lacking. Networks with tolerable delays are useful in the wireless networks where such types of problems exist like intermittent connectivity, long and variable delay, etc. DTNs solve this type of problem by using a store-carry-forward technique [2]. DTNs gain much attracted as it well suited for the challenging network areas which are Terrestrial Network, Military Network, Space Network, and Maritime Network, etc.

The routing options in the network with delay tolerance is quite tedious and complicated due to intermittent and sparse connectivity, Social Networks a property

N. Sonkar (✉) · S. Pandey · S. Kumar
National Institute of Technology, Raipur, India
e-mail: nidhisonkar055@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_72

is useful to ease the challenges of routing and for a node to select relay nodes. Facts of social characteristics are used in social networks like how human beings communicate, meet, and transfer their talks to each other [3]. Social Delay Tolerant Networks adopt the social behaviors of humans because they carry mobile phones and when two people meet their mobile phone can easily get communicate to each other that means one mobile can transfer data to other easily. There are multiple properties of Social DTN like Community [4], Centrality [5], Friendship [6], Similarity [7], Interest [8], and Selfishness [9], etc.

In this article, we are using the property of Friendship and Location-based Similarity for proposing new algorithms for routing in Networks with the delay tolerance. The problem occurs in Friendship based routing [10] is, if the source node would not meet with the friend of destination node then delivery would fail and no communication could be initialized. Therefore, we introduce the concept of location-based similarity with the friendship based routing, so usually when the data point at the source is not able to meet to the friend of destination then conveys the information to the node which is going to the similar location as the destination and if the relay node finds the friend of destination then it forwards the message to the destination otherwise it direct transmit message to the destination. Therefore, by the proposed routing scheme the delivery probability can be increased and the delay can be reduced.

The paper is arranged accordingly: The overview of previous work is presented in Sect. 2. The design of the proposed routing scheme is presented in Sect. 3. In Sect. 4, we discussed simulation and results. In Sect. 5, we conclude the paper by the conclusion and future work.

2 Related Works

The portion discusses about the general routing algorithms in Delay Tolerant Networks and then, the path identification based on the social behaviors that follows the put forth and the features of the network based on its social behaviors especially in Delay Tolerant Networks. In DTN routing Schemes firstly, we discuss Multicopy based routing schemes. The initial routing scheme in DTN is Epidemic [11] routing, put forth endures high bandwidth, buffer space, and energy even though it enjoys heightened delivery rate at minimized latency. Therefore, new routing schemes were put forth for bringing down the replicated counts with a heightened delivery ratio. Which are wait as well as spray [12], and spray and wait at multiperiods [13]. The First scheme is considered in a single copy based DTN routing scheme is PROPHET [14], in which the history of encountering of the node is considered for prediction of future movement of the node. For, e.g., the node visiting numerous locations there are probabilities of visiting the same node/location in the future. MaxProp [15] This routing scheme achieves better performance when limited resources are available.

In Social-based DTN routing Schemes many recent types of research have focused on social-based routing algorithms (mainly a human with mobile phones) to get the effective routing algorithm in Delay Tolerant Networks. We discuss them as follows:

Authors in [16], used both similarity and betweenness centrality to form novel algorithms for path identification that has better performance. In [17], Authors proposed a novel algorithm for identifying the paths named Bubble Rap in where every datapoint (node) have two rankings that are global and local. This algorithm got high attention due to the high delivery ratio. An epidemic with routing that is based on the community is provided in the [18]. The effect of the socially selfish routing algorithm is presented in [19]. The paper proffers a new routing that relies on social behavior algorithm to overcome the problem in the routing relying on the friendship. The algorithm uses the property of the social DTN friendship and similarity to propose a new algorithm in which location is calculated by Marcov model [20]. The methodology of the proposed routing algorithm is presented in the next section.

3 Proposed Methodology

Figure 1 shows the model for Friendship and Location-based routing, in this model when any node comes in the contact range of another node they share information with each other. The information includes Encounter time, contact history, previous transmission history, etc. For transmission node first, check for friendship and if friendship does not exist then only, the node goes for location-based similarity. That means it overcomes the problem of Friendship based routing that if any node doesn't have any friend they can also transfer messages by transferring the message to a node that is going to the location of friend of node or destination. The methodology to calculate friendship between two nodes and to predict the location of the node is defined as follows:

3.1 Friendship Between Nodes

Friendship is a concept of sociology which defines a close relationship between two persons. Delay Tolerant Network used this concept in nodes, two nodes are called as friends if they have long-lasting connectivity with regular meeting [3]. In sociology, if two people have common properties like similar interests, similar actions, and they meet frequently with each other then only they can be friends [21]. DTN is also inspired by sociology two nodes can be friends if they have similar contact history [22] or common interest [23]. In this article, we are using a similar contact history for defining the friendship between two nodes.

For analyzing the relationship between nodes we have to analyze the quality of the link between two nodes. To find the better link quality that gives the better relationship between nodes we are using three features of behavior of node: Longevity, high frequency, and regularity. That is the friendship between two nodes will be strongest if they have frequent and regular connectivity with long-lasting. Here the term frequent

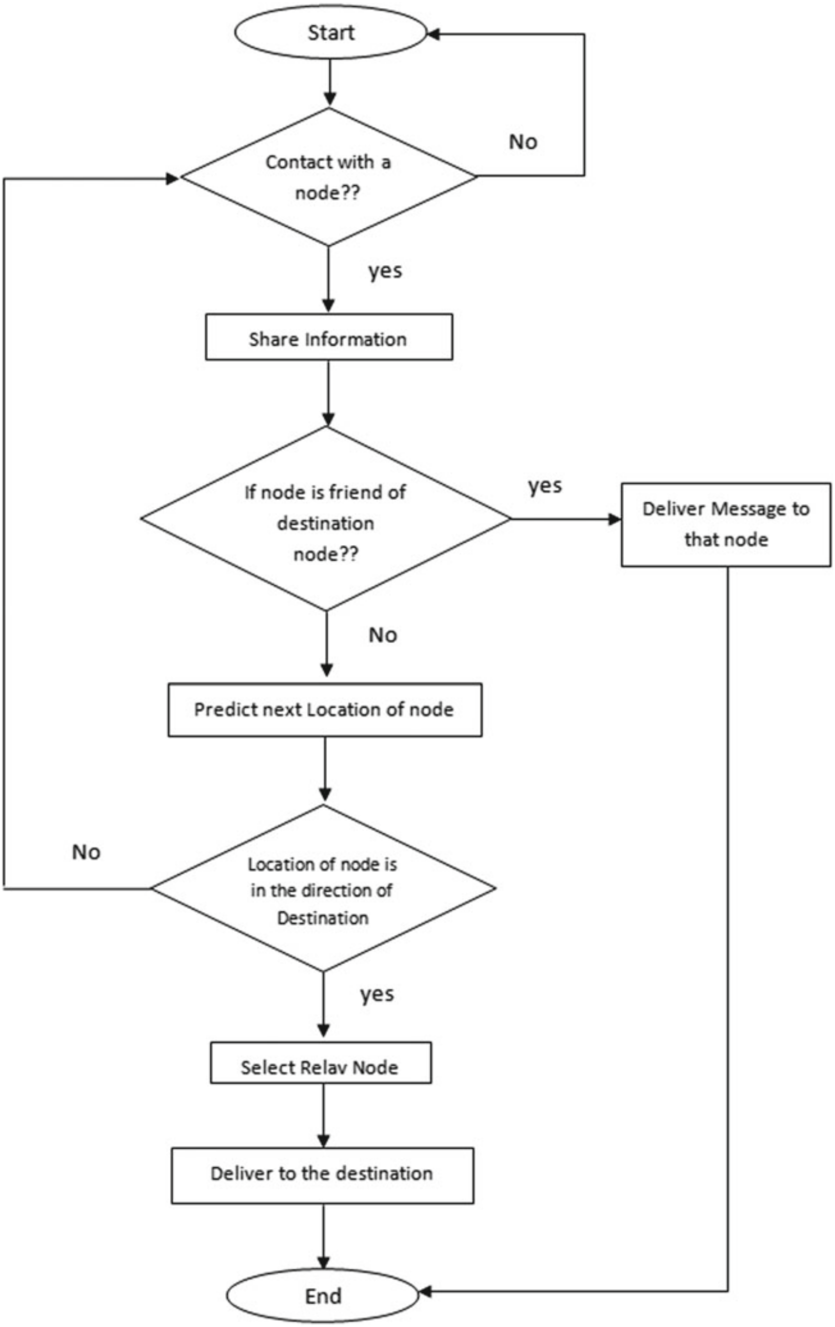

Start
Contact with a node??
No
yes
Share Information
If node is friend of destination node??
yes
Deliver Message to that node
No
Predict next Location of node
No
Location of node is in the direction of Destination
yes
Select Relav Node
Deliver to the destination
End

Fig. 1 Model for proposed methodology

and regular are different, two nodes not regular but frequent or regular but not frequent (like once in a week) and still called as friend but friendship is weaker in this scenario than the nodes meet regular and frequent both. All nodes can find link quality with other nodes by their contact history and forward the messages which have higher link quality than others.

3.2 Location-Based Similarity

The MobySpace or Ecludean virtual space has been proposed for taking decision for routing in DTN [24], in this method routing decision is taking place by selecting the node which has similar mobility pattern to the destination node. For characterization of the mobility of any node historic information about contacts is used that the node already had. We can calculate the probability by the history of each node that they can move in the location of the destination or not. For example, if a person goes to office every day at a particular time then we can get the probability that any day he can go to that location. Another example is, a student goes to school at a particular time from home to school and after school, he goes to tuition from home to tuition classes. İn this way can find the probability of going to particular location of any node that is similar to destination or not.

The limitations of this method are the prediction is only based on probability not accurate. For example, we only consider that a person every day arrived in particular location but if due to any reason he can not go to office in that day we could not send the data to that node. Therefore, we are extending this work of location prediction by Artificial Intelligence, so we can predict the location of node accurately.

3.3 Forwarding Strategy

In the proposed routing algorithm we are using a friendship based routing algorithm with location-based similarity for recovering the problems in friendship based routing scheme. The problem in friendship based routing is when any node does not have any friend they could not send data to the destination for this we consider that if any node wants to send message who does not have any friend also can send the data. For this, we are taking the location-based similarity scheme for routing, that is, when any node does not have any node they can send data by the node who are going the location of the destination. For forwarding strategy an algorithm has been proposed which is as follows:

Proposed routing algorithm

[Input: History of all nodes, Encounter Time, Destination node.]
[Output: Data transferred to the destination.]

```
Get information about nodes that comes in the range of contact.
Find the friendship between nodes and the destination.
if the node has a friendship with the destination
find the node that has the closest friendship
transfer data to that node
else
find the node that is going to the location of the destination
transfer the data which is having the highest probability
Deliver data to the destination.
End
```

4 Simulation and Results

ONE Simulator abbreviated as Opportunistic Networking Environment Simulator is designed especially for routing in Delay Tolerant Networks. We are using ONE simulator for simulation of the proposed strategy of routing to estimate the functions of that strategy of routing that was out forth is better than existing algorithms. The parameters used in the multi hop DTN routing algorithm for maritime networks are given in Table 1.

Figure 2 shows the impact of a number of nodes to probability, it is obvious that friendship and location-based routing gives higher delivery probability than friendship-based because when any node who want to send message to the destination and cannot find the node that is friend of destination who can also send the data with no delay by sending the data to the node that is going towards the destination, therefore, average latency is brought down by proposed routing algorithm as shown in Fig. 4. The minimized delay increases as the node count opportunity for transferring data increases, therefore, when the node count and buffer size increases, the delay decreases as shown in Figs. 3 and 5.

Table 1 Simulation parameters

Parameter	Value
Transmit range	100 m
Buffer size	20–100 MB
Node count	21–100
TTL (Message)	300 min
Size of message	100 b
Simulation time	5000 s
Number of locations	25

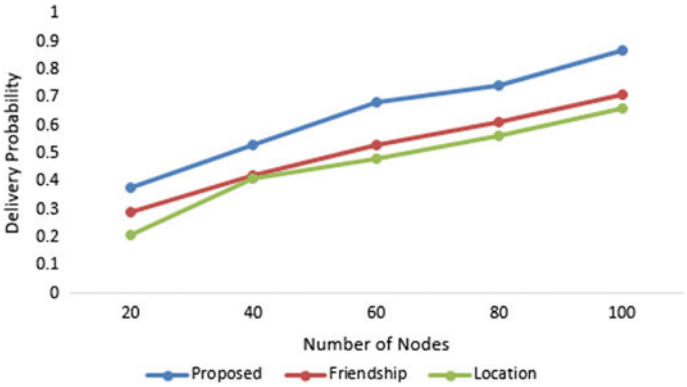


Fig. 2 Number of nodes versus delivery probability

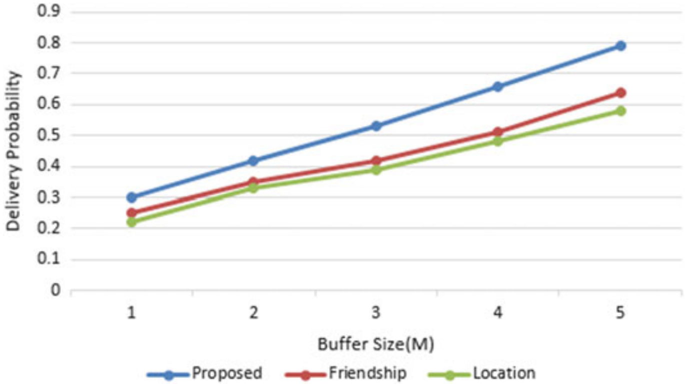


Fig. 3 Buffer size versus delivery probability

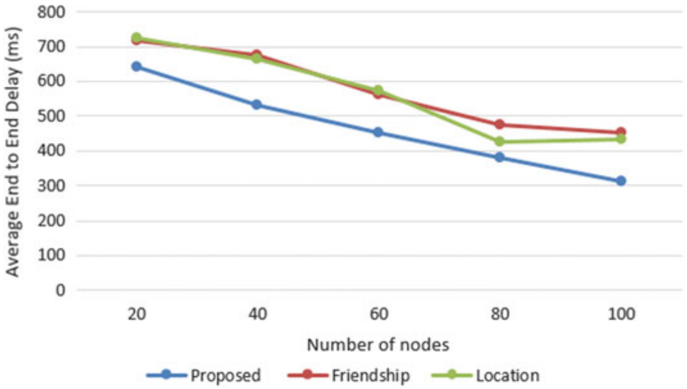


Fig. 4 Number of nodes versus average end to end delay

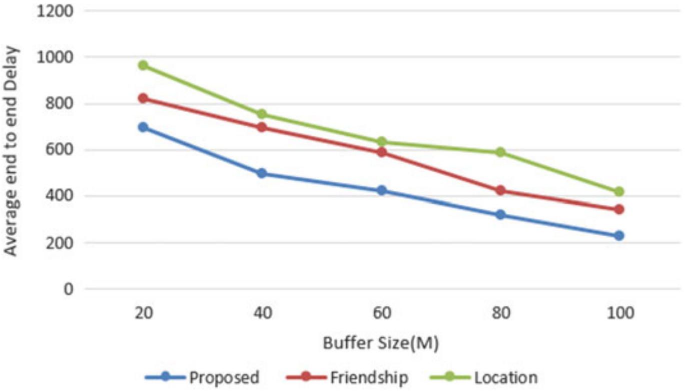


Fig. 5 Buffer size versus average end to end delay

5 Conclusion and Future Work

In this paper, we considered the problem in friendship based routing algorithm and proposed a new routing algorithm by adding the location-based similarity in friendship based routing, so that the node who does not have any friend can also send the message to the destination. All the nodes can transmit data to the destination, therefore, data delivery probability increased and delivery delay is decreased because all the nodes can transmit data immediately even they do not have any friends.

In the future, we can use Artificial Intelligence techniques to detect the location of nodes accurately and for finding friendship and we can use this algorithm in real physical nodes for finding the real accurate results.

References

1. Venkataraman V, Lam S Delay-tolerant networking —A tutorial
2. Fall K (2003) A delay-tolerant network architecture for challenged internets. In: Proceedings of the 2003 conference on applications, technologies, architectures, and protocols for computer communications—SIGCOMM '03, p 27
3. Zhu Y, Xu B, Shi X, Wang Y, Member S (2013) A survey of social-based routing in delay tolerant networks: positive and negative social effects. IEEE Commun Surv Tutor 15(1):387–401
4. Guo T, Yang Y (2015) Community based routing in social delay tolerant networks. In: 2015 ninth international conference on frontier of computer science and technology, pp 321–324
5. DEO Networks (2003) A routing mechanism based on social networks and betweenness centrality in
6. Bulut E, Szymanski BK (2012) Exploiting friendship relations for efficient routing in mobile social networks. IEEE Trans Parallel Distrib Syst 23(12):2254–2265
7. Rothfus D, Dunning C, Chen X (2013) Social-similarity-based routing algorithm in delay tolerant networks. In: 2013 IEEE international conference on communications (ICC), pp 1862–1866

8. Tasfe M (2017) Gossip : a social interest based routing algorithm for pocket switched network, pp 22–24
9. Socievole A, Caputo A, De Rango F (2019) Routing in mobile opportunistic social networks with selfish nodes, vol 2019
10. Bulut E, Szymanski BK (2010) Friendship based routing in delay tolerant mobile social networks
11. Vahdat A, Becker D (2000) Epidemic routing for partially-connected ad hoc networks. Technical Report CS-200006. Department of Computer Science, Duke University. Recuperado de. http://issg.cs.duke.edu/epidemic/epidemic.pdf
12. Spyropoulos T, Psounis K, Raghavendra CS (2005) Spray and wait: an efficient routing scheme for intermittently connected mobile networks. In: ACM SIGCOMM workshop on delay-tolerant networking, pp 252–259
13. Bulut E, Wang Z, Szymanski BK (2010) Cost effective multi-period spraying for routing in delay tolerant networks 18(5):1–14
14. Pathak S, Gondaliya N, Raja N (2017) A survey on PROPHET based routing protocol in delay tolerant network. In: 2017 International conference on emerging trends and innovation in ICT, ICEI 2017, pp 110–115
15. Paper C, Jensen DD (2014) MaxProp : routing for vehicle-based disruption-tolerant networks. Apr 2006
16. Daly EM, Haahr M (2007) Social network analysis for routing in disconnected delay-tolerant MANETs. In: Proceedings of the 8th ACM international symposium on mobile ad hoc networking and computing—MobiHoc ’07, p 32
17. Hui P, Crowcroft J, Yoneki E (2011) Bubble rap: social-based forwarding in delay-tolerant networks. IEEE Trans Mob Comput
18. Li F, Wu J (2009) LocalCom : a community-based epidemic forwarding scheme in disruption-tolerant networks. In: 2009 6th annual IEEE communications society conference on sensor, mesh and ad hoc communications and networks, pp 1–9
19. Li Q, Zhu S, Cao G (2010) Routing in socially selfish delay tolerant networks. In: 2010 Proceedings IEEE INFOCOM, pp 1–9
20. Jeon I, Lee K (2016) A dynamic Markov chain prediction model for delay-tolerant networks 12(9):0–6
21. Mcpherson M, Smith-lovin L, Cook JM (2001) Birds of a feather : homophily in social networks birds of a feather : homophily in social networks. Jan 2001
22. Bulut E, Szymanski BK (2010) Friendship based routing in delay tolerant mobile social networks. In: GLOBECOM 2010, 2010 IEEE Global …, pp 1–5
23. Zhang Y, Zhao J (2009) Social network analysis on data diffusion in delay tolerant networks, p 345
24. Conan V. DTN routing in a mobility pattern space, pp 276–283

Fake Account Detection Using Machine Learning

Priyanka Kondeti, Lakshmi Pranathi Yerramreddy, Anita Pradhan, and Gandharba Swain

Abstract Nowadays the usage of digital technology has been increasing exponentially. At the same time, the rate of malicious users has been increasing. Online social sites like Facebook and Twitter attract millions of people globally. This interest in online networking has opened to various issues including the risk of exposing false data by creating fake accounts resulting in the spread of malicious content. Fake accounts are a popular way to forward spam, commit fraud and abuse through an online social network. These problems need to be tackled in order to give the user a reliable online social network. In this paper, we are using different ML algorithms like Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF) and K-Nearest Neighbours (KNN). Along with these algorithms we have used two different normalization techniques such as Z-Score and Min-Max to improve accuracy. We have implemented it to detect fake Twitter accounts and bots. Our approach achieved high accuracy and true positive rate for Random Forest and KNN.

Keywords Data mining · Classification · Logistic regression · Support vector machine · K-Nearest neighbours · Random forest · Normalization

P. Kondeti (✉) · L. P. Yerramreddy · A. Pradhan · G. Swain
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur 522502, Andhra Pradesh, India
e-mail: priyankakondeti1999@gmail.com

L. P. Yerramreddy
e-mail: pranathi.yr@gmail.com

A. Pradhan
e-mail: anita.pradhian15@gmail.com

G. Swain
e-mail: gswain1234@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_73

1 Introduction

In this contemporary world, people are being dependent on Online Social Networks (OSNs). As many users are attracted and showing more interest to use OSN in their work-life or for personal uses. This gives an opportunity for the spammers to target people by collecting sensitive information by creating fake accounts. Fake accounts are being created in order to hide their identity and to accomplish their targets [1]. Bucket et al. [2] presented a supervised discretization technique known as Entropy Minimization Discretization (EMD) based on attributes, they have used the Naïve Bayes algorithm to evaluate the fake accounts in twitter. This technique can even be applied for all OSNs, for this process, they proposed their own dataset with 16 attributes. Finally, they presented three different evaluation criteria's before discretization and after that is an increase in accuracy, results from 85.5% to 90.41% and they showed that Naïve Bayes works perfectly for discrete values. To identify fake accounts in OSN Naman et al. [3] proposed a different model in a step by step process firstly they gathered the information and cleaned it. Then created some fabricated accounts. Therefore, they validated the data and then injected fake accounts by creating new attributes. At this stage, supervised ML techniques are applied and finally, results are evaluated. In this process, they analyzed, identified and abolished fictitious bot accounts. This model helped them to identify fake accounts created by humans from a real one.

These activities motivated the researchers to analyze the abnormal activities of Facebook and Twitter users by detecting and studying them. Furthermore, in recent years banks and financial providers in the U.S are even analyzing Twitter and Facebook accounts before granting the loan [4]. Whereas to create a Robust Fake account detection model Yeh-Chen and Shyhtsun [5] proposed a strategy to analyze user activity by collecting several popular pages which are at an ease to be attacked. They used collector filter accounts to analyze if any malicious activities like spam keywords, extreme promoting a particular company, etc. to verify whether they are among the selected group. Then the model is trained. They used three ML algorithms namely Random Forest (RF), C4.5 decision tree algorithm, Adaptive Boosting by giving a cluster of features as input for testing the model and a rank score is produced as output, which produces the probability to be a fake account. In this model, the RF classifier performed the best according to the correction rate and their model performed well in the real-world without any overfit problem.

In the present scenario, researchers are using various techniques to analyze fake accounts on OSN platforms by using various attributes. Some analysts detected fake accounts in OSN using a user profile. Some other analysts detected by using both sentiment analysis and user behaviour. Furthermore, some researchers used ego networks to analyze the clusters in the social networks and even their tweets. Some crawling tools are also used to extract the data which is available publicly [6]. Qiang et al. [7] proposed a new OSN user system known as Sybil Rank which is dependent on ranking the users using the social graph. In this model, social relationships are bidirectional which helps to detect fake accounts in large scale OSN's.

Mauro et al. [8] proposed a new approach according to empirical analysis and structure of typical social network interactions and their statistics to detect fake accounts created in OSN's. They analyzed from a dynamic point using social network graphs within the content of confidentiality threats. Kaur and Singh [9] proposed a wide range of approaches like supervised, semi-supervised and unsupervised methods. Besides this to detect anomalies in data mining and social networking domain they even analyzed according to cluster-based, proximity-based and classification-based networks. Yazan et al. [10] presented an Integro which is a robust and scalable defence machine by using distinct classification theory. They mainly analyzed the real accounts which accepted the fake requests and the process in the Integro system that takes place from user-level activities with the help of supervised machine learning algorithms. Finally, Integro uses probability in order to rank user accounts. Tsikerdekis and Zeadally [11] proposed a method using nonverbal behaviour in order to detect and identify deception in online social media. In this paper they used Wikipedia as an experiment and their method achieved a high detection accuracy than other methods. They even demonstrated how developers and designers had overcome these nonverbal data in analyzing the deceit by increasing the reliability in online communications. The proliferation of hoaxes, fake news and deceptive online data in Indonesia had cast a tremendous effect on Jakarta election which is already divided [12]. According to the statistics, the political advertising spending on Facebook by sponsor category between 2014 and 2018 have been increased. As per the findings, in the measured period, non-profits spent approximately US$ 2.53 millions sponsoring political advertising on Facebook, this demand for online social network giving opportunity for the hacker to spread fake news [13]. Twitter has also become an important outlet for administration and significant for state U.S. communications [14]. In CBC news Facebook shared that 955 million active monthly users in 8.7% users could be false accounts or facade. Fake accounts are being mainly used for businesses or for the spamming purpose only [15]. Estee and Jan [16] proposed a feature-based engineering model to detect bots in the social media platforms (SMPs) by extracting attributes such as follower-count and friend-count and showed an advanced successful score of 41.5% in identification of fake identities fabricated by human using SMPs. However, analyzing an account as genuine or fake in any social network is a complicated task and the various attacks like fake profiles, social engineering, fake news and profile compromise are increasing in the online social network. Hence there is a need to improve and implement new techniques. One of the well-known techniques used in recent days is data mining, which was user friendly in reading reports and significant approach, minimizing the errors and controlling in the standards of data sets.

In this paper, we are using a data mining approach which can be used in a wide range of areas including images, business marketing, transactions, banking, hospitals, medicine, insurance, monitoring video, satellites, e-mail messaging and repositories. Data mining is mainly used to extract the data from a collection of data and transforms it to a structure which is understandable for future use like Classification, Clustering, Regression, Association Rules, Prediction and Sequential patterns. Our approach is based on Classification with normalization, Association Rules and Prediction by using ML algorithms. There are eclectic algorithms but we have approached supervised learning algorithms such as Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), K-Nearest Neighbour (KNN) and by collecting datasets from Twitter users and online [19]. By using these datasets we performed preprocessing and classification to the data by improving the true positive rate in order to predict the fake accounts on Twitter.

2 Related Work

Cao et al. [1] created a system which used pipelining to group accounts into clusters; they implemented a cluster level model to detect fake accounts by giving cluster level features for ML algorithms as input. They proposed a generic pattern encoding algorithm to compute statistical features. Pipelining is based on three levels; firstly they implemented a Cluster Builder for cluster account score, secondly profile features were utilized for feature extraction and evaluate ML, thirdly account scorer is generated after training and evaluating ML models on new data. They used three algorithms like SVM, RF and LR. Among those three, RF gave the best result for all metrics 95% and even for the out-of-sample testing data 97%.

Using an ML algorithm Sarah et al. [4] implemented a new classification algorithm SVM-NN by combining SVM and NN (Neural Networks) along with the Management Information Base (MIB) baseline dataset to improve the efficiency for detecting bots and fake twitter accounts, they used dimension reduction and attribute selection methods. This new technique used a very tiny fraction of attributes even though they are able to classify correctly about 98% of the twitter accounts in their dataset.

Apart from traditional methods Myo and Nyein [6] proposed a new method by creating their own blacklist by data collection, preprocessing, extracting topic and keywords from Honeypot dataset. In order to show the difference between fake accounts and legitimate accounts, first, they have done feature extraction. They have extracted features from the user which is available publicly. Then they classified the data using decorate which is a specially designed artificial training example for meta-learner to build a diverse ensemble classifier. They have added a classified data to this ensemble to increase the rate of accuracy. This method is repeated until they have reached the maximum iteration and achieved their desired committee size. Twitter APIs are collected and this blacklist is used in the attribute extraction process. Finally, classification is done as real or fake. Their method has reached an accuracy of 95.4%, while the true positive value is 0.95.

In the previous papers, researchers used cluster-based features, network-based features, keyword and Profile analysis, classification algorithms due to crawling problems, network-based and profile features are difficult to extract. In this paper, we are implementing an account level detection by applying supervised learning methods to a twitter dataset beside this Z-score and Min-Max normalization is mainly used to improve the accuracy for SVM, LR, RF and KNN to detect the fake accounts.

3 Proposed Work

The proposed work comprises of four steps, (i) data classification, (ii) data preprocessing, (iii) data reduction or transformation and (iv) Algorithm selection. These steps are described below.

3.1 Data Classification

Data classification is characterized as the process for deciding the appropriate type, origin of data and appropriate resources for collecting data. In the data classification step, the data is selected from various Twitter accounts. We collected twitter datasets for analysis and to test our model, the dataset consists of different attributes such as name, status-count, friend-count and followers-count.

We selected these columns as feature attributes status-count, friends-count, followers-count, sex-code, favourites-count, Lang-code.

3.2 Data Preprocessing

We used machine learning algorithms in this process to convert and analyze the available raw data into feasible data. It is mainly used for better and accurate results. For instance, some algorithms like Random Forest do not support null values or they need a particular format. In such a case data preprocessing is a necessary step. Then we extracted two Comma-Separated Value (CSV) files fake and genuine users, we combined both files by sampling noise in the data and then feature labels were added as 0/1 to distinguish fake or real.

We selected particular columns as feature attributes status-count, friends-count, followers-count, listed-count, sex-code, favourites-count, Lang-code and then we have removed columns having more null values sex-code, Lang-code, and normalized the data for better accuracy.

In this method, we distributed the information for training and testing purpose at 80:20. To represent the values in the confusion matrix, the model is trained with x-train and y-train data, and then it is trained with test data for precision and recall values. Graphs are represented as Area under ROC Curve (AUC) and Receiver operating characteristic curve (ROC).

3.3 Data Transformation

It is a method of converting information from one format or structure into another format. For tasks such as data integration and management, data transformation is a crucial step to improve accuracy. In our model, we used two normalization techniques for data transformation.

Normalization in Data Mining

We are using two data normalization techniques such as Z-Score, Min-Max, it is mainly used when dealing with multiple attributes on a different scale and to scale the information into a smaller range, it is commonly applied for classification algorithms to improve the performance rate, so the attributes are normalized to bring on the same scale.

Z-Score: Z-Scores are mainly based on the mean value and standardized score these scores are linearly transformed data value with a mean of 0 and the scores have been given a common standard. It helps to understand the rate of a score as per the normal distribution of the data.

Min-Max: In this method, linear transformation is performed on original data, according to this minimum values of that feature is transformed as 0 whereas maximum values are transformed into 1 and remaining values have been changed into decimals between 0 and 1.

3.4 Algorithm Selection

K-Nearest Neighbour

K-Nearest Neighbour (KNN) is the type of supervised learning method. It is used for both pattern recognition along with classification. In KNN, a specific test tuple set is compared to the training data set already available which is identical to the test data set. It calculates the distance between the training data and the testing data using the Euclidean distance function. Class membership is the output of the KNN classification. Therefore, KNN classification has two stages, the first is the determination of the nearest neighbours and the next step is the determination of classes using the neighbours.

The working process of KNN classifier is defined below

1. Distance between the attributes is calculated from testing and training data sets.
2. Training data is sorted according to the distance values.
3. Obtain the neighbours (k) which are approximately close to the testing data.
4. Majority class of training data is added to the testing data.

The Euclidean distance between the training data set and testing data set is estimated using Eq. (1) where p_i stands for an element of training dataset and q_i stands for an element of testing dataset. This *D*(*p*,*q*) derives the smallest distance k and locate the corresponding data.

$$D(p,q) = \sqrt{\sum_{i=0}^{n} (p_i - q_i)^2} \tag{1}$$

Random Forest Algorithm

The Random Forest consists of a large number of individual trees which functions as an ensemble individual tree that divides a random forest class prediction, and the class with the most votes is our model's prediction.

Random forest (RF) is similar to bootstrapping algorithm along with Regression tree Classification and Decision Trees(CART). We have 1000 observations of 10 parameters in the entire population. RF attempts to create several CART models with different initial variables with samples. For example, it chooses randomly from the sample of 100 observations and randomly 5 are selected from initial variables to design a CART model. This procedure is repeated 10 times and each observation is analyzed finally. A final prediction is also a function of each prediction and this prediction is simply a mean value.

Support Vector Machines (SVM)

Support Vector Machine (SVM) is also a type of computer algorithm that can be trained to assign labels to the objects. It is a powerful tool for solving both classification and regression problems. It is one of the supervised learning methods and one of the best-known classification methods. SVMs are based on statistical learning theory, which is used to solve two-class binary problems without the loss of generality.

The main objective of SVM classifiers is to locate the decision boundaries like hyperplanes, which produces the separation of classes optimally and is mainly used to resolve the linear problems besides this it can be extended to handle nonlinear decision problems. SVM's features is a good generalization performance, lack of local minima and sparse solution distribution. It is based on the principle of Structural Risk Minimization (SRM) that minimizes the generalization error.

Logistic Regression

Logistic Regression (LR) is a type of linear algorithm, which is the method of relating dependent and independent variables using a logistic distribution functional form. The regression model can be mathematically designed by describing the likelihood of certain events Eq. (2). It obtains a linear relationship between the output and input. LR measures the probability of class inclusion for one of the data set's different categories. This is used for modelling the binary response data. If the response is in binary, it takes the form of success and indicates failure. Consider data, weights and class label 1/0.

$$P(c = \pm 1|d, a) = \frac{1}{1 + \exp(-c(a^T d + b))} \tag{2}$$

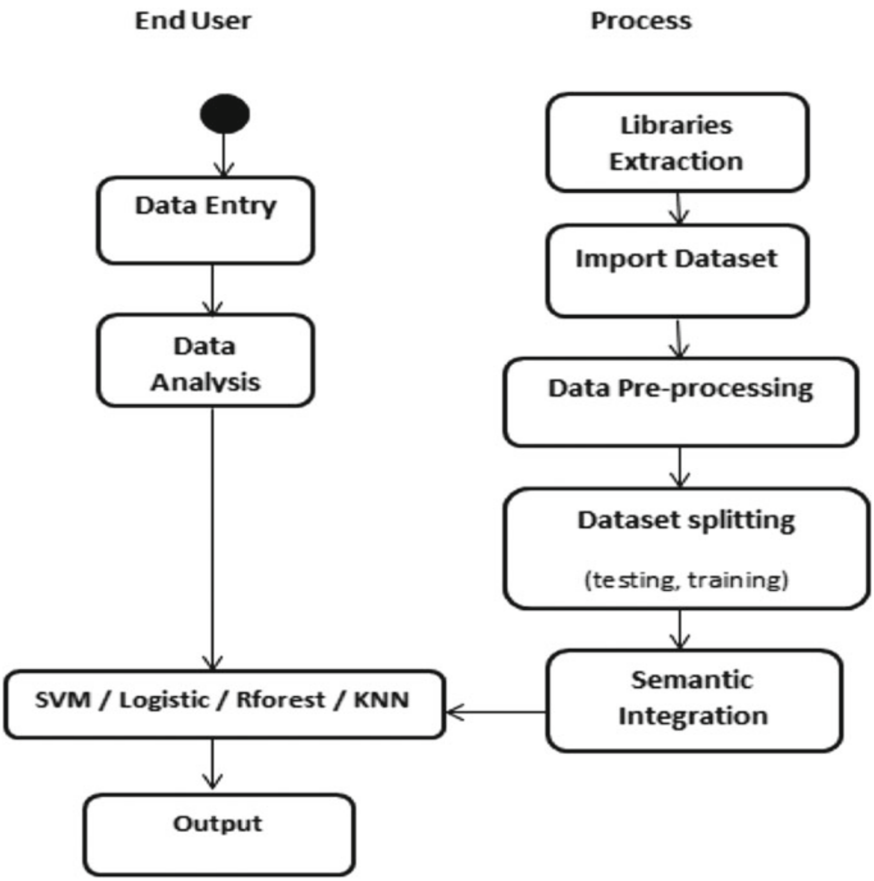


Fig. 1 Fake account detection strategy using data mining techniques

The proposed technique has been represented in the form of a flowchart in Fig. 1.

4 Results and Experimentation

The hardware used to implement this programme is a laptop with i5 processor, 500 GB Memory and 4 GB Ram. The Software used for this implementation is Anaconda, and the language used is Python. The dataset used in this program is taken from twitter online repository. The result of this technique is obtained and measured using the following metrics.

Table 1 Confusion matrix

	Class 1 Predicted	Class 2 Predicted
Class 1 Actual	TP	FN
Class 2 Actual	FP	TN

Confusion Matrix

In the area of machine learning, confusion matrix solves the statistical classification problem. Confusion Matrix is often used to discuss the functioning of the classification system of an algorithm on a collection of testing data that is defined for the true positive values. It is also called as an error matrix.

This allows ambiguity between groups to identify easily, e.g. one class is often mislabelled as another. Maximum performance measurements are obtained from the confusion matrix (Table 1).

Here,

- Class 1: Positive
- Class 2: Negative

Description of Terms

- Positive (P): positive values(for instance: is a ball).
- Negative (N): If the values are not positive (for example: is not a ball).
- True Positive (TP): If the prediction is positive, but the observed values are even positive.
- False Negative (FN): Examined as positive, and the predicted value is negative.
- True Negative (TN): If the examined values are negative, but predicted as negative.
- False Positive (FP): Examined values are negative, and predicted as positive.

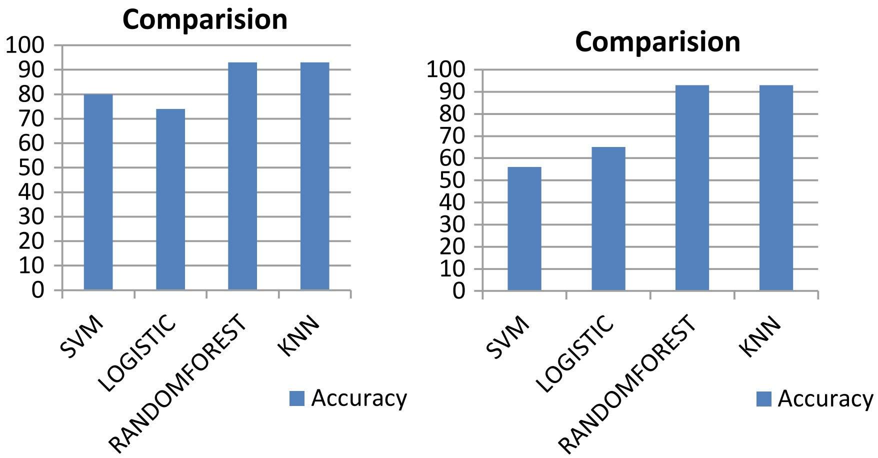
Accuracy for detecting fake accounts can be obtained by using *TP*, *TN*, *FP* and *FN* from Eq. (3).

$$\text{Accuracy Prediction} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

AUC and ROC curve is an efficiency calculation at various threshold rates for the classification issues. ROC is a curve of likelihood and AUC is the degree of separability factor. This shows the design value that can differentiate between classes. The model predicts accurately as 0's and 1's if the AUC is higher and it distinguishes better between accounts as fake or real.

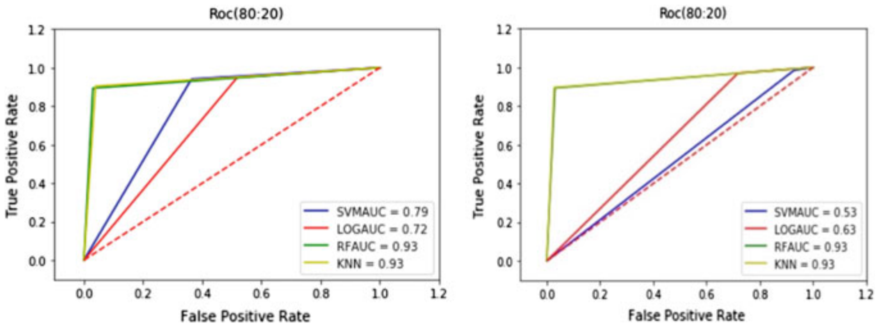
Based on the techniques described before, we assessed our strategies with ML classifiers, which includes Random Forest, Logistic Regression, KNN and SVM. In this experiment for testing the model we selected the datasets from twitter, these datasets are preprocessed and split into 80:20 cross-validation process this split data is used for testing and training the algorithms. We applied our technique to the testing data as per the analysis Random Forest and KNN performed well with high accuracy (98%). By applying Z-Score Normalization SVM and Logistic regression, accuracy rate had been increased as shown in Fig. 2.

To calculate the performance of the classifier, we generated a ROC curve, based on true positive and false-positive rate. The rank allocated to each account is based on the results of a particular classifier, so we tested the reliability by concentrating on a different range of ranks to see if our method works better by the rank created by our classification system. We used the ROC curve which shows the cut off for a test, whereas the best cut off has the highest true positive rate with a low false-positive rate, as shown in Fig. 3. ROC curve based on 80:20 split and account-level classification



(a) Z-Score Normalization **(b) Min-Max Normalization**

Fig. 2 Accuracy prediction using machine learning algorithms by account level



(a) Z-Score Normalization **(b) Min-Max Normalization**

Fig. 3 Comparison of supervised algorithms using ROC curve learning

Table 2 80:20 split account level testing performance

Z-Score normalization		Min-Max normalization	
Algorithm	AUC	Algorithm	AUC
KNN	0.932	KNN	0.934
Random Forest	0.931	Random Forest	0.931
Logistic Regression	0.715	Logistic Regression	0.626
SVM	0.788	SVM	0.528

worked well with RF and KNN by giving a 93% true positive rate with both types of the normalization techniques, however, SVM and Logistic performance are very low in Min-Max normalization when compared to Z-Score. On the other hand, we even focused on AUC. It can provide an integrated probability of performance with comparison to all possible classification thresholds. AUC is the rate at which the model is ranked random positive more than a random negative.

Table 2 shows the AUC precision for four algorithms at the account level, this shows that the identification of each algorithm is more accurate for every single account. When compared to the other algorithms we can observe that Random Forest and KNN are giving the highest accuracy.

5 Conclusions and Future Work

Finally, we conclude that our research has been done to analyze, detect and remove the fake accounts created on Twitter, but our method can be applied to other datasets from OSN platforms such as Facebook and LinkedIn. Due to the ease of ML algorithms, fake accounts can be analyzed using different ML classifiers. In this paper, we have used SVM, KNN, Random forest, logistic algorithms along with Z- Score and Min-Max normalization techniques to predict fake users. By using these techniques we have improved the accuracy to 98%. As future work, this can be extended by implementing feature selection for this method or to cluster the accounts into groups and develop an efficient model by analyzing more data to improve the accuracy for predicting the fake account in OSN.

References

1. Cao X, David MF, Theodore H (2015) Detecting clusters of fake accounts in online social networks. In: 8th ACM workshop on artificial intelligence and security, pp 91–101
2. Buket E, Ozlem A, Deniz K, Cyhun A (2017) Twitter fake account detection. In: IEEE 2nd international conference on computer science and engineering, pp 388–392
3. Naman S, Tushar, S, Abha T, Tanupriya C (2018) Detection of fake profile in online social networks using machine learning. In: IEEE international conference on advances in computing

and communication engineering. pp 231–234

4. Sarah K, Neamat E. Hoda MOM (2018) Detecting fake accounts on social media. In: IEEE international conference on big data. pp 3672–3681
5. Yeh-Cheng C, Shyhtsun FW (2018) FakeBuster: a robust fake account detection by activity analysis. In: IEEE 9th international symposium on parallel architectures, algorithms and programming. pp 108–110
6. Myo MS, Nyein NM (2018) Fake accounts detection on twitter using blacklist. In: IEEE 17th international conference on computer and information and information science. pp 562–566
7. Qiang C, Michael S, Xiaowei Y, Tiago P (2012) Aiding the detection of fake accounts in large scale social online services. In: 9th USENIX conference on networked systems design and implementation. pp 1–14
8. Mauro C, Radha P, Macro S (2012) Fakebook: detecting fake profiles in online social networks. In: IEEE international conference on advances in social networks analysis and mining. pp 1071–1078
9. Duraipandian M (2019) Performance evaluation of routing algorithm for manet based on the machine learning techniques. J Trends Comput Sci Smart Technol (TCSST) 1(1):25–38
10. Yazan B, Dionysios L, Georgos S, Jorge L, Jose L, Matei R, Konstatin B, Hassan H (2016) Integro: Leveraging victim prediction for robust fake account detection in large scale osns. Comput Secur, pp 142–168
11. Tsikerdekis M, Zeadally S (2014) Multiple account identity deception detection in social media using non verbal behaviour. IEEE Trans Inf Forens Secur 9(8): 1311–1321
12. How fake news and hoaxes have tried to derail jakarta's election. Internet draft(online)
13. Political advertising spending on facebook between 2014 and 2018 Internet draft. https://www.statista.com/statistics/891327/politicaladvertisingspending-facebookby-sponsor-category/.2018
14. Statista.twitter: number of monthly active users 2010–2018. Internet draft. https://www.statista.com/statistics/282087/number-of-monthlyactive-twitter-users/.2018
15. Cbc.facebook shares drop on news of fake accounts. Internet draft. http://www.cbc.ca/news/technology/facebook-shares-drop-onnews-of-fake-accounts-1.1177067.2012
16. Estee VDW, Jan E (2018) Using machine learning to detect fake identities: bots vs humans. In: IEEE access, pp 6540–6549
17. The list of email spam trigger words. http://blog.hubspot.com/blog/tabid/6307/bid/30684/The-Ultimate-List-of-Email-SPAM-Trigger-Words.aspx
18. Bayuk J (ed) (2010) Cyber forensics: understanding information security investigations. Springer's Forensic Laboratory Science Series, Humana Press. pp 59– 101
19. https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/user-object

Multi-objective Task Scheduling Using Chaotic Quantum-Behaved Chicken Swarm Optimization (CQCSO) in Cloud Computing Environment

G. Kiruthiga and S. Mary Vennila

Abstract Task scheduling is a challenging process with the increasing number of requests from the clients in a cloud system. Achieving efficient task scheduling with multiple objectives is much required in this modern era. A novel Chaotic Quantum-behaved Chicken Swarm Optimization (CQCSO) based task scheduling approach is presented in this paper. CQCSO is developed by applying chaotic theory and quantum theory to the standard Chicken Swarm Optimization to overcome its problem of premature convergence and local optima. CQCSO algorithm models the task scheduling as an optimization problem and solves it by formulating a multi-objective fitness function using task completion time, response time and throughput to ensure maximum Quality-of-service (QoS) satisfaction and minimum SLA violations. CQCSO identifies the task order and optimally schedules them to the suitable virtual machines with better performance. Experiments were conducted in CloudSim to evaluate the CQCSO approach and it provided efficient task scheduling than the prior existing algorithms.

Keywords Cloud computing · Task scheduling · Chaotic Quantum-Behaved chicken swarm optimization · Quality-of-service · Multi-objective problem

1 Introduction

Cloud computing paradigm has attracted attention in major scientific, mobile communication, medical and business fields of the recent big data revolution. Cloud computing provides an on-demand computing model to access the convenient shared resources including the networks, storage, applications, servers and services [1]. Cloud computing model has the ability to dynamically allocate adequate resources

G. Kiruthiga (✉) · S. Mary Vennila
PG and Research department of Computer Science, Presidency College, Chennai, India
e-mail: g_kiruthiga@yahoo.co.in

S. Mary Vennila
e-mail: vennilarhymend@yahoo.co.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_74

to each task from different users [2] and handle multiple tasks or jobs simultaneously from various regional users through scalable virtual resource allocation. This property enables the cloud users to obtain high performance services through the cloud optimization of the resource allocation to the users' tasks. As task scheduling is dependent on physical resource usage and has a direct influence on the QoS and customer satisfaction, the policymaking for efficient task scheduling becomes crucial in any cloud environment under certain specified constraints [3]. The vitality of the task scheduling scheme depends on the simultaneous balance between the users' necessities and the consumption of physical resources. Also, there occurs a converse relativeness in the pricing model for task processing time and task computation time [4]. The cloud clients often need superior service at affordable service charge while the service provider looks for providing better service with maximum profit. Yet, when the makespan of the workflow tasks is reduced, the cost will be increased as this process consumes more resources to reduce the task completion time. Hence the task scheduling problem remains as a NP-hard problem for which a good solution must be obtained to improve the performance at both the user-level and server-level based on good scheduling policy [5].

The primary intention of a good task scheduling policy is minimizing the cost and the completion time. To realize this objective, many researchers have contributed to the development of significant task scheduling using heuristic methods based on user requirements [6]. However, different users may have different requirements which are challenging to handle in dynamic cloud resources. The multi-objective task scheduling algorithms were employed for overcoming such challenges using optimization algorithms [7–9]. However, due to the poor search abilities and low convergence rate, most existing optimization algorithms return below-par performance [10]. This paper has focused on resolving this problem by developing a hybrid framework for multi-objective task scheduling. The major improvements of this paper are: (i) Considering throughput, response time and task completion time in cloud environment for formulating task scheduling problem as multi-objective optimization problem; (ii) The development of novel Chaotic Quantum-Behaved Chicken Swarm Optimization (CQCSO) algorithm by improving standard Chicken Swarm Optimization (CSO) [13] by using the chaotic theory and quantum mechanics for multi-objective task scheduling. Evaluations are made using the CloudSim framework over the Planet Lab tasks.

2 Related Works

Task scheduling can be based on single objective or multiple objectives. Single objective approaches can only optimize either the cost or time [11–13]. However, in an efficient cloud computing model, it is essential to consider many QoS objectives leading to multi-objective task optimization problem. Studies have been presented using optimization algorithms to obtain the best task scheduling based on multiple objectives. He et al. [14] introduced an adaptive task allocation strategy using an improved PSO algorithm based on energy, cost, processing time and transmission time. Reddy and Kumar [15] proposed a multi-objective task scheduling algorithm using whale optimization based on resource utilization, QoS and energy objectives. However, these models have less service availability and less searching capability and increases the overall execution time. Bindu et al. [16] proposed energy aware task scheduling using multi-objective GA by considering makespan and energy along with data transfer time as objectives. But, poor searching ability and premature convergence of GA negatively influences the performance.

Kaur and Kadam [17] developed a scheduling model using a selection and diversity enhanced multi-objective bacteria foraging optimization algorithm (MOBFOA). It considers makespan, resource usage cost and flow time as the objective parameters. Gomathi et al. [18] presented a novel Epsilon-fuzzy Dominance sort-based Composite discrete Artificial Bee Colony optimisation (EDCABC) considering makespan, cost and resources. Torabi and Safi-Esfahani [19] developed hybrid CSO and improved raven roosting optimization based dynamic task scheduling framework to minimize time and maximize throughput. Kumar and Venkatesan [20] developed hybrid genetic-ant colony optimization based scheduling by combining the ACO and GA. Jacob and Pradeep [21] proposed a task scheduling approach considering deadline violation rate along with cost and makespan using hybrid cuckoo particle swarm optimization (CPSO) that combines the cuckoo search optimization and PSO algorithm. Abdullahi et al. [22] developed a scheduling strategy using a combination of symbiotic organisms search (SOS) and chaotic optimization. Although these models solve the premature convergence problem, the computation complexity is high for very large workloads due to the extensive search processes.

From the literature study, it can be concluded that the existing multi-objective scheduling strategies have become the staple for efficient cloud task scheduling with minimized makespan and cost. However, those approaches are also limited by certain drawbacks like high computation complexity, larger time consumption and limited searchability of the considered optimization algorithms. These problems are considered in this research paper and an efficient multi-objective scheduling strategy is designed by using a hybrid optimization algorithm in the form of CQCSO.

3 Multı-objective Problem Formulatıon

Multi-objective scheduling problems must satisfy one or more conflicting objective functions based on multiple parameters. In the proposed CQCSO, the multi-objective task scheduling problem is modelled as an NP-hard optimization problem which is solved by task completion time, response time and throughput parameters. These three parameters are particularly selected in CQCSO as it directly impacts the makespan and cost. CQCSO considers response time and completion time together in order to avoid the congestion problems. These problems are mainly due to lack of prior consideration of the completion time of current tasks and due to the long waiting time. When considering these two parameters, the congestion problems can be minimized. Mathematically, the problem of multi-objective optimization can be defined as

$$F(x) = w_1 \times CT + w_2 \times RT + w_3 \times Th \tag{1}$$

where CT is the completion time of tasks, RT is the response time, Th is the throughput of the VMs and w_1, w_2, w_3 denote weights assigned to the parameters such that $w_1 + w_2 + w_3 = 1$ and $w_1 \geq 0, w_2 \geq 0, w_3 \geq 0$. For minimizing completion time and response time but improving the throughput, the weights are assigned accordingly. The three objective parameters can be calculated based on the time and successful task execution. For a system model with task set $T = \{t_1, t_2, \ldots t_n\}$ with n tasks in task queue and resources set $R = \{r_1, r_2, \ldots r_m\}$ with m resources in resource pool (VMs), the parameters can be computed as:

$$CT_{ir} = \sum_{i=1}^{N} Ft_i - St_i, \quad 1 \leq i \leq N; 1 \leq r \leq M \tag{2}$$

$$RT_{ir} = \sum_{i=1}^{N} Subt_i - Wt_i, \quad 1 \leq i \leq N; 1 \leq r \leq M \tag{3}$$

$$Th_{ir} = \sum_{i=1}^{N} \frac{\text{Succ tasks}}{\text{Total time}}, \quad 1 \leq i \leq N; 1 \leq r \leq M \tag{4}$$

where CT_{ir} denote completion time of task i on r-th VM, RT_{ir} denote response time of task i on r-th VM, Th_{ir} denote throughput of the resource r, Ft_i represent finishing time of task execution on r-th VM, St_i represent starting time of task execution on r-th VM, $Subt_i$ denote task submission time, Wt_i represent the task waiting time, and Succ tasks are the successfully completed tasks on the resource.

4 CQCSO Based Multi-objective Task Scheduling

The dynamic task scheduling is performed for an independent set of tasks generated from the Planet Lab workload traces. First, the standard Chicken Swarm Optimization is discussed followed by the suggested changes to it using the chaotic theory and quantum theory.

4.1 Applying Quantum Theory and Chaotic Search Process to CSO

CSO is modelled on the social hierarchy characteristics of the farm chickens [10]. In the farms, the chickens flock together as a group with the rooster acts as the leader of this group. Each group selects the leader rooster based on its ability to search the best food source while other hens and chicks are the members. Chicken Swarm Optimization is modelled based on this hierarchy. In the proposed CQCSO, the standard position equations will be modified using the quantum theory and chaotic search equations. First, the position of the rooster, hens and chicks will be updated based on a Gaussian distributed attractor point to improve the convergence rate of the algorithm. The position update equation based on the attractor is given as

$$p_{i,j} = \varphi_j^t L_{\text{best},j}^t + \left(1 - \varphi_j^t\right) G_{\text{best},j}^t \tag{5}$$

$$\varphi_j^t = \frac{\varepsilon_1 u_1}{\varepsilon_1 u_2 + \varepsilon_1 u_2} \tag{6}$$

where $L_{\text{best},j}^t$ represent local best solution vector element after t iterations, $G_{\text{best},j}^t$ denote global best solution vector element at t; φ_j^t represent convergence factor formed by positive constants $\varepsilon_1, \varepsilon_2$ and uniformly distributed random numbers u_1, u_2, u_3.

β is the contraction–expansion coefficient that influences the algorithm conquering rate and it is estimated as

$$\beta = \beta_0 + \frac{(t\text{max} - t) \times (\beta_1 - \beta_0)}{t\text{max}} \tag{7}$$

where t and $t\text{max}$ are initial and maximum iterations; β value is initially 1.0 which decreases until 0.5. Mean best (M_{best}) location is computed as the average of the local best ($L_{\text{best},j}^t$) solutions. İt is estimated using

$$M_{\text{best}}^t = \frac{1}{N} \sum_{i=1}^{N} L_{\text{best},i}^t \tag{8}$$

The original attractor point $p_{i,j}$ is represented as the mean and the standard deviation. This value will be equivalent to the average length from mean best to the local best of the quantum element. Therefore, a new attractor point is given by

$$Np_{i,j}^t = N\left(p_{i,j}^t, M_{\text{best}}^t - L_{\text{best},i}^t\right) \tag{9}$$

Finally, the new position update equation will be given as

$$x_{i,j}^{t+1} = Np_{i,j}^{t} \pm \beta \left| M_{\text{best},j}^{t} - x_{i,j}^{t+1} \right| \ln(1.0/u_{i,j}) \tag{10}$$

In this equation, the chaos theory is applied to replace the uniformly distributed random number $u_{i,j}$. It is done by employing the Logistic map concept as

$$\theta_{n+1} = \xi \theta_n (1 - \theta_n); \quad \theta \in (0, 1), 0 < \xi < 4 \tag{11}$$

Here ξ is a regulator constraint and θ_n is a chaotic factor. During the preliminary stage, θ_0 must be $\theta_0 \in (0, 1)$ and $\theta_0 \notin (0.25, 0.50, 0.75)$. The logistic map is chaotic only if $\xi = 4$. Depending upon all these changes, the final equations for updating position will be

$$x_{i,j}^{t+1} = Np_{i,j}^{t} \pm \beta \left| M_{\text{best},j}^{t} - x_{i,j}^{t+1} \right| \ln(1.0/\theta_{i,j}) \tag{12}$$

where $\theta_{i,j}$ is calculated as in Eq. (11).

Based on these update equations, the positions of the rooster, hen and chicks will be modified to form the proposed CQCSO to improve the convergence rate and avoid local optima problem.

4.2 Chaotic Quantum-Behaved Chicken Swarm Optimization

The chicken initial population of size X is defined. *RX, HX, CX and MX* denote the roosters, hens, chicks and the mother hens, respectively. The chicken with high fitness are only assigned as leader roosters, while the worst fitness chickens will become chicks. All X chickens are represented by their locations $x_{i,j}^{t} (i \in [1, \ldots, X], j \in [1, \ldots, D])$ at time step t %G (t is a smaller portion of G time-stamp) and search food in a D-dimensional space. The optimal position of the chickens are modified using a Gaussian distributed attractor point $p_{i,j}$. The location of the rooster is updated using

$$x_{i,j}^{t+1} = \left[Np_{i,j}^{t} \pm \beta \left| M_{\text{best},j}^{t} - x_{i,j}^{t+1} \right| \ln(1.0/\theta_{i,j}) \right] \times \left(1 + \text{Randn}\left(0, \sigma^2\right)\right) \tag{13}$$

where $\sigma^2 = \begin{cases} 1, & if\ f_i \le f_s, \\ \exp\left(\frac{(f_s - f_i)}{|f_i| + \varepsilon}\right), & \text{otherwise}, \end{cases} \quad s \in [1, X], s \neq i$

where $x_{i,j}^{t}$ is the best position until previous iteration t, $\text{Randn}\left(0, \sigma^2\right)$ denote Gaussian distribution with null mean and standard deviation σ^2. *s is* a rooster's index and f is the fitness value based on Eq. (1). The location of the hens is updated by

$$x_{i,j}^{t+1} = \left[Np_{i,j}^{t} \pm \beta \left| M_{\text{best},j}^{t} - x_{i,j}^{t+1} \right| \ln(1.0/\theta_{i,j}) \right]$$

$$+\ \left[S1 \times \theta_{i,j} \times \left(x^{t}_{r1,j} - x^{t}_{i,j}\right) + S2 \times \theta_{i,j} \times \left(x^{t}_{r2,j} - x^{t}_{i,j}\right)\right] \qquad (14)$$

Here $S1 = \exp(f_i - f_{r1})/(abs(f_i) + \varepsilon))$ and $S2 = \exp(f_{r2} - f_i)$. $r1 \in [1, \ldots, X]$ denote rooster index, while $r2 \in [1, \ldots, X]$ denote randomly selected hen's index $r1 \neq r2$. The location of the chicks is updated by

$$x^{t+1}_{i,j} = \left[Np^{t}_{i,j} \pm \beta \left| M^{t}_{\text{best},j} - x^{t+1}_{i,j} \right| \ln(1.0/\theta_{i,j})\right] + FL \times \theta_{i,j}\left(x^{t}_{m,j} - x^{t}_{i,j}\right) \qquad (15)$$

where $x^{t}_{m,j}$ denote *i*-th mother hen's position ($m \in [1, X]$. $FL(FL \in (0, 2))$ indicate uniform random distribution metric. It is set between 0 and 2 to improve the swarm diversity and obtain a better convergence rate.

Algorithm 1 summarizes the proposed CQCSO for task scheduling. The proposed CQCSO algorithm begins by initializing the chicken population. Then the task scheduling problem is mapped into the CQCSO. The tasks are assigned as chickens and the VMs are assigned as hierarchy groups. Then the fitness is estimated for each chicken and the positions are updated. Finally, the best scheduling is achieved at the end of the maximum iterations. The time and computation complexity of this algorithm is also lesser.

Algorithm 1: Proposed CQCSO based task scheduling

```
Input: Set of Tasks, set of VM
Output: Allotment of Task to VMs
Initialize a swarm of X chickens (tasks), groups (VMs) and associated parameters;
$f_{ij} = F(x)$ using Equation (1)
Compute the X chickens' fitness values, t=0;
While (t < $t_{max}$)
    If (t % G = = 0)
        Sort chickens based on fitness;
        Arrange chickens in hierrachical order;
        Categorize the chicken swarm into various classes;
        Assign relationship among chicks and hens;
    End if
    For i = 1 : X
        If i == rooster, Update rooster location using (13);
        Else if i == hen, Update hen location using (14);
        Else if i == chick, Update chick location using (15);
        End if
        Assess the properties of new location;
        If $L^{t+1}_{best} > L^{t}_{best}$,
            $G_{best}$ becomes updated location;
        End if
    End for
End while
```

Thus CQCSO algorithm is much suitable for the cloud task scheduling. It is very useful in real-time cloud-based applications like emails, social media, web services, online marketing, cloud healthcare, IT consolidated services, etc. The use of CQCSO based scheduling minimizes the waiting time and enables hassle-free services from the customers' point-of-view. This will also help the service providers serve more customers at a time and increase their customer pool along with the reduction of service maintenance cost.

5 Performance Evaluatıon

The performance of the CQCSO based task scheduling strategy is evaluated using CloudSim 3.0.3 simulator over the independent tasks generated from the Planet lab workload traces. Planet lab workload consists of CPU utilization of VMs from more than 500 servers around the world. The experiments are conducted on Intel ® core i5 processor with 1.8 GHz frequency CPU, 8 GB RAM and Windows 10 operating system using Eclipse and JDK 1.8. The CloudSim simulation settings are given in Table 1.

As the three objective parameters contemplated in this work, performance comparisons are made in terms of task completion time, response time, throughput, CPU utilization and Bandwidth utilization. Table 2 depicts the experimental results of the proposed CQCSO algorithm when the number of VMs is set as 20 and numbers of tasks are varied from 25 to 100.

Table 1 CloudSim settings

Entity type	Parameter	Value
Data center	Number of datacenters	1
	Type of data center	Heterogeneous
	Link delay (milliseconds)	10–100
	Bandwidth (Gbps)	1–10
Host	Number of host	5
	Number of cores	1–4
	Host RAM (MB)	4096
	Host Storage (MB)	1000000
	Host bandwidth (bps)	10000
VM	Number of VMs	20–100
	CPU (MIPS)	1000-10000
	RAM (MB)	512
	Bandwidth (bps)	1000
	Number of cores per VM	1
Task	Number of tasks	1000
	Task length (MI)	200–1000
	Task size	200–600
	Number of iterations	5–25

Table 2 Performance results of CQCSO (VM = 20)

No. of tasks	Task completion time (ms)	Throughput (bps)	Response time (seconds)	CPU utilization (%)	Bandwidth utilization (%)
25	24.15	44.337	0.01458	11.20	0.667
50	49.64	129.994	0.0275	12.23	1.531
75	74.43	160.588	0.0390	16.70	3.083
100	99.50	187.905	0.05295	26.149	5.82

Table 3 Performance results of CQCSO (VM = 100)

No. of tasks	Task completion time (ms)	Throughput (bps)	Response time (seconds)	CPU Utilization (%)	Bandwidth utilization (%)
200	199.07	2256.244	0.10568	17.975	8.652
400	399.11	3652.310	0.2142	21.444	19.703
600	588.27	4345.155	0.3207	28.763	37.831
800	799.74	4988.534	0.4252	37.950	64.565
1000	984.03	5001.0626	0.5254	47.548	93.9

The results obtained for 20 VMs scenario shows that the proposed CQCSO has linear performance in the dynamic cloud environment. The values of the performance metrics have increased consistently towards increasing task count. When the system performance metrics are consistent varying, the algorithm is effective. This demonstrates that the proposed CQCSO has achieved the desired objective of efficient task scheduling for a smaller workload.

Table 3 depicts the performance values of the proposed CQCSO when the number of VMs is set as 100 and numbers of tasks are varied from 200 to 1000.

Similar to Table 2, the results obtained for 100 VMs scenario in Table 3 shows that all performance metrics are consistently increasing. This validates that the CQCSO can enable better task scheduling even for a larger workload.

To further validate CQCSO, the simulation results of the CQCSO is assessed and equivated with other optimization algorithms based on task scheduling models. The existing task scheduling models considered for performance comparison are ACO [9], GA [16], GA-ACO [20] and CPSO [21]. Comparisons are made for these methods in a scenario with the 20 VMs and varying task count from 25 to 100. Figure 1 shows the Task completion time comparison of the considered scheduling models for 20 VMs. It is evident from the plot that the efficiency of the CQCSO outperforms the other compared models with reduced task completion time. This can significantly reduce the cost and overall execution time of the system.

Figure 2 shows the Response time comparison of the optimization-based task scheduling algorithms for 20 VMs. From the plot, it is understood that the CQCSO has minimal response time than the other scheduling models. This performance

Fig. 1 Task completion time

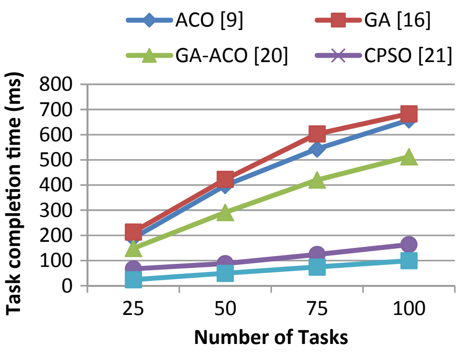
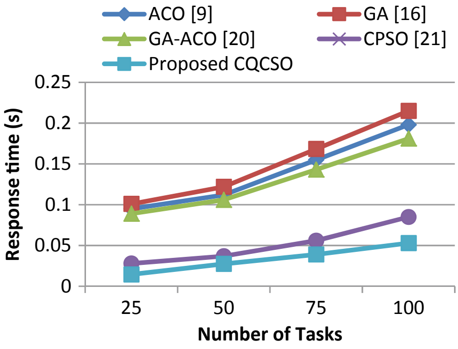
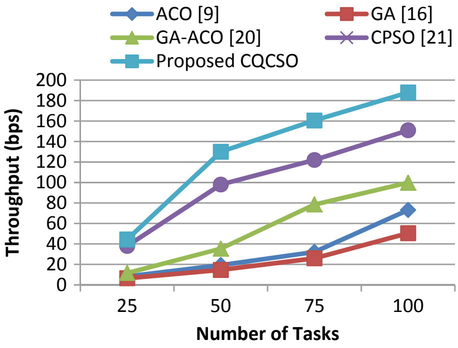


Fig. 2 Response time



directly reduces the delay of execution and the resources wasted for queuing. The major reason for this positive change is the improved convergence rate of CQCSO. Likewise, Fig. 3 depicts the throughput values of the optimization-based scheduling models for 20 VMs. From the figure, it can be justified that CQCSO achieves improved throughput. The better resource utilization of the proposed approach is

Fig. 3 Throughput



evident from the improved throughput rate and it is due to effective task allocation of CQCSO with minimum cost and lesser computation complexity. It can be concluded from the simulation results that the proposed CQCSO achieved optimal task scheduling in most scenarios in the cloud environment.

6 Conclusıon

This paper aimed at developing an efficient task scheduling strategy using an advanced multi-objective optimization algorithm. The proposed CQCSO algorithm has been developed by integrating the chaotic theory and quantum theory into the standard chicken swarm optimization. This proposed algorithm was intended to improve the task scheduling based on multiple parameters namely task completion time, response time and throughput with better QoS. The experimental results were obtained and compared with that of existing optimization-based task scheduling algorithms. This comparison results showed that the CQCSO has better performance with minimized task completion time, reduced response time and increased throughput than the existing algorithms. In the future, the proposed algorithm can be improved by integrating other feasible optimization concepts. Also, the inclusion of other objectives like energy will also be examined.

References

1. Furht B, Escalante A (2010) Handbook of cloud computing, vol 3. Springer, New York
2. Ergu D, Kou G, Peng Y, Shi Y, Shi Y (2013) The analytic hierarchy process: task scheduling and resource allocation in cloud computing environment. J Supercomput 64(3):835–848
3. Elzeki OM, Rashad MZ, Elsoud MA (2012) Overview of scheduling tasks in distributed computing systems. Int J Soft Comput Eng 2(3):470–475
4. Ilavarasan E, Thambidurai P (2007) Low complexity performance effective task scheduling algorithm for heterogeneous computing environments. J Comput Sci 3(2):94–103
5. Sindhu S, Mukherjee S (2011) Efficient task scheduling algorithms for cloud computing environment. In: International conference on high performance architecture and grid computing. Springer, Berlin, Heidelberg, pp 79–83
6. Guo L, Zhao S, Shen S, Jiang C (2012) Task scheduling optimization in cloud computing based on heuristic algorithm. J Netw 7(3):547
7. Jang SH, Kim TY, Kim JK, Lee JS (2012) The study of genetic algorithm-based task scheduling for cloud computing. Int J Control Autom 5(4):157–162
8. Al-Maamari A, Omara FA (2015) Task scheduling using PSO algorithm in cloud computing environments. Int J Grid Distrib Comput 8(5):245–256
9. Zuo L, Shu L, Dong S, Zhu C, Hara T (2015) A multi-objective optimization scheduling method based on the ant colony algorithm in cloud computing. IEEE Access 3:2687–2699
10. Meng X, Liu Y, Gao X, Zhang H (2014) A new bio-inspired algorithm: chicken swarm optimization. In: International conference in swarm intelligence. Springer, Cham, pp 86–94
11. Netjinda N, Sirinaovakul B, Achalakul T (2014) Cost optimal scheduling in IaaS for dependent workload with particle swarm optimization. J Supercomput 68(3):1579–1603

12. Tawfeek M, El-Sisi A, Keshk A, Torkey F (2015) Cloud task scheduling based on ant colony optimization. Int Arab J Inf Technol (IAJIT) 12(2):129–137
13. Abdullahi M, Ngadi MA (2016) Symbiotic organism search optimization based task scheduling in cloud computing environment. Future Gener Comput Syst 56:640–650
14. He H, Xu G, Pang S, Zhao Z (2016) AMTS: adaptive multi-objective task scheduling strategy in cloud computing. China Commun 13(4):162–171
15. Reddy GN, Kumar SP (2017) Multi objective task scheduling algorithm for cloud computing using whale optimization technique. In: International conference on next generation computing technologies. Springer, Singapore, pp 286–297
16. Bindu GH, Ramani K, Bindu CS (2018) Energy aware multi objective genetic algorithm for task scheduling in cloud computing. Int J Internet Protoc Technol 11(4):242–249
17. Kaur M, Kadam S (2018) A novel multi-objective bacteria foraging optimization algorithm (MOBFOA) for multi-objective scheduling. Appl Soft Comput 66:183–195
18. Gomathi B, Krishnasamy K, Balaji BS (2018) Epsilon-fuzzy dominance sort-based composite discrete artificial bee colony optimisation for multi-objective cloud task scheduling problem. Int J Bus Intell Data Min 13(1–3):247–266
19. Torabi S, Safi-Esfahani F (2018) A dynamic task scheduling framework based on chicken swarm and improved raven roosting optimization methods in cloud computing. J Supercomput 74(6):2581–2626
20. Kumar AS, Venkatesan M (2019) Multi-objective task scheduling using hybrid genetic-ant colony optimization algorithm in cloud environment. Wireless Pers Commun 107(4):1835–1848
21. Jacob TP, Pradeep K (2019) A multi-objective optimal task scheduling in cloud environment using cuckoo particle swarm optimization. Wireless Pers Commun 109(1):315–331
22. Abdullahi M, Ngadi MA, Dishing SI, Ahmad BIE (2019) An efficient symbiotic organisms search algorithm with chaotic optimization strategy for multi-objective task scheduling problems in cloud computing environment. J Netw Comput Appl 133:60–74

ARIMA for Traffic Load Prediction in Software Defined Networks

Sarika Nyaramneni, Md Abdul Saifulla, and Shaik Mahboob Shareef

Abstract Internet traffic prediction is needed to allocate and deallocate the resources dynamically and to provide the QoS (quality of service) to the end-user. Because of recent technological trends in networking SDN (Software Defined Network) is becoming a new standard. There is a huge change in network traffic loads of data centers, which may lead to under or over-utilization of network resources in data centers. We can allocate or deallocate the resources of the network by predicting future traffic with greater accuracy. In this paper, we applied two machine learning models, i.e., AR (autoregressive) and ARIMA (Autoregressive integrated moving average) to predict the SDN traffic. The SDN traffic is viewed as a time series. And we showed that the prediction accuracy of ARIMA is higher than the AR in terms of Mean Absolute Percentage Error (MAPE).

Keywords SDN traffic · Internet traffic prediction · Autoregressive · Autoregressive integrated moving average

1 Introduction

Due to the recent technological trends like cloud computing, the massive growth of e-commerce, mobile app-based ecosystems, the Internet of Things, etc., the amount with which new data is generated is overwhelming. Organizations should be ready to cater to the changes in the network traffic. If we forecast the network traffic then the network resources can be allocated or deallocated accordingly. In traditional networks, the logic resides in physical switches hardcoded into them and

S. Nyaramneni (✉) · M. A. Saifulla · S. M. Shareef
SCIS, University of Hyderabad, Gachibowli, Telangana, India
e-mail: sarikanyaram@gmail.com

M. A. Saifulla
e-mail: saifullah@uohyd.ac.in

S. M. Shareef
e-mail: smsshareef@outlook.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_75

because of this, it's a very difficult task to scale existing networks for highly dynamic network traffic, security, and privacy policies. As the network devices like switches, routers, firewalls, etc., keep on increasing, coupled with changing security and privacy policies the network becomes more complex.

In traditional networks, the task becomes complicated because of the limited set of device configuration commands. As the networks become more complex the task becomes more complicated which requires external tools and ad hoc methods to configure the network devices. SDN paradigm addresses these limitations by separating the data plane and control plane and making the control plane centralized. This centralized view of the network facilitates easy maintenance and up-gradation of the network [1–3]. In the initial step, we understand the type of statistics that could be collected in SDNs and how new information can be obtained from the collected data. In the next step, we employ an effective Machine Learning (ML) technique for traffic load prediction in SDN Data Centers [4, 5].

The organization of the paper follows as Related Work in Sects. 2 and 3 discusses the prediction technique used in this paper. Section 4 describes the experiments and analyzes their performance. Section 5 concludes the paper.

2 Related Work

Moayedi and Masnadi applied the ARIMA model to predict and detect the anomalies in traditional network traffic. They simulated the network with the random process, ARIMA process without trend, ARIMA process with trend, and ARIMA process with anomalies and they applied autocorrelation function and partial autocorrelation function on these [6]. Balaji, Yong Zeng, Kaushik Deka, and Medhi have proposed a framework to provide bandwidth dynamically by developing a Seasonal autoregressive Conditional Heteroskedasticity (ARCH) based model for the traditional network [7]. Faisal Iqbal, Md. Zahid, etc., were searched for a predictor that has higher accuracy and lower complexity of computation and lower consumption of power. And they concentrated on three different classes of predictors like classic time series, artificial neural networks, and wavelet transform-based predictors and they evaluated these different predictors by applying on real network traces (CAIDA traces, University of Auckland Traces, Bellcore Research Traces) [8]. Boutaba, Salahuddin, etc., summarized the different time series forecasting, as well as non-time series forecasting applied on network traffic [9].

3 Prediction Technique

Time series is a collection of data at a certain period of time. Time series data can be used for the forecast. Forecasting can be done on time series data and non-time series data [10]. The time series data will consist of a sequence of data (observation) with respect to time. The analysis of the time series data gives a good understanding of the data and provides the components that can be used to predict the data. Mainly three components are considered for prediction, they are, Level, which is average of the observations, Trend, which says the observations are increasing or decreasing, Seasonality, which means if there are any repeating traits in the given time series. Apart from these three components, there is another component, i.e., Noise, which can be outliers or unusual values. The series is aggregate of all these four components, where level and noise will be present in all the time series data and the other two are optional. The model can be seen as a combination of all these four components additively or manipulatively or even a complex combination of these components.

3.1 Autoregressive Integrated Moving Average (ARIMA)

The time series forecasting technique used in this paper is ARIMA (Autoregressive integrated moving average), the variant of the Autoregressive moving average (ARMA) [6]. ARMA has two parts autoregression and moving average. In Linear regression the output is modeled based on the linear combination of input values, we can use previous values like values at previous two timestamps (*t*–1) and (*t*–2) to predict the value at next timestamp (*t*), like

$$W(t) = x_0 + x_1 * W(t-1) + x_2 * W(t-2) \tag{1}$$

here $x_{0,}\ x_{1,}\ x_2$ are coefficients of the AR model.

As the model uses the previous time stamps of the same input variable to predict future value it is called autoregression. Moving average is used to smooth the data by removing the noise from the data, it's a simple method frequently used in time series forecasting. It calculates new values using the original observations in time series, the window size is the observations it uses to calculate the average and the window moves and again average is calculated, hence it's called moving average.

$$\text{newobs}(t) = \text{avg}(\text{obs}(t-2), \text{obs}(t-1), \text{obs}(t)) \tag{2}$$

So, autoregression along with Moving average is called ARMA. This method does not consider the trend or seasonality in data, so, it is suitable for the data where we don't have any trend or seasonality.

ARIMA models the future value as a linear function (ARMA) of differenced (I) observations at prior time stamps. So it combines autoregression (AR), Moving average (MA) and with a preprocessing step to difference, the observations called integration (I). This method is useful for data with trends [6]. The general form of the ARIMA model is ARIMA (*p*, *d*, *q*), for observable random variable *yt* as follows:

$$y_t = \sum_{i=0}^{p+d} \Phi_i y_{t-1} + \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} + \varepsilon_t \tag{3}$$

where $\{\Phi_i\}_{i=1}^{p+d}$ is an autoregressive parameter, $\{\theta_k\}_{k=1}^{q}$ is the moving average parameter, p order of autoregressive, d order of differencing, q order of moving average.

4 Experiments and Performance Analysis

We generated and captured SDN traffic for this experiment by using a mininet emulator but it is not a real-world dataset, and as SDN dataset is not publicly available, we downloaded and used a traditional dataset, i.e., 20031207-000000-0.gz which is of 24 hours continuous traffic trace released by the University of Waikato for this experiment [11]. After converting cumulative statistics of SDN traffic to timely statistics the behavior of SDN traffic and the traditional traffic remains the same.

In this paper, we applied autoregressive (AR) and ARIMA (Autoregressive integrated moving average) two-time series forecasting models on the dataset with different sampling intervals, i.e., 1 second, 1 minute, and 10, 15, 20, and 30 minutes to predict the SDN traffic. We predicted the traffic in terms of bytes. We measured the prediction accuracy of these two models using the Mean Absolute Percentage Error (MAPE). MAPE should be less than 20% for the good prediction model. MAPE can be calculated with the formula

$$\text{MAPE} = \frac{100}{N} \sum_{i=1}^{N} \left(\frac{A_t - P_t}{A_t} \right) \tag{4}$$

Here A_t is actual and P_t is predicted data and N is the number of inputs.

Before going into the experiment it is good practice to verify the dataset used for the experiment is predictable or not, this can be done by using lag plot, if the lag plot grows diagonally for the given time series data then that time series is predictable.

Figure 1 represents the lag plot for the dataset with a 1-second sampling interval, which we used for the experiment. This plot says that the dataset is predictable because all the points have grown diagonally in the plot. Now we can apply the prediction models on the selected dataset.

Figure 2 shows the autoregressive (AR) model prediction accuracy for the dataset. Here x-axis denotes the time in seconds and bytes denoted by the y-axis, and the red line is the predictions, blue line is the actual traffic, we can see the prediction accuracy of AR is very low. Figure 3 represents the prediction accuracy of ARIMA on the dataset with a 1-second sampling interval, and here also the time in seconds is denoted

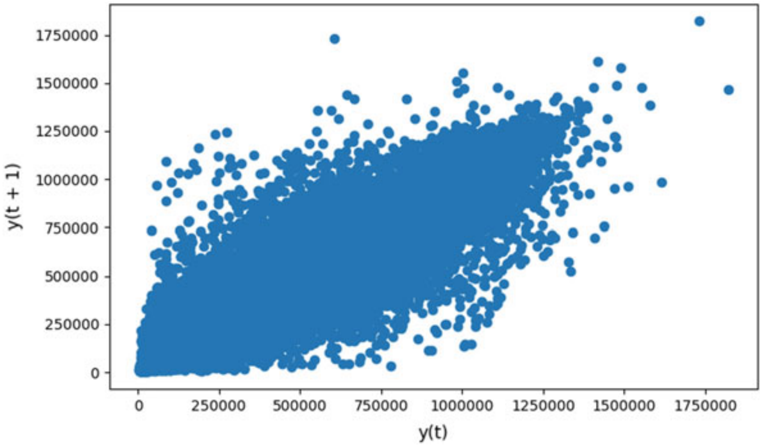


Fig. 1 Lag plot for 1-second sampling interval

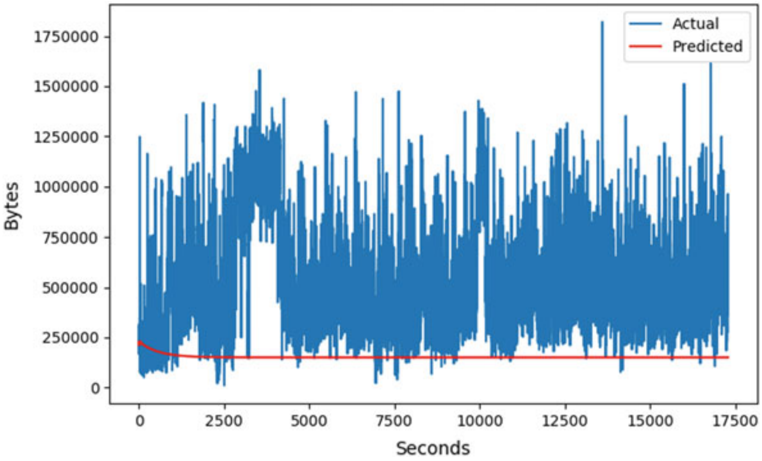


Fig. 2 AR result plot for 1-second sampling interval

by the *x*-axis and the *y*-axis denotes the bytes and Fig. 3 shows that the prediction accuracy of ARIMA model is more than the AR model.

Now, the same dataset is resampled to a 1-minute sampling interval, such that time series is having a minute gap between readings instead of seconds. From Fig. 4 (lag plot for the dataset with 1-minute sampling interval) even though the dataset is resampled to 1 min still it is predictable.

AR model prediction accuracy is shown in Fig. 5, here we can observe that there is a substantial amount of error in predicted values the same as sampling interval of 1 second, however, it is less in comparison. After that, we applied ARIMA to predict SDN traffic. Figure 6 shows the prediction accuracy of ARIMA, here we can observe

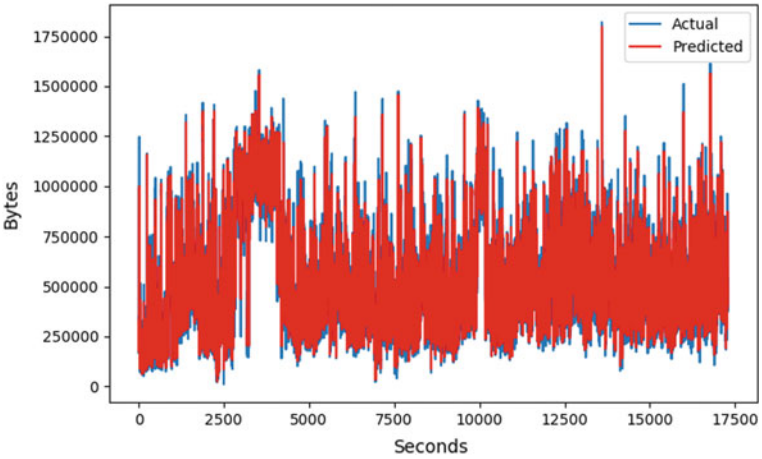


Fig. 3 ARIMA result plot for 1-second sampling interval

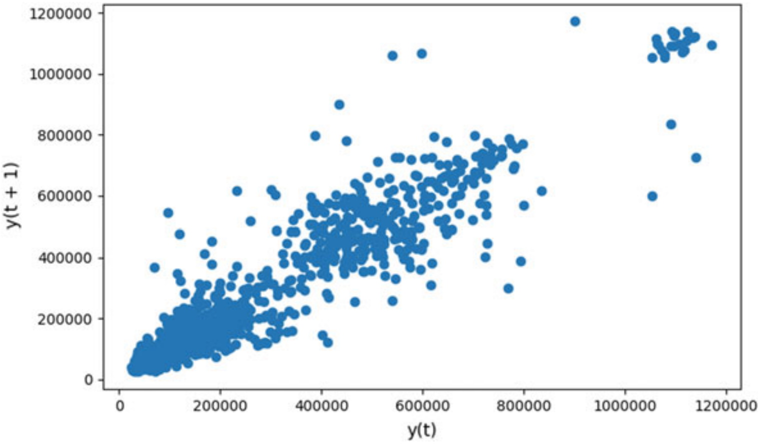


Fig. 4 Lag plot for 1-minute sampling interval

that the prediction accuracy of ARIMA is better than AR.

Again we resampled the dataset to 30 minutes sampling interval and Fig. 7 shows the lag plot for the resampled dataset, in this plot, there are some outliers which reduce the accuracy of prediction. Still, it can be predictable because more points are on the diagonal.

Figure 8 shows the autoregressive (AR) model's prediction accuracy for the dataset. Here the time in minutes is represented by the x-axis and bytes represented by the y-axis, and the red line is the predictions and the blue line is the actual traffic. Figure 9 represents the prediction accuracy of ARIMA on the dataset with 30 minutes sampling interval here also the time in minutes is indicated by the *x*-axis

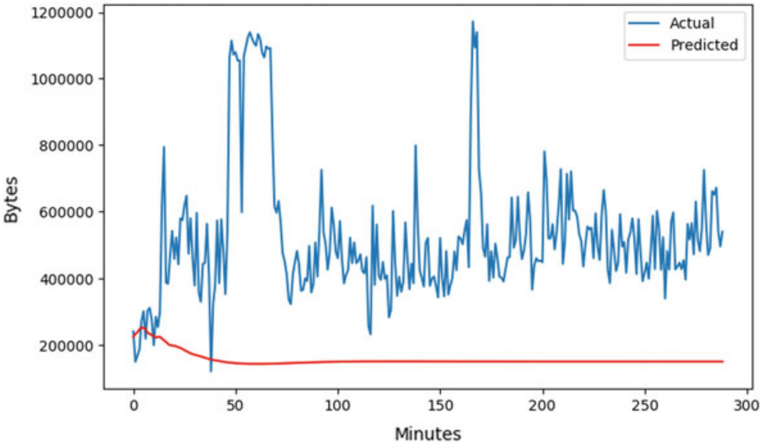


Fig. 5 AR result plot for 1-minute sampling interval

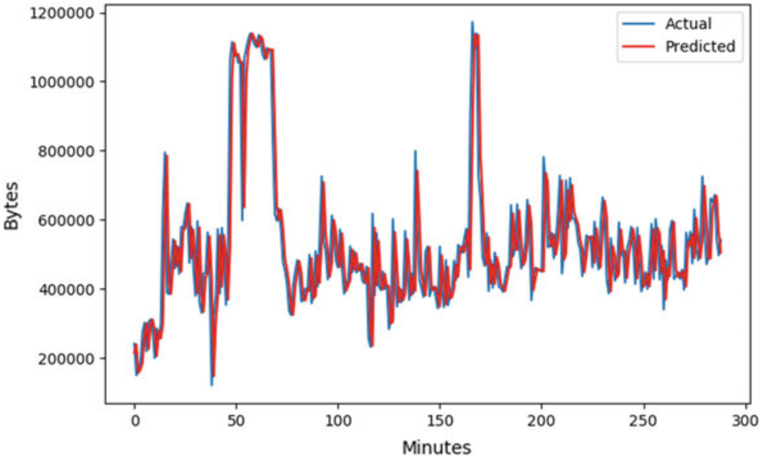


Fig. 6 ARIMA result plot for 1-minute sampling interval

and the throughput in terms of bytes indicated by the *y*-axis. From Figs. 8 and 9 we can say that the prediction accuracy of AR model is lesser than ARIMA. Likewise, we applied these two models on the remaining 10, 15, 20 minutes sampling intervals.

Table 1 describes the measures of prediction accuracy of AR and ARIMA in terms of Mean Absolute Percentage Error (MAPE) when we apply on the dataset with three different sampling intervals (seconds, 1 minute, 10, 15, 20, 30 minutes). From this table, we observed that the MAPE value of ARIMA is less than 20% for the 1, 10, and 15 minutes sampling intervals of the time series and the AR model MAPE value is large (MAPE >20% is not considerable).

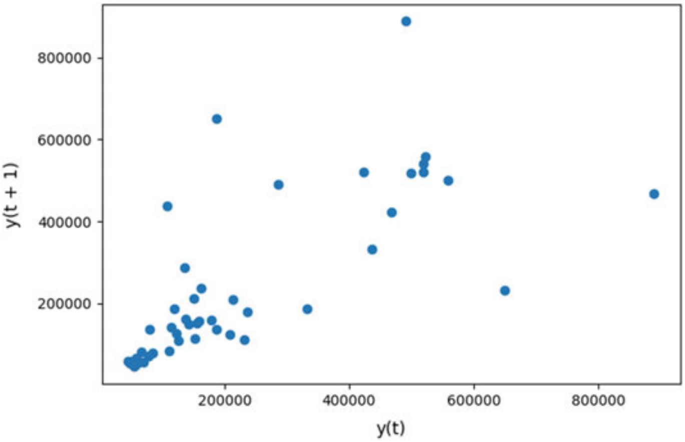


Fig. 7 Lag plot for 30-minutes sampling interval

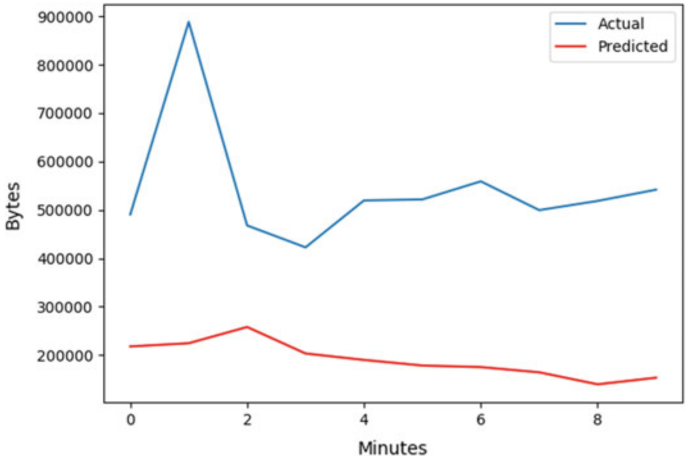


Fig. 8 AR result plot for 30-minutes sampling interval

5 Conclusion

In this paper, we applied and compared two machine learning models namely, autoregressive (AR) and Autoregressive integrated moving average (ARIMA) to predict the SDN traffic with different sampling intervals. According to the Table 1 AR model is not preferable for the SDN traffic prediction (because MAPE is >20%) So, among these two models, ARIMA is preferable for the SDN traffic prediction (MAPE is <20% for 1 Minute, 10 Minutes, and 15 Minutes sampling intervals). In order to improve the accuracy of prediction, we will apply the other prediction models in the future.

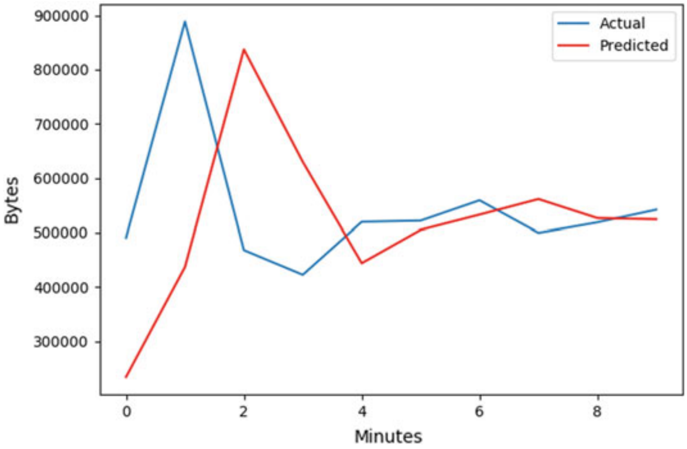


Fig. 9 ARIMA result plot for 30-minutes sampling interval

Table 1 MAPE values of AR and ARIMA

Sampling interval of time series	MAPE (Mean Absolute Percentage Error)	
	Autoregressive (AR) model	Autoregressive Integrated Moving Average (ARIMA) model
1 Second	65.414	24.671
1 Minute	66.726	18.569
10 Minutes	69.398	18.372
15 Minutes	66.221	19.159
20 Minutes	67.869	21.645
30 Minutes	63.713	27.072

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

1. Hamad DJ, Yalda KG, Okumus IT (2018) Getting traffic statistics from network devices in an SDN environment using OpenFlow. http://itas2015.iitp.ru/pdf/1570195931.pdf
2. Braun W, Menth M (2014) Software-defined networking using openflow: protocols, applications and architectural design choices. Dir Open Access J (DOAJ)
3. Akyildiz IF, Lee A, Luo M, Wang P, Chou W (2014) A roadmap for traffic engineering in SDN-OpenFlow networks. ELSEVIER Comput Netw 71:1–30
4. Feng H, Shu Y (2005) Study on network traffic prediction techniques. In: Proceedings. 2005 international conference on wireless communications, networking and mobile computing

5. Parsaei MR, Sobouti MJ, et al (2017) Network traffic classification using machine learning techniques over software defined networks. J Adv Comput Sci Appl (IJACSA) 8(7):79–99
6. Moayedi HZ, Masnadi-Shirazi MA (2008) Arima model for network traffic prediction and anomaly detection. In: 2008 international symposium on information technology
7. Krithikaivasan B, Zeng Y, Deka K, Medhi D (2007) ARCH-based traffic forecasting and dynamic bandwidth provisioning for periodically measured nonstationary traffic. IEEE/ACM Trans Netw 15(3):683–696, June 2007
8. Iqbal MF, Zahid M, Habib D, John LK. Efficient prediction of network traffic for real-time applications. J Comput Netw Commun 2019. Article ID 4067135. https://doi.org/10.1155/2019/4067135
9. Boutaba R, Salahuddin MA, Limam N, Ayoubi S, Shahriar N, Estrada-solano F, Caicedo OM (2018) A comprehensive survey on machine learning for networking: evolution, applications and opportunities. J Internet Serv Appl 9(16):99
10. Brownlee J (2019) Classical time series forecasting methods. 02 Jan 2019. https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/
11. https://wand.net.nz/wits/waikato/1/waikato_i.php. Accessed 23 Oct 2019

Improved Logistic Map Based Algorithm for Biometric Image Encryption

Mahendra Patil, Avinash Gawande, and D. Shelke Ramesh

Abstract Tumultuous maps are ordinarily favored for the age of arbitrary numbers in encryption calculations because of high arbitrariness, capriciousness, aperiodic, and affectability. These calculations are completely required upon the underlying qualities and bear high correspondence to cryptographic calculations. Right now, the proposed encryption calculations are dependent on improved strategic guide for biometric picture encryption. The proposed calculated guide is improved to conquer its lacuna concerning haphazardness, irregular dispersion, unbound, and little parameter space. Besides the proposed calculated guide produces arbitrary numbers dependent on ongoing factors, haphazardly chose qualities and the last numbers are diffused arbitrarily through XOR activity. Trial results on different biometric unique mark pictures exhibit better encryption parameters and around level histogram of the encoded pictures that expand trouble during unapproved unscrambling. The proposed calculation can be viably considered for multi-biometric unique mark picture encryption in government and nongovernment associations because of its less intricacy and less encryption time.

Keywords Calculated map · Encryption algorithm · Biometric images · Encryption time

M. Patil (✉)
Faculty of Engineering, Pacific Academy of Higher Education and Research University, Udaipur, India
e-mail: apmahendra@yahoo.com

A. Gawande
Sipna College of Engineering & Technology, Amravati, Maharashtra, India

D. Shelke Ramesh
Shivajirao S. Jondhale College of Engineering Dombivali (E), Mumbai, India

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_76

1 Presentation

Expanding quantities of biometric pictures are acquired, put away, transmitted, and checked in different associations. The utilizations of biometric pictures, for example, iris, face, and unique mark incorporate day-by-day participation checking frameworks, brilliant passage frameworks, access to savvy individual correspondence frameworks, banking, UIDAI, and so on. Because of the significant uses of biometric pictures, encryption of these pictures is critical from the security point of view that may bring about loss of individual and monetary data. In this manner, it is important to build up a safe, however, less mind-boggling encryption calculation that gives better encryption parameters and takes less encryption/unscrambling time. An encryption calculation has been extensively planned by examined researchers/designers for different types of information prior and is subject to enthusiasm for a long time due to its need. Along these lines, extensive investigation has occasioned in standard calculations or strategies like Data Encryption Standard (DES), Advanced Encryption Standard (AES), RC4, and RC6. Actually, it is for the most part seen that the use of these calculations, for example, DES, AES, RC6, and so on for picture encryption was not helpful with required parameters. Subsequently, it is comprehended that the picture encryption calculations stay divergent as contrasted and information encryption plans for the explanation that pictures information are ordinarily huge in size, likeness in information, needs extra period for encryption, and likewise needs a proportionate amount of period for unscrambling. Computational overhead like complex numerical tasks, trigonometry, logarithms, changes makes the good old encryption calculation unsatisfactory for present picture encryption and decoding [1–3]. Various arrangements built up on spatial or time and recurrence or change space strategies are prescribed for picture encryption [4–7]. Spatial or time area strategy utilizes antiquated encryption calculations despite the fact that recurrence space utilizes changes, for example, DCT, DWT, DFT, and FFT. Anyway, the basic thought is to revise pixel areas in the pictures to be encoded by turbulently produced change orders utilizing different methods [8–10]. Two noteworthy highlights of prevailing encryption systems or calculations are dissemination and disarray. Thusly, it is constantly needed to devise a plan or strategy or algorithm that can include additional dispersion and disarray properties. Likewise, it is tentatively seen that basic highlights of biometric pictures are the significant bottleneck in the usage of customary encryption plans. Consequently, these customary frameworks are inadmissible for picture encryption. In this way, the motivation is to devise a plan to scramble biometric pictures skillfully without relinquishing fundamental highlights. To improve the security of biometric pictures, a biometric-based cryptosystem approach is vital that consolidates cryptography and biometrics is basically required. This component through a biometric cryptosystem is comprehended to improve the security of biometric pictures during its handling. Basically, the measure is the quantity of pixels change rate (NPCR) and bound together normal evolving power (UACI) in pixels between two information and scrambled pictures [11]. Thirdly, entropy of the info and encoded pictures that propose a range of pixels' esteems secured, for instance, 0–255 (8) and 0–128 (7). The proposed encryption calculations are dependent on improved strategic guide for biometric picture encryption. It was tentatively seen that the proposed strategic guide based encryption calculation takes high keyspace, entropy roughly equivalent to eight (8) for 8-piece dark pictures and around zero connection on a level plane, vertically and askew contiguous pixels. Another noteworthy bit of leeway of the proposed encryption calculation is the perfect qualities for NPCR, UACI, and entropy it offers. Furthermore, it is likewise tentatively seen that the histogram of the encoded picture is practically level. It obviously builds the trouble in the unscrambling process for an unapproved individual or who doesn't have a decoding key. The proposed calculation or methodology can be applied for multi-biometric picture encryptions in government and nongovernment association because of its less multifaceted nature and less encryption time. The paper is organized as follows: Segment I and II presents and audits encryption calculations applied for biometric pictures separately. Segment III portrays the proposed improved strategic guide based encryption and decoding methodology in the points of interest. Trial results are discussed in Area IV and concluded in Sect. 5.

2 Related Work

Numerous analysts, researcher, architects, and designers have recognized that "1D disordered mapping calculation" have been determined to have numerous security glitches because of its straightforward game plan and poor cryptographic execution. Enormous advances in cryptanalysis recommend 2D sine or calculated confused mapping. Turbulent maps are generally favored for the age of arbitrary numbers in encryption calculations because of high irregularity, flightiness, aperiodic, and affectability. These calculations are completely reliant upon the underlying qualities and bear high correspondence to cryptographic calculations. It is without a doubt necessitated that strategic guide be improved to beat its lacuna concerning arbitrariness, irregular conveyance, unbound, and little parameter space. In this manner, "2D sine regulated mapping calculations" [10] that are impacted after the strategic and sine maps are important to daze lacunas of a strategic guide. A calculation dependent on two independent disorganized premise capacities with beginning states can altogether utilize wanted disarray and dispersion. These arbitrary numbers acquired through disorganized maps are utilized for adjusting pixel areas and differing the estimations of pixels. Along these lines, new pixel plan either its area or esteem or both in the scrambled biometric picture will be altogether unique as for input picture that makes encoded biometric picture hard to split. To additionally add the security to the information of biometric picture, select OR (XOR) tasks and vertical/level turns can be acquainted that makes it safe with different differential assaults. Another "2D calculated iterative riotous guide" with "unbounded breakdown (ICMIC) course mapping dependent on course tweak coupling model" [12, 13] was illustrated. The presentation assessment of exploratory outcomes delineated that the calculation has the advantages of hyper turbulent conduct, wide confusion range, and high multifaceted nature. A picture square encryption system built upon turbulent maps portions the picture into squares and scrambled with a selective or activity and disorderly window [14–16] which obviously demonstrates that it has a huge keyspace, and the resultant encoded pictures have homogeneous histograms. A few calculations scramble the information biometric picture utilizing changes, for example, Arnold change to rearrange pixel positions. It likewise adds security to the information picture and results have demonstrated better clamor assault properties. Anyway, the great calculated guide calculation with wanted disarray and dissemination properties may smother the upsides of scramblers.

3 Improved Logıstıc Map Based Encryption Algorıthm

Right now, 2D improved calculated guide based methodology for encryption of unique mark pictures is discussed next to the unscrambling system. Strategic maps are for the most part alluded as the wormhole model. A straightforward 1D strategic guide numerical articulation can be expressed as

$$x(i + 1) = \mu x(i)(1 - x(i)) \tag{1}$$

where $x \in (0, 1)$ and $\mu \in (0, 4)$ are characterized as strategic guide parameters. The estimations of calculated guide parameters are characterized inside the endorsed run that makes the created numbers aperiodic, irregular, and they don't combine. Though the calculated guide parameters outside the recommended range might be deterministic and unite too scarcely any qualities. We have 1D strategic guide by presenting new parameter $y \in (0, 1)$ alongside x that adds greater security to the calculation. The number arrangement produced through these three parameters x, y, and μ is irregular, aperiodic, and doesn't combine. It is important that the estimations of the parameters should be in the endorsed run.

3.1 Encryption Procedure

1. Let B be the 8-piece dark scale input biometric unique mark picture of size $p \times q$. The stages expanded in the encryption of the biometric unique mark picture are as per the following. Reset the clamorous parameters μ, $x(0)$, and $y(0)$. Reset N is the quantity of redundancies, $I = 0$.
2. Acquire irregular number succession utilizing clamorous improved strategic guide, reiterations, and overhead parameters as follows:

$$x(i + 1) = \mu x(i)(1 - x(i)) \tag{2}$$

$$y(i + 1) = (\mu + \mu\, y(i))x(i)\,(1 - x(i)) \tag{3}$$

$$x(i) = (\text{floor}(x(i) * 256)) \bmod m \tag{4}$$

$$y(i) = (\text{floor}(y(i) * 256)) \bmod m \tag{5}$$

We have set the estimation of parameters $x(0)$, $y(0)$, and μ to 0.23, 0.564, and 3.89, which are called calculated parameters and $m \times m$ is the size of the information biometric picture. These strategic parameters are constrained by $y(i)$ and μ together. In this way, the created yield grouping is aperiodic, unpredictable, and arbitrary. At long last piece selective OR activity is applied to make it progressively secure.

$$z(i) = x(i)\ \text{XOR}\ y(i) \tag{6}$$

3. Horizontal stage move, jth section of the information biometric unique finger impression picture to nth segment of the yield picture, where n is gained from the jth estimation of the irregular number z.
4. Vertical stage move, jth line of a level plane permutated input biometric unique finger impression picture to nth column of the yield picture where n is gained from the jth estimation of the irregular number z. Obtained $Ec(p, q)$ is the encoded picture through flat and vertical changes.
5. Gray-level dissemination utilizing two XOR tasks

$$E1(p, q) = \text{bitxor}(Ec(p, q), z(m - 1)) \tag{7}$$

$$E(p, q) = \text{bitxor}(E1(p, q), z) \tag{8}$$

6. Repeat the procedure for N number of reiterations, $I = I + 1$. The resultant scrambled picture (E) can be put away and transmitted with no issues. The arbitrary vector z shapes the key of the encryption that can't be produced until the strategic parameters are known. Little change in strategic parameters totally changes the arbitrary arrangement and doesn't accomplish precise unscrambling.

3.2 Decryption Procedure

The complete procedure of decoding is expressed with the required advancements. Let E be the 8-piece dark scale scrambled biometric unique finger impression picture of size $p \times q$.

1. Reset the disorganized parameters μ, $x(0)$, and $y(0)$.
2. Reset N is the number of reiterations, $I = 0$.

3. Acquire arbitrary number succession utilizing tumultuous improved calculated guide, reiterations, and overhead parameters as follows:

$$x(i + 1) = \mu x(i)(1 - x(i)) \tag{9}$$

$$y(i + 1) = (\mu + \mu y(i))x(i)(1 - x(i)) \tag{10}$$

$$x(i) = (\text{floor}(x(i) * 256)) \bmod m \tag{11}$$

$$y(i) = (\text{floor}(y(i) * 256)) \bmod m \tag{12}$$

We have set the value of parameters $x(0)$, $y(0)$, and μ to 0.23, 0.564, and 3.89 which are called as logistic parameters and $m \times m$ is the size of the input biometric image. These logistic parameters are controlled by $y(i)$ and μ jointly. Thus, the generated output sequence is aperiodic, complex, and random. Finally, bitwise XOR operation is applied to make it more secure.

$$z(i) = x(i)\ \text{XOR}\ y(i) \tag{13}$$

1. Gray-level diffusion using two XOR operations:

$$E1(p, q = \text{bitxor}(E(p, q), z) \tag{14}$$

$$Ec(p, q = \text{bitxor}(E1(p, q), z(m - 1)) \tag{15}$$

2. Vertical stage move, jth column of the info biometric unique mark picture to nth line of the yield picture where n is gained from the jth estimation of the arbitrary number z.
3. Horizontal stage move, jth section of the vertically permutated input biometric unique mark picture to nth segment of the yield picture where n is gained from the jth estimation of the irregular number z.
4. Obtained $D(p, q)$ is the scrambled picture through flat and vertical stages.
5. Repeat the procedure for N number of reiterations, $I = I + 1$.

The total methodology for encryption and decoding of the biometric picture through the disorderly calculated guide is laid out.

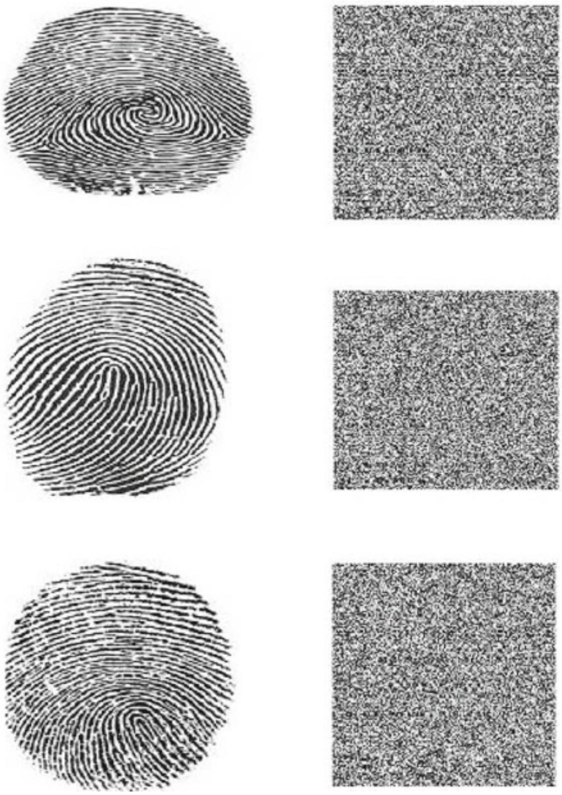
4 Experimental Results

Investigational results for visual testing, entropy assessment, pressure attack, and noise ambush are jumped on the mixed picture using muddled determined guide based picture encryption plan with various degrees of perplexity and scattering (Fig. 1).

4.1 Visual Testing

No pictographic closeness between the info and scrambled pictures was seen that totally shows the goal of encryption calculation. The deliberate estimations of NPCR and UACI are portrayed in Table 1.

Fig. 1 Visual testing



(a) Input Images

(b) Encrypted Images

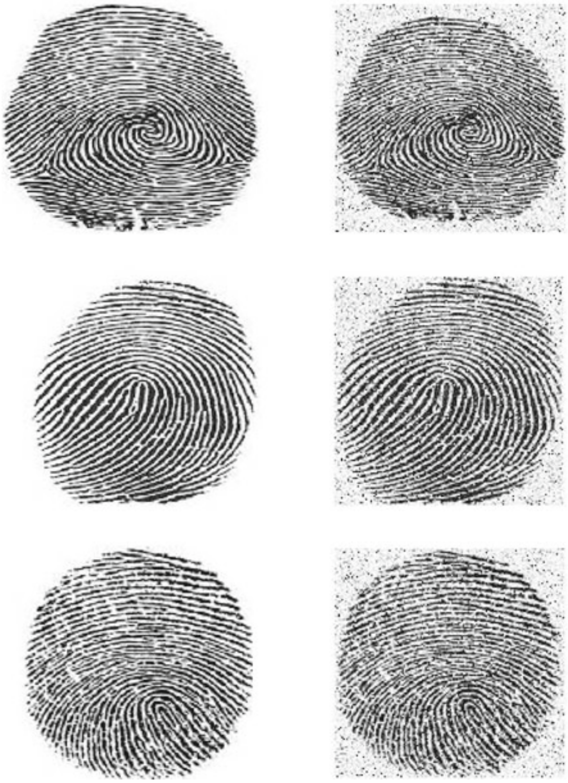
Table 1 NPCR and UACI values

Fingerprint image	NPCR	UACI
1	99.56	27.16
2	99.57	28.24
3	99.58	28.10

4.2 Noise Attack

Right now, salt and pepper commotion of thickness 0.1 was brought into the scrambled picture utilizing the confused strategic guide. Figure 2 portrays the boisterous encoded input picture and its decoded partner pictures. The investigational results totally exhibit that the clamor assault characterized as salt and pepper or drive commotion impacts the encoded picture which is portrayed in the picture shown in Table 2.

Fig. 2 Noise testing



(a) Input Images (b) Decrypted Images

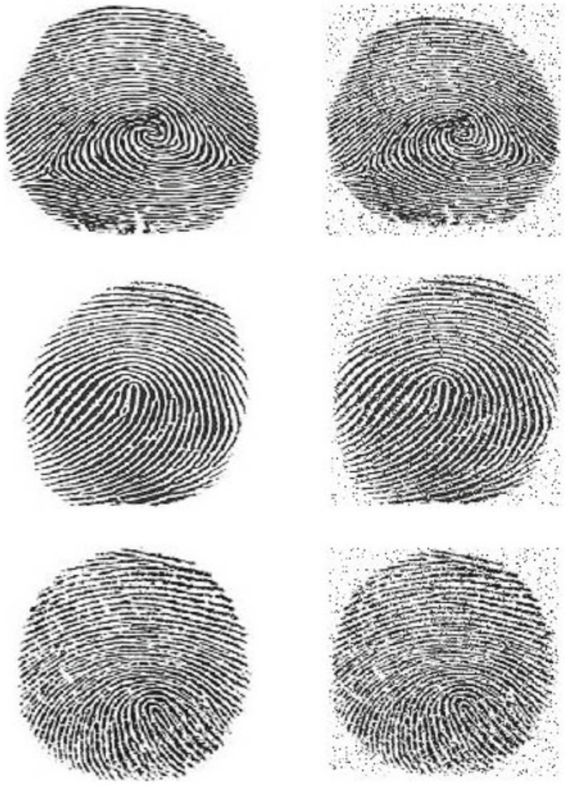
Table 2 Noise and compression attack

Fingerprint image	Noise attack	Compression attack
	Correlation coefficients	
1	0.91	0.9
2	0.91	0.89
3	0.9	0.89

4.3 Compression Attack

Right now, a special scrambled biometric unique mark picture was compacted to JPEG/JPG position and some time ago unscrambled utilizing the confused calculated guide. Figure 3 shows the info biometric unique mark picture that was scrambled utilizing tumultuous calculated guide and the unscrambled picture got in the wake of coding and interpreting utilizing the JPEG/JPG calculation. In this manner, it is totally shown that pressure doesn't extensively upset the unscrambling procedure and

Fig. 3 Compression testing



(a) Input Images (b) Decrypted Images

Table 3 Entropy values

Fingerprint image	Input image’	Encrypted image
1	6.47	7.99
2	6.28	7.99
3	6.5	7.99

Table 4 Comparison of biometric image encryption methods

Algorithms	NPCR	UACI	Entropy
Rubik’s cube [17]	98.77	26.4	7.94
chaotic sine map [18]	99.61	32.67	7.99
Logistic chaotic map	99.57	27.83	7.99

is very viable. Table 2 presents estimations of connection coefficients acquired among the first biometric unique finger impression picture and deciphered accordingly with the presentation of commotion and pressure assaults (Table 3).

4.4 Entropy Analysis

Comparative assessment of three significant encryptions techniques, for example, Rubik’s Cube Principle, Sine Chaotic Map, and proposed Logistic Chaotic Map is portrayed in Table 4. Results shows that the proposed disordered calculated guide based picture encryption conspire for unique mark biometric pictures is better in terms of all encryption parameters.

5 Conclusion

The research work finalized the clamorous calculated guide based picture encryption plot for unique mark biometric pictures right now. To improve the encryption gain investigation recreations are analyzed by using the parameters such as pressure, testing, assault commotion, and clamor. The calculation utilizing disordered strategic guide is for the most part appropriate for biometric unique finger impression picture encryption. The proposed calculation is incredible to pressure assault and delights all picture encryption parameters, for example, NPCR, UACI, and entropy.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Zhu Z, Zhang W, Wong K-W et al (2011) A chaos-based symmetric image encryption scheme using a bit-level permutation. Inf Sci 181:1171–1186
2. Wen J, Chang X-W (2015) A modified KZ reduction algorithm. Proc IEEE Int. Symp. Inf. Theory 7:451–455
3. Bhatnagar G, Wu QJ (2012) Chaos-based security solution for fingerprint data during communication and transmission. IEEE Trans Instrum Meas 61(4):876–887
4. Rostami MJ, Shahba A, Saryazdi S et al (2017) A novel parallel image encryption with chaotic windows based on logistic map. Comput Electr Eng 62:384–400
5. Ye GD, Huang X (2017) An efficient symmetric image encryption algorithm based on an intertwining logistic map. Neurocomputing 251:45–53
6. Sui L, Duan K, Liang J (2015) Double-image encryption based on discrete multiple-parameter fractional angular transform and two-coupled logistic maps. Opt Commun 343:140–149
7. Chen G, Mao Y, Chui CK (2004) A symmetric image encryption scheme based on 3D chaotic cat maps. Chaos Solitons Fractals 21(3):749–761
8. Sun J, Liao X, Chen X, Guo S (2017) Privacy-aware image encryption based on logistic map and data hiding. Int J Bifurc Chaos 27(5). Art. no. 1750073
9. Li C, Xie T, Liu Q, Chen G (2014) Cryptanalyzing image encryption using chaotic logistic map. Nonlinear Dyn 78(2):15451551
10. Kori P, Dubey P, Richhariya (2015) Double phase image encryption and decryption using logistic tent map and chaotic logistic Map. IJSART Int J Res Technol 11:33–39
11. Majumdar S (2014) An analytical survey on different secured image encryption techniques. IJCAT Int J Comput Technol 8:396–403
12. Chatterjee S, Roy S, Das AK, Chattopadhyay S, Kumar N, Vasilakos AV (2018) Secure biometric-based authentication scheme using chebyshev chaotic map for multi-server environment. IEEE Trans Dependable Secure Comput 15(5):824–839
13. Yap W-S, Phan RC-W, Yau W-C, Heng S-H (2015) Cryptanalysis of a new image alternate encryption algorithm based on chaotic map. Nonlinear Dyn 80(3):1483–1491
14. Tong X-J et al (2016) A joint color image encryption and compression scheme based on hyper chaotic system. Nonlinear Dyn 84(4):2333–2356
15. Zhu C (2012) A novel image encryption scheme based on improved hyperchaoticsequences. Opt Commun 285(1):29–37
16. Pareek NK, Patidar V, Sud KK (2006) Image encryption using chaotic logistic map. Image Vis Comput 24:926–934
17. Patil MV, Gawande AD, Dilendra (2019) Biometric image encryption algorithm based on modified Rubik’s cube principle. Springer lecture notes in computational vision and biomechanics, vol 30, pp 1865–1873
18. Patil MV, Gawande AD, Dilendra (2019) Biometric image encryption based on chaotic sine map and information entropy. Springer lecture notes on data engineering and communications technologies, vol 38, pp724–732

Data Security System for A Bank Based on Two Different Asymmetric Algorithms Cryptography

Md. Ashiqul Islam, Aysha Akter Kobita, Md. Sagar Hossen, Laila Sultana Rumi, Rafat Karim, and Tasfia Tabassum

Abstract Recently, a strong security system is very important for a safe banking system. To prevent hacking of important information of bank and client, a secure banking system is a must. This paper deals with a strong security system using a hash function and two different asymmetric algorithms (DSA and RSA) at a time, it will enhance data security. We are using RSA and DSA encryption algorithm to secure our system from unauthorized access. In RSA and DSA, we have to generate two keys called Public and Private Keys. If we use the signer's private key for encryption, then we have to use the signer's public key for decryption. The system will verify by confirmation and certificate and the sender will be sent an OTP via mobile phone of the receiver to confirm the authentication. This is the most efficient data security system to save the bank from hacktivism.

Keywords Cryptography · Asymmetric algorithm · DSA · RSA · OTP

Md. Ashiqul Islam (✉) · A. A. Kobita · Md. Sagar Hossen · L. S. Rumi · R. Karim · T. Tabassum
Daffodil International University, Savar, Dhaka, Bangladesh
e-mail: ashiqul15-951@diu.edu.bd

A. A. Kobita
e-mail: ayshakobita@gmail.com

Md. Sagar Hossen
e-mail: sagar15-1504@diu.edu.bd

L. S. Rumi
e-mail: lailasultanarumi@gmail.com

R. Karim
e-mail: rafat15-867@diu.edu.bd

T. Tabassum
e-mail: tabassum.tasfiaahmed@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_77

1 Introduction

In this modern era, we are exchanging our important information, files over the Internet very fast and easily. It is the easier way to exchange our information via internet and it is the easiest path hacker can follow to stealing our information to achieve their motive. If we talk about the banking system, the picture is the same. So, a strong security system is must everywhere. At present, we need to ensure data security and confidentiality of the client's data in order to maintain the client's confidence in the banking system. Whenever we talk or think about a secure and strong security system, the first thing that comes to our mind is Cryptography.

Cryptography is the method of transforming plain text into incomprehensible text or code, which cannot be readable by any unauthorized use without the sender and receiver. Cryptography plays a fundamental role to build up data security and secrecy of data from third-party access. Cryptography ensures some individual security conditions before giving access to any user in any system.

Authentication: The mechanism of verifying a valid system user.

Privacy/confidentiality: Assuring that no one else can learn about the data or information except the valid receiver.

Integrity: Ensuring that the recipient receives the intact and unedited message of the sender.

Non-repudiation: The process is used to confirm that the sender actually sent this message and the valid recipient receives it.

In different types of Cryptographic algorithms, we are going to use Public-Key Cryptography (PKC) and Hash Function along with One-Time Password (OTP) to design the security system.

Public-Key Cryptography (PKC)

Public-key cryptography or Asymmetric cryptography is the method of encrypting and decrypting data using the public key and private key, respectively [1]. This paper includes two popular asymmetric algorithms RSA (Rivest–Shamir–Adleman), DSA (Digital Signature Algorithm) [2].

RSA (Rivest–Shamir–Adleman)

The RSA algorithm is the foundation of a cryptographic system that is used for security reasons. It allows asymmetric encryption and it is popular for secure data transmission process over a wireless network like the Internet. RSA is the most popular encryption algorithm that is used in the banking security system.

DSA (Digital Signature Algorithm)

To provide strong data security, DSA is used to prove user authentication. DSA is an asymmetric key algorithm and it creates a digital signature. It helps to secure sensitive data by issuing a distinctive identity to its signers. The sender uses the Hash function to generate a message digest using a private and public key pair, signs it digitally with his signature, and sends it. The receiver has to use the sender's public key to decrypt the signature [3]. It helps to prove the authenticity of the sender and data.

Hash Function

Hash function is an arithmetical conversion that is used for data encryption and decryption [4].

One-Time Password (OTP)

To provide two-step verification, we use One-Time Password (OTP) in our system. It will strengthen our security more.

2 Related Work

Some authors presents a comprehensive survey on the utilization of blockchain technology and provided distributed security services. These services include entity authentication, confidentiality, privacy, provenance, and integrity assurances [4, 2, 5]. At this point, we should take a serious look at banking security [6]. And [7] Sankalp Jagga has discussed in detail the comparison of authentication technique and encryption technique on banking security. Many encryption techniques already have privacy insured with different algorithms like Digital Signature Algorithm (1024 bits), Hashing, RSA Digital Signature (1024 bits) in different models [1, 8, 9]. At present, many writers are thinking about developing a mobile security system as we keep mobile as not only personal communication but also the companion to bank email. So mobile can be secured by signature authentication. Vagner Schoaba [10, 11] developed a system that can secure mobile with a digital signature. And cryptographic techniques developed a lot of applications in the banking industry. Arpan says [12] that implementation of data security makes business management more efficient. Even, OTP can be more secure in our system [3, 13].

3 Proposed System Design

Here we are going to present our recommended work in which we are presenting a data security system for a bank based on two different asymmetric algorithm cryptography to protect important data and sensitive information from leaking. In our proposed work, the system use the following techniques or algorithm (Fig. 1):

- Hash Function,
- Asymmetric Encryption using RSA and DSA,
- One-Time Password (OTP).

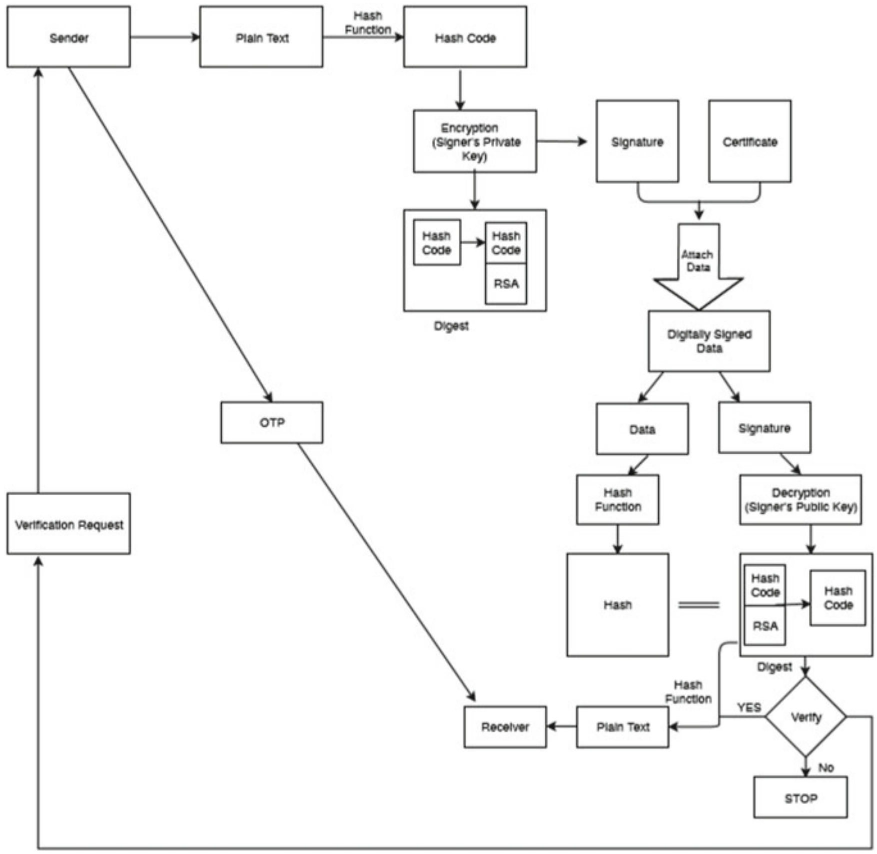


Fig. 1 Proposed system [2]

Hash Function

Hash function converts a plain text into a hash code which is unreadable for an ordinary person [4]. It uses fixed-length hash value for data encryption or decryption instead of keys. It ensures data integrity that the data has not been changed by any intruder or an unauthorized user.

Asymmetric Encryption using RSA and DSA

Asymmetric encryption used two different keys, Public Key and Private Key, for data encryption and decryption between the sender and receiver. Using two different keys for encryption and decryption the strong security is ensured. In an encryption process, RSA is used to encrypt the hash code one more time and attach a digital signature with it and then send it to the specific receiver. A digital signature confirms the identity of the valid sender.

One-Time Password (OTP)

One-time password is used to verify that the receiver of the message or information is the authentic one. In this, the sender sent a code to the receiver and the code changes every time for security reasons.

4 Methodology

4.1 Algorithm for the System

Step 1: Start the procedure.

Step 2: Send plain text and generate a hash code using hash function and encrypt it using the signer's private key.

Step 3: Digest the hash code and encrypted data using RSA Algorithm. It can produce a digital signature and certificate.

Step 4: Attach the signature and certificate to generate digitally signed data.

Step 5: Then digitally signed data is divided into data and signature. Data is produced hash function to hash code and signature are decrypted using the public key and hash code is generated.

Step 6: If the data hash code and signature hash code are the same, then it sends to the receiver otherwise stop the procedure.

Step 7: The receiver sends a confirmation/verify request to the sender to authenticate the verification process.

Step 8: Sender checks the verification request and sends OTP to the receiver mobile phone and authenticate the system.

Step 9: Using an OTP, receiver will get the plain text.

Step 10: End Procedure.

4.2 System Operation

In the proposed methodology, two different asymmetric algorithms along with hash function and One-time password are combined in order to make our security system more stronger so that the sensitive information and messages can be passed in a secure way without accessing of any unauthorized use and messages will remain safe from alteration.

We are using RSA and DSA encryption algorithm to make our system more secure from the third-party user or hacker. In RSA, we have to generate two keys called Public and Private keys. If the sender uses a public key for encryption then the receiver will use the private key for decryption [11]. On the other hand, if the sender uses a private key for encryption then the receiver will use the public key for decryption. On the other hand, DSA is used to generate the signature in private and can be verified in public.

In the initial step, the sender will send a plain text using a hash function and it will convert the text into hash code. The RSA algorithm is implemented with the hash code obtained in the encryption process. It is also called the message digest process.

Then the message digest will be signed digitally using the signer's private key and adds certificate with it and then the attached data will be sent to the receiver.

As soon as the digitally signed data reaches the receiver, it will be divided into two parts: data and signature. By using the hash function on the data, the hash code is obtained. On the other hand, the hash code we get from the decryption of signature using signer's public key will be the same. When the hash code is the same, it will generate the plain text and send it to the receiver. At the same time, the receiver will run a verification process. It will send a verification request to the sender. Instead of his request, the sender will send one-time password. Thus, the sender will be verified one more time.

5 Advantages

(a) The first and most important advantage of our system is that it will provide a safe and secure data transmission over an insecure wireless network.
(b) Due to the use of prime number factoring in RSA, it is difficult to split the RSA algorithm. So using this algorithm will keep our data secure from alteration.
(c) Our system will make sure that any unauthorized user or hacker cannot access the system.
(d) As we know, one-time password turns invalid in a moment so it will be almost impossible for hackers to obtain the sensitive data and use it.

6 Future Scope

At present, it must ensure data security in any sector. Hackers are constantly becoming more and more active. In line with them, we should also design and implement a more strong security system. Network and security give us a huge opportunity to work further with our system.

We will try to update our system or will implement a new security system, where we will use four individual hash functions and a device named RSA ID, which is used to create different codes and the code will be used in related hash method [14].

We can think about implementing a new security system not only for securing our banking system but also for large companies, ministry, etc. We can use UTII (unique thumb Impression Identity) with cryptography along with the one-time password to make a strong security system.

Table 1 Comparative study of existing system and proposed system [5]

S. No.	Conventional system	Proposed systems
1	In banking security system based on symmetric encryption, a single key is used and if any unauthorized person can obtain that key, he will have access to read and change the document	In our proposed system, we used asymmetric encryption which generates two different keys for encryption and decryption and it is difficult to obtain two keys at a time which makes our system more secure
2	Password-based security system is not that much strong and secure because the password can be predicted	In our proposed system, we used digital signature along with one-time password which makes the system more stronger
3	In a conventional banking security system, symmetric encryption is used with a hash function that does not make the system secure because of symmetric keys	We used hash function along with two different asymmetric encryption in the proposed system that provides much better security system to secure data

7 Results and Discussion

Our system is going to give far more better performance than other conventional systems (Table 1).

8 Conclusion

Our goal was to implement a system through which we could securely exchange information over an insecure network. Data sharing over the Internet or wireless network is so fast. So it has become popular and is increasing day by day. But sometimes there are incidents like data theft and leaked confidential information of office [15], company as well as the bank. So our paper presented a data security system for the bank that will secure data transmission and will keep confidential and sensitive information secure from hackers. Using cryptography along with a one-time password makes our system more secure and reliable than other conventional systems. It provides two-step authentication, the confidentiality of data as well as data integrity. So we are very successful in fulfilling our goals or objectives.

References

1. Tariq Banday M (2011) Easing PAIN with digital signatures. Int J Comput Appl (0975–8887) 29(2)
2. Stallings W (2005) Cryptography and network security, 4th edn. Prentice Hall, pp 58–309
3. Mahto D, Yadav DK (2015) Enhancing security of one-time password using elliptic curve cryptography with biometrics for e-commerce applications. In: 2015 Third International Conference on Computer, Communication, Control and Information Technology (C3IT). IEEE, pp 1–6

4. Salman T (Student Member, IEEE), Zolanvari M (Student Member, IEEE), Erbad A (Member, IEEE), Jain R (Fellow, IEEE), Samaka M (Member, IEEE) (2019) Security services using blockchains: a state of the art survey. IEEE Commun Surv Tutor 21(1)
5. Harmouch Y, Kouch R (2017) A fair comparison between several ciphers in characteristics, safety and speed test. In: Europe and MENA cooperation advances in information and communication technologies, vol. 520. ISBN: 978-3-319-46567-8
6. Dutta M, Ashiqul Islam Md, Mamun MH, Psyche KK, Al Mamun M Bank vault security system based on infrared radiation and GSM technology. In: International conference on intelligent data communication technologies and internet of things ICICI 2019, pp 120–127
7. Jagga S, Sharma P (2014) Banking authentication technique. Int J Inf Comput Technol 4(13):1305–1314. International Research Publications House. ISSN 0974-2239. http://www.irphouse.com
8. Jansma N, Arrendondo B (2004) Performance comparison of elliptic curve and rsa digital signatures. nicj.net/files
9. Akgül A, Çavuşoğlu Ü, Zengin A, Pehlivan I (2017) The design and implementation of hybrid RSA algorithm using a novel chaos based RNG. Chaos Solitons Fractals 104:655–667
10. Schoaba V, Sikansi FEG, Branco LC (2011) Digital signature for mobile devices: a new implementation and evaluation. Int J Futur Gener Commun Netw 4(2)
11. V. S. Miller, “Use of elliptic curves in cryptography,” in Advances in Cryptology CRYPTO 85 Proc., ser. Lecture Notes in Computer Science, H. Williams, Ed. Springer Berlin Heidelberg, 1986, vol. 218, pp. 417–426
12. Kar AK Cryptography in the banking industry. Frontiers 1(1). https://www.researchgate.net/publication/269405090. Business
13. Mahto D, Yadav D (2015) Enhancing security of one-time password using elliptic curve cryptography with finger-print biometric. In: 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), March 2015, pp 1737–1742
14. Dixit P, Gupta AK, Trivedi MC, Yadav VK (2018) Traditional and hybrid encryption techniques: a survey. Netw Commun Data Knowl Eng 4. ISBN: 978-981-10-4599-8
15. Saraswat V, Sahu RA, Sharma G, Kuchta V, Markowitch O (2019) Public-key encryption with integrated keyword search. J Hardw Syst Secur 3(1):12–25

Digital Signature Authentication Using Asymmetric Key Cryptography with Different Byte Number

Md. Sagar Hossen, Tasfia Tabassum, Md. Ashiqul Islam, Rafat Karim, Laila Sultana Rumi, and Aysha Akter Kobita

Abstract Nowadays, each and every system needs proper security. Only proper security can save important documents. There are different types of security systems that people use for safety. Digital Signature is one kind of a security system. Digital Signature is the public key primitive of different types of message authentication. Digital signature is one kind of technique that converts the handwritten signature in digital data. It is one kind of cryptographic value that is calculated from the data and a secret private key which is only known by the signer. In this paper, we want to make the digital signature more secure by using the ED25519 algorithm, which is an asymmetric algorithm. In this algorithm, the signature will be converted into different byte number, which will make the security system more strong.

Keywords Cryptography · Asymmetric algorithm · Digital signature · Primitive · Message authentication

Md. Sagar Hossen (✉) · T. Tabassum · Md. Ashiqul Islam · R. Karim · L. S. Rumi · A. A. Kobita
Daffodil International University, Savar, Dhaka, Bangladesh
e-mail: sagar15-1504@diu.edu.bd

T. Tabassum
e-mail: tabassum.tasfiaahmed@gmail.com

Md. Ashiqul Islam
e-mail: ashiqul15-951@diu.edu.bd

R. Karim
e-mail: rafat15-867@diu.edu.bd

L. S. Rumi
e-mail: lailasultanarumi@gmail.com

A. A. Kobita
e-mail: ayshakobita@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_78

1 Introduction

In the modern age, the wireless networking system is the network that can be easily used in the security system. Wireless networking has been enjoying fast development. Security is the most concerning part of the digital area. For each and every step of life, we need security. Security refers to the steps that are taken to protect a place or any type of information.

The Internet offers low cost for a security system, because of using wireless network people can easily secure their data. The wireless network has two types of the domain-specific network, which can easily play an important role in the security system.

Mainly wireless network is commonly used in the security system because of easy implementation of a security system. Wireless network contains a different type of node, where we can easily process the data and send the data to the authorized user [1].

Besides, the advantage of the wireless network the security system has some disadvantages, because it has some security issue, though every security system uses wireless network the hackers can easily find a different way to break the systems. Nowadays, the risks to users of wireless technology have increased as the security service has become popular for making these security systems strong we need to focus on a different type of symmetric and asymmetric algorithm.

In this paper, we work to make the bank security system more strong by using an asymmetric algorithm. We work with a digital signature using the asymmetric algorithm.

Digital signature is the most secure service that is used everywhere. This security system is a little critical offered by encryption. In traditional key management systems, the binding between the public key and the identification of the signer is obtained via a digital certificate [2].

Asymmetric algorithm is dealing with public and private keys to encrypt and decrypt. One key of the pair can be shared with everyone which is known as public key, another is the secret key that is known as a private key. One key is used in encrypting the message and another key from the one used to encrypt the message is used for decrypting. SSH, Open PGP, SSL, etc., are some protocols that rely on asymmetric cryptography for digital signature function. Rivest–Shamir–Adleman is the most used in this algorithm, which is embedded in the SSL protocols.

In RSA, we used the ED25519 algorithm. It is a part of RSA. The important thing is that it works fast and converts the data into a byte number [3]. It is a public-key signature system with a different type of activation part.

2 Related Work

Secure routing protocol [4–6] used the digital signature scheme based on network IDs in the wireless sensor network to improve the security and efficiency of the network. Hash value also analyzed and compared it with the bit rate. SHA and MD5 techniques [7] are utilized based on the deep analysis scheme with public-and private-key cryptographic techniques and it is applied in the wireless sensor network [8]. In WSN, security can be more strong by hashing on the plaintext and encrypted by well pairing multiple sender broadcast authentication hindrance [9, 10]. Some authors try to construct a network without infrastructure by using an ad hoc network and a single ad hoc network. They ensure their wireless security [11–13]. We are mainly focusing on banking security to produce safe transaction [14].

3 Proposed Work

In the modern age, the digital signature is widely used in business and financial industries as well as in the bank security system. The digital signature is nothing but cryptographic tools that are mainly used in the security system. But sometimes the digital signature fails to verify the signature (Fig. 1).

In this paper, we analyze some methods to solve the problem. We study the asymmetric algorithm, which is totally different from the symmetric algorithm. We want to propose some faster security system, for this reason, we use the ED25519 algorithm which is a part of the asymmetric algorithm. In ED25519, people can easily secure their information. Especially for the banking security system, it is more sweetly organize than other security system because the working process to be faster. Customer who deals with the bank is always in a hurry, so if the authentication system becomes faster, the working process becomes more faster. For this purpose, we try to make a faster security system using the ED25519 algorithm. We also use the certificate confirmation system and one-time password which will make our security system more effective.

4 Methodology

In asymmetric encryption, there is a different type of algorithm. Among them, the ED25519signing algorithm is the most recommended public-key algorithm in the security system. This algorithm is implemented using the Twisted Edwards Curve which offers a better security with faster performance compared to DSA OR ECDSA. Nowadays, RSA is the most widely used public algorithm [15]. If we compare it with ED25519 then it is slower.

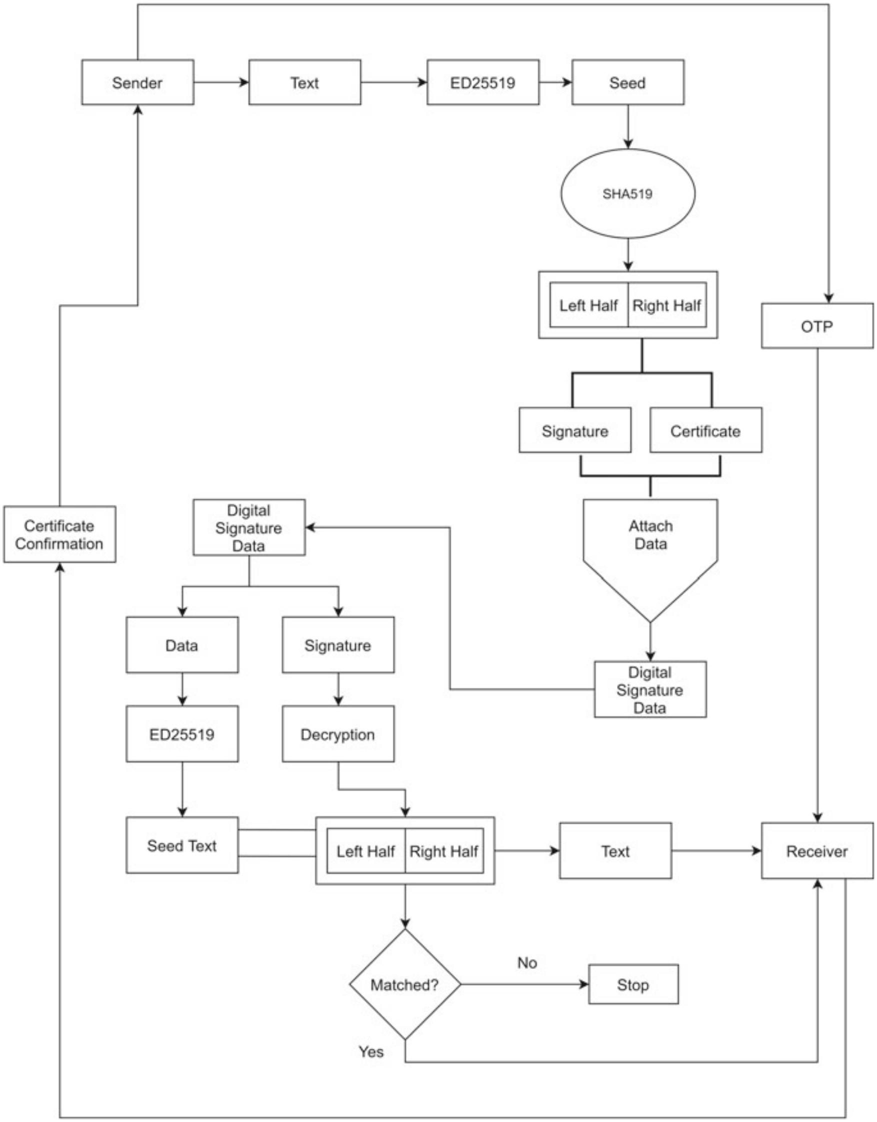


Fig. 1 Proposed system [11]

Digital signature is the best algorithm because of the fast signing process. There is also uniformly generated random seed that contains byte. In this algorithm, hashed SHA512 is also used. In the first step, the sender sends a plain text then it will encrypt by the ED25519algorithm. Then the plain text turns into a hash then it turns into a random seed which contains 32 bytes, the seed is then hashed using ShA256, which gets 64 bytes, which is then split into left part (contain 32 bytes, private key), right part (32 bytes). Then the plain text turns into non-determination signature (64 bytes). First of all, digitally signed data will be divided into two parts; in the first part there is data where we used ED25519 algorithm. The data will be turned into hash on another part we will find a signature, where we will decrypt it. After decryption the signature will be converted into two parts with a different byte number, if the seed byte number matches with signature byte number then the signature will be turned into plain text, using ED25519 then the signature will be verified with a signature certificate and sends a request to the sender [16]. Sender sends a verification code to the receiver phone number by using a one-time password. And by using ED25519 also the receiver obtains the plain text.

5 Algorithm for the System

The below system design is described by the following algorithm:

Step 1: Initiate the process.

Step 2: Send plain text and encrypt it using the ED25519 algorithm then generate a random seed using a hash function.

Step 3: Convert the random seed into 32 bytes using the SHA512 algorithm, then it can generate a signature and certificate.

Step 4: Attach the signature and certificate to generate digitally signed data.

Step 5: Then digitally signed data is divided into data and signature. Data is produced random seed using ED25519 algorithm and signature are decrypted the random seed in 32 bit.

Step 6: If the data random seed and signature random seed are matched, then the message is sent to the receiver otherwise the procedure is stopped.

Step 7: The receiver will be sending a confirmation/verify request to the sender to authenticate the verification process.

Step 8: Sender will check the verification request and sends the OTP to the receiver mobile phone and authenticate the system.

Step 9: Using the OTP, the receiver will get the plain text.

Step 10: End Procedure.

6 Advantages

i. For making a strong security system, we have selected the ED25519, because of some specific advantage of it. In the EdDSA signing algorithm, ED25519 has some particular instantiation which made it more flexible and different from other algorithms [17].

ii. The best thing is that it has slightly smaller keys with 32 bytes, and signature size with 64 bytes. Because of its small size, it can work 30% faster than other algorithms. It also has the ability of fast batch verification. So by using this, we can verify the signature easily, which saves our time, no secret branch condition is needed in ED25519 and eventually, no secret memory access is needed [17, 18]. So the whole signing process is more secure.

7 Future Scope

In the future, we intend to add some more algorithm in the system by which can easily notify if any hacker wants to break the system and also use some carve by which the security system can be easily understood by the sender and user, to control user access according to his access. So the security system can work faster and it is difficult to hack the system for hacker.

8 Comparative Study

See Table 1.

9 Conclusion

The main purpose of this paper is to introduce a security system based on wireless networking using the ED25519 algorithm in the digital signature [10]. The full authentication process will be done in an asymmetric way. The main contribution of this research work is to authenticate a security framework, which provides better security. In most of the security systems, a symmetric algorithm is a large security system and is slow. But in this paper, we used a symmetric algorithm which is really a faster security system. In this paper, the unauthorized problem will be easily solved. Even a strong security system is performed in our approach so it is a little bit difficult for the hacker to hack the security system [13]. We conclude not only that it is important to include a strong integrity check into an RSA encryption, but also that this integrity check must be performed in the correct step with proper notification

Table 1 Comparative study between existing system and proposed system [3]

Proposed system	Existing system
1. It can perform 8%faster 2. Private key length is 32 bytes (256 bits = 251 variable bits + 5 predefined) 3. No public key recovery is allowed in this algorithm 4. In safe curve security, 11 of 11 tests passed	1. It can perform 8% slower 2. Private key length is 32 bytes (256 bits) 3. Public key recovery is possible 4. In safe curve security, 7 of 11 tests passed

and confirmation code. We have also believed that we find a faster strong security system.

References

1. Ibriq J, Mahgoub I (2004) Cluster-based routing in wireless sensor networks: issues and challenges. In: Proceedings of SPECTS'04
2. Hess F (2003) Efficient identity based signature schemes based on pairings. Lecture notes in computer science, vol 2595. Springer, Berlin
3. Heinzelman WR, Chandrakasan A, Balakrishnan H (2000) Energy efficient communication protocol for wireless microsensor networks. In: Proceedings of 33rd Hawaii international conference on system sciences (HICSS'00)
4. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) Wireless sensor networks: a survey. Comput Netw 38:393–422
5. Romer K, Mattern F (2004) The design space of wireless sensor networks. Wirel Commun 11(6). IEEE
6. Abbasi AA, Younis M (2007) A survey on clustering algorithms for wireless sensor networks. Comput Commun 30
7. Gupta P, Kumar S (2014) A comparative analysis of SHA and MD5 algorithm. In: (IJCSIT) Int J Comput Sci Inf Technol 5(3)
8. Yasmin R, Ritter E, Wang G (2010) An authentication framework for Wireless Sensor Networks using identity-based signatures. In: 2010 10th IEEE international conference on computer and information technology (CIT 2010). IEEE, Piscataway, New Jersey, USA, pp 882–889
9. Boneh D, Franklin M (2001) Identify-based encryption from the weil pairing. In: Proceedings of CRYPTO'01. LNCS, vol 2139
10. Barreto P, Kim H, Bynn B, Scott M (2002) Efficient algorithms for pairing-based cryptosystems. In: Proceedings of CRYPTO'02
11. Karl H, Willig A (2016) Protocols and architectures for wireless sensor networks. Wiley 56(1):1–44
12. Taluk S. (2014). "A Taxonomy of Wireless Micro-Sensor Network Models," SIGMOBILE Mob. Comput. Commun, Rev., Vol. 6, No. 2. pp. 28–36
13. Yu S, Zhang B, Li C, Mouftah H (2014) "Routing protocols for wireless sensor networks with mobile sinks": a survey. IEEE Commun Mag 52(7):150–157
14. Dutta M, Ashiqul Islam Md, Mamun MH, Psyche KK, Al Mamun M Bank vault security system based on infrared radiation and GSM technology. In: International conference on intelligent data communication technologies and internet of things ICICI 2019, pp 120–127
15. Perrig A, Szewczyk R, Wen V et al (2003) SPINS: security protocols for sensor networks. In: Proceedings of Mobicom'03
16. Carman DW (2005) New directions in sensor network key management. Int J Distrib Sens Netw 1(1)
17. Oliveira LB, Ferreira A, Vilaca MA et al (2007) SecLEACH-on the security of clustered sensor networks. Signal Process Inf Process Data Manag WSNs 87(12)
18. Banerjee P, Jacobson D, Lahiri SN (2007) Security and performance analysis of a secure clustering protocol for sensor networks. In: Proceedings of 6th IEEE international symposium on network computing and applications

Digital Signature Authentication for a Bank Using Asymmetric Key Cryptography Algorithm and Token Based Encryption

Rafat Karim, Laila Sultana Rumi, Md. Ashiqul Islam, Aysha Akter Kobita, Tasfia Tabassum, and Md. Sagar Hossen

Abstract Nowadays security system is being with more important issues. In modern science, technology is updated day by day and we are getting insecure in our daily life. Through this project, a digital signature authentication security system has been designed to protect the bank from unauthorized access. In this paper, we are producing the most efficient and productive security system based on digital signature authentication using the asymmetric key cryptography algorithm and token-based encryption. In this project, we are using public and private keys for encryption and decryption data with the hash function and also provide digitally signed data, RSA algorithm to encrypt the data. The receiver will send a certificate and confirmation request to the sender to verify the certificate and the sender will send an OTP via phone through the Internet to authenticate the receiver. All of these works are producing a good and valuable security system in the banking sector.

Keywords Cryptography · Asymmetric keys · Hash functions · RSA · OTP · Data security

R. Karim (✉) · L. S. Rumi · Md. Ashiqul Islam · A. A. Kobita · T. Tabassum · Md. Sagar Hossen
Daffodil International University, Savar, Dhaka, Bangladesh
e-mail: rafat15-867@diu.edu.bd

L. S. Rumi
e-mail: lailasultanarumi@gmail.com

Md. Ashiqul Islam
e-mail: ashiqul15-951@diu.edu.bd

A. A. Kobita
e-mail: ayshakobita@gmail.com

T. Tabassum
e-mail: tabassum.tasfiaahmed@gmail.com

Md. Sagar Hossen
e-mail: sagar15-1504@diu.edu.bd

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_79

1 Introduction

Nowadays security is most important and is involved in every field in our day-to-day life. Security is important in fields like banking security, office information security, online security, social media security, etc. The computer is a major material and also like a human part of our daily work. We all are doing our work on a computer with the Internet while communicating with others or transfer any file, any important work file. When we transfer any file sometimes our information may be hacked. These days most of the systems are based on a static password, this type of password is cushy and easy to guess for the attack. Sometimes we used the same password in every account like (email, bank account). Yet hackers has been apply many techniques such as poke nose to detect and steal the password. So, we need secure communication and secure transformation. Data security is the most significant for data communication and data transformation on the Internet over the connected path. In our banking security system, we also need secure data communication and transformation for banking authentication and also for client's data for the banking system.

So, in this paper, we want to introduce and describe a new authentication method for banking security system based on cryptography. Cryptography is the most effective for security purpose, but cryptography is continuously based on research.

Cryptography

It is a method of assuring or protecting information and communications through the use of plain text or code with Encryption and Decryption, which cannot be reached any unauthorized user without sender and receiver or only those for whom the information is intended can be read and processed. It is built-up protection for data security and secure data from third-party access. Cryptography gives some code or any condition before giving any access to any user in any system. After fulfilling the condition or code, sender and receiver may access any system.

Authentication

The process is given four identities. Asymmetric algorithm (RSA), hash function, digital signature, and token-based encryption. Using private key and public key for encryption and decryption data with hash algorithm also provides digitally signed data through the Internet.

Privacy/confidentiality

Ensuring that no one can read the message, data, or any information except the authentic receiver.

Integrity

Assuring that the recipient receives unaltered and unedited data any way from the original of the sender.

Non-repudiation

These mechanisms prove that the sender really sends this message and receiver receives the unedited original message or data. We are using three types of cryptographic algorithms:

DSA (Digital Signature Algorithm)

DSA is an asymmetric key and it is creating a digital signature. It is a signature that is a mathematical technique used to validate the accuracy and rectitude of a message, software, or digital document. It provides more strong data security. It is used to prove sender and receiver authentication. The sender uses a hash function to generate message digest using private and public keys with encryption and decryption signs it digitally with the user's signature and send it a receiver. The receiver has to use the sender's public key to decrypt the signature.

Phone or Email Authentication

After digital signed an OTP code is provide on sender phone number or email. Then the sender provides that code verification on the receiver phone number or email.

2 Related Work

Internet banking has come to the modern age as an arms race between financial institutions and public network attackers. Banks can take clear advantage of the solutions of some authors. Alain presented and offered high security against common attacks [1]. If every system has two or more factoring systems in their account then attack will be reduced and in [2–4] it is shown cryptography and encryption are more secure in data encryption. RSA method is implemented in FPGA and the result is much better [5]. Digital signature identities encrypted information with a private key [6–9] that's why information will be so much secured. Decryption and verification in 8 steps are shown in [7]. Arpan says [10] that implementation of data security on business management can be more efficient. Even, OTP can be more secure in our system [11, 12] and authentication will be fluent and fast. Pranav Kumar describes in their paper about the mechanism of secure cryptography on Wi-Fi connectivity [13].

3 Proposed Work

In this paper, we are introducing the proposed work with four different ways of authentication approaches to secure the banking system. For example, gold or money in bank lockers. In our proposed work, the four authentication techniques are the following:

- Asymmetric encryption with private key digest with hash algorithm,
- Digital signature verification,
- Token-based validation or phone, email verification,
- Asymmetric decryption with public key digest with hash algorithm.

4 Methodology

4.1 Algorithm for the System

Step 1 Start the procedure.
Step 2 Send data by encrypting it using a hash function and private key.
Step 3 To generate a signed data using hash code and RSA algorithm.
Step 4 Send the signed data via the Internet to the receiver.
Step 5 The receiver decrypts the signed data in hash code by using a public key.
Step 6 Receiver sends a confirmation/verification request to the sender.
Step 7 Sender checks the verification request and send the OTP to the receiver's mobile phone and authenticate the system.
Step 8 Using the OTP, the receiver will get the plain text.
Step 9 End Procedure.

4.2 System Operation

In our day-to-day life, security is the main matter in every field. So for this security purpose, many people used bank lockers to secure their property like gold, money. So in our work, we want to make a banking security system that is based on cryptography. In this cryptography-based security system, we applied asymmetric algorithm (RSA), hash function, digital signature, and token-based encryption. Using private key and public key for encryption and decryption data with a hash algorithm, also provides digital singed data through the internet.

From Fig. 1, it can be seen that the sender sends data or massage. Then the owner's private key is encrypted with hash code. The RSA algorithm is implemented with a hash code obtained in the encryption process. It is also called the digest process. Encrypted data or digest is then assigned to digitally signed data. After assigning a digital sign, this data is passed through the Internet for decryption. Hash code is decrypted with public key and provides plain text. Sender sends a code through the network on phone. On the other hand, a digital signed is verify the authentication process and send certificate acknowledgement to the sender and receiver. Receiver send a request to the sender to provide a verification code. Sender send the verification code through network and receiver receives the code. After all verification and send code processes are finished, the receiver will receive a plain text. All of these work together in a security system for providing more data security.

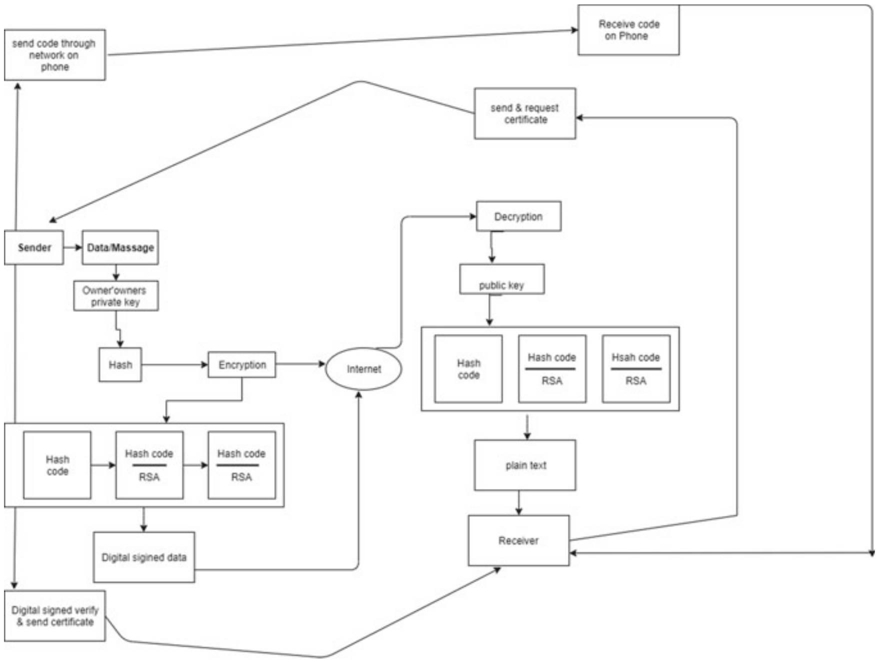


Fig. 1 Proposed system

5 Advantages

(a) In our proposed work, we used asymmetric (private) key with encryption, hash algorithm, digital signature algorithm, decryption data with public key which makes the connection more secure.

(b) We do not only use the phone verification but also asymmetric and trusted encryption and decryption data with digital signature verification certificate security based on the banking technique.

(c) In our proposed work, we use a digital signature instead of a password with combination asymmetric (private and public) key encryption and decryption data to digest with hash algorithm to verify the digital signature using token-based encryption which makes it more secure.

(d) This process ensures that the recipient receives altered and unedited data, which is sent by the sender, and this data is original.

(e) Any other user cannot access any information without permission of the sender and receiver. Strong security are approach to protect the information.

6 Conclusion

Security is an extensive point or issue in any system like online/offline payment system, office information transfer system, office employee system, banking security system as there are various threats, hacker can access it easily if we get mistake which affect the security system increase risk. This proposed authentication technique for the banking security system is very secure to protect user, and banks from any attacker or hacker gain. The access to confidential information of user like encryption data, digital signature, phone number, email, etc. can not be find by any hacker in easy way. This system implements an effective authentication to reduce fraud, hacker, and picklock and make strong customer authentication necessary to enforce security to assist any type of institution that needs a security system [14]. Using cryptography along with digital signature and phone verification code makes our system more secure and predictable than other prevalent systems. It accommodates very strong authentication, confidentiality of data as well as data integrity. So we are very hopeful about our goals in this system.

7 Future Work

This paper discussed an extended new banking authentication procedure for banking securely. In this, system two keys are used one by the receiver and one by the sender. If you need to transfer any data 1st need to know encryption data, digest with hash algorithm and verified it with digital signature. An OTP code is send to the sender phone to verification it then again decryption data digest it with hash algorithm, last stage receiver got the phone verification code from sender then receiver get the plain text [15].

At present, it is a must to ensure data security in any sector. Hackers are more interrogate. In this time, any hacker cannot hack the system. On the other hand, if a hacker found the data, the hacker cannot understand because this data was encrypted, decrypted, and verified, so the hacker didn't hack the system on time. In this system, we should also design and implement a more strong security system. Network and security give us an extensive opportunity to work further with our system. If we work on this system, the bank is provided with most powerful security system and also achieves the client's faith [14]. This is a unique Banking Authentication system that provides more security than other older security systems. In this modern creation, we need to be more updated. We will try to update our system or will implement a new security system, where we will use individual hash function and RSA (Rivest–Shamir–Adleman) DSA (Digital Signature Algorithm) method, which is used to create different codes and the code will be used in the related hash method. So this system has given an updated security system to save our bank locker.

References

1. Hiltgen A, Kramp T, Weigold T (2006) Secure internet banking authentication. Article in IEEE Security and Privacy Magazine, Apr 2006
2. Ezeofor CJ, Ulasi AG (2014) Analysis of network data encryption & decryption techniques in communication systems. Int J Innov Res Sci Eng Technol (An ISO 3297: 2007 Certified Organization) 3(12), Dec 2014
3. Stallings W (2005) Cryptography and network security. 4th edn. Prentice Hall, pp 58–309
4. Brodney A, Asher J (2009) Tales of the encrypted. http://library.thinkquest.org/28005/flashed/index2.shtml
5. Iana GV, Anghelescu P, Serban G (2011) RSA encryption algorithm implemented on FPGA. In: International conference on applied electronics, pp 1–4
6. Rivest RL, Shamir A, Adleman L (1978) Methods for obtaining digital signatures and public key cryptosystems. Commun ACM 21:120–126
7. Tariq Banday M (2011) Easing PAIN with digital signatures. Int J Comput Appl (0975 – 8887) 29(2), Sep 2011
8. Kazmirchuk S, Anna I, Sergii I. Digital signature authentication scheme with message recovery based on the use of elliptic curves. In: International conference on computer science, engineering and education applications ICCSEEA 2019: Advances in computer science for engineering and education II, pp 279–288
9. Takahashi K, Matsuda T, Murakami T, Hanaoka G, Nishigaki M (2019) Signature schemes with a fuzzy private key. Int J Inf Secur 18(5):581–617, Oct 2019
10. Kar AK. Cryptography in the banking industry. © Bus Front 1(1). https://www.researchgate.net/publication/269405090
11. Mahto D, Yadav DK (2015) Enhancing security of one-time password using elliptic curve cryptography with biometrics for e-commerce applications. In: Computer, communication, control and information technology (C3IT), 2015 third international conference on. IEEE, pp 1–6
12. Mahto D, Yadav D (2015) Enhancing security of one-time password using elliptic curve cryptography with finger-print biometric. In: Computing for sustainable global development (INDIACom), 2015 2nd international conference on, March 2015, pp 1737–1742
13. Singh PK, Vij P, Vyas A, Nandi SK, Nandi S. Elliptic curve cryptography based mechanism for secure wi-fi connectivity. In: International conference on distributed computing and internet technology ICDCIT 2019, pp 422–439
14. Dutta M, Ashiqul Islam M, Hasan Mamun M, Kaem Psyche K, Al Mamun M. Bank vault security system based on infrared radiation and GSM Technology. In: International conference on intelligent data communication technologies and internet of things ICICI 2019, pp 120–127
15. Prakasha K, Gowda P, Acharya V, Muniyal B, Khandelwal M. Enhanced authentication and key agreement mechanism using PKI. In: International conference on applications and techniques in information security ATIS 2018, p 40

Accuracy Analysis of Similarity Measures in Surprise Framework

Sanket Kamta and Vijay Verma

Abstract Recommender Systems (RS) are growing technologies, which can be very useful for consumers in finding items of their interest on the web. Collaborative filtering(CF), a popular approach of Recommender Systems, recommends items to users based on other users with the same taste. The key step of the collaborative filtering method is to compute the similarity between users and items. The success of any recommender model hugely depends on how accurately the notion of similarity has been modelled. There are many open-source Python frameworks, which are useful in building and experimenting with RS models in the industry as well as academics such as Surprise, Python-recsys, Case Recommender, Polara, Spotlight, and RecQ. This work provides a brief introduction to the Surprise, a Python library for Recommender System, explaining its architecture, implementation and main features. Furthermore, a comparative study of the Surprise framework with other related frameworks is provided to demonstrate the fact why it is better than other frameworks in terms of implementing and handling new and complex recommender models. We have evaluated the accuracy of various built-in similarity measures provided by the Surprise framework using the real-world benchmark MovieLens datasets (100 k and 1 M). Thereafter, the accuracy of the recommendation is measured in terms of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The empirical results demonstrate that Pearson-baseline outperforms other built-in similarity measures available in the Surprise framework.

Keywords Recommendation · Collaborative filtering · Framework · Similarity measures

S. Kamta (✉) · V. Verma
Department of Computer Engineering, National Institute of Technology, Kurukshetra, Haryana, India
e-mail: iamsanketkamta@gmail.com

V. Verma
e-mail: vermavijay1986@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_80

1 Introduction

A Recommender System is a software which is used for suggesting useful items to active users [1]. These recommendations can be in many forms, such as which games to play, which movies to watch or which places to visit. The growth of e-commerce websites has offered users with a lot of choices. RS helps those users to decide which item to select from the ocean of items available on the web. The prediction process done by the RS is dependent on the user's desire and constraints. RS collects the data about the users by analyzing their previous transactions and feedback for certain items.

RS has come up as an important area of research since the mid-1990s [2]. There has been a lot of advancement in the field of recommender system in today's world. RS is used in these famous internet sites like Amazon [3], YouTube [4], Netflix [5], IMDB [6] helping the users find a suitable item that best matches their needs. A lot of dedicated symposium and workshops are held in this field, ACM Recommender System (RecSys) [7] being one of them.

There are various RS techniques depending on the address domain, the knowledge used and the prediction algorithm [8]. Content-based filtering is used to recommend items to an active user based on the user's preferences made in the past. Collaborative filtering is used to recommend items to an active user based on the other users with whom it shares the same taste. CF, the most popular RS technique, is often said as "people-to-people correlation". Demographic filtering is used to recommend items to an active user based on a certain common personal attribute such as sex, age, country, etc. Community-based filtering is used to recommend items to an active user based on the choice of its social friend. Hybrid RS combines the above techniques to make recommendations such that the advantage of any technique is used to fix the disadvantage of other technique. A general classification of RS is shown in Fig. 1.

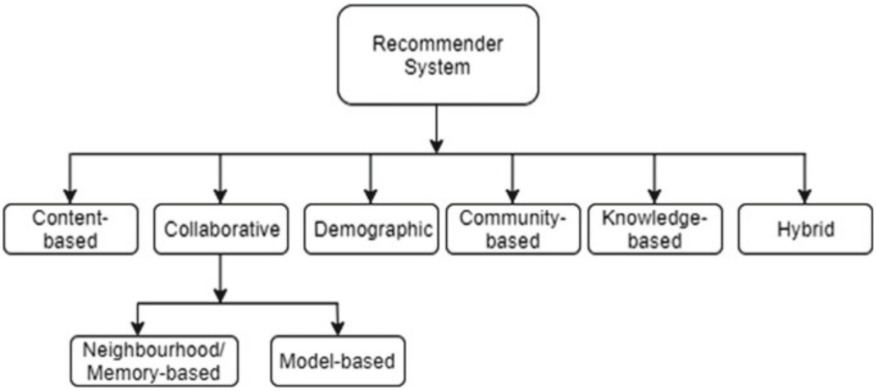


Fig. 1 General classification of recommender system

Out of the above-mentioned RS techniques, CF is still the most popular one as it is flexible across different domains and can produce more serendipitous recommendations [9, 10]. Also, there are a lot of researches going on to enhance the accuracy of the prediction done by the RS based on collaborative filtering. In CF, finding the similarity between users and items plays an important role and improving it may lead to a more accurate prediction. In recent times, prediction algorithms use a lot of invariants such as normalizing the ratings, removing bias, using fancier aggregation, etc., to further improve the accuracy of predictions. As a result of which, RS developers are facing the problem of exponentially larger design space, which adds to the complexity of the standard prediction algorithm.

In our work, we have presented the design, implementation, features and evaluation of Surprise [11], a Python library for Recommender System which has a number of integrated similarity measures, datasets and prediction algorithms to tackle the problem of larger design space. We have also presented a comparative study of why Surprise is better than other CF frameworks based on Python. Later, we have conducted some experiments to analyse the accuracy of various built-in similarity measures provided by Surprise.

2 Related Frameworks

The rapid growth of RS has raised significant challenges to the existing frameworks. There are various existing frameworks on the Python environment, which are useful in making recommendations based on collaborative filtering. We will brief some of the most common ones—Python-recsys [12], Case Recommender [13], Polara [14], Spotlight [15] and RecQ [16].

Python-recsys is used for implementing recommendation models based on the two famous prediction algorithms, SVD [17] and Neighbourhood SVD. For recommendation, we first need to calculate similarity which is done by the Cosine similarity index. Besides prediction, we can even evaluate the accuracy of the model by the built-in evaluation metrics, MAE and RMSE. For simplicity, two famous datasets, MovieLens and Last.fm are also provided in this library. The framework can be downloaded from [18]. For more information, we refer the interested reader to the online documentation and examples available at [19].

Case Recommender is an open-source Python framework published in Pypi (a repository of software for Python), under the MIT licence. It helps in constructing high-performance and customized recommendation models using the different varieties of integrated prediction and evaluation approaches. It is structured in a form that can support content-based, collaborative filtering and hybrid approaches of RS. It aims at being one of the useful tools in the field of research and education, rather than large-scale commercial operations. For more information about this framework, the interested reader may visit [13].

Polara is a fast and flexible framework, which supports both Python 2 and Python 3 environment. This framework basically aims at recommending *top-n items* to users, taking the concept of feedback polarity into consideration. Feedback polarity is an idea where the negative feedbacks by the users are also considered as sensitive data. These negative feedbacks, in turn, are used to recommend items to users who have not given any positive feedbacks (only negative feedbacks), that is how the cold-start problem can be handled. The idea of feedback polarity was introduced in the paper. It analyses the RS model deeply to curb any irrelevant recommendations. It also gives us the freedom to create our own RS model by importing the RecommenderModel class from the library. It makes use of various other Python packages like *Pandas, Numpy, Scipy* and *Numba*, to achieve better performance. The installation guide, along with various examples, is provided under the page [20].

Spotlight can be used to build shallow RS models (one with fewer features) as well as deep RS models (one with a large number of features). Basically, it can be considered as a tool for exploring and prototyping new recommendation techniques. On the set of built-in datasets provided, we can perform implicit and explicit factorization to build a sequence RS model where the model recommends items based on the sequence of previous items a user interacted in the past. An example of a sequence RS model is YouTube recommendation. The authors of this framework have provided details about it in the documentation [15].

RecQ is a Python 2.7.× framework for recommender systems in which a number of the state-of-the-art recommendation models are already implemented. It is platform-independent, which means it can run on any platform (Windows, Linux, MacOS). We can also visualise the input datasets with this framework directly. It can perform faster computation because of the support of *Numpy*, *Pandas* and *Matplotlib* library. The flexibility in its design helps researchers implement new recommendation algorithms quite easily. The documentation of the above framework can be found in [21].

Surprise has turned out to be far more superior than other frameworks in terms of dataset handling, the number of integrated prediction algorithms, evaluation and similarity measures and reliability in backend support. We have shown the core features of previous frameworks and their comparisons with Surprise in Table 1.

Table 1 Comparing Surprise to existing software frameworks for recommender systems

Frameworks	Backend	Integrated datasets	Integrated similarity measures	Integrated evaluation metrics	Integrated prediction algorithms
Python-recsys [12]	Divisi2	MovieLens, Last.fm	Cosine	MAE,RMSE	SVD [17], SVD neighbourhood
Case Recommender [13]	×	×	×	MAE, RMSE, Precision, Recall, NDCG	Matrix factorization [22], SVD, SVD ++ [23], CoRec [24], UserKNN [25], ItemKNN [26]

(continued)

Table 1 (continued)

Frameworks	Backend	Integrated datasets	Integrated similarity measures	Integrated evaluation metrics	Integrated prediction algorithms
Polara [20]	TensorFlow	MovieLens	×	Precision, Recall, NDCG	SVD
Spotlight [15]	PyTorch	MovieLens 100 k, 1 M, 10 M, 20 M Goodbooks 10 k	×	RMSE, MRR, Precision, Recall	×
RecQ [16]	TensorFlow	Ciao, Epinions, Douban, LastFM	PCC, Cosine	RMSE, MRR, Precision, Recall, MAP	SlopeOne [27], SoRec [28], SVD, SVD++, PMF [29]
Surprise	SciKits	MovieLens 100 k, 1 M Jester	PCC, Cosine, MSD, Pearson-baseline	MSE, RMSE, MAE, FCP	KNNBasic [30], KNNWithMeans, SVD, SVD++, SlopeOne, CoClustering

3 Surprise Framework

The Surprise is a Python library for Recommender System or preferably, a Python library for rating prediction algorithm. SurPRISE (Simple Python Recommendation System Engine) was developed at the University of Toulouse III by Nicolas Hug. The Surprise framework is freely available under the BSD 3 Clause licence at the URL: http://surpriselib.com/.

The Surprise library was introduced in 2017 for quick and easy prototyping and giving better control over the experiments. In Listing 1, we have shown a simple implementation of how we can download dataset and perform prediction for any user in Surprise.

```
from surprise import KNNWithMeans
from surprise import Dataset
d = Dataset.load_builtin('ml-1M') #downloading dataset
tset = d.build_full_trainset() # building trainset
algo = KNNWithMeans() # using integrated prediction
algorithm
algo.train(tset) # fitting data
algo.predict('Bob', 'Big Bang Theory')
```

Listing 1. Implementation of simple prediction in Surprise

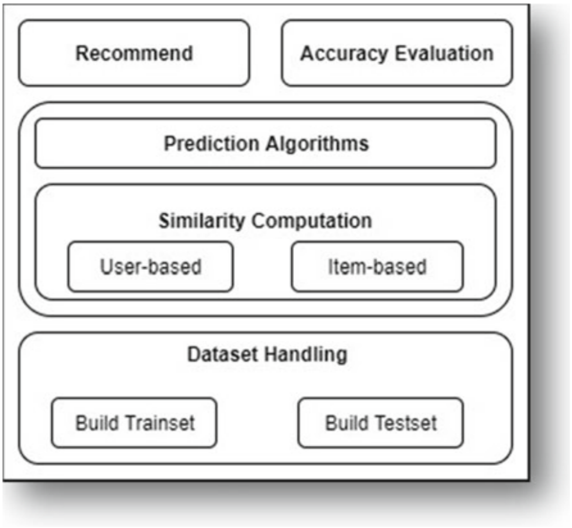


Fig. 2 Architecture of Surprise library

3.1 Architecture

In this section, we have described the architecture of Surprise framework. We have presented Surprise in a hierarchical form dividing each feature in the form of modules. For making predictions, the dataset is first loaded and separated into two classes—train set and test set. On top of that we perform some similarity computation and prediction algorithms to either perform recommendation or evaluate the accuracy of prediction. The modular architecture of Surprise is shown in Fig. 2.

3.2 Salient Features

Surprise has a wide variety of features, which differentiates and overpowers it over other frameworks. While we offer a vignette of some of the main features of interest in Surprise, this paper is by no means comprehensive. For more information, the interested reader may refer to the online documentation at the URL: https://surprise.readthedocs.io/en/stable/index.html.

- *Easy dataset handling*

Surprise provides some built-in datasets like MovieLens 100 k, MovieLens 1 M and Jester. We can use the Dataset.load_buildin() method to load the built-in datasets. We can also load custom datasets from files by first defining the path of the file and then passing it in the Dataset.load_from_file(file_path, Reader) method. This library also gives us the freedom to load datasets from the Panda's data frame. We need to define the rating scale parameter and then pass it into the Dataset.load_from_df() method.

- *Built-in similarity measures*

Surprise provides a bunch of built-in similarity measures like cosine, Pearson, MSD, person-baseline. An implementation of the cosine similarity is shown in the below code snippet.

```
sim_option = {'name': 'pearson',
              'user_based': False,
              'min_support': 10}
algo = KNNBasic(sim_option = sim_option)
```

The user–user similarity is defined by 'user_based': True whereas item–item similarity is defined by 'user_based': False.

- *Built-in prediction algorithm*

It provides a number of prediction algorithms such as SVD, KNN, etc. An illustration of KNNWithMeans prediction function is shown in the below code snippet. The algorithm is first imported and then provided with some parameters like the K-nearest neighbour and similarity measure. Finally, the algorithm is given the test set or train set to predict the result.

```
algo = KNNWithMeans(k=10, sim_options=sim_options)
test_pred = algo.test(testset)
```

- *Custom algorithms are easy to implement*

We can even create our own prediction algorithm. Any algorithm is nothing but a class derived from AlgoBase and having an estimate method. The predict method calls the estimate method. In our case, it will always predict the rating as 3.

```
class Predictor(AlgoBase):
       def estimate(self, user, item):
              return 3
algo = Predictor()
algo.train(trainset)
pred = algo.predict('Bob', 'Big Bang Theory') # will
call estimate
```

- *Built-in evaluation measures*

Surprise provides a bunch of evaluation measures such as MAE, RMSE, FCP, etc. We can evaluate the algorithm's performance by first importing the accuracy module and passing algorithm to it.

```
accuracy.rmse(test_pred, verbose=True)
accuracy.mae(test_pred, verbose=True)
```

4 Experiments

Experiments have been conducted to determine the accuracy of various ready-to-use similarity measures in the Surprise library using the MovieLens datasets. For recommendation purposes, we have utilized the traditional K-nearest neighbour algorithm (implemented as KNNWithMeans in the framework). The dataset has been randomly divided into train set and test set for evaluating the accuracy in terms of MAE and RMSE.

4.1 Datasets

We used two stable benchmark datasets from MovieLens. MovieLens is a web-based RS that is used for recommending movies to users, based on their previous ratings and reviews for movies using CF. It was introduced in 1997 by GroupLens Research, a research lab in the Department of Computer Science and Engineering at the University of Minnesota. The brief description of the MovieLens datasets along with the sparsity in each is given in Table 2. The term *Sparsity* implies the portion

Table 2 Various MovieLens datasets

Name	Year	Details	Sparsity
MovieLens 100 K	1998	Number of ratings:100,000 Number of movies:1700 Number of Users:1000	0.937
MovieLens 1 M	2000	Number of ratings:1 million Number of movies:4000 Number of Users:6000	0.957

of the user–item matrix that has not been rated by the user. Both the datasets shown in the table is very sparse.

4.2 Evaluation Metrics

There are various evaluation metrics to determine the accuracy of the models. In our experiment, we evaluated the accuracy of various similarity measures by using the integrated evaluation metrics—the MAE and RMSE. The MAE value is calculated by first summing the absolute errors of the *N* corresponding ratings–prediction pairs and then averaging the sum. Formally,

$$\text{MAE} = \frac{\sum_{i=1}^{N} \left| r_i - \hat{r}_i \right|}{N}$$

RMSE value is the square root of the average of squared differences between prediction and actual observation. Formally,

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (r_i - \hat{r}_i)}{N}}$$

A smaller value of evaluation metrics, in our case MAE and RMSE, indicates better accuracy.

The similarity measure is the measure of how much alike two elements are. In our case, the similarity measure is defined as the closeness of two users in the given dataset. There are a lot of ways to determine the similarity, but we have selected four of the integrated similarity measures provided in the framework. Table 3 shows the various similarity measures we used in our experiment.

4.3 Results and Discussion

In our experiment, we have utilized the user-based CF recommendation with KNN, and the similarity measure modules have been chosen from the Cosine Vector, Pearson Correlation, Mean Squared Difference, and Pearson-baseline one after the other. The values of MAE and RMSE are evaluated by varying the neighbourhood size *k* of an active user from 10 to 50 (at an interval of 10).

Figure 3 demonstrates MAE value against different neighbourhood sizes for (a) MovieLens 100 k (b) MovieLens 1 M. Similarly, Fig. 4 depicts RMSE value against different neighbourhood sizes for (a) MovieLens 100 k (b)MovieLens 1 M. It can be observed that as we increase the value of *k* (number of neighbours), the accuracy of recommendation increases. It can even be visualized that Pearson-baseline results

Table 3 A list of similarity measures

Author's	Methods	Formula
Breese et al. [31]	Cosine Vector (CV)	$\mathrm{CV}(u, v) = \frac{\sum_{i \in I_{uv}} r_{ui} r_{vi}}{\sqrt{\sum_{i \in I_u} r_{ui}^2} \sqrt{\sum_{j \in I_v} r_{vj}^2}}$
Resnick et al. [32]	Pearson Correlation (PC)	$\mathrm{PC}(u, v) = \frac{\sum_{i \in I_{uv}} (r_{ui} - \overline{r_u})(r_{vi} - \overline{r_v})}{\sqrt{\sum_{i \in I_{uv}} (r_{ui} - \overline{r_u})^2} \sqrt{\sum_{i \in I_{uv}} (r_{vi} - \overline{r_v})^2}}$
Shardanand and Maes [33]	Mean Squared Difference (MSD)	$\mathrm{MSD}(u, v) = \frac{\lvert I_{uv} \rvert}{\sum_{i \in I_{uv}} (r_{ui} - r_{vi})^2}$
	Pearson_baseline	$\text{Pearson\_ baseline}(u, v) = \frac{\sum_{i \in I_{uv}} (r_{ui} - b_{ui})(r_{vi} - b_{vi})}{\sqrt{\sum_{i \in I_{uv}} (r_{ui} - b_{ui})^2} \sqrt{\sum_{i \in I_{uv}} (r_{vi} - b_{uj})^2}}$

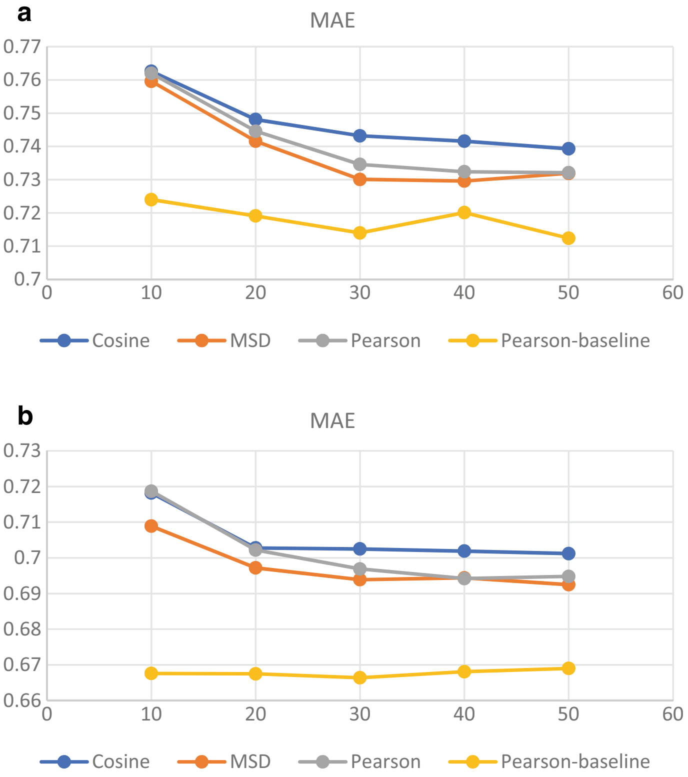


Fig. 3 **a** MAE on MovieLens 100 K **b** MAE on MovieLens 1 M

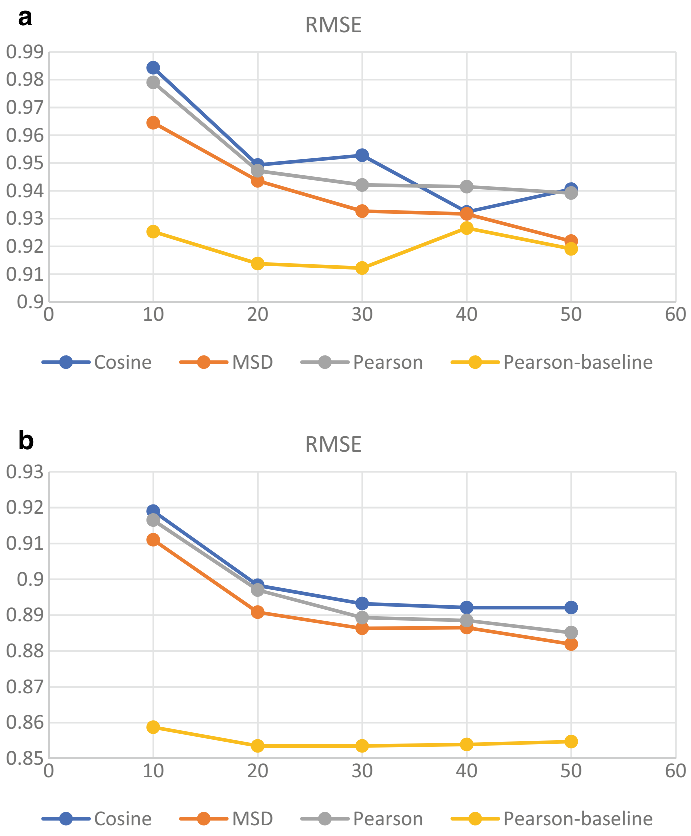


Fig. 4 **a** MAE on MovieLens 100 k **b** RMSE on MovieLens 1 M

in better accuracy (in terms of MAE and RMSE) than other similarity measures. Furthermore, in Pearson-baseline, we can observe that as the value of *k* crosses 30, the accuracy of the algorithm drops or in other words, neighbourhood size of 30 may be useful for an accurate recommendation.

5 Conclusion

In this work, we have summarized the various Python frameworks which are useful for building RS models based on collaborative filtering. We have shown why Surprise framework is better than other frameworks in terms of implementing and handling new and complex recommender models. We even conducted an experiment to analyse the accuracy of various built-in similarity measures present in the Surprise framework using the two most popular RS evaluation metrics, MAE and RMSE. From the results obtained by experiment, we can say that Pearson-baseline similarity measure outperforms other similarity measures by a wide margin. Well, in future, we will further optimize the similarity measures to obtain a more accurate recommendation.

References

1. Ricci F, Rokach L, Shapira B (2015) Recommender systems handbook
2. Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans Knowl Data Eng 17:734–749. https://doi.org/10.1109/TKDE.2005.99
3. Amazon. https://www.amazon.in/
4. Youtube. https://www.youtube.com/
5. Netflix. https://www.netflix.com/in/
6. IMDB. https://www.imdb.com/
7. ACM RecSys. https://recsys.acm.org/
8. Bobadilla J, Ortega F, Hernando A, Gutiérrez A (2013) Recommender systems survey. Knowl Based Syst 46:109–132. https://doi.org/10.1016/j.knosys.2013.03.012
9. Su X, Khoshgoftaar TM (2009) A survey of collaborative filtering techniques. https://doi.org/10.1155/2009/421425
10. Chakraborty J, Verma V (2016) A survey of diversification techniques in Recommendation Systems. In: Proceedings of 2016 international conference data mining and advanced computing SAPIENCE 2016, pp 35–40. https://doi.org/10.1109/sapience.2016.7684161
11. Hug N (2017) Surprise, a Python library for recommender systems. http://surpriselib.com
12. Celma O Pyrecsys.(sf). Ocelma Recuper. http://ocelma.net/software/python-recsys/build/html/index.html
13. Case Recommender Documentation. https://github.com/caserec/CaseRecommender
14. Frolov E, Oseledets IV (2016) fifty shades of ratings: how to benefit from a negative feedback in Top-N recommendations tasks. CoRR abs/1607.0
15. Spotlight Documentation. https://maciejkula.github.io/spotlight/
16. Yu J, Gao M, Li J, Yin H, Liu H (2018) Adaptive implicit friends identification over heterogeneous network for social recommendation. In: Proceedings of the 27th ACM international conference on information and knowledge management, pp 357–366
17. Pryor MH (1998) The effects of singular value decomposition on collaborative filtering. Dartmouth College, Hanover, NH, USA
18. Python-recsys Installation. https://github.com/ocelma/python-recsys
19. Python-recsys Documentation. http://ocelma.net/software/python-recsys/build/html/
20. Polara. https://github.com/evfro/polara
21. RecQ. https://github.com/Coder-Yu/RecQ
22. Koren Y, Bell R, Volinsky C (2015) MF techniques. 199–207. https://doi.org/10.1002/9781118770368.ch6
23. Kumar R, Verma K, Sunder B, Rastogi S (2014) Social popularity based SVD++ recommender system. Int J Comput Appl 87:33–37. https://doi.org/10.5120/15279-4033
24. da Costa AF, Manzato MG, Campello RJGB (2018) CoRec: a Co-training approach for recommender systems. In: Proceedings of the 33rd annual ACM symposium on applied computing. ACM, New York, NY, USA, pp 696–703
25. Subramaniyaswamy V, Logesh R (2017) Adaptive KNN based recommender system through mining of user preferences. Wirele Pers Commun 97:2229–2247. https://doi.org/10.1007/s11277-017-4605-5
26. Baltrunas L, Ricci F (2009) Context-based splitting of item ratings in collaborative filtering. In: Proceedings of the third ACM conference on recommender systems. ACM, New York, NY, USA, pp 245–248

27. Lemire D, Maclachlan A (2005) Slope one predictors for online rating-based collaborative filtering. In: Proceedings of 2005 SIAM International Conference Data Mining, SDM 2005, pp 471–475. https://doi.org/10.1137/1.9781611972757.43
28. Ma H, Yang H, Lyu MR, King I (2008) SoRec: social recommendation using probabilistic matrix factorization. In: Proceedings of the 17th ACM conference on information and knowledge management. ACM, New York, NY, USA, pp 931–940
29. Salakhutdinov R, Mnih A (2009) Probabilistic matrix factorization. In: Advanced in Neural Information Processing Systems 20—Proceedings of 2007 Conference, pp 1–8
30. Soucy P, Mineau GW (2001) A simple KNN algorithm for text categorization. In: Proceedings 2001 IEEE international conference on data mining, pp 647–648
31. Breese J, Heckerman D, Kadie C (1998) Empirical analysis of predictive algorithms for collaborative filtering John. In: Proceedings of Fourteenth Conference Uncertain Artificial Intelligence, pp 43–52
32. Resnick P, Iacovou N, Suchak M, Bergstrom P, Riedl J (1994) GroupLens: an open architecture for collaborative filtering of netnews. Acta Hepatol Jpn 58:183–190. https://doi.org/10.2957/kanzo.58.183
33. Shardanand U, Maes P (1995) Social information filtering: algorithms for automating "word of mouth". Conf Hum Factors Comput Syst—Proc 1:210–217

Computational Method for Cotton Plant Disease Detection of Crop Management Using Deep Learning and Internet of Things Platforms

Bhushan V. Patil and Pravin S. Patil

Abstract Cotton is a major crop from income point of view in India. Cotton crops are damaged due to early fall off leaf or leaf will get infected due to diseases. Due to sudden change in climatic conditions, plant diseases occurred either scorching temperature in the crop filed or some pesticides will be required within a time. There are multiple systems to detect and restrain the diseases on a cotton leaf through soil monitoring in classification and identification of numerous diseases like bacterial blight, Alternaria, and many more. After disease detection, will be provided to the farmers using various machine learning algorithms and IoT-based system. In this paper, the main focus is on a new deep learning method, which investigates to automatically identify a diseased plant from leaf images of the cotton plant and IoT-based platform in collecting various sensor data for detecting climatic changes. The deep CNN model is developed to perform cotton plant disease detection using infected and healthy cotton leaf images by collecting images through the complete process used in training and validation for image preprocessing; augmentation and fine-tuning. Different test cases were accomplished to check the performance of the created model and make this new system economical and independent. This newly created system gives accuracy as efficient as possible for cotton plant disease detection and restrains by improving crop production, this paper provides an innovative path to researchers for developing a cotton plant disease identification system.

Keywords Deep learning (DL) · Internet of things (IoT) · Leaf disease · Convolutional neural network (CNN)

B. V. Patil (✉)
Electronics and Telecommunication Engineering, R. C. Patel Institute of Technology, Shirpur, India
e-mail: patilbhushan007@gmail.com

P. S. Patil
Electronics and Telecommunication Engineering, SSVPS's BSD College of Engineering, Dhule, India
e-mail: pspatil777@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_81

1 Introduction

Distinct diseases in cotton crops is a serious alert that affects financial as well as commercial influence. The disease should be correctly diagnosed and identified early for further action to save the crop. Advancements in technologies are leading as we are surrounded by the huge volume of smart sensors and intelligent system (IS), interconnected through Internet and cloud platforms. Cotton is one of the most crucial crops in Maharashtra. Numerous diseases obstruct the growth of the plant in fields which may cause a massive loss in the quality of products. Cotton crops are damaged due to early fall off leaf or leaf will be infected due to diseases. Plant diseases are generally affected by various climatic conditions like scorching temperature in the crop filed and also some pesticides will be required within a time. There are multiple systems to detect and restrain the diseases on cotton leaf through soil monitoring in classification and identification of numerous diseases like bacterial blight, Alternaria, and many more [1]. After disease detection with its possible solutions, it is provided to the farmers using various machine learning algorithms and IoT-based system. To face problems like Maintainable Environmentalism, unwanted reduction, and topsoil optimization; requires numerous and heterogeneous variables gathering agricultural data and perform analysis for developing production techniques.

Deep learning is a very popular technique due to the accuracy of results when trained with huge data. It takes more time for training as compared to testing phase due to a number of parameters. Different deep learning model architectures such as AlexNet, VGG Network, GoogleNet, ResNet are the ones that use hundreds even thousands of these residual layers to create a network and then train, ResNeXt is the current technique for object recognition, RCNN (Region-Based CNN) used to solve the object detection problem, YOLO (You Only Look Once). Deep Learning (DL) is also named as deep structured learning or hierarchical learning, i.e., one of the parts of machine learning methods is basically worked on artificial neural networks. Learning can be differentiated into supervised, semi-supervised, or unsupervised. The word “deep” in “deep learning” depicts data is transferred through a number of layers. The main advantages of deep learning are to extract appropriate features automatically.

There are innumerable deep learning models like Convolutional Neural Network, AlexNet, GoogleNet, VGG (Visual Geometry Group) generally applied for plant and leaf disease detection. There are various tools to experiment with Deep Learning. The most popular are Theano, TensorFlow, Keras, Caffe, PyTorch, TFLearn, Pylearn2, Cuda, OverFeat, etc.

Figure 1 shows an overview of the proposed system.

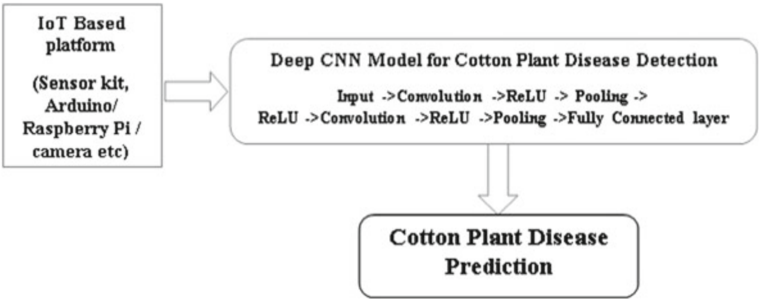
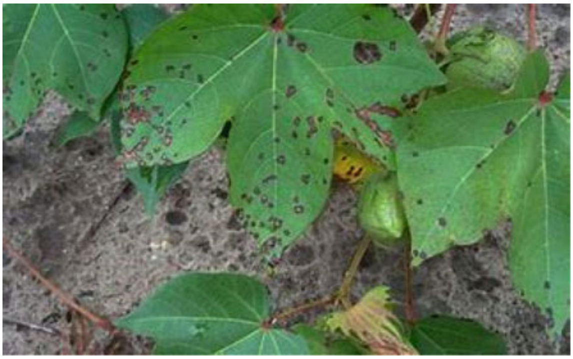

IoT Based platform
(Sensor kit, Arduino/ Raspberry Pi / camera etc)
Deep CNN Model for Cotton Plant Disease Detection
Input ->Convolution ->ReLU -> Pooling -> ReLU ->Convolution ->ReLU ->Pooling ->Fully Connected layer
Cotton Plant Disease Prediction

Fig. 1 Overview of the proposed system

Fig. 2 Bacterial-blight-infected leaf



2 Diseases of Cotton Leaf

2.1 *Bacterial Blight*

(i) It is one of the bacterial diseases caused by Xanthomonas Campestris Pv. Malvacearum bacteria.
(ii) Figure 2 shows infected leaf with bacterial blight [2].
(iii) It starts with dark green color and the border of red to brown or dark brown to black color.

2.2 *Alternaria*

(i) It is one of the types of diseases caused by fungal of Alternaria macrospora [2, 3].
(ii) Figure 3 shows the infected leaf.
(iii) Alternaria disease observed on the below part of leaves than the upper part with confusion of symptoms with spots of bacterial blight.

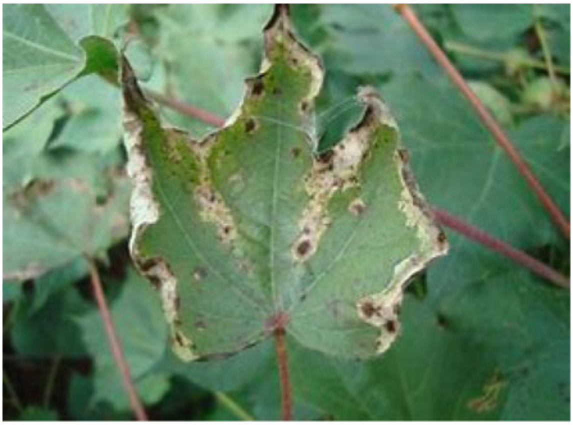


Fig. 3 Alternaria-infected leaf

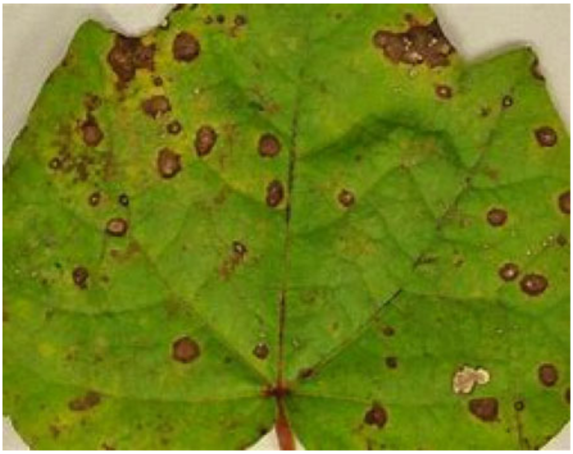


Fig. 4 Cercospora-infected leaf

2.3 *Cercospora*

(i) The infected leaf has red spots in unpredictable shapes with different shades of color (yellowish, purple, dark brown) spread out up to 2 cm.
(ii) Figure 4 shows a Cercospora-infected leaf.
(iii) The lesion region, Blackish leaf spot shows up through delicate veins that influence more aromatic, seasoned leaves of developing plants [2, 3].

2.4 *Grey Mildew*

(i) It's an irregular fungal disease about 110 mm diameter with light lustrous spots [2, 3].

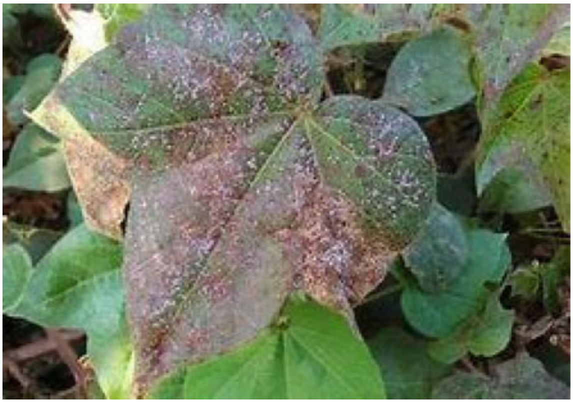


Fig. 5 Grey Mildew leaf

(ii) Figure 5 shows infected leaf.
(iii) The infection appears in whitish spots.

2.5 *Fusarium Wilt*

(i) It caused due to Fusarium oxysporum [2, 3].
(ii) Figure 6 shows Fusarium Wilt infected leaf.
(iii) The infected plant is shadier green and show down as the yellow dark color of the leaves.

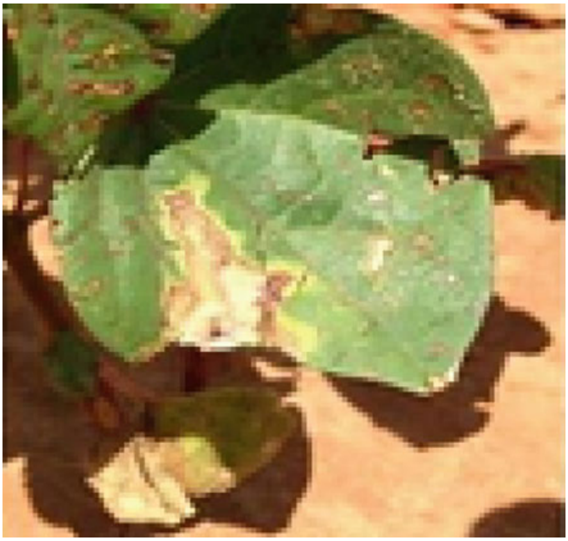


Fig. 6 Fusarium Wilt infected leaf

3 An Artificial Neural Network (ANN)

Figure 7 shows an artificial Neural Network model (ANN), which is based on biological neural networks. Each circular node represents an artificial neuron and a connection from the input of another to the output of one artificial neuron is represented by an arrow [4].

There are mainly two categories of Artificial Neural Network (ANN): (i) Feed Forward ANN and (ii) Feedback ANN.

Feed Forward ANN:

In this type, information flow in network is unidirectional. The information sends through a unit to another unit not possible without feedback loops (Fig. 8).

Feedback ANN:

This kind of category requires feedback loops that are used in content-addressable memories (Fig. 9).

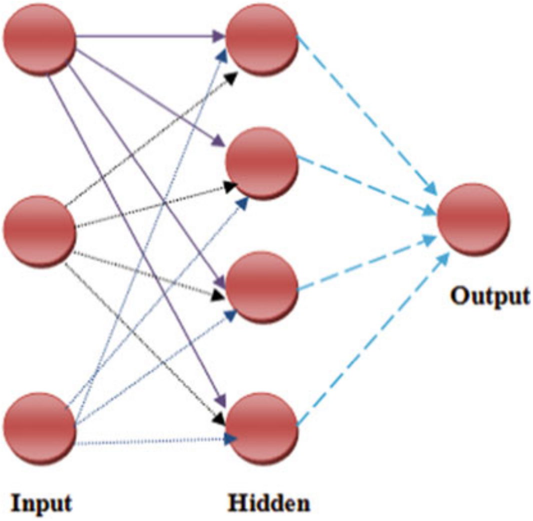


Fig. 7 Schematic representation of ANN

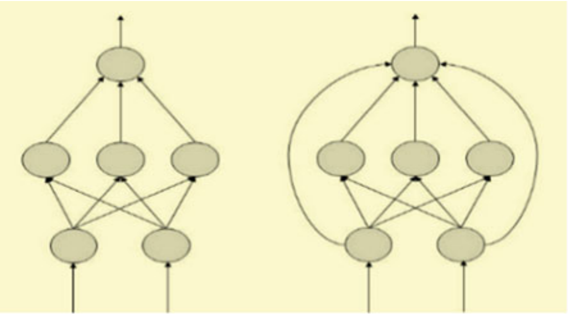


Fig. 8 Feed forward ANN (*Source* Computational science with Suman kumar swarnkar blogspot)

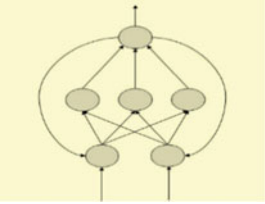


Fig. 9 Feedback ANN (*Source* Computational science with Suman kumar swarnkar blogspot)

Fig. 10 Phases for CNN

Deep CNN Model for Cotton Plant Disease Detection

Input ->Convolution ->ReLU -> Pooling ->
ReLU ->Convolution ->ReLU ->Pooling ->Fully Connected layer

This method describes the approach using the deep convolution neural network in detecting plant diseases, which enables the model to differentiate the infected leaves with healthy leaves or from the environment through deep CNN.

CNN technique is supervised type, which combines convolutional layers, ReLU Layer, fully connected layers, pooling layers, and activations layers.

A. Convolution Layer:

Convolution is the starting layer which is generally used to extract features from the input. It applies a convolution operation to input. This layer applies different filter layers to create a feature map (**kernel)** (Fig. 10).

B. Rectified Linear Unit (ReLU):

It is generally used to increases non-linearity on the feature map in deep learning models as an activation function. For negative input values setting back to returns 0.

This can be represented as

$$f(x) = x+ = \max(0, x) \tag{1}$$

where $x \rightarrow$ Input to neuron (A ramp function)

C. Pooling Layer:

Pooling layer is used to reduce the dimensionality, and it reduces different parameters when the image is large, and it controls the overfitting problems [5].

D. Fully Connected Layer:

This is the final layer where actual classification happens, where each neuron in the input is connected to each neuron in the output, here we add an artificial neural network to combine features and attributes which predict classes [5].

E. Train CNN with TensorFlow:

For numerical computation, TensorFlow is a software library to code for creating a convolutional neural network. Mathematical operations are represented through nodes while the multidimensional data arrays are represented through graph edges (tensors) communicated between them.

4 Literature Survey

The different reviews of papers are shown in Table 1,

5 Methodology

This paper proposed a cotton plant disease recognition system using deep learning having different steps as follows: Collecting dataset, pre-processing dataset, training the Convolutional Neural Network model (CNN) for identification and detecting types of leaf diseases, validation of model through the result (Fig. 11).

Dataset:

All the images were collected from day-to-day survey from the IoT-based system camera and sensor implemented on crop field and infected survey areas that we used for training hence to differentiate the leaves from the surrounding, then train the deep neural network.

The main objective of this study is to chance for the network by increasing appropriate features when using more augmented images. The main purpose of applying augmentation is to reduce overfitting during the training stage. In image augmentation, numerous transformation techniques such as affine transformation, prospective transformation, rotation are used.

Image Preprocessing and labeling:

Images in the dataset may be in different formats, quality, and resolution, for better feature extraction. Hence, the images need to be preprocessed, less than 500 pixels will not be considered as valid images for the dataset. The rest will be resized to 256 × 256 in order to reduce the time for training [11].

Neural Network Training:

In this step, to train the deep convolutional neural network (CNN) to make an image classification model, there are numerous deep learning frameworks like Python library Theano [12] and machine learning library that extends Lua, Torch7 [13]. Also, apart from this, there is Caffe, an open-source deep learning framework [14], containing reference CaffeNet model. Also for the prediction of the model, we need to compute the F1 score for the test's accuracy. Classification model's performance

Table 1 Reviews of different research papers

Sr. No.	Year of publication	Abstract
1	2016	This paper [4] explains the CNNs generation of achieved results of image classification, that describe the development of the recognition method for plant disease through the use of deep convolutional networks depending upon leaf image classification
2	2018	It [6] focused on deep learning methodologies to develop convolution neural network models for plant disease detection. Several model architectures were trained and reached a maximum success rate for the identification of plant–disease combinations. High success rate ensures the model is an early warning tool and in the future, it prolonged to operate in real-time crop growing to carry an integrated plant disease identification method
3	2019	Each and every country serves agriculture automation as the primary concern and prominent subject. With increasing in population at a fast rate, the need for food also increases tremendously, his paper [7] discusses a detailed survey, which the author has proposed and a system is focused that is implemented using IoT methodology on the botanical farm
4	2017	Crop diseases are today considered as the main threat in supply for food. Increasing demand of smartphone technology all over the world, it becomes technically feasible to influence image processing techniques for the identification of plant conditions from a simple photo, the main focus was on [8] by using deep convolution scheme for distinct classes to classify diseases status and crop species
5	2016	A disease on crop leads to the main issue in food security, and also identification is a difficult task in various parts due to inefficient infrastructure. Today through a mishmash of increasing global smartphone penetration with advancement the author author propos[5] focusses and discusses disease identification with the help of a deep convolutional neural network
6	2019	In this [9], [10] the author proposed the Kinship verification system using a CNN, as kin and non-kin classifier categories to examine whether image pair likely belongs to a unique category or not. Through this algorithm, various problems like tracking the images, the arrangement of images from a group using a machine are solved. Also, his work focused on illumination variations by proposing an adaptive version of the CLAHE algorithm to address the challenge of the illumination variations

is evaluated through the metric F1 score. In Keras, to compute the F1 score for each epoch, the F1 score reaches the best score as a 1, and the worst score as a 0 (Fig. 12).

Performed Tests: Test set for prediction of the leaf as healthy/unhealthy with its disease to evaluate the performance of the classifier.

Fine-Tuning: Fine-tuning helps to increase the accuracy of prediction.

Equipment Used: Training of the CNN performed in NVIDIA Graphics Processing Unit GPU mode such as GTX 1080ti, Tools such as Anaconda Python,

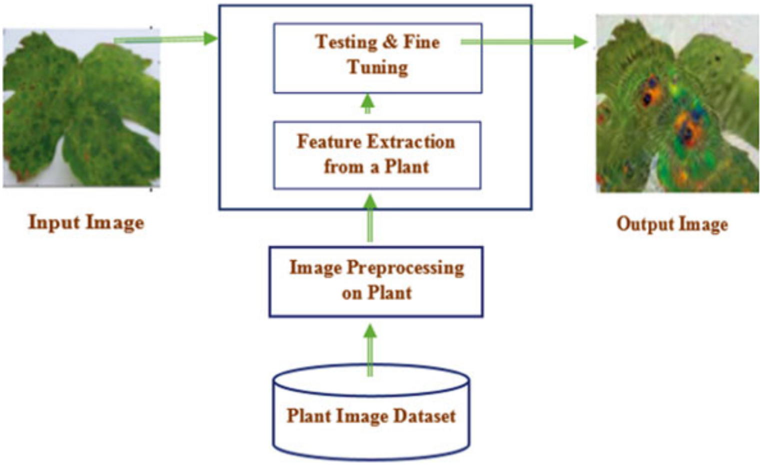


Fig. 11 Architecture of plant disease detection system using deep learning

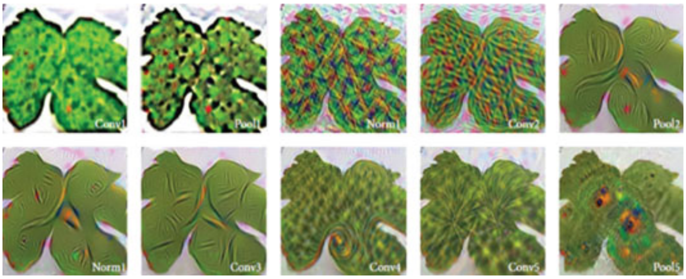


Fig. 12 Output layer images (*Source* Hindawi Publishing Corporation Computational Intelligence and Neuroscience Article ID 3289801)

and libraries such as OpenCV, Caffe integrated with CudNN. You will train the model for 10 k iterations, you should see the accuracy to be around 98.0%.

6 Conclusion

This model developed a new computational method that will identify the process of disease detection through images of the cotton plant and IoT-based platform in collecting various sensor data for detecting climatic changes. Using deep CNN model, the detection and classification of cotton plant is performed through the complete process from pre-processing to fine-tuning. Different test cases are accomplished to make new systems economical as well as independent by constantly checking the performance of the created model. So this new system with accuracy and efficiency will turn to be an improvement in the production of the crop. Smart farming using IoT platform with the mixing of sensors automates the irrigation system by monitoring field conditions anywhere to make this system efficient to one improve crop production. In the future, through this new methodology, the prediction of diseases on the cotton plant will be time-consuming to make it as effective.

References

1. Plant village Cotton. https://www.plantvillage.org/en/topics/cotton
2. Adhao AS et al (2017) Machine learning regression technique for cotton leaf disease detection and controlling using IoT. In: IEEE international conference on electronics, communication and aerospace technology ICECA, pp 449–454
3. Texas plant disease handbook. http://plantdiseasehandbook.tamu.edu/industryspecialty/fiber-oilspecialty/cotton
4. Sladojevic S et al (2016) Deep neural networks based recognition of plant diseases by leaf image classification. Comput Intell Neurosci. 11 pp. Article ID 3289801. Hindawi Publishing Corporation. http://dx.doi.org/10.1155/2016/3289801
5. Prasanna Mohanty S, Hughes D, Salathé M (2016) Using deep learning for image-based plant disease detection, pp 1–6 https://arxiv.org/ftp/arxiv/papers/1604/1604.03169
6. Ferentinos KP et al (2018) Deep learning models for plant disease detection and diagnosis, Elsevier- ScienceDirect-. Comput Electron Agric 145:311–318. https://doi.org/10.1016/j.compag.2018.01.009
7. Jha K et al (2019) A comprehensive review on automation in agriculture using artificial intelligence, Elsevier- ScienceDirect. Artif Intell Agric. 1–12.https://doi.org/10.1016/j.aiia.2019.05.004
8. Cortes E (2017) Plant disease classification using convolutional networks and generative adversarial networks. Stanford University ecortes@stanford.edu report
9. Yashwant Patil H et al. Deep learning based kinship verification on KinFaceW-I Dataset. In: TENCON 2019-IEEE region 10 conference-Kochi, India (TENCON). https://doi.org/10.1109/TENCON.2019.8929460
10. Yashwant Patil H et al (2015) A survey on image quality assessment techniques, challenges and databases. Int J Comput Appl 975: 34–38. www.ijcaonline.org (0975 – 8887). National Conference on Advances in Computing (NCAC 2015)
11. Kulkarni O (2018) Crop disease detection using deep learning. In: IEEE fourth international conference on computing communication control and automation (ICCUBEA- 2018). 978-1-5386-5257-2/18
12. Bergstra J, Bastien F, Breuleux O et al (2011) Theano: deep learning on GPUs with python. In: Proceedings of the NIPS 2011, big learningworkshop, Granada, Spain
13. Collobert R, Kavukcuoglu K, Farabet C (2011) Torch7: a Matlab-like environment for machine learning, BigLearn, NIPSWorkshop EPFL-CONF-192376
14. Jia Y, Shelhamer E, Donahue J et al (2014) Caffe: convolutional architecture for fast feature embedding: In Proceedings of the ACM conference on multimedia (MM'14), ACM, Orlando, Fla, USA, pp 675–678

Foot Ulcer and Acute Respiratory Distress Detection System for Diabetic Patients

M. S. Divya Rani, T. K. Padma Gayathri, Sree Lakshmi, and E. Kavitha

Abstract Health care and wellness management for a diabetic are one of the most promising information technology in the field of medical science. A healthcare monitoring system is necessary to constantly monitor diabetic patients' physiological parameters. Hence the major scope of this proposed project work is to develop a smart health monitoring system that overcomes many complications in diabetic patients by periodically monitoring patients' heartbeat rate, SPO2 (Peripheral capillary oxygen saturation) level, foot pressures, etc. Therefore, the IoT concept is used and sensors are connected to the human body with a well-managed wireless network that periodically monitors the physiological parameters of the body to avoid high risks in diabetic patients. Continuous health monitoring remotely works because of the integration of all components with wearable sensors and implantable body sensors networks that will increase the detection of emergency conditions at risk. Also, the proposed system is useful to operate remotely because of inbuilt Wi-Fi in the system.

Keywords Diabetic Foot Ulcer (DFU) · SPO2 (Peripheral capillary oxygen saturation) · ECG (electrocardiogram) · Acute Respiratory Distress Syndrome (ARDS)

M. S. Divya Rani (✉) · T. K. Padma Gayathri · S. Lakshmi · E. Kavitha
Department of TCE, Sir MVIT, Bangalore, India
e-mail: divyarani.9934@gmail.com

T. K. Padma Gayathri
e-mail: tkpadmag@gmail.com

S. Lakshmi
e-mail: sreeluperuru@gmail.com

E. Kavitha
e-mail: dr.kavitha_te@sirmvit.edu

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_82

1 Introduction

Acute Respiratory Distress Syndrome (ARDS) is a kind of respiratory failure due to the rapid onset of extensive inflammation in the lungs showing symptoms like rapid breathing, shortness of breath, and bluish skin coloration [1]. A literature review was conducted to identify the different causes of renal failure and cardiac arrest in most of the diabetic patients. It was found that the risk of death due to cardiac arrest is Pulmonary Edema that has been increased among diabetic patients which are associated with cardiovascular disease. Pulmonary Edema is a fluid collection in the air spaces and tissues present in the lungs that decrease the process of gas exchange which leads to respiratory failure. Pulmonary Edema is also due to either failure of the heart to eliminate blood effectively from the pulmonary system or damage to the lung parenchyma due to kidney-related problems in the case of diabetic patients. Pulmonary edema, especially critical, can lead to serious respiratory failure conditions or cardiac arrest in most of the diabetic patients. Congestive heart failure which is mainly due to the heart's incapability to push the blood out of the pulmonary system at an adequate rate resulting in increased heart pressure and pulmonary edema [2].

Diabetic Foot Ulcer (DFU) is also one of the major problems of diabetes mellitus, and maybe the key element of the diabetic foot. Being a natural process, wound healing is usually taken care of by the body's inherent mechanism of action that works consistently most of the time, Diabetes mellitus is one of the metabolic complications in type I and II diabetic patients that slow down the wound healing process. Peripheral neuropathy is the most common problem of diabetes and its early discovery through the provision of education and appropriate foot care may reduce impairment especially in diabetes-based foot complications. Most common among the neuropathies are chronic sensorimotor distal symmetric polyneuropathy and autonomic neuropathies.

Hence, a new secure IoT-Based Modern Healthcare monitoring system for diabetic patients is proposed, to give flexibility and fast operational speed to get expected outcomes. With wearable sensors and embedded sensor networks, the detection of emergencies increases to reduce the high risk of deaths in diabetic patients [3–7]. Therefore, a new secure IoT-based new technology for constant healthcare monitoring system for diabetic patients Type I and Type II is proposed.

2 Proposed System

Figure 1 shows the system architecture. The major scope of this proposed system is to develop a smart health monitoring system that overcomes many complications in diabetic patients by periodically monitoring patients' heartbeat rate, SPO2 (peripheral capillary oxygen saturation) level, foot pressures, etc. In this proposed technology, various hardware were used to accomplish the tasks namely Arduino

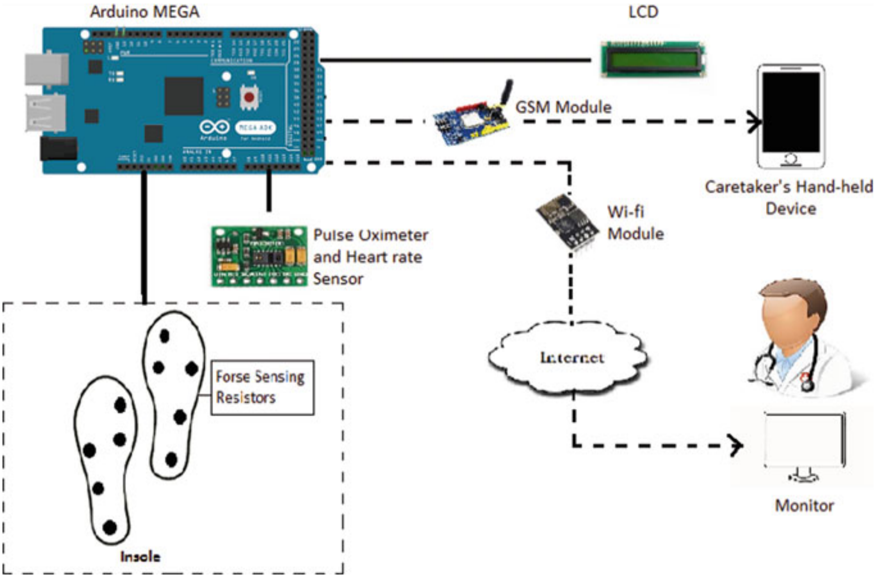


Fig. 1 System architecture

Mega, pulse oximeter or photoplethysmography sensor, pressure sensors, etc. and more sensors also can be used to detect various vital parameters. The proposed smart system comprised of four main parts. The first part is for detecting the blood oxygen level to recognize the indication of pulmonary edema, second being the detection of electrocardiogram commonly referred to as ECG or EKG (heartbeat detection) to indicate the cardiac arrest and the third part is to detect the pressure from the pressure sensors to detect the foot ulcers under normal or abnormal conditions of patients and finally to provide the detected data for remote monitoring. Remote monitoring of the measured vital parameters of the patient enables a caretaker or health specialist to observe a patient's health progress in any emergency and nonemergency condition.

Remote viewing of the data enables a doctor or health specialist to monitor a patient's health progress away from hospital premises in any emergency and nonemergency condition. Arduino Mega is a prototyping device that is connected with sensors, sensors are coupled with human bodies and this prototype is connected with software systems by using a wireless connection.

2.1 Implementation

There are two main sections: Transmitter section and the Receiver Section. The transmitter section comprises an insole, which consists of pressure sensors known as Force-Sensing Resistor sensors placed on the 5 areas of rubber insole to detect the pressure at various areas of each foot. The pulse oximeter and heart rate sensor collects data when the patient places a finger on it which subsequently detects the blood oxygen saturation level and heart rate of a diabetic patient. Continuously the data is sent to the Arduino Mega 2560. The data transmitted is encoded and sent serially over air interface through RF module, i.e., GSM and Wi-Fi, the measured vital parameters of the patients are displayed on the display unit for patient view and in case of emergency, an SMS alert is sent to caretakers mobile. The collected sensor data is connected to the Wi-Fi module that pushes data to the ThingsBoard server via the MQTT protocol. Hence, the data is investigated using a built-in, customizable dashboard. ThingsBoard database is used because the system formats the recorded data which is flexible to access. When abnormal data has been recorded, an alert indication through messages will be sent to the doctor's mobile, and hence critical situations can be handled to avoid the risks. With the help of an RF module fixed at the transmitter side, the data is further transferred to the receiver section.

The receiver section is RPM (Remote Patient Monitoring) system and the technology is used for observing patients' health outside the conventional health centers. The remote healthcare system offers healthcare specialists or doctors to access the digital data of heart rate, SPO2 and foot pressure sensor to a centralized view of all the diabetic patients allowing them to investigate further, creating customized precautions according to a diabetic patient's health condition and ailment and hence alerts and reminders that trigger the Doctor and Caretaker in case of medical emergencies.

2.2 *Block Diagram*

The block diagram of the detection system is shown in Fig. 2. The sensors are placed on a rubber foot insole, which is a thin material especially manufactured for diabetic patients. The sensors on the rubber insole and all the sensors that measure heart rate and SPO2 are connected to the Arduino Mega 2560 as shown in the Fig. 3. These

Fig. 2 System block diagram

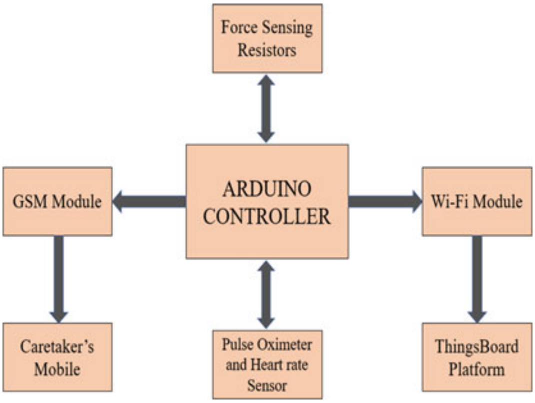


Fig. 3 Arduino Mega 2560

signals are interfaced by PC using Arduino IDE. The application program that is running on Arduino Mega 2560 is written using the C++ programming language, which is quite simple and easy to access.

There are two functions to be performed, one is calculating the foot pressures and the other being measurement of heart rate and SPO2. The keypad embedded at the transmitter allows the patient to select any one function at a time from the available two functions. When key 1 is pressed, foot pressure measurement is originated and when key 2 is pressed a heart rate/pulse oximeter is activated. The data is measured and processed in the main controller and hence compared with the specified thresholds according to the patient weights. Further, the Arduino is connected to the GSM module, which is used to send alert messages to the caretaker's mobile phone in case of both critical and noncritical conditions.

The transmitter also has a Wi-Fi Module that sends the data to the Thingsboard server via MQTT protocol by using Pub–Sub-Client library for Arduino. The data is investigated using a built-in customizable dashboard. The application that is running on Arduino is written using the Arduino SDK. The data stored in ThingsBoard can be observed by the doctor from anywhere.

The link to the dashboard can be privately shared with doctor to monitor the data. The same data is also available on the LCD for immediate view by the patient.

The proposed system is a smart health monitoring system, which uses the latest IoT technology for monitoring the vital parameters of the diabetic patients. The main component of this project uses an Arduino Mega board, which is interfaced with a Wi-Fi through the serial communication port. The main sensors which are used in this project are FSR for foot pressure and MAX30100 sensor for measuring the heart rate and SPO2 level. The measured values are displayed on the LCD display in the system itself, and are sent serially to the Wi-Fi Module. The Wi-Fi is programmed in order to work with the ThingsBoard.

Data is acquired from FSR sensors and MAX30100 by the Arduino controller.

The data is processed in the Arduino controller and compared with the specified thresholds. Arduino is connected to the GSM module, which is used to send alert messages to the caretakers' or patient's phone in case of emergency.

Wi-Fi module sends the data to the dashboard.

The data stored in ThingsBoard can be viewed by the doctor from anywhere.

The link to the dashboard can be privately shared to monitor the data. The same data is also available on the LCD for immediate view.

3 System Components

3.1 Required Hardware Components

- **Arduino Mega**

The Arduino Mega 2560 is a microcontroller board which enables the following [8]:

- **MAX3100**

Figure 4 represents MAX30100, it is pulse oximetry that measures the blood oxygen saturation level which is called a SPO2 level [9].

- **FSR**

Figure 5 represents a Force-Sensing Resistor (FSR). FSR is a sensor whose resistance changes when a force, pressure or mechanical stress is applied [10].

Most FSRs are either a circular or rectangular sensing area. The square FSR is suitable for broad-area sensing, while the circular sensors can provide more precision to the location being sensed.

- **ESP 8266-01**

ESP8266-01 is an impressive, low-cost Wi-Fi module suitable for adding wireless accessibility through the Internet to an existing controller project via UART serial connection shown in Fig. 6 [11]. Wi-Fi module is used to access the cloud ThingsBoard through the Internet.

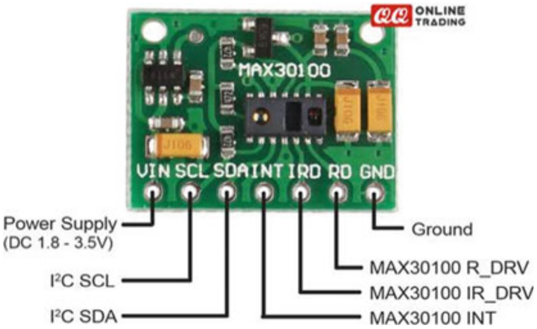


Fig. 4 HR/SPO2 measurement device

Fig. 5 Round FSR

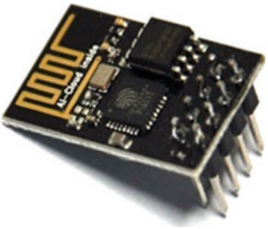


Fig. 6 Wi-Fi module

Fig. 7 GSM module

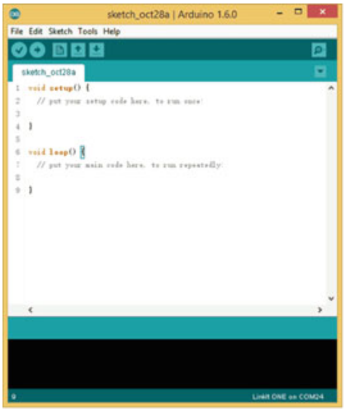
- **GSM SIM 800C**

Figure 7 represents SIM800C that is an RF module that can transmit SMS, Voice, and data information with low power consumption. In our project, we are using GSM SIM 8000C to send SMS alert messages to the caretaker.

3.2 Required Software Components

- **Arduino IDE**

The Arduino Integrated Development Environment (IDE) is a cross-platform application that is written in the programming language Java. This software is open-source software that is generally used as an editor and compiler for Arduino Mega 2560. It



```
void setup() {
  // put your setup code here, to run once:

}

void loop() {
  // put your main code here, to run repeatedly:

}
```

Fig. 8 Arduino IDE

is easily accessible for operating systems like MAC, Windows, Linux, and runs on the Java Platform. Arduino IDE shown in Fig. 8 supports both C and C++ languages.

- **ThingsBoard dashboard**

ThingsBoard is an open-source tool for IoT applications mainly incorporated for real-time data collection, visualization, processing, and device management [12].

4 Hardware Setup

The experimental setup which is designed for acquiring the foot pressure signal and HR/SPO2 levels is shown in Fig. 9. The pressure from the sensors is obtained by voltage ranging from 0 to 5 v output. The voltage can be converted to a pressure unit also. The voltage recorded will be the same for the normal person without abnormalities based on the sensors we positioned on the rubber insole.

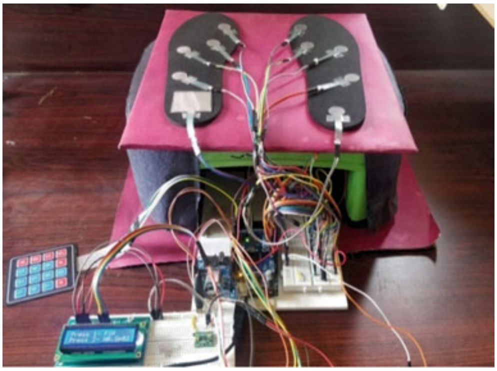


Fig. 9 Position of the pressure sensors

Figure 10 shows the hardware demonstration of this work making a nondiabetic person weighing 52 kg to stand on the sensors placed on the foot insole. The FSR1 to FSR5 is placed on the left insole and the FSR6 to FSR10 is placed on the right insole. The readings are shown in Fig. 11 further it is sent to caretakers' mobile, which indicates the reading of voltages in millivolts with respect to FSR. In our project, for 52 kg nondiabetic patients these readings will be the optimized threshold value. If suppose the patient is with diabetics and suffering from acute foot ulceration disease, the output voltage will vary with respect to each FSR. Hence, we conclude in this work that the voltage variations with respect to each position of the sensors will enable us to find the abnormalities in diabetic patients suffering from foot ulcers. Due to the foot ulcer, the pressure in the particular area of the foot increases further increasing the voltage levels. Hence foot ulceration can be detected and the same has been reported to caretaker's mobile as shown in Fig. 11.

If the diabetic patient is suffering from kidney-related issues, the heart rate and SPO2 level are detected by placing the finger on the device when there is a discomfort condition faced by them. Figure 12 shows the record of heartbeat and SPO2 levels on the LCD Display. The normal heartbeat is within 60–100 beats per minute and

Fig. 10 Complete hardware

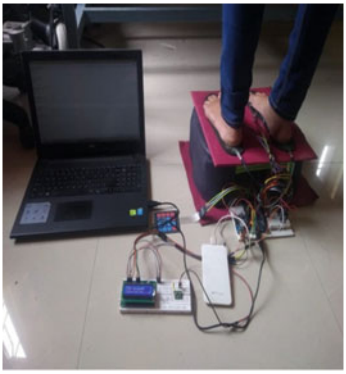
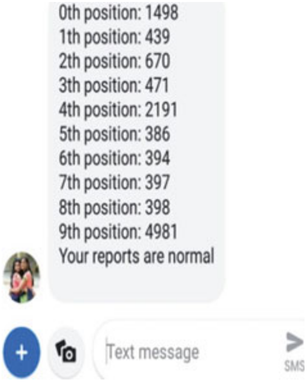


Fig. 11 Message alert on mobile



0th position: 1498
1th position: 439
2th position: 670
3th position: 471
4th position: 2191
5th position: 386
6th position: 394
7th position: 397
8th position: 398
9th position: 4981
Your reports are normal

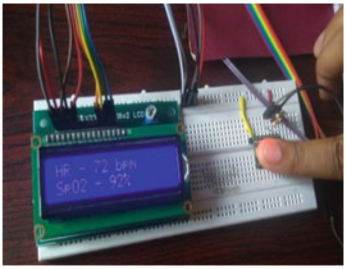
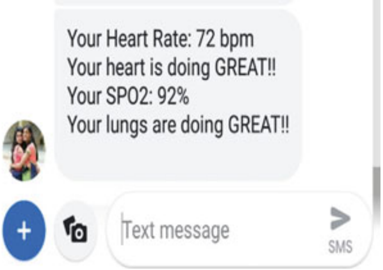


Fig. 12 SPO2 and HR measurements on LCD

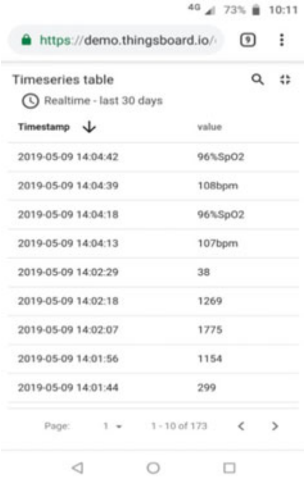
the normal SPO2 level is 90–100%. The recorded value is for a healthy person and the same has been sent to a caretaker's mobile which is shown in Fig. 13.

Figure 14 shows the continuously sent data from the insole and HR/SPO2 sensors to the dashboard, which is monitored by a doctor who treats the patient. If there is an emergency, the patient will be asked to visit the hospital immediately.



Your Heart Rate: 72 bpm
Your heart is doing GREAT!!
Your SPO2: 92%
Your lungs are doing GREAT!!

Fig. 13 Message alert on mobile



Timeseries table

Realtime - last 30 days

Timestamp ↓	value
2019-05-09 14:04:42	96%SpO2
2019-05-09 14:04:39	108bpm
2019-05-09 14:04:18	96%SpO2
2019-05-09 14:04:13	107bpm
2019-05-09 14:02:29	38
2019-05-09 14:02:18	1269
2019-05-09 14:02:07	1775
2019-05-09 14:01:56	1154
2019-05-09 14:01:44	299

Page: 1 ▾ 1 - 10 of 173

Fig. 14 Message alert on Thingsboard

5 Conclusions

The proposed work helps a diabetic patient prone to the situation leading to the development of a risky life-threatening condition of deaths due to Pulmonary Edema that can be monitored to predict and alert in advance any indication of the body status. The diabetic foot ulcer can be monitored at an earlier stage and reduces any foot amputation possibilities in diabetic patients. There is a significant reduction of hospitalizations as patients suffering from chronic diseases are on a remote healthcare monitoring program, hospital admissions can be reduced greatly. The proposed smart health monitoring system leads to increased healthcare team productivity, enabling more evidence-based care and more efficient diabetic patient care management. Acute care discharge planning is enhanced using remote health care management solutions. Further, the system can be improved by implanting many FSRs on foot insole to get accuracy in the sensing area.

References

1. Coons JC, Murali S (2011) Pharmacotherapy for acute heart failures syndromes. Health Syst Pharm 11–35
2. Wakai A, Mc Cabe A (2013) Nitrates for acute heart failure syndromes. Cochrane Database Syst Rev. https://doi.org/10.1002/14651858.cd005151
3. Patil SL, Madhuri A, Thatte U, Chaskar M (2009) Development of planter foot pressure distribution system using flexi force sensors. Sens Transducers J 108(9):73–79
4. Malvade PS, Joshi AK, Madhe SP (2017) IoT based monitoring of foot pressure using FSR sensor 0635–0639. https://doi.org/10.1109/iccsp
5. Sudarvizhi D Ms, Nivetha M, Priyadharshini P, Swetha JR (2019) Identification and analysis of foot ulceratıon using load cell technıque. IRJET 06(03):7792–7797
6. Nithya AN, Premkumar R, Dhivya S, Vennila M (2013) A real time foot pressure measurement for early detection of ulcer formation ın diabetic patient using labview. In: International conference on design and manufacturing, pp 1302–1309
7. Chin YF, Huang TT (2013) Development and validation of a diabetes foot self care behavior scale. J Nurs Res 21:1–12
8. Mega 2560 datasheet (2015). http://www.alldatasheet.com
9. Max 30100 datasheet (2015). http://www.alldatasheet.com
10. FSRAdafruit (2018). https://www.cdn-learn.adafruit.com.pdf
11. Wi-Fi module for Aurdino (2017). https://www.arduino.cc
12. Open source IOT platform (2017). https://www.thingsboard.io

Cluster-Based Prediction of Air Quality Index

H. L. Shilpa, P. G. Lavanya, and Suresha Mallappa

Abstract The present world is facing the crucial issue of air pollution, which is threatening the human race and the environment equally. This situation requires effective monitoring of air quality and recording the pollution levels of different pollutants SO_2, CO, NO_2, O_3 and particulate matters ($PM_{2.5}$ and PM_{10}). To achieve this, we need efficient prediction and forecasting models which not only monitors the quality of the air we breathe in but also forecast the future to plan accordingly. In this direction, various data mining methods are adopted for analysing and visualizing concentration levels of air pollutants using Data Mining on big data and data visualization. In this work, we have explored the partition-based clustering technique to extract the patterns from the air quality data. We have compared prediction methods, Auto-Regressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) on representatives of each cluster and also done forecasting for the year 2018. The results have proven that ARIMA works better than LSTM.

Keywords Air quality index · ARIMA · Air pollution · Big data clustering · Forecasting · LSTM · Prediction · Particulate matters

1 Introduction

We are in a scenario where the entire humanity is under the threat of air pollution for which we are directly or indirectly responsible. Air pollution is creating havoc as it damages the life on this planet. Human beings, animals, forests and water bodies are

H. L. Shilpa (✉)
P.E.S. College of Engineering, Mandya, India
e-mail: ly.shilpahl@gmail.com

P. G. Lavanya · Suresha Mallappa
Department of Studies in Computer Science, UoM, Mysore, India
e-mail: lavanyapg@gmail.com

Suresha Mallappa
e-mail: sureshasuvi@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_83

all affected by air pollution. Air pollution is also posing a hazard to the environment as it is the root cause for the depletion of the ozone layer which acts as a shield to the earth and absorbs the sun's ultraviolet radiation. The quality of the air affects the quality of our life. It is continuously changing from place to place and from time to time. Hence the U.S. Environmental Protection Agency (EPA) and local air quality agency are working to provide information about the quality of the outdoor air in an easy way such that even a common man can understand and take necessary action. Air Quality Index (AQI) is one such important tool in this direction as it tells us how good or bad the air we breathe is. By creating this awareness, local officials also give us information to safeguard ourselves from the ill effects of air pollution.

Gases and meteorological parameters data are usually recorded by automatic stations at regular time intervals. Meteorological data is typically multivariate that often consists of many dimensions. The AQI is calculated for five major air pollutants regulated by the Clean Air Act: Ground level Ozone (O_3), Partical Pollution ($PM_{2.5}$ and PM_{10}), Carbon Monoxide (CO), Sulfur Dioxide (SO_2), and Nitrogen Dioxide(NO_2). For each of these pollutants, EPA has established national air quality standards to protect public health [1]. This work proposes clustering-based analysis and visualization of air quality data which is proven to be an efficient method for knowledge discovery. Historical air quality data from 2003 to 2017 has been mined using the clustering technique to group states of the USA on the basis of the Air Quality Index in respective states. Prediction of air quality using ARIMA and LSTM techniques has been performed on the cluster wise data choosing one state as a representative from each cluster. The prediction result obtained for the year 2017 is compared with the actual data and the result shows that cluster wise prediction gives nearer value to the actual data. Forecasting also has been performed for the year 2018 and compared with EPA actual data which shows the efficiency of the techniques used.

2 Related Work

As Air Quality Index is very important for society, research in this area is also increasing and has a lot of scope. A study about air pollution caused by industries in Nigeria is studied in [2]. They chose two industrial plants in Nigeria, Ife steel plant and Ibadan Asphalt, for their analysis and implemented an AQMS application using a decision tree classification algorithm and did the prediction. In their study, they considered three pollutants: Particulate Matter (PM_{10}), Sulphur Dioxide (SO_2) and Nitrogen Dioxide (NO_2). A study on daily air quality data of Delhi at ITO (a busiest traffic intersection place) by considering Respirable Suspended Particulate Matter (RSPM), SO_2, NO_2 and Suspended Particulate Matter (SPM) pollutants is done in [3]. Initially, they have performed prediction using Multi-Linear Regression (MLR), but because of the multi-colinearity problem they worked on prediction technique using Principal Component Analysis for four different seasons from 2000 to 2006. They observed and predicted values in the years 2000–2005 and for the year 2006.

A study on varying trends of ambient air quality and the levels of related air pollutants are analysed based on the database monitored at Bahadurshah Jafar Marg near ITO intersection Delhi, which is carried out in [4]. They predicted three pollutants NO_2, SO_2 and RSPM using Seasonal ARIMA. The study suggested that since the prediction varies with the daily traffic, the prediction will be improved by including the traffic volume and industrial emission data. The air quality in Malaysia is predicted for the year 2016 using the pollutants CO and NO_2 in [5]. Pollutant data from the selected four monitoring stations from the years 1996–2006 are analysed using ARIMA and SARIMA models to forecast the trend of the pollutants. The forecasting models show that the predicted values are within the range prescribed by the country's organizations NAAQS and DOE. KNN classifier is used to predict the air quality index using the parameters sulphur dioxide, nitrogen monoxide, nitrogen dioxide, carbon monoxide and ozone in [6]. The data used is recorded for 29 days in June 2009. To have an accurate prediction more data is required.

The study [7–9] is based on annual data collected from www.epa.gov with NO_x, SO_2, VOC, CO, PM_{10} and $PM_{2.5}$ and they found that the SOFM model gives better prediction when compared to other models. Models are implemented with limited historical data and can be improved with input from multiple data sources. In [10], six different methods are used to measure the air quality index in Nagpur and Maharashtra of India from May to October 2014 with reference to PM_{10}, $PM_{2.5}$, SO_2 and NO_2.

In [11], prediction of $PM_{2.5}$ concentration from meteorological data is done through statistical models based on relevant machine learning models. The study is carried out for the two locations in Quito. The models are based on wind speed, wind direction and precipitation. A method is proposed in [12] to know the health effect because of traffic flow and meteorological conditions in Wroclaw. The study is done considering two years of data between 2015 and 2016. Random Forest based partition model is used to model the regression relationship between the NO_2, NO_X and $PM_{2.5}$ pollutants, meteorological conditions and temporal conditions.

From all the above studies, it is observed that the study of Air Quality is the need of the hour as necessary steps have to be taken by the Government and Public to address the problem. Another important observation is though many people have addressed the problem using different prediction and forecasting techniques, the data considered is limited to a specific area and comparatively less. Hence, we were motivated to carry out clustering-based prediction and forecasting using a large amount of data, i.e. daily data collected for a period of 15 years. We also noted that there is no comparative study between ARIMA and LSTM models. As we observe from the clustering of the data, the Government can take some of the common measures to reduce the pollution level in all the states which fall on the same cluster instead of each state. This helps in reduced expenditure while controlling air pollution. K-means with DTW algorithm is used for clustering because of its optimal match between two time series. For the analysis of air quality, we used two techniques ARIMA and LSTM. ARIMA is one of the better models to analyse the time series data and can be done using forecasting. It works on a single variable. Since ARIMA captures linear relationships we use the LSTM model for non-linear associations and have carried out the analysis.

3 Proposed Methodology

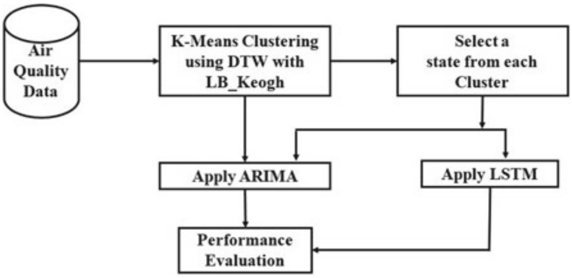
In this work, we propose a clustering-based method for air quality prediction and forecasting. Initially, we downloaded the data from the Historical Air Quality dataset and perform ARIMA on the entire dataset to predict air quality using a 70:30 ratio for training and testing, respectively. We evaluate the performance of our prediction using different metrics as explained in Subsection 3.6. Further, we explore the efficacy of partition-based clustering technique on the data and perform K-means clustering. Instead of the normally used Euclidean distance, Dynamic Time Warping (DTW) with LB_Keogh is used as it has been proved comparatively better. Clustering groups the states based on the air quality index. A state is chosen as the representative of the cluster and prediction is performed on the four states representing four clusters. The results obtained are evaluated using different metrics. The overall flow of the proposed methodology is given in Fig. 1.

3.1 Time Series Modelling

"A time series is a set of observations measured sequentially through time" [13]. Time series is a collection of data points that are collected at constant time intervals. It is a dynamic or time-dependent problem with or without increasing or decreasing trend, seasonality. Time series modelling is a powerful method to describe and extract information from time-based data and helps us to make informed decisions about future outcomes. The forecasting method analyses the sequence of historical data in a period of time to establish the forecasting model.

Depending on the number of variables, the analysis becomes either univariate which uses only one variable or multivariate which uses more than one variable [14]. Here, we use a single variable $PM_{2.5}$ for the analysis of time series. In our work, we make use of two pieces of knowledge—factors influencing pollution and Air Quality Index(AQI) observed every year between 2003 and 2017. With respect to these concepts, we determine the similarity between time series of multiple states and the time series of the 5479 days from the year 2003 to 2017. We have worked with $PM_{2.5}$ daily data as this has been seen to be the cause for lung diseases in the USA.

Fig. 1 Workflow of the proposed method


Air Quality Data
K-Means Clustering using DTW with LB_Keogh
Select a state from each Cluster
Apply ARIMA
Apply LSTM
Performance Evaluation

As per the study of "American Lung Association", more than 4 in 10 Americans live with unhealthy air according to 2018 "State of the Air" report [15]. In this work, we have used the ARIMA model for prediction of $\text{PM}_{2.5}$ pollutant and we have also compared the ARIMA model with the LSTM model. The predicting of future point for the series in ARIMA is such that

- the auto-regressive part gives the pattern of growth/decline in the data;
- the "integrated" part gives the rate of change of the growth/decline in the data;
- the moving average deals with the noise between consecutive time points.

This model was introduced by George P. Box and Gwilym Jenkins, hence it is also called a Box Jenkins method. They proved that non-stationary data could be made stationary by "differencing" the series [16]. Forecasting using ARIMA for a stationary time series is similar to a linear regression equation. The predictors depend on the parameters (p, d, q) of the ARIMA model.

Number of AR (Auto Regressive) terms(p): The value of p represents the order of AR model. An AR model is a Linear Regression of the current value of the series against one or more prior values of the series. For the p^{th} order Eq. (1) has the form

$$X_t = \delta + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots + \varphi_p X_{t-p} + \varepsilon_t \tag{1}$$

Where,

$$\delta = \left(1 - \sum_{i=1}^{p} \varphi_i\right)\mu \tag{2}$$

Here, $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$: past time series; $\varphi_1, \varphi_2, \ldots, \varphi_p$: unknown parameters relating X_t to $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$; ε_t: error term of the model; μ: the process mean. Number of MA (Moving Average) terms (q): In a prediction equation, lagged forecast errors are given by MA terms. It improves the current forecast. For qth order, Eq. (3) has the form

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{3}$$

Here, $\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_{t-p}$: the past random shocks; $\theta_1, \theta_2, \cdots, \theta_q$: unknown parameters relating X_t to $\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_{t-p}$; μ: the mean of the series.

MA estimate fitting is more difficult than AR models as the error terms are not observable. They also have a less obvious interpretation.

The input data series for ARIMA requires to be stationary which means constant mean, variance, autocorrelation through time. The non-stationary data has to be differenced in order to make the data stationary. The difference is usually one or two ($\text{d} \leq 2$) for practical purposes. The non-stationary data after differencing one or more times gives an Integrated(I) model which is stationary [17].

We check the stationarity of time series using (a) Plotting Rolling Statistics, where the moving average or moving variance is plotted to see its variance with time, (b) Dickey–Fuller Test, a well–known statistical test to check stationarity.

An important concern is to find the optimal parameters for ARIMA model. To determine the order, we use two plots Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). ACF/PACF is used to determine the ‘p’ and ‘q’ values.

3.2 K-Means Clustering

The squared error between μ_k and the points in the cluster c_k is defined as in Eq. (4).

$$J(C_k) = \sum_{x_i \in C_k} \left\| x_i - \mu_k^2 \right\| \tag{4}$$

The objective of K-means is to minimize the sum of the squared error over all the K clusters [18]. The number of clusters for this data was empirically obtained using the elbow method.

3.3 Dynamic Time Warping (DTW)

K-means clustering technique uses different distance methods to calculate the distance between centroids and the data points. In this work, we use DTW distance which is a very robust technique for measuring time series similarity when compared with Euclidean distance. DTW is a popular shape-based similarity measure for time series data. It can also support non-equal length time series and hence is ideal to be used with time series data. As it is computationally expensive, many lower bound measures have been used along with DTW to reduce the cost. One such lower bounding measure is LB_Keogh [19], which has been used here. Suppose we have two time series, a sequence $Q = q_1, q_2, \cdots, q_i, \cdots, q_n$ and a sequence $C = c_1, c_2, \cdots, c_j, \cdots, c_m$.

In order to obtain an optimal path, the path that gives a minimum cumulative distance at (n, m) has to be chosen. The distance is defined as in Eq. (5):

$$D_{\mathrm{DTW}}(Q, C) = \min_{\forall w \in P} \sqrt{\sum_{k=1}^{K} d_{w_k}} \tag{6}$$

If there are ties when selecting the minimum, the algorithm randomly chooses any neighbour producing different paths but the warping distance will still remain the same.

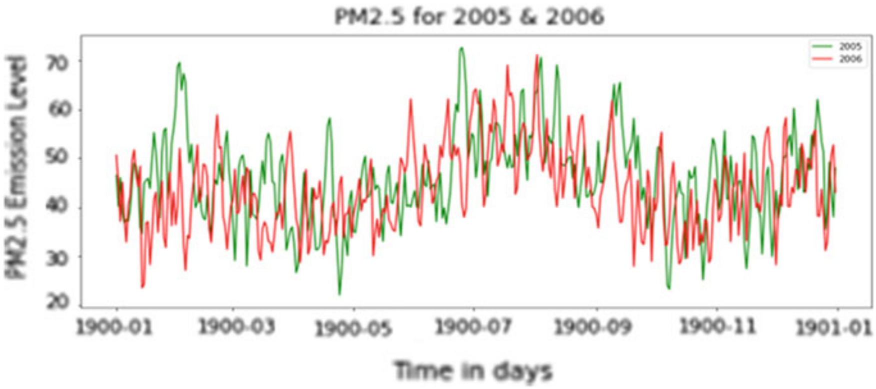


Fig. 2 Time Series for the year 2005 and 2006

We considered two time series for the years 2005 and 2006, and applied both Euclidean and DTW distance methods to time series. The Euclidean distance results were 215.63 and the DTW distance results were 101.93. The result of LB_Keogh distance reduces, even more, i.e. it gives 47.514. So from this, we conclude that DTW with LB_Keogh is better than Euclidean distance. The graph in Fig. 2 shows the time series for the years 2005 and 2006.

3.4 Long Short-Term Memory (LSTM)

The reason to develop the LSTM is to prevent long-term dependency problem, i.e. preventing the model from “remembering” the data over a long period of time. This method is based on Recurrent Neural Networks (RNN) with additional features to memorize the data. In RNNs repeating module has one layer, but in LSTM it has four layers. The four layers are: three sigmoid layers and a ‘*tanh*’ layer. The four layers are themselves used in a “recurrent” way using gates to allow/disallow information through them.

The cell state is a key of LSTM. It moves straight throughout the chain without modifying the information. The LSTM gates allow optionally the information to pass through and are used to add or delete the information to the cell state. The gates are composed of Sigmoid neural network layer and a multiplication operation. The output is either 0 or 1. If the output is 0, ‘let nothing through’ or if it is 1, ‘let everything through’.

To control and protect the cell state in LSTM, three gates (forget, input and output) were used.

Below steps describe the LSTM working procedure.

Step 1: ‘Forget layer gate’: the number yields between 0 and 1. 1 indicated ‘completely keep this’ and 0 indicated ‘completely get rid of this’

Step 2: Store the new information in the cell state. It has two parts. One part is called ‘input gate layer’ which decides the value to update and another part ‘*tanh*’ layer creates a vector of new candidate values.

Step 3: Combine the above two parts to update the state

Step 4: The output is based on cell state. First, run a sigmoid layer which decides what parts of the cell state are going to output. Then, put the cell state to through ‘tanh’ (the values to be between −1 and 1) and multiply it by the output of the sigmoid gate, so that we can output only the parts we decide.

3.5 Metric Measures

The different metric measures are given in Eqs. (6)–(9).

Mean Squared Error (MSE): It measures the average squared error of predicted value. The equation is

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_i\right)^2 \tag{7}$$

Root Mean Squared Error (RMSE): It is a square root of Mean Squared Error. The equation is given by

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_i\right)^2} \tag{8}$$

Mean Absolute Error (MAE): It measures an average of absolute differences between the actual values and the predictions. The equation is

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|X_i - \bar{X}_i\right| \tag{9}$$

Mean Absolute Percentage Error (MAPE): It is a measure of prediction accuracy of a forecasting method. The formula is

$$MAPE = \frac{100}{n} \times \sum_{i=1}^{n}\left|\frac{X_i - \bar{X}_i}{X_i}\right| \tag{10}$$

Here, n: Total number of data points; X_i: Actual observation time series; $\bar{X}_i$: Predicted time series

4 Implementation and Results

4.1 Data Collection

For this work, we gathered data from historical air quality data from the Kaggle website. It contains the values which are measured by Environment Protection Agency (EPA) [20]. To access the data from EPA, it is necessary to have Google credentials. The data is accessed using BigQuery, which is a RESTful web service that provides us with an interactive way to handle massive datasets stored in Google storage. We have to use our Google cloud credentials to access the same and use queries to retrieve data.

To fix the number of years of data that has to be considered for prediction, we practically tested on California state $PM_{2.5}$ (24-hours average) AQI data and calculated RMSE value using the ARIMA model by passing datasets in different ratios as training and test data which is given in Table 1. The graph for the same is given in Fig. 3, which shows as the number of years of data increases, the prediction error rate is very less. So we chose 15 years of data for prediction.

Table 1 RMSE value recorded for different ratios of data

Data in Months/Years	90–10	80–20	70–30	60–40	50–50
01-12-2017 to 31-12-2017 (1 Month)	27.071	31.805	27.839	26.115	24.72
01-09-2017 to 31-12-2017 (4 Months)	23.948	18.831	18.213	17.11	16.265
01-06-2017 to 31-12-2017 (7 Months)	19.327	17.06	17.06	14.947	13.773
2017 (1 Year)	17.936	15.254	13.771	10.33	12.021
2016–2017 (2 Years)	15.149	12.612	11.294	9.691	10.477
2013–2017 (5 Years)	11.879	10.425	9.946	9.691	9.558
2008–2017 (10 Years)	10.378	9.662	9.692	9.874	9.852
2003–2017 (15 Years)	10.378	9.662	9.778	9.858	9.831

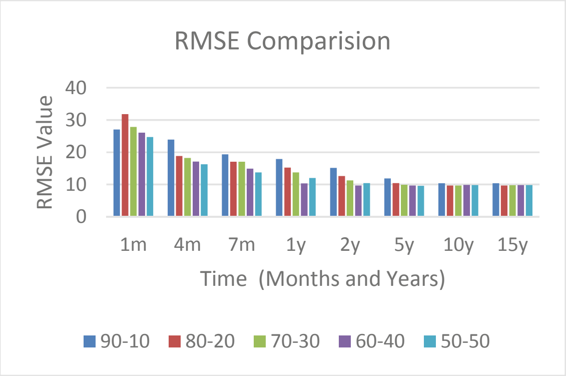


Fig. 3 RMSE value comparison for different ratios for different time

4.2 Experiment Results

We initiated our work with the very well-known time series forecasting static model—ARIMA. We tested this model on data with the $PM_{2.5}$ (24-hour average) AQI value collected from 2003 to 2017. The data contains 12,29,732 rows, which is daily data collected for all counties from 35 states of the USA. We have taken the date wise mean of the entire data from 2003 to 2017. The model is trained with 70% of data and 30% of data is tested. After applying the ARIMA model, we obtained an RMSE of 5.261. Figures 4 and 5 show the time series and the predicted graph, whereas Table 2 gives the measured metrics. While this value is small, there is a significant issue to be considered. For this, the data is the mean value of the pollutant level for a particular date across all the states in the USA, i.e. 5479 data points. For a particular

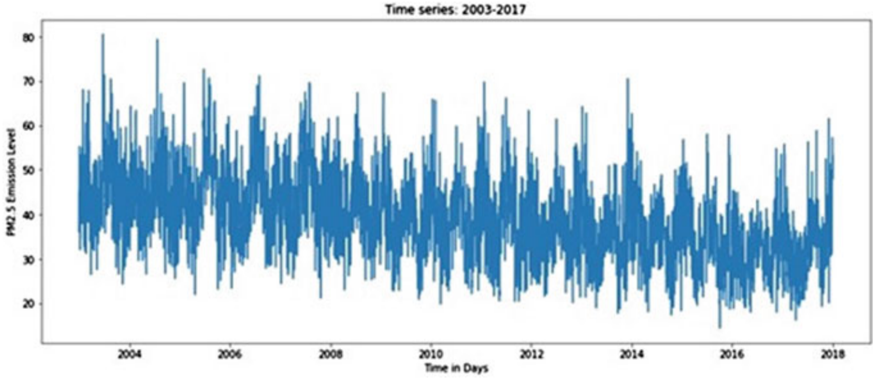


Fig. 4 Time series data from 2003 to 2017

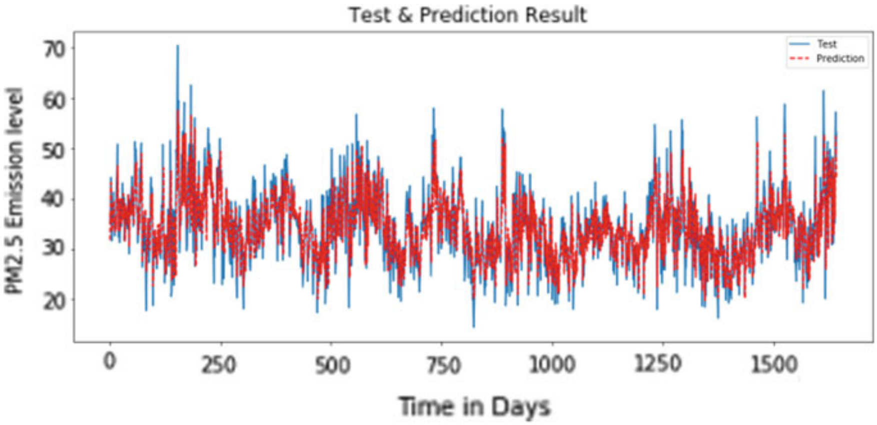


Fig. 5 Prediction using ARIMA for 70–30 ratio

Table 2 Resultant metric measures

Model	MSE	RMSE	MAE	MAPE
ARIMA	27.677	5.261	4.070	0.237
LSTM	33.753	5.810	4.494	0.138

date, every state will result in the same prediction. This provided a motivation to rethink our prediction method.

There are two ways through which Particulate Matter (PM) emissions are generated. The first one is by directly releasing the particles into the atmosphere (primary PM). Human-made sources of $PM_{2.5}$ are more important than natural sources like emissions from road vehicles, industrial emissions, population exposure close to roadside, etc. The second one is by releasing its precursors, such as nitrogen oxides (NO_x and NO_2) and sulphur dioxide (SO_2). These are called secondary particles. We understand that air pollution in one state is affected by population, topology, climate and industrial activities in adjacent states [21]. By merely fitting a regression model on the entire data across the U.S, we destructively allow unrelated states to influence the regression of pollution in another state. Therefore, to overcome such a deterrent influence, we delve into clustering air pollution series of individual states. The resulting clusters combine those time series that are similar, i.e. they would be topologically, topographically and climatically similar. This effectively deals with the spatio-temporal behaviour of the data.

Figure 6 shows the time series of cluster 3 with states list, Fig. 7 shows the US state map and Fig. 8 shows the time series prediction of cluster 1.

We can observe from Table 3 that there is an increase in RMSE value when compared to the overall date wise mean of 2003–2017 data of all states except in cluster 2. The change in RMSE value in each cluster is because of the PM level on that time series in the grouped states. We note some of the main reasons for particulate matter pollution in different clustered states. In California, population is one of the

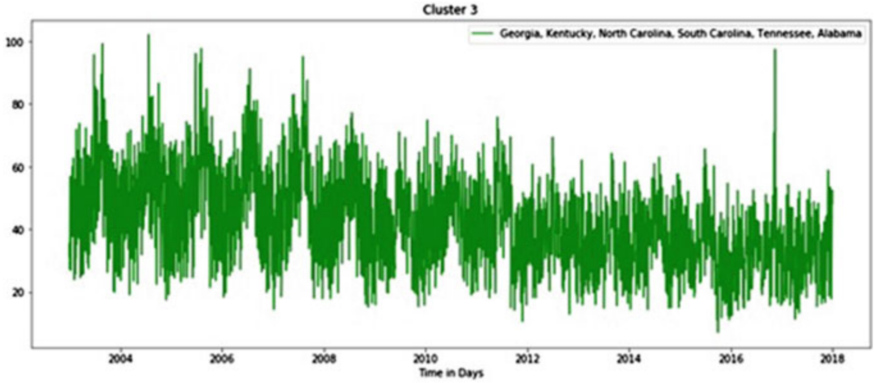
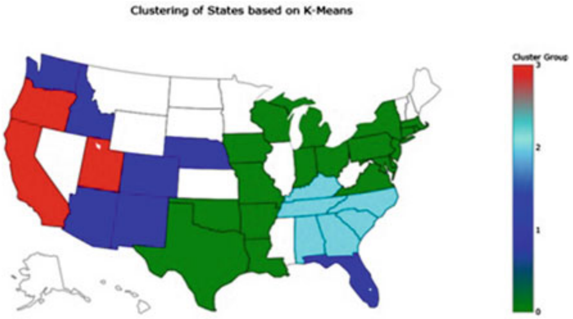


Fig. 6 Cluster 3 time series

Fig. 7 US state map showing clustered states



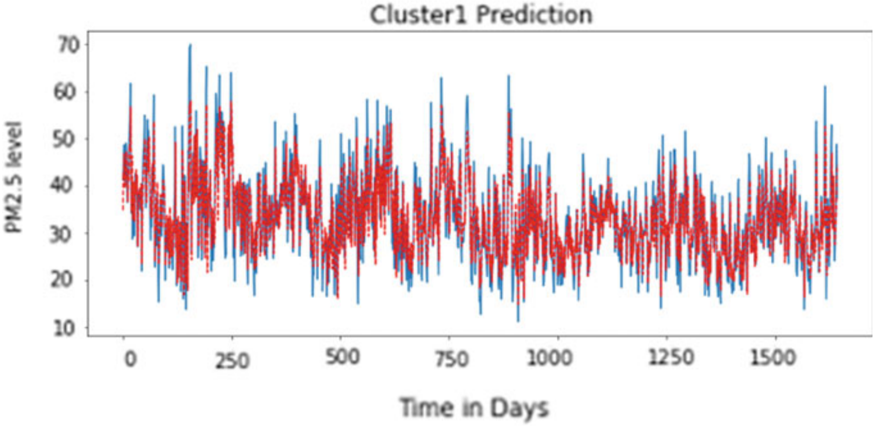


Fig. 8 Prediction result using ARIMA of cluster 1 with 70% training data and 30% test data

Table 3 Clustered states and error measures of each cluster with the ration 70:30

Cluster No.	States on each cluster	MSE	RMSE	MAE	MAPE
1	Connecticut, Delaware, Indiana, Iowa, Louisiana, Maryland, Massachusetts, Michigan, Missouri, New Jersey, New York, Ohio, Oklahoma, Pennsylvania, Rhode Island, Texas, Virginia, Wisconsin, Arkansas	32.454	5.697	4.445	0.145
2	Colorado, Florida, Idaho, Nebraska, New Mexico, Washington, Arizona	22.150	4.706	3.375	0.122
3	Georgia, Kentucky, North Carolina, South Carolina, Tennessee, Alabama	42.069	6.486	4.990	0.163
4	Oregon, Utah, California	63.286	7.955	5.522	0.197

important reasons for the increase in PM level which affects neighbouring states. In Alabama, which belongs to cluster 3, one of the reasons for PM pollution is coal and oil-fuelled power plants.

The cluster is done based on time series, so the states falling in each cluster are having similar $PM_{2.5}$ levels at that time. We select a state from each cluster as a representative and perform prediction using ARIMA for those four states, viz., California, Colorado, Michigan and New Jersey. This prediction would be more useful as it is specific to each cluster which contains states with similar values of $PM_{2.5.}$ The value of $PM_{2.5}$ for different states lying in cluster 3 is shown in Fig. 9.

We performed prediction for the year 2017 using different historical data for California and Colorado states and the values for different metric measures are shown in Tables 4 and 5. The prediction was done using historical data for 1 year, 2 years, 5, 10 and 14 years of data. It can be observed that the value of RMSE decreases with an increase in the amount of data. This proves that we can give a better prediction if we have more historical data.

The prediction plot of Georgia state (a representative state of cluster 3) for the year 2017 using 14 years of data is shown in Fig. 10.

We compared the ARIMA model with another well-known model of LSTM by performing prediction for the year 2017 using data from the years 2003 to 2016. The different measures which are given in Table 6 show that the ARIMA model performs better in prediction when compared with MSE and RMSE measures of the LSTM model.

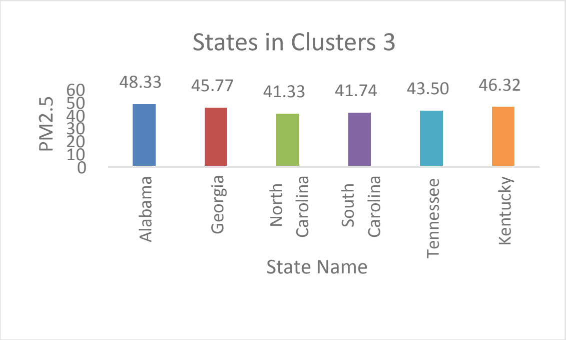


Fig. 9 Cluster 3 states $PM_{2.5}$ level

Table 4 Error measures for prediction for 2017—California State

State	Trained year(s)	MSE	RMSE	MAE	MAPE
California	2016	108.406	10.412	7.746	0.455
	2015–16	107.615	10.374	7.72	0.455
	2012–16	107.199	10.354	7.718	0.457
	2007–16	106.465	10.318	7.669	0.459
	2003–16	105.581	10.275	7.692	0.456

Table 5 Error measures for prediction for 2017—Colorado State

State	Trained year(s)	MSE	RMSE	MAE	MAPE
Colorado	2016	129.071	11.361	7.998	0.579
	2015–16	125.011	11.181	7.848	0.588
	2012–16	124.105	11.14	7.875	0.585
	2007–16	123.678	11.121	7.986	0.58
	2003–16	121.474	11.022	7.913	0.591

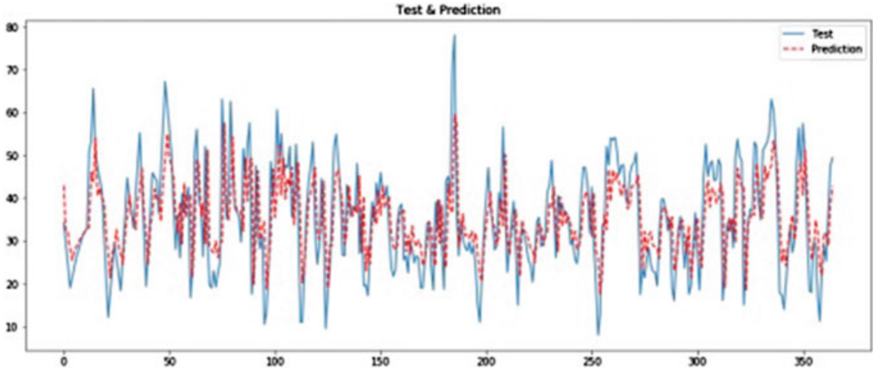


Fig. 10 Prediction plot of Georgia state for 2017 with training data from 2003 to 2016

Table 6 Comparison between ARIMA and LSTM models

State name	ARIMA model				LSTM model			
	MSE	RMSE	MAE	MAPE	MSE	RMSE	MAE	MAPE
California	105.581	10.275	7.692	0.456	114.815	10.714	7.965	0.213
Colorado	121.474	11.022	7.913	0.591	121.723	11.033	7.754	0.356
Georgia	88.663	9.416	7.249	0.411	396.705	14.641	12.249	0.406
New Jersey	139.405	11.807	9.385	0.526	486.589	18.641	15.13	0.608

The mean value of the predicted API for the year 2017 is compared with the actual data is shown in Table 7. We can observe that when we considering the complete data of all 35 states, we get a value of 34.391 which is nearer to 38.972 which is the actual value. But if the same is compared with different representative states, there is a less difference which proves the significance of cluster-based prediction.

AQI Forecasting using ARIMA model
The daily forecasting for the next 365 days, i.e. for the year 2018 is done based on previous days' $PM_{2.5}$ values by taking the seasonal difference of 365 days for 2003–17. We implemented monthly predictions for the next 12 months, by taking the monthly mean data of previous years from 2003 to 2017. Here the seasonal

Table 7 Comparison of predicted values with actual values

	Using ARIMA	Actual
Overall	34.391	38.972
California	41.748	42.254
Colorado	28.371	27.089
Georgia	35.229	35.301
New Jersey	35.585	31.679

difference is for 12 months. In California and Colorado states, PM level is low in April (33.677 and 18.761) and high in December(58.973 and 33.094). In Georgia and New Jersey states, it’s low in October (28.911, 24.216) and high in July (35.658, 38.396). The plot for forecast AQI using daily and monthly data for California state is shown in Fig. 11a and b.

The mean AQI value for the four representative states for the year 2018 is given in Table 8. The forecasting done for the year 2018 for California state is compared with the EPA dataset. The mean AQI values for different seasons for California state are shown in Table 9. It shows that the prediction using ARIMA is almost near to the actual EPA—2018 data.

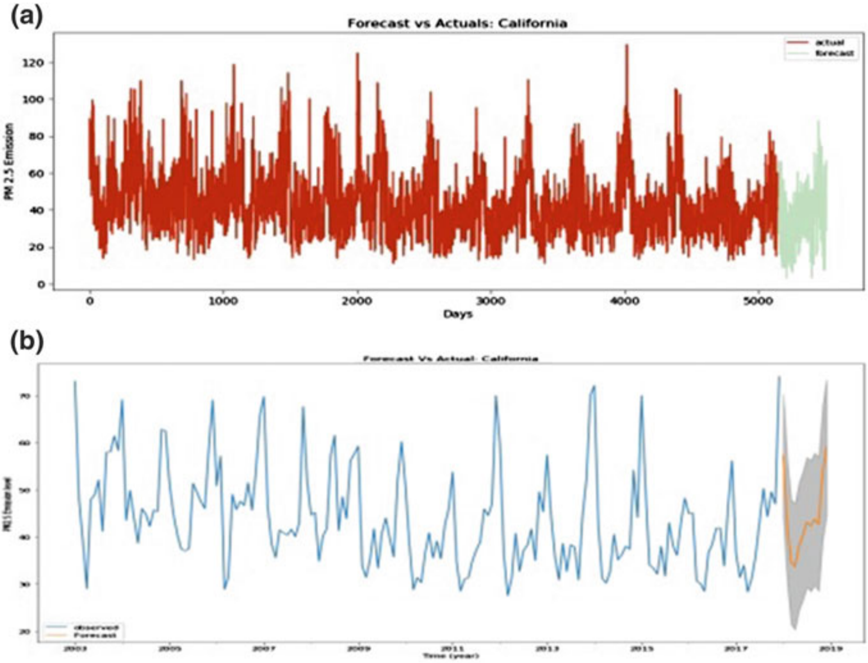


Fig. 11 Forecast AQI for California state using **a** daily data and **b** monthly data

Table 8 Mean of forecast values of PM for the year 2018

State	PMAQI for the year 2018
California	38.857
Colorado	26.231
Georgia	36.365
New Jersey	31.187

Table 9 Comparison of forecast and actual data for California state

Season	EPA Actual—2018	Forecast using ARIMA for 2018
Winter	41.695	37.331
Spring	30.244	32.133
Summer	48.692	40.265
Autumn	44.377	45.742
Annual mean	41.241	38.857

5 Conclusion

This work presents a clustering technique based air quality prediction model and also a comparative study of two well-known prediction methods, ARIMA and LSTM. We propose a model that groups all the states that follow similar pollutant level behaviour using K-means clustering. A representative state of the clusters has been chosen and prediction for that state is carried out which is observed to be better compared to the prediction using the complete data. It is also observed that prediction improves with the increase in the amount of data. We have presented a comparison between two prediction models, ARIMA and LSTM and observe that for our dataset ARIMA technique performs slightly better compared to LSTM. Using ARIMA model, we forecast the $PM_{2.5}$ pollutant level for the year 2018 for California, Colorado, Georgia and New Jersey states. We verified our model by comparing our forecasted result of California state with EPA actual data for 2018 and our model gives the nearest value when compared with EPA actual data. This further emphasizes the cluster-based prediction and forecasting of air quality data. The work can be further extended to work with other pollutants and by including additional variables like temperature, humidity, wind speed, etc., or by excluding some unwanted seasonal values. Further other alternative algorithms such as machine learning algorithms, fuzzy techniques, etc., can be used to increase the prediction accuracy.

References

1. https://airnow.gov/index.cfm?action=aqi_broucher.index

2. Ofoegbu EO, Fayemiwo MA, Omisore MO (2014) Data mining industrial air pollution data for trend analysis and air quality index assessment using a novel back-end AQMS application software. Int J Innov Sci Res 11(2):237–47s
3. Kumar A, Goyal P (2011) Forecasting of air quality in Delhi using principal component regression technique. Atmos Pollut Res 2(4):436–44 Elsevier
4. Kaushik I, Melwani R (2007) Time series analysis of ambient air quality at ito intersection in Delhi (India). J Environ Res Dev 2(2)
5. Mohd ZI, Roziah Z, Marzuki I, Muhd SL (2009) Forecasting and time series analysis of air pollutants in several area of Malaysia. Am J Environ Sci 5(5):625–32
6. Dragomir EG (2010) Air quality index prediction using K-nearest neighbour technique. Bull PG Univ Ploiesti Ser Math Inf Phys LXII 1:103–108
7. Delavar MR, Gholami A, Shiran GR, Rashidi Y, Nakhaeizadeh GR, Fedra K, Hatefi Afshar S (2019) A novel method for improving air pollution prediction based on machine learning approaches: a case study applied to the capital city of Tehran. ISPRS Int J Geo-Inf 8(2):99
8. Lee MH, Rahman NH, Latif MT, Nor ME, Kamisan NA Seasonal ARIMA for forecasting air pollution index: a case study. Am J Appl Sci 94):570–578. ISSN 1546-9239 © 2012 Science Publications
9. Barai SV, Dikshit AK, Sharma S Neural network models for air quality prediction: a comparative study. In: Soft computing in Industrial applications 2007. Springer, Berlin, pp 290–305
10. Nigam S, Rao BP, Kumar N, Mhaisalkar VA (2015) Air quality index-a comparative study for assessing the status of air quality. Res J Eng Technol 6(2):267
11. Kleine Deters J et al (2017) Modeling PM2. 5 urban pollution using machine learning and selected meteorological parameters. J Electric Comput Eng
12. Kamińska JA (2019) A random forest partition model for predicting NO2 concentrations from traffic flow and meteorological conditions. Sci Tot Environ 651:475–483
13. Chatfield C (2000) Time-series forecasting", Chapman and Hall/CRC
14. Abdel-Aal RE (2008) Univariate modeling and forecasting of monthly energy demand time series using abductive and neural networks. Comput Ind Eng 54(4):903–17 Elsevier
15. State of the Air report (2018). https://www.lung.org/our-initiatives/healthy-air/sota/key-findings/
16. Box GEP, Jenkins GM, Reinsel GC (1994) Time series analysis, forecasting and control, 3rd edn. Prentice Hall, Englewood Clifs, NJ
17. Brockwell P J., Davis RA, Calder MV (2002) Introduction to time series and forecasting, vol 2. Springer, New York
18. Jain AK (2010) Data clustering: 50 years beyond K-means. Pattern Recogn Lett 31:651–666 Elsevier
19. Keogh E, Ratanamahatana CA (2005) Exact indexing of dynamic time warping. Knowl Inf Syst 7(3):358–86 Springer
20. https://www.kaggle.com/epa/epa-historical-air-quality
21. Clark LP, Millet DB, Marshall JD (2014) National patterns in environmental injustice and inequality: outdoor NO_2 air pollution in the United States. PLoS ONE 9(4):e94431

A Memetic Evolutionary Algorithm-Based Optimization for Competitive Bid Data Analysis

Pritam Roy

Abstract One of the most difficult decision-making problems for buyers is to identify which suppliers to provide contracts to of the biddable items for a competitive event after considering several conditions. This is an example of pure integer linear programming (ILP) problem with cost minimization where all the decision variables, i.e., the quantity of the biddable items to be awarded to suppliers, are always nonnegative integers. Normally for solving ILP, Gomory cutting plane or Branch and Bound technique using the simplex method are applied. But when the problem to be solved is highly constrained and a large number of variables are involved, finding a feasible solution is difficult and that can result in poor performance by these techniques. To address this, an improved memetic meta-heuristic evolutionary algorithm (EA) such as shuffled frog leaping algorithm (SFLA) is utilized to find the optimum solution satisfying all the constraints. The SFLA is a random population-based optimization technique inspired by natural memetics. It performs particle swarm optimization (PSO) like positional improvement in the local search and globally, it employs effective mixing of information using the shuffled complex evolution technique. In this paper, a modified shuffled frog leaping algorithm (MSFLA) is proposed where modification of SFLA is achieved by introducing supplier weightage and supplier acceptability to improve the quality of the solution with a more stable outcome. Simulation results and comparative study on highly constrained and a large number of items and suppliers' instances from bidding data demonstrate the efficiency of the proposed hybrid meta-heuristic algorithm.

Keywords Competitive bidding · Integer linear programming (ILP) · Modified shuffled frog leaping algorithm (MSFLA) · Memetic algorithm (MA)

P. Roy (✉)
SAP Ariba, Bengaluru, India
e-mail: pritam.roy@sap.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_84

1 Introduction

The rapidly growing market of e-commerce and e-procurement has opened lots of possibilities among buyers and sellers [1, 2]. In online sourcing projects, especially in case of Request For Proposal (RFP) events, after all the suppliers completed the competitive bidding or after the bidding process ends, one of the most important and difficult tasks for buyers is to evaluate the bids [3] and provide the contract to the supplier for those materials or line items.

The main purpose of bid evaluation is to determine the best bid for a line item among the bids submitted on or before the bid closing date and time specified in the online sourcing event. The best bid is not always the lowest priced bid, the optimum bid is more complex with a combination of goods, works or services, where value for money may not be achieved by using only the lowest price as the contract award criterion. The ideal bid for a line item is completely dependent on the buyer's requirements and how the buyer wants to provide the contracts of the line items. In order to determine the best responsive evaluated bid accurately in accordance with the buyer requirements and other possible constraints, we can formulate and solve this by converting it to a pure integer linear programming (ILP) problem with cost minimization. The basic linear programming problem can be expressed as

$$objective\, func : maximize\, or\, minimize\, c^T x$$
$$subject\, to\, Ax \leq b\, and\, x \geq 0$$

where A is the matrix of coefficient, x is the vector of variables, c and b are vectors of coefficient. The objective or fitness function is to maximize or minimize the expression. The inequalities $Ax \leq b\, and\, x \geq 0$ are the constraints which have to be satisfied while optimizing the objective function.

Mostly, for solving ILP problems, Gomory cutting plane or Branch and Bound technique are used using the simplex method, but it requires a lot of time. When the problem is highly constrained and a large number of variables are involved, in that case it is very difficult to get the optimum solution and performance impact is there. Traditional numerical methods may not be cost-effective since the computing time needed for a solution is greatly dependent on the dimension and the structure of the matter. In some approaches, a neural network is used for solving the integer programming problem, but it also fails to provide the optimum result [4–6].

In this paper, modified SFLA is applied to find the best possible outcome for the bid analysis problem. In order to satisfy the buyer requirements with supplier data (i.e., supplier weightage, and other constraints like quantity allocation or sometimes supplier provides less price per quantity if the contract is given to that supplier with more than the cut-off quantity for an item), we have implemented mutation like a process where we are modifying a quantity variable randomly to meet the criteria and getting the optimum outcome easily. The proposed algorithm is run with several different sets of parameter configurations and the outcome is very exciting as it provides very accurate results compared to other methodologies or algorithms [14].

2 Bid Evaluation Problem and Converting it to ILP

We have converted all the available bid data to a matrix or array form for the computational purpose.

From the buyer side, while creating a sourcing event, they are providing the line items with the required number of quantities (QL_i), for line item ranges *1–n*

$$QL_i = \{\, QL_1 \; QL_2 \; QL_3 \; \ldots . \; QL_n \,\}$$

There may be *s* number of suppliers who will bid for those line items and after the bidding is over, the final representations of the bids will be

$$\begin{gathered} S_1 = \{\, L_1S_1 \; L_2S_1 \; L_3S_1 \; \ldots . \; L_nS_1 \,\} \\ S_2 = \{\, L_1S_2 \; L_2S_2 \; L_3S_2 \; \ldots . \; L_nS_2 \,\} \\ \ldots\ldots \\ S_s = \{\, L_1S_s \; L_2S_s \; L_3S_s \; \ldots . \; L_nS_s \,\} \end{gathered}$$

As quantities are always integer values and our final goal is to find the allocated quantities for each supplier for each line item, the problem is pure integer linear programming (ILP) which is actually an np-hard problem. The optimum bid evaluation result will be how the buyer will allocate the contracts to the supplier, which means what quantity of each line item should be allocated to the suppliers.

Outcome result will be

$$\begin{gathered} L_1 = \{\, QL_1S_1 \; QL_1S_2 \; QL_1S_3 \; \ldots\ldots . \; QL_1S_s \,\} \\ \text{subject to } \textstyle\sum_{i=1}^{i=s} QL_1S_i = QL_1 \\ L_2 = \{\, QL_2S_1 \; QL_2S_2 \; QL_2S_3 \; \ldots\ldots . \; QL_2S_s \,\} \\ \text{subject to } \textstyle\sum_{i=1}^{i=s} QL_2S_i = QL_2 \\ \ldots\ldots\ldots\ldots\ldots \\ L_n = \{\, QL_nS_1 \; QL_nS_2 \; QL_nS_3 \; \ldots\ldots . \; QL_nS_s \,\} \\ \text{subject to } \textstyle\sum_{i=1}^{i=s} QL_nS_i = QL_n \end{gathered}$$

The main objective is to minimize the cost for all the line items subject to all the constraints.

$$Minimize \sum_{i=1}^{i=n} \sum_{j=1}^{j=s} (QL_iS_j * L_iS_j)$$

Buyer can define based on what conditions they want to evaluate the best bid, that may be based on supplier weightage (some suppliers provide good quality products compared to other suppliers) or they may want to distribute the allocation through 3–4 suppliers rather than going with only one supplier, etc. We can represent those conditions as constraints.

Constraint based on the quantity is $min(QL_1) \leq QL_1S_i \leq max(QL_1)$ and $\sum_{i=1}^{i=s} QL_1S_i = QL_1$ for line item L_1.

This upper bound and lower bound of the quantity can be changed based on the buyer requirements. The default value for $min(QL_1) = 0\, and\, max(QL_1) = QL_1$. If the buyer wants to provide the allocation such that all the suppliers should allocate minimum 5 quantities and at most 20 quantities, then the values will be $min(QL_1) = 5\, and\, max(QL_1) = 20$, but $\sum_{i=1}^{i=s} QL_1S_i = QL_1$ should always satisfy.

3 Shuffled Frog Leaping Algorithm (SFLA)

The concept of Shuffled Frog Leaping Algorithm (SFLA) was coined by Muzaffar Eusuff and Kevin Lansey [7–9]. SFLA is a memetic evolutionary algorithm (EA), which involves memes that carries important information from one generation to another. It is inspired by the natural evolution of frogs. After a certain number of memetic evolution steps, all the frogs are mixed again and shuffled among the memeplexes and we are performing a global search procedure to improve the solution outcome [11, 12]. Local search and shuffling process continue until all the global iterations are completed. The detailed description of the algorithm is mentioned below.

Step 1: improve only the frog with the worst fitness in each local iteration. The improvement function of the frog with the worst fitness (F_w) is

$$D_w = rand() * (F_b - F_w)$$

where D_w is the alteration in the frog position, F_b is the best frog and F_w is the worst frog. *rand() is the* random value generator between 0 and 1

$$new\ F_w = F_w + D_w$$

Step 2: the above process yields an improved solution, which replaces the worst frog.

Step 3: perform the local search procedure for all the memeplexes.

Step 4: if local searches are completed then combine all the frogs and shuffle between the memeplexes.

Step 5: one global operation consists of completing all the local iterations within each memeplex.

Step 6: either get the best solution (best fitness value) from F, if all the given global iterations are finished or shuffle the frogs, go to step 4 and perform the steps accordingly.

Step 7: end of the algorithm.

4 Proposed Modified SFLA to Solve ILP Problem

In this paper, we have made some modification in the global search procedure of SFLA and implemented it to deal with the ILP problem to find the optimum bid allocation to suppliers. The proposed solution will solve this problem for each line item one by one, and finally sum it for all the line items.

4.1 Representation of a Frog

Shuffled frog leaping algorithm is a kind of hybrid genetic algorithm (GA) where instead of genes we are working on memes to pass the information. Here, in the algorithm, each frog is referred to as a solution vector (like memotypes in terms of memetic evolution) and the population of frogs means different solution vectors with different dimensions. Hence, in the formulated ILP problem, each frog (F) is a possible solution of allocated quantity for each supplier participated in the event, and each dimension or meme of a frog (M) represents the quantity allocated to a single supplier (S). The dimensions of the solution vector (length of the memotype) equal M with index range from *1 to s* considering there are s suppliers present and the population size is equal to p (number of frogs or solution vectors) with index range from *1 to p*. We have generated the population in such a way that each frog (F) must satisfy the condition: $\sum_{i=1}^{i=s} MS_i = QL_1$ where QL_1 is the total quantity for line item L_1

$$\begin{aligned} L_1F_1 &= \left\{ MS_1 \; MS_2 \; MS_3 \; \ldots\ldots \; MS_s \right\} \\ L_1F_2 &= \left\{ MS_1 \; MS_2 \; MS_3 \; \ldots\ldots \; MS_s \right\} \\ &\quad \ldots\ldots. \\ L_1F_n &= \left\{ MS_1 \; MS_2 \; MS_3 \; \ldots\ldots \; MS_s \right\} \end{aligned}$$

4.2 Fitness Function

Our main purpose is to minimize the total cost considering all the buyer requirements. We can represent the fitness function as $\sum_{i=1}^{i=s} (MS_i * L_1S_i)$ where L_1S_i is the bid price provided by the supplier S_i for line item L_1.

4.3 Modified Local Iteration Process

Each individual frog represents a possible solution to the bid evaluation problem. We are dividing the whole population of frogs into m number of memeplexes, so that each memeplex has (*p/m*) number of frogs. In SFLA local search is completed with improving the fitness value of the worst frog, F_w within default local iteration for each memeplex.

$$F_w : L_1 F_w = \left\{ MS_1\ MS_2\ MS_3 \ldots\ldots MS_s \right\}$$
$$F_b : L_1 F_b = \left\{ MS_1\ MS_2\ MS_3 \ldots\ldots MS_s \right\}$$

We will randomly select one meme (MS_x) out of all the s memes present in F_w to perform the improvement in the fitness value. Suppose, we have selected MS_x meme. Next step will be finding the positive or negative difference (ΔMS_x) between the MS_x value of F_w and F_b and we will improve it using the following way:

$$\Delta MS_x = ceiling(rand() * abs(MS_x(F_b) - MS_x(F_w)))$$
$$new\ MS_x = MS_x \pm \Delta MS_x$$

where *rand()* will be any value between 0 and 1. As we have to maintain the overall quantity constraint, we will add or subtract the delta changes with MS_x meme with the F_w and accordingly we will have to perform the opposite operation on any of the other randomly selected memes of F_w.

4.4 Modified Global Iteration Process

We will perform the above steps for all the memeplexes and then merge all the populations together in one global iteration. As the local iteration process of the algorithm only improves the performance of the worst frog, F_w by determining how worse it is from the best frog, F_b, these conditions may lead to the local enslavement and at the same time, affect the convergence of SFLA [13]. For removing these local dependencies, we propose another modification to the algorithm.

Initially after getting all the supplier bid data, we will find the mean value (μ) of the bid from all those data and save it for further operation.

$$\mu = \sum_{j=1}^{j=s} L_i S_j / s \quad \text{for line item } L_i$$

1. After the global iteration completes, we will get the optimum solution frog $\left(F_{\mathrm{bg}}\right)$ and select one meme or quantity variable $\left(MS_{\mathrm{xg}}\right)$ randomly.
2. Get the supplier bid value for that line item ($L_1 S_x$), and check whether the value is less than or greater than the mean bid value (μ).

3. If the supplier bid value is less than the mean value then, find another quantity variable $\left(MS_{\mathrm{yg}}\right)$ from that frog for which the supplier bid value is higher than supplier bid value $\left(L_1S_y\right)$.
4. Now interchange the quantities between them, i.e., supplier with lower bid value should get the maximum number of quantities allocated and those extra allocated quantities should come from the supplier with the higher bid value.

$$\begin{gathered}\text{If } L_1S_x \leq \mu \text{ and } L_1S_y > L_1S_x \text{ then,}\\ \Delta q = MS_{yg} - min(QL_1)\\ new\ MS_{xg} = MS_{xg} + \Delta q\\ new\ MS_{yg} = min(QL_1)\end{gathered}$$

5 Evaluation of the Proposed MSFLA Algorithm

5.1 Simulation Setup

There are three main parameters that affect the convergence speed and the quality of the solution obtained from the MSFLA algorithm:

1. Population Size (F)
2. Number of Memeplexes (m)
3. Number of Global iterations (g)

Based on the observations we have set, the parameters and the proposed algorithm are tested using different bidding data with an increasing level of complexity in buyer conditions and for all the cases, we have received a satisfactory solution.

5.2 Experimental Results and Analysis

After running the algorithm multiple times with the parameters mentioned in Table 1, we come to a conclusion that for the proposed MSFLA algorithm, increasing the population size in accordance with the supplier range and number of line items

Table 1 Parameter setup for the algorithm

Parameter name	Value
Population size (F)	500
Number of memeplexes (m)	50
Number of global iterations (g)	20

considerably improves the optimal outcome and stabilize the solution. But also, in case of huge bidding data, we have to maintain a suitable population size, otherwise it will impact the performance. The final experimental result is given below:

1. Population size or number of the frogs (F) should be given according to the number of suppliers participating. Otherwise, it might iterate for a long time without any significant improvement. If the supplier is less, we can provide a minimum population size to get the result.
2. After the modification of the existing SFLA algorithm, it provides optimum solutions more accurately than the original algorithm. The local search is basically based on the fitness value, i.e., we are improving the outcome based on minimum cost function validating all the buyer constraints. The global search is based on individual supplier bid data, and it is comparing between two suppliers and how they have placed the bid for the same line item. Ultimately, for a local iteration, we are considering optimizing the whole cost, whereas in global iteration we are analyzing each supplier bid data to get the best outcome.

6 Conclusion

In this paper, we have discussed the advantages and effectiveness of the Modified Shuffled Frog Leaping Algorithm (MSFLA) to solve the bid evaluation problem after converting the problem into pure integer linear programming (ILP). Modifying the traditional shuffled frog leaping algorithm by adding additional attributes in the global iteration process provides a much better outcome compared to other algorithms. The proposed modified algorithm is also verified with different supplier bid data and for different parameter setups. Experimental results exhibit consistency and performance efficiency for the custom scenarios, proving the optimized solution.

References

1. Hasker K, Sickles R (2010) eBay in the economic literature: analysis of an auction marketplace. Rev Ind Org 37:3–42
2. Jank W, Zhang S (2011) An automated and data-driven bidding strategy for online auctions. INFORMS J. Comput. 23:238–253
3. Liu SL, Lai KK, Wang SY (2000) Multiple criteria models for evaluation of competitive bidss. IMA J Manag Math 11(3):151–160
4. Ghasabi-Oskoei H, Mahdavi-Amiri N (2006) An efficient simplified neural network for solving linear and quadratic programming problems. Appl Math Comput J 452–464. Elsevier
5. L.R. Arvind Babu and B. Palaniappan, "Artificial Neural Network Based Hybrid Algorithmic Structure for Solving Linear Programming Problems", International Journal of Computer and Electrical Engineering, Vol. 2, No. 4, August 2010, pp 1793-8163
6. Nasira GM, Ashok Kumar S, Balaji TSS Neural network implementation for integer linear programming problem. Int J Comput Appl 1(18):0975–8887
7. Eusuff MM, Lansey K, Pasha F (2006) Shuffled frog-leaping algorithm: a memetic metaheuristic for discrete optimization. Eng Optim 38(2):129–154

8. Zhang X, Hu X, Cui G, Wang Y, Niu Y (2008) An improved shuffled frog leaping algorithm with cognitive behavior. In: Proceedings of 7th world congress intelligent control and automation 2008
9. Elbeltagi E, Hegazy T, Grierson D (2005) Comparison among five evolutionary based optimization algorithms. Adv Eng Inf 19(1):43–53
10. Kennedy J, Eberhart RC (1995) Particle swarm optimization. In: Proceedings of IEEE conference on neural networks, vol 4, pp 1942–1948
11. Karthiban MK, Raj JS (2019) Big data analytics for developing secure internet of everything. J ISMAC 1(02):129–136
12. Roy P (2014) A new memetic algorithm with GA crossover technique to solve Single Source Shortest Path (SSSP) problem. In: INDICON 2014
13. Roy P, Roy P, Chakrabarti A Solving network-constrained non smooth economic dispatch problems through a gradient-based approach. Appl Soft Comput 13(11):4244–4252. Elsevier
14. Wang Z, Zhang D, Wang B, Chen W (2019) Research on improved strategy of shuffled frog leaping algorithm. In: 2019 IEEE 34rd Youth academic annual conference of Chinese association of automation (YAC)

Tunable Access Control for Data Sharing in Cloud

S. Sabitha and M. S. Rajasree

Abstract Attribute-Based Encryption (ABE) suffers communication and computation overhead due to the linearly varying size of the ciphertext and the secret key, depending on the number of attributes in the access policy. This paper proposes a multilevel attribute-based access control scheme for secure data sharing in the cloud to reduce the overhead. It produces a constant size ciphertext and a compact secret key to efficiently utilize the storage space and reduce the communication cost. This method flexibly shares ciphertext classes among the randomly selected users with a specific set of attributes. All other ciphertext classes outside the set remain confidential. It allows dynamic data updates and provides access control of varying granularity, at user-level, at file-level, and attribute-level. Granularity levels can be chosen based on applications and user demands. This scheme tackles user revocation and attribute revocation problems, and prevents forward and backward secrecy issues. It allows the data owner to revoke a specific user or a group of users. It is very useful for secure data storage and sharing.

Keywords Cloud computing · Attribute-based encryption · Access control · Key-aggregate cryptosystem · Data sharing

1 Introduction

Cloud computing is a modern computing paradigm, which supplies resources as services over the Internet. It allows the use of computing resources as of low cost and on-demand basis. With the rapid growth of cloud computing, data outsourcing and sharing have been increasing dramatically. Outsourced data in the cloud face

S. Sabitha (✉)
College of Engineering Trivandrum, Trivandrum, Kerala, India
e-mail: sabitha@cet.ac.in

M. S. Rajasree
APJ Abdul Kalam Technological University, Trivandrum, Kerala, India
e-mail: rajasree40@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_85

many challenges such as security and privacy. When sensitive data and computation get outsourced and shared, it becomes out of control of the data owner. Hence servers or other users might be able to get valuable information from the outsourced data.

Let us consider the scenario for selected data collaboration and sharing: project manager of an IT company encrypts and outsources projects and guidelines to the cloud, then share some selected projects among the development team, testing team, marketing team, and HR team. All the above team members are allowed to decrypt and read the data, but only the selected members of the development team are allowed to modify the project.

In this paper, we propose a ciphertext-policy hierarchical attribute-based access control with constant size ciphertext and constant size compact key for data sharing in the cloud. It can resolve the problem specified in the above scenario. The proposed scheme delegates the decryption rights of the selected set of ciphertext classes among the randomly selected members from the set of users having the same set of attributes and data access privileges. All the members of the development team have the same set of attributes and data manipulation privilege so that ABE scheme allows all the members to modify the data. But for the above scenario, it is required to delegate the decryption rights of the selected set of files to the selected members of the development team. With the help of key-aggregate cryptosystem and hierarchical attribute-based encryption (HABE), the proposed system allows only selected members of the development team to modify the data only if they possess the required compact key and their data manipulation request satisfied with the claim policy of the ciphertext. The scheme verifies user's data manipulation request against the claim policy of the ciphertext to allow them to modify the data using the attribute-based signature(ABS). Claim policy verification method permits the authorized users with data write privilege to modify the ciphertext.

Partial signing and partial decryption verify the signature and decrypt the ciphertext before permit the users to modify the data. It reduces the computation overhead on users by delegating partial computation to the CSP. The proposed scheme generates constant size ciphertext to efficiently utilize the storage space. It generates constant size compact key for the set of secret keys that were used to encrypt the outsourced data. Hence, authorized users who have the corresponding compact key and required attributes would be able to update the ciphertext. User-level, as well as file-level access control, is incorporated along with the attribute-based access control. The user-level access control helps to prevent the revoked user to access data.

This paper is organized as follows. Section 2 describes the work related to the hierarchical attribute-based encryption schemes. System model is discussed in Sect. 3. Section 4 describes the modified CP-HABE scheme for data sharing to enhance performance. Performance analysis is described in Sect. 5. Section 6 summarizes the work.

2 Related Works

Hierarchical attribute-based encryption (HABE) improves scalability and reduces the key management overhead. Wang et al. [1] proposed hierarchical attribute-based encryption in 2011 by combining HIBE and CP-ABE to achieve fine-grained access control, full delegation, and high performance for data sharing. Later in 2012, Wan et al. [2] proposed a scalable and flexible hierarchical attribute-based solution (HASBE) for data sharing in the cloud which, however, could not support write operation on the outsourced ciphertext. Deng et al. [3] proposed a ciphertext-policy hierarchical ABE with short ciphertexts in 2014. In 2015, Teng et al. [4] proposed a hierarchical attribute-based access control scheme with constant size ciphertext. Its hierarchical authorization structure reduces the risk of a single authority in attribute-based access control scheme.

Dong et al. [5] proposed a scalable and secure data collaboration scheme called SECO, using hierarchical identity-based encryption (HIBE) in 2015, in which cloud cannot verify the write permission of the user. Identity-based signature (IBS) is used in this scheme for authenticity verification. Huang et al. [6] proposed a secure data collaboration method using HABE in 2017. Attributes in this scheme are represented as {Assistant Professor, Associate Professor, Professor, Principal}. It reduces decryption and signature generation overhead on the user by partially delegating the decryption and signature generation capability to the CSP without disclosing the original data to the untrusted CSP. The scheme supports write operation on ciphertext by the authorized users with the help of ABS. But, user revocation and attribute revocation problems continue to be unresolved. It does not support user-level and file-level access control. User selection among the authorized recipient group is not permitted in the key-aggregate public key cryptosystem proposed by Chu et al. [7] in 2014. It is a compact secret key scheme for scalable data sharing in the cloud. It aggregates a set of secret keys as a compact single key for delegating the decryption rights of any set of ciphertext classes. However, access control is not based on attributes.

3 System Model

This paper proposes a secure and flexible collaborative data sharing with constant size ciphertext and constant size compact key. It delegates the decryption rights of the selected set of ciphertext classes among the randomly selected users from the authorized users with a specific set of attributes. It provides multilevel, tunable access control over the shared data. The proposed scheme utilizes a hierarchical structure. The system model of the proposed scheme is shown in Fig. 1. It is the modified form of CP-HABE scheme. It consists of root authority, domain authorities, cloud service provider, data owner, and users. The format of the outsourced ciphertext is shown in Fig. 2, where ID is the identification number of a file, DEK is the symmetric

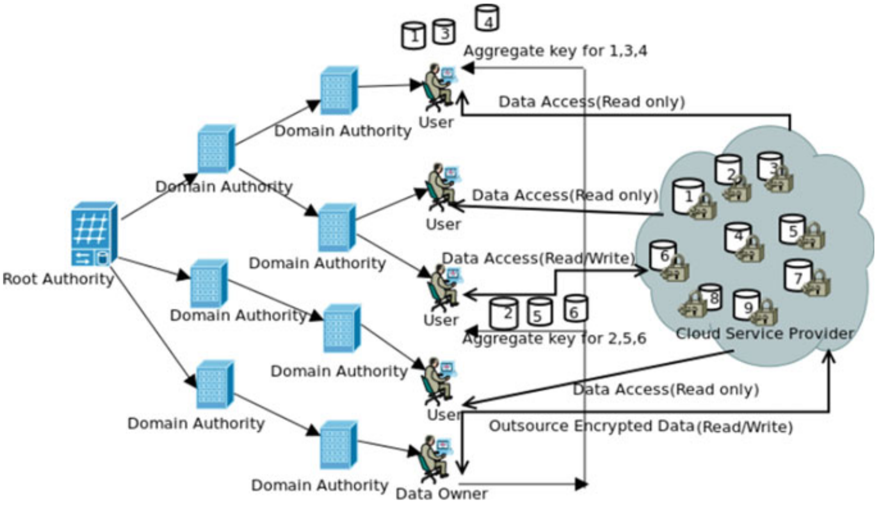


Fig. 1 System model of the modified CP-HABE scheme for data sharing

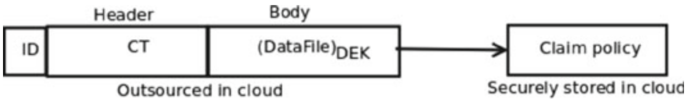


Fig. 2 Outsourced file format

key used to encrypt the data file, CT is the ciphertext of $DEKs$ using ciphertext policy hierarchical attribute-based encryption and key-aggregate encryption. The claim policy is an access structure used to verify the user's data write privilege [8].

4 Construction of the Enhanced CP-HABE Scheme for Data Sharing

Let us consider the scenario of collaboration and sharing of a selected data set: project manager of an IT company encrypts and outsources projects and guidelines to the cloud. Then share some selected projects among the development team, the testing team, the marketing team, and the HR team. All the above team members are allowed to decrypt and read the data, but only selected members of the development team are allowed to modify the project.

The proposed scheme delegates the decryption rights of the selected set of ciphertext classes among the randomly selected members from the set of authorized users having the same set of attributes and data access privileges. All the members of the development team have the same set of attributes and data access privileges such that the ABE scheme allows all the members to modify the data. But for the above scenario, it is required to delegate the decryption rights of the selected set of files to the randomly selected members of the development team. With the help of key-aggregate cryptosystem and CP-HABE, the proposed system allows only selected members of the development team to modify the data. The selected members should possess the required compact key and, their data modification request should be satisfied with the claim policy of the ciphertext. In this scheme, verification of the user's data modification request is done using an attribute-based signature (ABS).

The hierarchical structure of the proposed scheme distributes the key management burden and risk of central authority to the domain authorities. The proposed method is an enhanced ciphertext-policy hierarchical attribute-based encryption (CP-HABE) scheme with constant size ciphertext and constant size compact key. It delegates the decryption rights of the selected set of ciphertext classes among the randomly selected users. Partial decryption and partial signing enhance the performance of the scheme. User revocation and attribute revocation are efficiently managed.

The framework of the modified CP-HABE scheme is defined here. It creates a constant size ciphertext and compact key. User-level, file-level, and attribute-level access controls are incorporated in the new scheme. Since the security and access control level can be enlarged or minimized depending on the requirements, it provides multilevel tunable access control over the shared data. The proposed scheme maintains forward and backward secrecy. Eight phases of operations are involved in this scheme:

Phase1(System setup): This phase contains a setup algorithm.

- $Setup(1^d) \rightarrow PK, MK_0$: The root authority executes the *Setup* algorithm. It uses security parameter d to generate a public key, PK and a master key, MK_0

Phase2(Domain setup): It contains 2 algorithms: *CreateDA* and *Agr-Setup*.

- *CreateDA*$(PK, MK_0, A) \rightarrow MK_l$: The root authority runs *CreateDA* algorithm using the public key PK, its master key MK_0 and a set of attributes A to generate the master key MK_l for the next level domain authority.
- *Agr-Setup*$(1^\lambda, n) \rightarrow param$: The lower level domain authority who manages the user is responsible for executing *Agr-Setup* algorithm. It takes security parameter λ and a number of ciphertext classes n, to generate the system parameter *param*.

Phase3(Key generation): *KeyGen, Agr-KeyGen, and Agr-Key* are the 3 algorithms in this phase.

- $KeyGen(PK, MK_l, S) \rightarrow SK$: The lower level domain authority which manages the user executes the algorithm. It uses the public key PK, the master key MK_l, and the user's attribute set S to compute the secret key SK.
- *Agr-KeyGen()*→ *AMSK, APK*: The data owner executes the algorithm to generate an aggregate master key $AMSK$ and an aggregate public key APK.
- *Agr-Key(AMSK, P)*→ K_{agr}: The data owner runs this algorithm to create an aggregate key K_{agr}, which delegates the decryption power of a set of ciphertext class indices P.

Phase4(Encryption): This phase contains *Encrypt* algorithm:

- *Encrypt(PK, DEK, APK, i, τ)* $\rightarrow CT$: The algorithm uses the public key PK, the symmetric key DEK which was used to encrypt the shared data, the aggregate public key APK, access policy τ, and index i of the ciphertext class to generate the ciphertext CT. The data owner encrypts DEK to generate the ciphertext CT.

Phase5(Data re-encryption): This phase consists of 3 algorithms: *KEKGEN, Re-encrypt, and HdrMsg*.

- *KEKGEN()* $\rightarrow$ *KEK tree*: CSP runs the algorithm to create KEK tree for the universe of users.
- *Re-encrypt* $(CT, G_i) \rightarrow CT'$: CSP runs the algorithm to re-encrypt the ciphertext CT. It takes ciphertext CT and group G_i to create a re-encrypted ciphertext CT'.
- *HdrMsg()* $\rightarrow$ *Hdr:* CSP runs this algorithm to deliver attribute group keys to valid users through Hdr.

Phase6(Decryption): This phase contains 2 algorithms: ParDec and Decrypt.

- *ParDec(CT', PAK)* $\rightarrow CT_p$: CSP runs the algorithm to generate partial ciphertext CT_p. It takes re-encrypted ciphertext CT' and partial attribute key PAK.
- *Decrypt* $(CT_p, SK', K_{agr}, S, i) \rightarrow DEK$: User executes this algorithm to decrypt the partial ciphertext CT_p. The algorithm takes the user's updated secret key SK', aggregate key K_{agr}, set of ciphertext classes, S and index i of the ciphertext class for the decryption.

Phase7(Writing data): It contains 3 algorithms: *ParSign, Sign, and Verify*

- *ParSign(PAK)* $\rightarrow PST$: CSP runs the algorithm after receiving partial attribute key PAK from the user and outputs a partial signature, PST.
- *Sign(PST, SK)* $\rightarrow ST$: The user runs the algorithm to create a signature, ST by using partial signature PST from CSP and his own secret key SK.
- *Verify(ST, GAK)* $\rightarrow T/F$: CSP executes the algorithm to allow the requested user to update the outsourced data. On successful verification of the signature, the user is allowed to update the data. This algorithm takes the signature ST and a global attribute key GAK for verification. It returns "true/false" depending on the verification of the signature.

Phase8(Ciphertext and key update): This phase contains 2 algorithms: *CipherUpdate and KeyUpdate* algorithm.

- *CipherUpdate(CT)* $\rightarrow CT'$: CSP executes the algorithm using the ciphertext CT to create an updated ciphertext CT'.
- *KeyUpdate(SK, Hdr)* $\rightarrow SK'$: User runs the algorithm using the secret key SK and header message Hdr to create an updated key SK'.

4.1 Correctness

Theorem 1 *Our access control scheme is correct. Users can decrypt the ciphertext to get back DEK only if their attributes satisfy the access structure in the ciphertext and the ciphertext classes belong to the aggregate key.*

Proof If and only if user's attributes satisfy the access structure then, $e(g, g)^{(r_l+\delta)\beta\psi q_y(0)} = e(g, g)^{(r_l+\delta)\beta\psi s}$. The correctness is as follows:

$$
\begin{aligned}
&C \cdot \frac{e(K_{agr} \cdot \prod_{j\in p, j\neq i} g_{n+1-j+i}, C_1)}{e(\prod_{j\in p} g_{n+1-j}, C_2)} \\
&= C \cdot \frac{e(\prod_{j\in p} g^{\varepsilon}_{n+1-j} \cdot \prod_{j\in p, j\neq i} g_{n+1-j+i}, g^t)}{e(\prod_{j\in p} g_{n+1-j}, (g^{\varepsilon} \cdot g_i)^t)} \\
&= C \cdot \frac{e(\prod_{j\in p} g^{\varepsilon}_{n+1-j}, g^t) e(\prod_{j\in p, j\neq i} g_{n+1-j+i}, g^t)}{e(\prod_{j\in p} g_{n+1-j}, g^{\varepsilon t}) e(\prod_{j\in p} g_{n+1-j}, g^t_i)} \\
&= C \cdot \frac{e(\prod_{j\in p} g_{n+1-j+i}, g^t)/e(g_{n+1}, g^t)}{e(\prod_{j\in p} g_{n+1-j+i}, g^t)} \\
&= DEK \cdot \frac{e(g, g)^{\alpha\beta s} \cdot e(g_1, g_n)^t}{e(g_{n+1}, g^t)} \\
&= DEK \cdot e(g, g)^{\alpha\beta s} \cdot e(g_{n+1}, g^t)/e(g_{n+1}, g^t) \\
&= DEK \cdot e(g, g)^{\alpha\beta s}
\end{aligned}
$$

$$
\begin{aligned}
e(\widetilde{C}, \widetilde{D})/(A)^{1/\psi} &= \frac{e(g^s, g^{(\alpha+r_l+\delta)\beta})}{e(g, g)^{((r_l+\delta)\beta\psi s)^{1/\psi}}} \\
&= \frac{e(g^s, g^{(\alpha+r_l+\delta)\beta})}{e(g, g)^{(r_l+\delta)\beta s}} \\
&= \frac{e(g^s, g^{\alpha\beta} \cdot g^{(r_l+\delta)\beta})}{e(g, g)^{(r_l+\delta)\beta s}} \\
&= \frac{e(g^s, g^{\alpha\beta}) \cdot e(g^s, g^{(r_l+\delta)\beta})}{e(g, g)^{(r_l+\delta)\beta s}} \\
&= \frac{e(g, g)^{\alpha\beta s} \cdot e(g, g)^{(r_l+\delta)\beta s}}{e(g, g)^{(r_l+\delta)\beta s}} \\
&= e(g, g)^{\alpha\beta s}
\end{aligned}
$$

$$
Then \quad \frac{C \cdot e(K_{agr} \cdot \prod_{j\in p, j\neq i} g_{n+1-j+i}, C_1)/e(\prod_{j\in p} g_{n+1-j}, C_2)}{e(\widetilde{C}, \widetilde{D})/(A)^{1/\psi}}
$$

$$
= DEK \cdot e(g, g)^{\alpha\beta s}/e(g, g)^{\alpha\beta s} = DEK
$$

Theorem 2 *User's collaboration request is permitted only if their attribute-based signature satisfies the claim policy and ciphertext class indices belong to the aggregate key.*

Proof If and only if user's ABS satisfies the claim policy, then $e(g, g)^{(r_l+\delta)\beta P_y(0)} = e(g, g)^{(r_l+\delta)\beta t}$. Only if authorized user's aggregate key is satisfied with the ciphertext class indices, then correctness is as follows:

$$\begin{aligned}
&\frac{e(g, S') \cdot e(K_{agr}, g)}{e(H(R), S) \cdot (I)^{1/t} \cdot e(\prod_{j \in p} g_{n+1-j}, g^{\varepsilon})} \\
&= \frac{e(g, H(R)^{\mu} \cdot g^{(\alpha+r_l+\delta)\beta}) \cdot e(\prod_{j \in p} g^{\varepsilon}_{n+1-j}, g)}{e(H(R), g^{\mu}) \cdot e(g, g)^{(r_l+\delta)\beta t/t} \cdot e(\prod_{j \in p} g_{n+1-j}, g^{\varepsilon})} \\
&= \frac{e(g, H(R)^{\mu} \cdot g^{(\alpha+r_l+\delta)\beta}) \cdot e(\prod_{j \in p} g^{\varepsilon}_{n+1-j}, g)}{e(H(R), g^{\mu}) \cdot e(g, g)^{(r_l+\delta)\beta t/t} \cdot e(\prod_{j \in p} g^{\varepsilon}_{n+1-j}, g)} \\
&= \frac{e(g, H(R)^{\mu} \cdot g^{(\alpha+r_l+\delta)\beta})}{e(H(R), g^{\mu}) \cdot e(g, g)^{(r_l+\delta)\beta}} \\
&= \frac{e(g, H(R)^{\mu}) \cdot e(g, g^{(\alpha+r_l+\delta)\beta})}{e(H(R), g^{\mu}) \cdot e(g, g)^{(r_l+\delta)\beta}} \\
&= \frac{e(g^{\mu}, H(R)) \cdot e(g, g)^{\alpha\beta} \cdot e(g, g)^{(r_l+\delta)\beta}}{e(H(R), g^{\mu}) \cdot e(g, g)^{(r_l+\delta)\beta}} \\
&= e(g, g)^{\alpha\beta}
\end{aligned}$$

5 Performance Analysis

5.1 Comparison

The proposed scheme is compared with various data sharing and collaboration schemes as shown in Table 1. Comparisons are in terms of access control, data collaboration, Write/Read access, full delegation, partial decryption, partial signing, signature verification, signature scheme, user revocation, attribute revocation, ciphertext size, and selective data sharing. All the schemes are based on attribute-based encryption, except the scheme by Chu et al. [7], which is based on key-aggregate cryptosystem. In Table 1, "M-W-M-R" represents multiple write and multiple read operation while "1-W-M-R" represents single write and multiple read operation.

5.2 Experimental Evaluation

The proposed scheme is implemented on a 2.2GHz Intel Core-i7 processor with 8GB of RAM running 64-bit Linux Kernel version 3.8.0. The algorithms are implemented using a CP-ABE toolkit [15] and a pairing-based cryptographic (PBC) library [16]. The experiments are performed by varying the number of attributes in the secret

Table 1 Comparison with other related schemes

Schemes	Huang [6]	Hur [9]	Hur [10]	Ruj [11]	Li [12]	Jahid [13]	Teng [14]	Chu [7]	[Proposed scheme]
Access control	HABE	CP-ABE	ABE	DABE	ABE	CP-ABE	HABE	Aggregate key	HABE
Data collaboration	Yes	No	No	Yes	Yes	No	No	No	Yes
Write/read access	M-W-M-R	1-W-M-R	1-W-M-R	M-W-M-R	M-W-M-R	1-W-M-R	1-W-M-R	1-W-M-R	M-W-M-R
Full delegation	Yes	No	No	No	No	No	Yes	Yes	Yes
Partial decryption	Yes	No	Yes	No	No	No	No	No	Yes
Partial signing	Yes	No	No	No	No	No	No	No	Yes
Signature verification	Yes	No	No	Yes	No	No	No	No	Yes
Signature scheme	ABS	No	No	ABS	No	No	No	No	ABS
User revocation	No	Yes	No	Yes	Yes	Yes	No	No	Yes
Attribute revocation	No	Yes	No	No	No	Yes	No	No	Yes
Constant size Ciphertext	No	No	No	No	No	No	Yes	Yes	Yes
Selective data sharing	No	No	No	No	No	No	No	No	Yes

key and the access policy from 10 to 100. The worst-case scenario of the proposed scheme is analyzed. The AND gate is the conjuncture for the worst case. Most of the computationally expensive operations in the proposed scheme are delegated to the CSP without disclosing original data. The proposed scheme is compared with Huang et al. [6] and Bethencourt et al. [17]. The decryption time on the user is negligibly small in the proposed scheme. The decryption time on the CSP against the number of attributes in the secret key of the proposed scheme is same as that of Huang et al. [6]. From the experiments, it is clear that our scheme can be used for resource-constrained devices.

6 Conclusion

This paper proposes a multilevel tunable access control mechanism for data sharing in the cloud. It is a flexible and on-demand access control mechanism. The proposed scheme creates a constant size ciphertext and a constant size compact key to improve the performance of the system. The scheme adopts a hierarchical attribute authority structure that reduces the key management overhead by distributing the overhead among the domain authorities. The proposed work supports user-level, file-level, and attribute-level access control. It gives multilevel, tunable security over the shared data. It also offers the user revocation and attribute revocation capability. Partial decryption and partial signing reduce the computational overhead on the user. The system has been experimentally analyzed and its performance against other related works has been compared in the best-case and the worst-case scenarios. The security analysis of the proposed system has been carried out and it has been concluded that the shared data is fully collusion-secure and confidential. Comparison and analysis show that the scheme is efficient and enables secure, selective data sharing in the cloud. It can be utilized in resource-constrained mobile devices.

References

1. Wang Guojun, Liu Qin, Jie Wu, Guo Minyi (2011) Hierarchical attribute-based encryption and scalable user revocation for sharing data in cloud servers. Comput Secur (Elsevier) 30:320–331
2. Wan Z, Liu J, Deng RH (2012) HASBE: A hierarchical attribute-based solution for flexible and scalable access control in cloud computing. IEEE Trans Inf Forensics Secur 7(2):743–754
3. Deng H, Wu Q, Qin B, Domingo-Ferrer J, Zhang L, Liu J, Shi W (2014) Ciphertext-policy hierarchical attribute-based encryption with short ciphertexts. Inform Sci 275:370–384
4. Teng W, Yang G, Xiang Y, Zhang T, Wang D (2016) Attribute-based access control with constant-size ciphertext in cloud computing. IEEE Trans Cloud Comput. (99):1
5. Dong X, Yu J, Zhu Y, Chen Y, Luo Y, Li M (2015) Seco: secure and scalable data collaboration services in cloud computing. Comput Secur 50:91–105
6. Huang Q, Yang Y, Shen M (2017) Secure and efficient data collaboration with hierarchical attribute-based encryption in cloud computing. Future Gener Comput Syst 72:239–249

7. Chu C-K, Chow SSM, Tzeng W-G, Zhou J, Deng RH (2014) Key-aggregate cryptosystem for scalable data sharing in cloud storage. IEEE Trans Parallel Distrib Syst 25(2):468–477
8. Zuo C, Shao J, Liu JK, Wei G, Ling Y (2018) Fine-grained two-factor protection mechanism for data sharing in cloud storage. IEEE Trans Inf Forensics Secur 13(1):186–196
9. Hur J, Noh DK (2011) Attribute-based access control with efficient revocation in data outsourcing systems. IEEE Trans Parallel Distrib Syst 22(7):1214–1221
10. Hur J (2013) Attribute-based secure data sharing with hidden policies in smart grid. IEEE Trans Parallel Distrib Syst 24(11):2171–2180
11. Sushmita R, Milos S, Amiya N (2014) Decentralized access control with anonymous authentication of data stored in clouds. IEEE Trans Parallel Distrib Syst 25(2):384–394
12. Ming L, Yu S, Yao Z, Kui R, Wenjing L (2013) Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption. IEEE Trans Parallel Distrib Syst 24(1):131–143
13. Jahid S, Mittal P, Borisov N (2011) Easier: encryption-based access control in social networks with efficient revocation. In: Proceedings of the 6th ACM symposium on information, computer and communications security (ASIACCS'11), pp 411–415
14. Wei T, Geng Y, Yang X, Ting Z, Dongyang W (2015) Attribute-based access control with constant-size ciphertext in cloud computing. IEEE Trans Cloud Comput 99:1–11
15. Ciphertext-policy attribute-based encryption toolkit (2014)
16. Lynn B (2012) The pairing-based cryptography (PBC) library
17. Bethencourt J, Sahai A, Waters B (2007) Ciphertext-policy attribute-based encryption. In: Proceedings of the IEEE symposium on security and privacy, pp 321–334

Methodology for Implementation of Building Management System Using IoT

Ankita Harkare, Vasudha Potdar, Abhishek Mishra, Akshay Kekre, and Hitesh Harkare

Abstract In recent times the need for automation has grown many folds. The need of the hour is to control and display the working of systems utilized in everyday life with a click of a button. This paper presents the designing and implementation of the parking management system (PMS) as a part of a building management system (BMS). Also, the paper suggests the methodology for the effective working of the building management system. The implementation of the communication protocol will be both wired and wireless for effective working. The user will be able to see the display of the status of occupancy immediately as the data will be updated through the Internet periodically. The objective is to reduce the search time required to search for any vacant slot for parking and indicating it, using red or green LED indicators and thus sending all the real-time information to a master server for monitoring purposes or to a display at the entrance. The ultrasonic parking system has been successfully checked and its design has been verified which successfully detects the presence or absence of vehicles under it and indicates it by blinking red/green LED, respectively. PMS is planned and strategised for interconnection using internet/intranet connectivity.

Keywords Building management system · Parking management system · Internet · Automation · Ultrasonic

A. Harkare (✉) · V. Potdar · A. Mishra
Department of Electronics and Communication Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur 440010, India
e-mail: harkareah@rknec.edu

A. Kekre
Techwalnut Innovations LLP, Nagpur, India
e-mail: akshay@techwalnut.com

H. Harkare
Technodeal Enerpower Pvt. Ltd,, Nagpur, India
e-mail: harkare.hr@technodeal.co.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_86

1 Introduction

The management of larger building projects creates a lot of pressure on the owner and requires manpower to manage the same [1]. The important management fields in BMS include light control, water management and parking management. Parking has become one of the major problems in countries like India and finding a place for the same is another major issue. The parking management practices currently in use are inefficient and have proven unsuccessful in easing out the parking process [2]. Drivers have to spend most of their time in searching for the vacant parking slots in parking that are almost full of vehicles. The actual problem lies not with the inadequate supply but with ineffective management. Similarly, electricity and water management are major concerns in larger areas where wastage of precious resources is increasing and left unattended. The methodology for lighting control and water management is an important part of the building management system and is discussed in detail. This paper is further divided into three sections. Section 2 discusses the methodology adopted to implement the BMS effectively. Section 3 discusses the wired model for small spaces and Sect. 4 discusses the wireless model for larger spaces and uploading the desired status using the Internet of Things.

2 Methodology Adopted for BMS

The entire system indicates various sections of a building that can be automated under the building management system. The methodology adopted includes Parking system, Water management and Leakage control, Automatic light control, Safety and Security. Under these heads, the applications taken into account under each section include occupancy or vacancy determination of all the parking slots in the parking area of a building and hence display the status at the entrance via IoT [1]. For wireless light control: lights of a room can be automatically controlled wirelessly depending on the presence and absence of a human in that room, thus reducing power consumption. For Water Management and control: water saving can be done by automating plumbing problems including automatic leakage detection and alerting. For Safety and Security: access control uses PIR sensors and alert the concerned personnel regarding any malicious activity.

The designed PMS system applies the Internet of Things as a tool to control, monitor and synchronize various individual systems mentioned under BMS. The master controllers are required in order to collect the data from the corresponding subsystem and send it to the server. The method focuses on the minimization of wiring requirements by using WSN, i.e., Wireless Sensor Networks. Figure 1 explains the block diagram of implementing BMS for a given space. As is explained from Fig. 1, each subsystem of BMS is controlled individually by its own Master Controller the access to which is controlled by the main server. All the subblocks work simultaneously and hence the requirement to have individual master controllers will suffice the

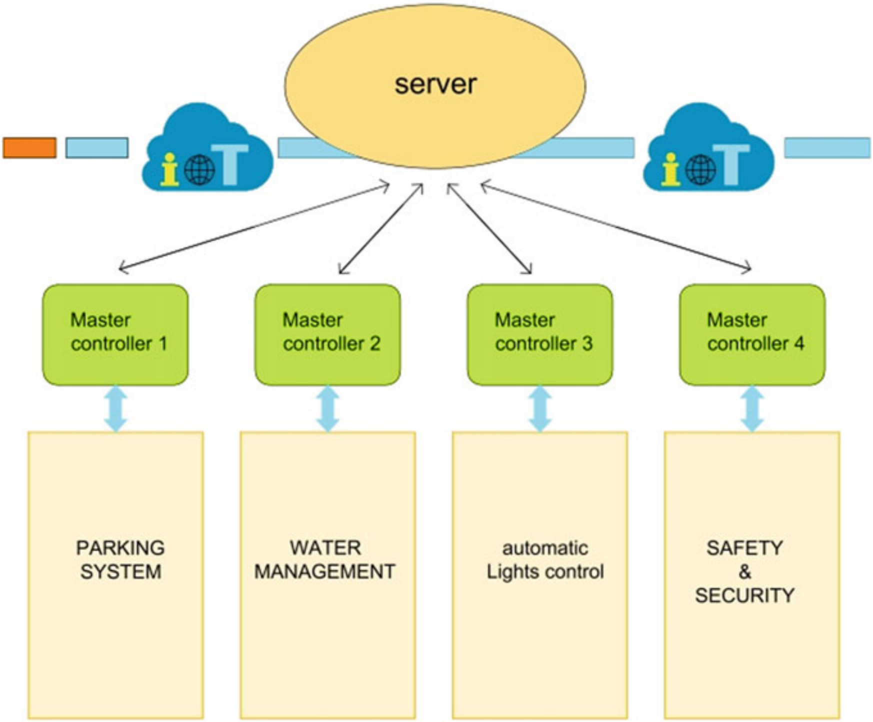


Fig. 1 Buiding management system using IoT [1, 11]

purpose. All the data and status of working of each subblock is periodically sent to the main server which can be accessed by the user at any given time from a distant device with Internet facility [1, 3]. The subblocks will be explained taking into account the parking management system (PMS). The rest of the paper will consider PMS as an example to explain the subblocks' working.

The three main sections of the parking management system comprise i. Slave device with sensor, ii. a Master controller and iii. the Main server. The suggested parking guidance and information system, or car park guidance system, presents drivers with dynamic information on parking within controlled areas [2, 4]. It is a technology that helps to find vacant spaces in the parking area of a building, car location thus enhancing user experience. It includes LED indicators like red to indicate occupied slot and green to indicate vacant parking slot, thus sending all the real-time information to a master server for monitoring purposes. The objective is to reduce the search time required to search for any vacant slot for parking. It is an ultrasonic sensor-based parking system that detects the presence or absence of a vehicle under it depending on the sensor's output. Figure 2 represents the block diagram of the slave device.

The basic subblock of the slave device comprises i. Ultrasonic sensor, ii. Microcontroller unit and iii. Communication unit as shown in Fig. 2. Since ultrasonic sensor

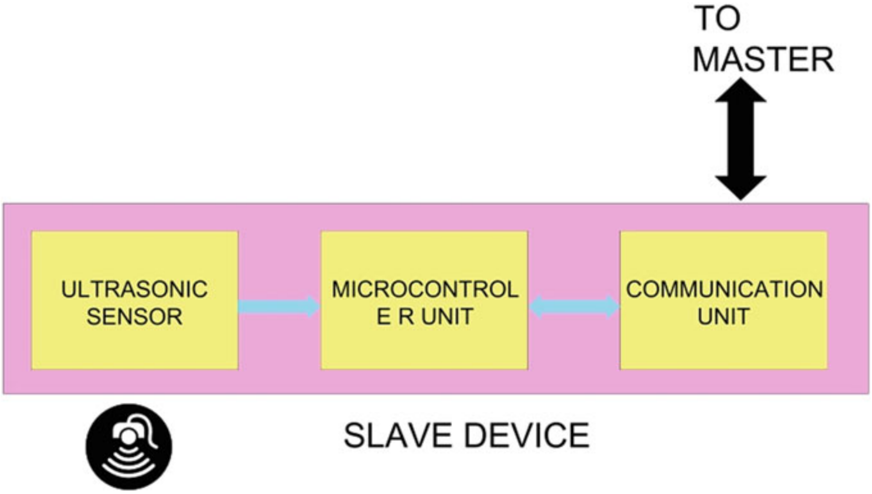


Fig. 2 Slave device

works on the principle similar to radar and sonar waves, it generates high-frequency sound waves continuously. For the presence of any vehicle under the slave device, the sound waves hit the vehicle and return to the sensor. This data is further given to the microcontroller unit which then processes the data and determines that a vehicle is present or not and passes the result to the communication unit. The task of the communication unit is to communicate the obtained data to the master controller. The master controller acts as the heart of the system. Status from all the slave devices in the entire parking area is sent to one master controller which processes the result and then the result of all the occupied and vacant slots is sent to the main server wirelessly, that is, via IoT. The main server contains data of the number of vacant and occupied slots in the entire working system, which can be displayed at the entrance of the parking area of a building. Figure 3 shows the block diagram of master–slave communication with the main server. RS-485 is used as a communication channel for communication between slave devices and the master controller.

2.1 Parking Management System

Figure 4 shows the block diagram of the parking management system. The system is divided into wired and wireless models.

The wired model is constructed using the RS-485 network. It further consists of a model using max-485 and a model using isolated RS-485 which is best suitable for the multistoreyed parking areas. The wireless model is constructed using ESP Wi-Fi devices and is best suitable for small and medium parking areas. Sections 3 and 4 will describe the wired and wireless models in detail [5].

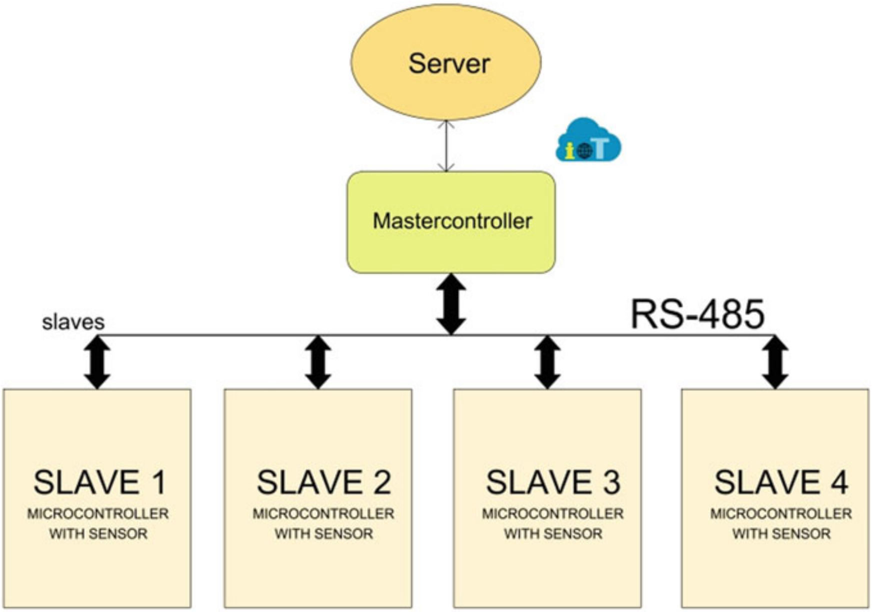

Server
IoT
Mastercontroller
slaves
RS-485
SLAVE 1
MICROCONTROLLER WITH SENSOR
SLAVE 2
MICROCONTROLLER WITH SENSOR
SLAVE 3
MICROCONTROLLER WITH SENSOR
SLAVE 4
MICROCONTROLLER WITH SENSOR

Fig. 3 Master–slave communication with the main server

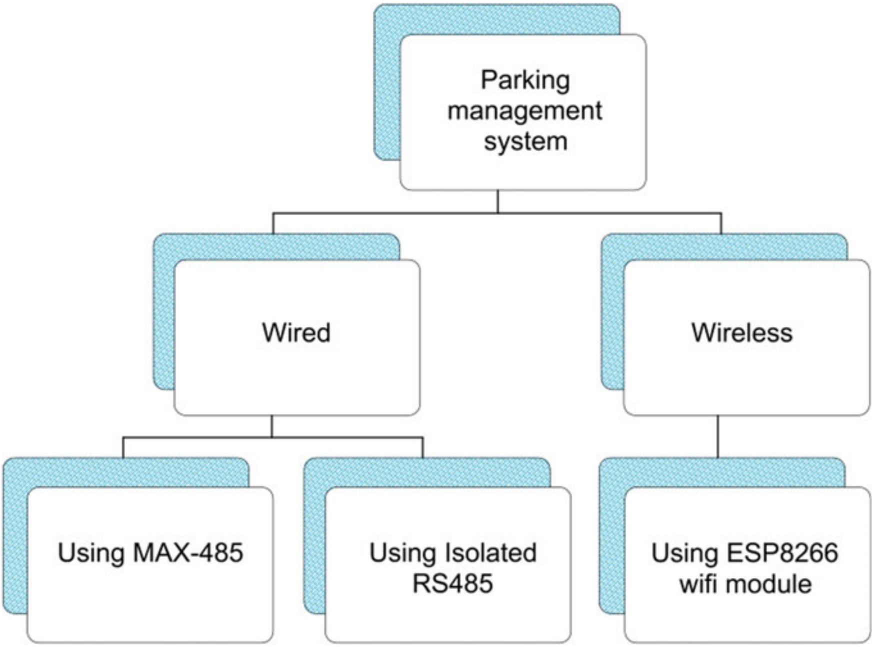


Fig. 4 Parking management system

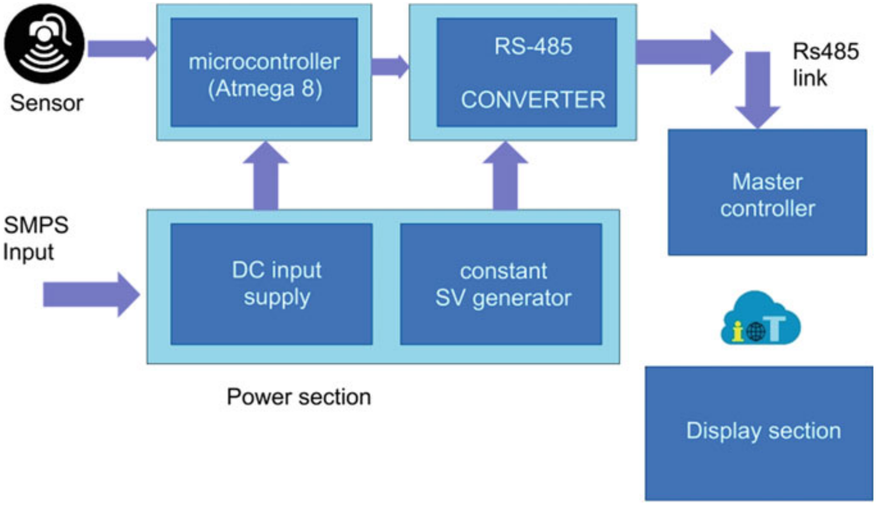


Fig. 5 Wired model of communication

3 Wired Parking Model

The wired parking management system consists of RS-485 as the means of communication used. Multiple slaves communicate with the master using RS-485 communication. Multiple device (up to 32) communication at half-duplex is allowed by RS-485 on a single pair of wires. Data is transmitted differentially (that allows high noise immunity and long-distance communication) on two wires twisted together, referred to as a “twisted pair” [6, 7] (Fig. 5).

4 Wireless Parking Model

It is basically a Wireless Sensor Network which uses Wi-Fi. Each parking location consists of a Wi-Fi module interfaced with an ultrasonic sensor and LED lights. The ultrasonic sensor detects the presence of the vehicle in the parking slot using the principle of distance measurement. If the vehicle is present, the LED lights turn red and green otherwise. The status of the parking will be updated to the server at the same time.

The HTTP protocol is used to log the data into the server. Each device acts an HTTP client. The server collects the data from the clients and displays the status of every parking location on a webpage. The server will also perform crucial tasks like changing Wi-anFi configurations like Router SSID, Password, HTTP Host IP and HTTP port. Figure 6 shows the slave to server communication. The wireless parking model has two operating modes: i. Configuration Mode and ii. Normal Mode [8].

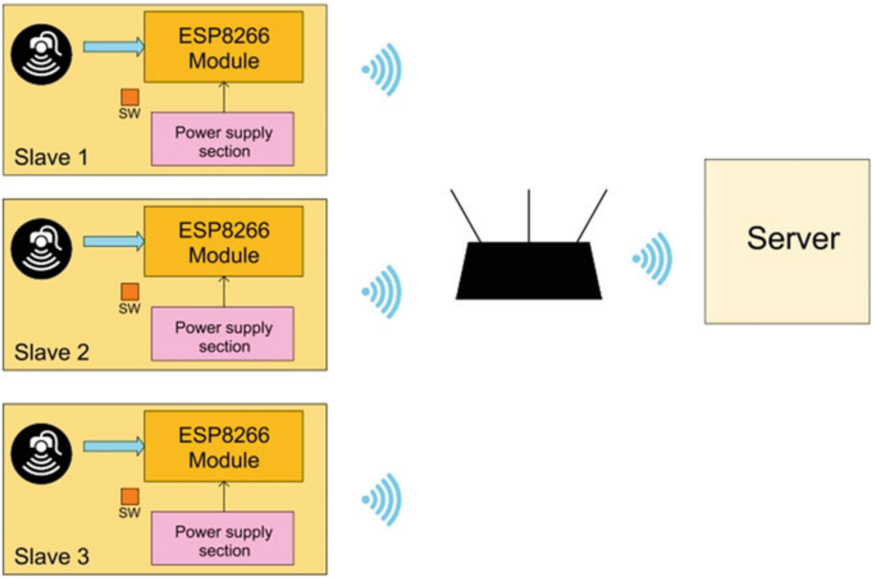
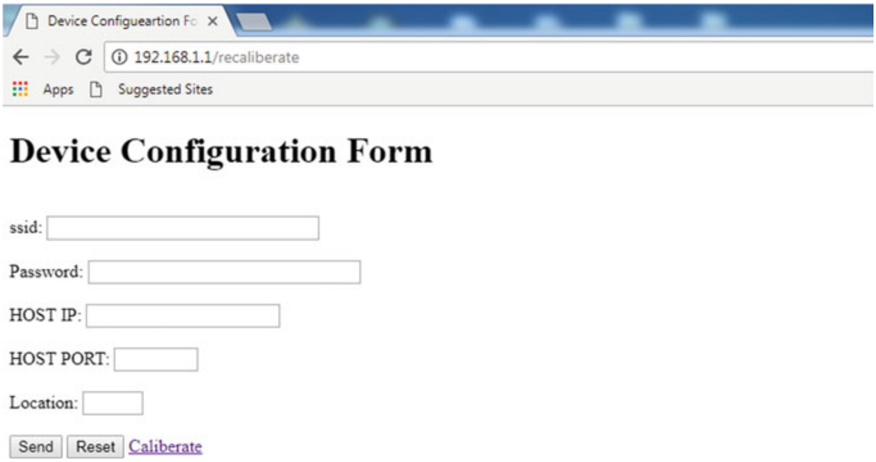


Fig. 6 Slave server communication

The configuration mode is used to change the network settings of the device. With the help of this configuration page, one can easily force the device onto the desired network. The changed configurations can also be stored in the database. Figure 7 shows the device configuration form for the user.



Device Configuration Form

ssid:

Password:

HOST IP:

HOST PORT:

Location:

Fig. 7 Device configuration form

In the normal mode, the device continuously uploads data on the server. The ultrasonic sensor provides the real-time data (presence or absence of car) every 1 s and sends it to the server. The server stores all the values covered at every specific second in the database on the server host computer. An HTML page picks up the values every second from the database and displays it on the page/display section.

In this way, the wired and wireless models together contribute to the effective working of the parking management system.

5 Results

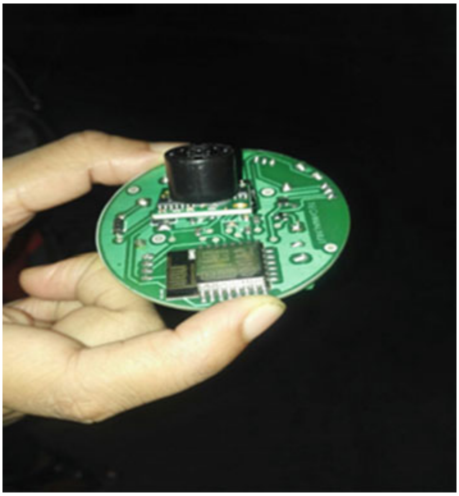
The parking slot module consisting of the ultrasonic sensor was designed to be fit at the roof of each parking slot. Figure 8 shows the module that was designed keeping in mind the compactness of the space. Each parking slot with the module gives red or green light depending upon the occupancy of the space. For example, the display section shown below indicates the presence or absence of a car in a particular parking slot.

Red—slot occupied.

Green—slot vacant.

Figure 9a shows the parking space with green and red LED lights as will be visible to the driver and Fig. 9b shows the display on the mobile device regarding the parking space occupancy. This Parking Management System can be further customized based on the client’s requirements.

Fig. 8 Ultrasonic sensor module for each parking slot designed in association with Techwalnut Innovations LLP



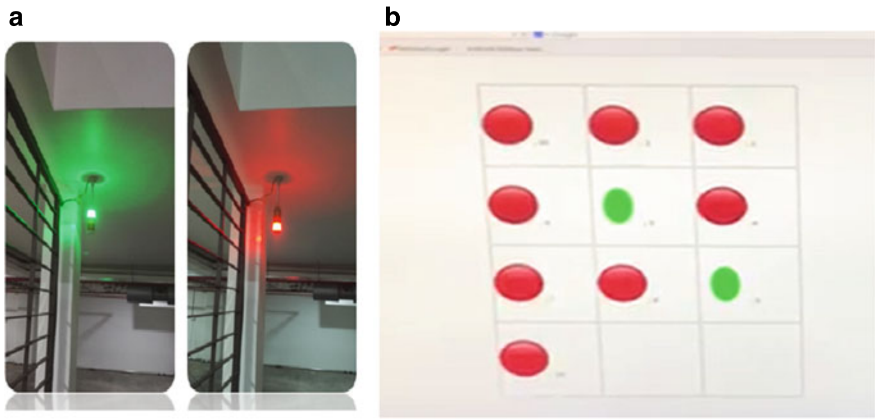


Fig. 9 **a** Parking space installations. **b** Display demo page

6 Future Scope

The wireless model developed for parking management system can be used for many other applications in a building management system. The basic model for wireless data transfer remains the same, only the sensor required might change according to the applications. The applications that can be included are water management system and lighting control using the same model. For miniaturization, the sensor can be MEMS-based [9] installed at the ground plane of the parking lot and will reduce the cost of installation. This sensor will work based on the vibrations caused by the moving vehicle in the slot [10].

7 Conclusion

An effective and robust parking management system is designed and implemented. The methodology of building management system is discussed and different controls can be designed for automatic and easy working of a building. The system is installed and works efficiently in commercial complexes.

Acknowledgments We appreciate the involvement of Techwalnut Automation and Software Solutions Pvt. Ltd, Nagpur, and Technodeal Enerpower Pvt. Ltd, Nagpur, for their involvement in the project and providing necessary kits and components for designing of the BMS unit. Due to their active involvement, the device was tested on a real site and the verification and testing were thus proved to be successful in real-time situations. The paper is the extension of the thesis “Building Management System” by Vasundhara Potdar and Abhishek Mishra under the guidance of Prof. A.H. Harare.

References

1. ShengweiWang J (2002) Integrating building management system and facilities management on the internet. Autom Constr (Elsevier) 11(6):707–715
2. Chen C-T, Lin I-C, Lee C-C, Wu S-K, Chen H-C, Wu S-W (2011) Parking lot management system. National Chin-Yi University of Technology, Taichung (TW), US Patent 8,502,6,98B2
3. Robert J, ThomasNeal A, AndersonStuart G, Behar DMA (2006) Building management system. Honeywell International Inc., Morristown, NJ (US), US Patent 7,567,844,B2
4. Knibbe EJ, Eindhoven (1994) Building management system. U.S. Philips Corporation, New York, US Patent 5,565,855,A
5. Zeitman S (1997) Parking management system. Shlomo Zeitman, Raanana, Israel, US Patent 5,940,481,A
6. Rodriguez JF, Diaz EJ, Cleaver DA (2012) Interactive valet parking management system. KLEVER LOGIC, INC., Austin, TX (US), US Patent 2,012,023,296,5,A1
7. Kumar NM (2019) Energy and power efficient system on chip with nanosheet fet. J Electron 1(01):52–59
8. Manea F, Cepışcă C (2007) PHP + APACHE + TESTPOINT = An original way for having remote control over any type of automation. U.P.B. Sci. Bull. Ser C 69(2)
9. Kumar A, Balpande SS (2014) Energy scavenging from ambient vibrations using MEMS device. Int J Sci Prog Res (IJSPR) 05(01). ISSN: 2349-4689
10. Kumar A, Balpande SS, Anjankar SC (2016) Electromagnetic energy harvester for low frequency vibrations using MEMS. Procedia Comput Sci 79:785–792 (Elsevier)
11. Sharma V, Dixit S, Sridharan A, Roychoudhury SS, Schristopher J, Justin B, SeifiMark J, Freund G, Thareja A (2018) Systems and methods for enhancing building management system interaction and visualization. Johnson Controls Technology Company, Auburn Hills, MI (US), US Patent 1,027,804,8,B2

A Brief Understanding of Blockchain-Based Healthcare Service Model Over a Remotely Cloud-Connected Environment

Subhasis Mohapatra and Smita Parija

Abstract Blockchain technology is in the limelight due to its immutable and anonymous data recording without a centralized authority. In this paper, we present a blockchain-based healthcare model to demonstrate a large-scale blockchain. Blockchain-based model covers numerous sectors like financial, IoT, health care, etc. The blockchain technology has proved its authenticity as a virtually incorruptible cryptographic data storage which provides to the end user in a secure way. The blockchain model is composed of a series of machines that can give on-demand service over the safety net. The access point of heterogeneous medical records over the network is maintained by a hub of computers which operates as a pseudo-anonymous system. The requisite of medical sensitive data is a value from all aspects that range from patient symptoms information, medical certificate, transaction detail, etc. The beauty of the blockchain model is its decentralization by which it can fight against a single point attack. The requisite of sensitive data is growing stronger day by day which needs more sophisticated service for devices collecting personal information through web-based applications or cloud. This paper highlights a comprehensive insight on accessing structured medical data through blockchain. By prototyping, this model will identify, extract and analyze the real-life data metric point in further implementation. A medical certificate was counterfeited by a certain group of people; by using this they are creating a headache for the government and different services for a physically challenged person are misused by them inappropriately. Electronically medical records are using age-old technology for storing; they need modification and strict secure regulation to showcase our technological up-gradation. In this paper blockchain, the pattern-based model, makes medical data management more transparent for service seekers and providers. The giant step of blockchain is going to revolutionize the future e-healthcare the way it interacts with and stores data. But it's not going to happen in the very next day. It can face certain smaller problems which

S. Mohapatra (✉)
Department of Computer Science Engineering, Adamas University, Kolkata, India
e-mail: mohapatra.subhasish@gmail.com

S. Parija
Department of Electronics, CV Raman Engineering College, Bhubaneswar, Odisha, India
e-mail: smita.parija@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_87

can be mitigated at its nascent stage slowly. So this model hopes to solve the regulatory hurdles. Blockchain is emerging as the plateau of productivity for distributing information in health care. In this paper showcases a model that incorporates cryptographic evidence of work over multiple locations; this strong espionage among data blocks checks data against theft. It could influence a convergence point over an interoperable network and can decrease central administrative control. An electronic health record that uses blockchain is an integration of clinical data, medical certificate, etc.; it establishes in its entirety, medical data gathered across a range of smart devices, and hospital databases over the cloud and other sources could all be securely encrypted to individual demands for full chain access. Every service interaction is added with intrinsic security and by implementing healthcare blockchain.

Keywords E-Healthcare · Cloud computing · Blockchain

1 Introduction

The proliferation of blockchain technology and its adoption by various organizations enlightens the path for creating tamper-proof and decentralized technology scalable across the various networks. Blockchain can replace the conventional centralized model in health care due to its high interoperability and personal data privacy. The necessity behind data privacy is to protect personal credentials which follows data privacy right of a citizen to disclose his personal information. This paper elaborates on the structural drawback behind the conventional healthcare model and strongly supports blockchain-based architecture for future service delivery to the citizen. Generally authentication service requirement proof is effective for the following services, i.e. identity verification in medical certificate delivery, symptom analysis from pathos logical data, the image of endoscopy and disease data for medication by using blockchain. Here in this paper, a user must authenticate before access is granted to the service seeker [1–3]. The emergence of blockchain is extending its application to any facet of real-life use cases where the systematic review is still needed for more research and better understanding. Health care is the one industry where well-defined search methodologies are used to access health records to extract citizen information for analyzing the characteristics of a scientific database for a systematic review of relevant records. The plethora of applications for patient data dissemination over cloud is a remarkable advancement in the healthcare sector. Blockchain algorithm is reforming the traditional healthcare practices to a more reliable means. The potential nature of blockchain witnesses security, reliability in the healthcare domain. It is adding an imperative clinical decision and distributed transaction security to a more reliable means. In terms of data-intensive blockchain is adding reliability, scalability to secure the healthcare domain in an effective way. Blockchain is reforming vastly the electronic health records on a regular basis; due to the sensitive nature of data it needs regular, imperative transaction diagnosis for achieving security and privacy. Health care is a sector where a large amount of data is exchanged every hour so a centralized trusted third party constitutes the security domain but at a certain point, if a trusted third party malfunctions or fails maliciously it can lead to huge loss in terms of money besides the fact that a resource developing nation like India is a resource-intensive country where these things can affect its gross domestic product. So a judicious decision can integrate the economy.

1.1 Motivation

The application of blockchain in the healthcare domain showcases the medical data management platform for appropriate access grant to service seekers and providers. Adhering to this we organize peer-to-peer flow of data access between the service seeker and provider. Blockchain authenticates the service flow between the registered patient and service delivery admin, i.e. here blockchain document updates the ledger appropriately. Technology catches one's attention because of its safer immutable control over all the flow of data, where each data carries a unique hash. Resilient behavior can prevent against single-point failure. It is gaining popularity day by day due to its transparency and accountability by maintaining transaction history and peer-to-peer (P2P) interaction without any third party access. Blockchain is fascinated over other organizations including health care due to its timestamp-based immutable linking.

Here in this model, we introduce medical record validation that can ensure high data security. In this blockchain model, it grants access to anyone who is willing to participate. It can set up a node (user) within blockchain for the transaction (record information) by verifying the record based on consensus protocol. The data stored in each block depends on the type of blockchain. This model can track the transaction inside the blockchain ledger. It can remove cheating, cyber attacks or other electronic crimes related to the healthcare perspective. The future application of this model looks bright for any facet of the healthcare industry. The data which are preserved in health care are highly sensitive and their management is still cumbersome. This model enables a platform for others to access records by granting access to patients and practitioners.

1.2 Contribution

Blockchain is defined as a decentralized ecosystem where transaction history is maintained across a peer-to-peer network of a data-stored personal computer coined as a node [4–6]. Blockchain nodes are bound together by the cryptic method. This current is trend getting massive limelight in every sector including health care the blockchain can replace middleman in health care and has tremendous potential to transform medical data to be more vibrant, secure and disintermediated. Data managed by medical organizations is categorized as follows: patient health information, electronic medical record, data collection from IoT devices, medical insurance claim, etc. The blockchain-based solution allows the citizen to prove the authenticity of any document more frequently; it validates the transaction. This is an efficient technology that can help citizens against medical data breaches in the health industry [7–9]. Blockchain first worked on a cryptic secured data hub which was described by Stuart Haber in 1991; consequently, it gave rise to Merkel tree-based approach (by Haber and Stornetta) that improved efficiency in the transaction (integrate several documents into one block). The conceptualization is carried forward by Satoshi Nakamoto in 2008 by implementing a hash cash-based method, the first block is known as genesis block and subsequent blocks are linked by the previous block. Blockchain stands upon 3 pillars: miners (who assemble a block of the transaction by selecting from a transaction pool), blockchain (data structure for the chronological order of a mined block) and the client (who buys goods from other vendors). Satoshi Nakamoto first experimented this technique in the financial sector popularly known as Bitcoin and due to its highly cryptic decentralization feature, it is penetrated to every sector with great momentum. The reliance of blockchain was implemented first in the Secure hash algorithm, i.e. Bitcoin cryptocurrency; the code is open-source accessible to all to modify as per industry standards; the chronological shift in blockchain algorithm is adding efficacy for a systematic review [10–12]. The scope of utility in the blockchain route is a secure channel for data transfer. The blockchain has gained popularity due to its decentralized, digital public ledger jointly maintained by multiple parties securely by using cryptography.

2 Proposed Model

Here in this proposed model, an electronic health record will be maintained in the cloud where all specific information regarding medical data is maintained for substantial use.

STEP1—A medical-based repository will be maintained in a hospital where patient information of a person is restored in a separate database that includes NAME, UNIQUE-ID, ADDRESS, DISEASE INFORMATION, CONTACT NO., SPOUSE NAME and BIOMETRIC DATA.

In the pathological database, we maintain pathological data of a patient that has categories as a blood sample, urine sample, semen sample, patient name and ID given as in Fig. 1.

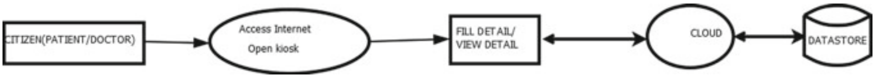


Fig. 1 Phase-1 diagram of restoring data over cloud

In the medical certificate database if any person is applying for any handicapped-certificate or death-certificate from the concerned hospital, it is maintained; here, we use the fields person name, ID, date of issue, concerned authority and one cryptic hash value is established for blockchain navigation.

In the image database, all such images of a patient with patient ID, and X-ray, endoscopy data are maintained for further reference by a researcher and a doctor.

STEP2—A blockchain is maintained for P2P (peer-to-peer) cryptic network establishment between patient and doctor, analytic researcher. It can check against tampering of the medical certificate, and leaking of identity in doctor referral for patient personal information. A researcher must not use the data exorbitantly, his identity and transaction history will be maintained for further analysis [13, 14]. The beauty of blockchain lies in its technological establishment.

STEP 3—In this step a cloudlet-based metadata creation will speed up the data delivery to the respective end user; cloudlet is used for mobility enhancement in a cloud-based architecture. It represents the middle tier of a 3-tier hierarchy: mobile *device, cloudlet and cloud*. A cloudlet can be viewed as a *data center in a box* whose goal is to *bring the cloud closer* [7, 15]. Here doctor/researcher can use mobile devices for accessing patient information, certificate authentication and metadata will act as a cloudlet between cloud data storage and end user. The blockchain centric health data retrieval model is given in Fig. 2.

STEP 4—In this step doctor and the researcher can use the handheld device, monitor, to access this kiosk that is a conglomeration of a set of hospitals that is using this blockchain technology over the Internet. The usual flow of cryptic data is given in Fig. 3.

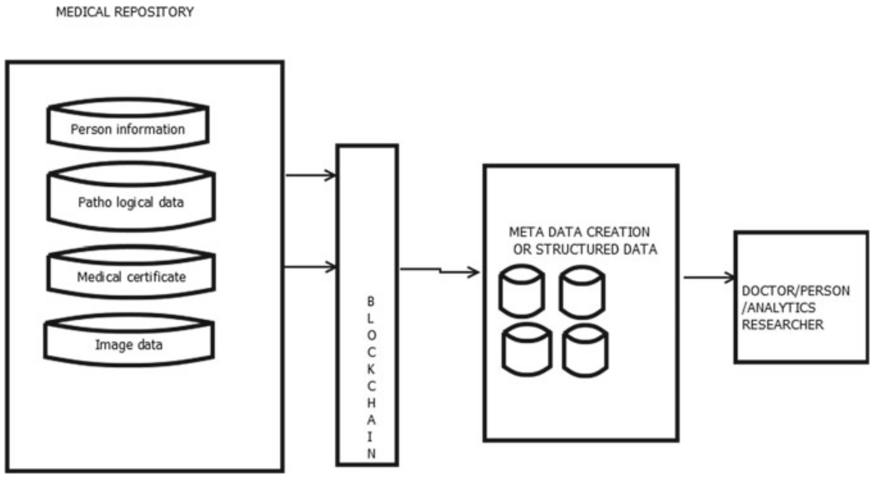


Fig. 2 Medical resource delivery model through blockchain

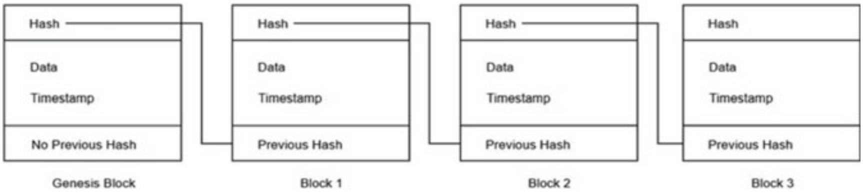


Fig. 3 Blockchain schematic flow

3 Conclusion

This paper provides a holistic overview of blockchain technology and the schematic framework of its extant application in the healthcare ecosystem. It delivers its potential application for innovation in health care. This futuristic model will help from the various aspects, from an administrative as well as research point of view. The data processing for clinical and nonclinical data maintains a proof legitimate record. Since blockchain serves as a source of transparent information, immediate implementation of this model can create accuracy and fidelity. The frequent exchange of health data over the Internet creates a potential security threat for citizens so blockchain-centric infrastructure is the only remedy for all such hurdles. This novel decentralized architecture boosts the healthcare domain by fostering security audits of clinical data repository. For the extant adoption of this technology, we need to address its scalability and regulatory framework implementation for a resource-intensive industry like health care. Finally, the regulatory challenges lie in its determinants of user understanding and attributes of a blockchain application that spread across multiple healthcare jurisdictions. Blockchain has tremendous potential of adding persistency, anonymity and suitability to healthcare. This era of blockchain technology is structuring trust relationships, integrating collaboration in peer to peer maintenance of efficiency, promoting data sharing and assisting the government's penetrating supervisory capacity over healthcare transactions.

References

1. Mettler M (2016) Blockchain technology in healthcare the revolution starts here. In: Proceedings of the 2016 IEEE 18th international conference on e-health networking, applications and services (Health com), Munich, Germany, pp 520–522
2. Kuo TT, Kim HE, Ohno-Machado L (2017) Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 24:1211–1220
3. Roman-Belmonte JM, De la Corte-Rodriguez H, Rodriguez-Merchan ECC, Corte-Rodriguez H, Carlos Rodriguez-Merchant E (2018) How blockchain technology can change medicine. Postgrad Med 130:420–427
4. PayPal, Inc—PayPal Reports Second Quarter 2018 Results. https://investor.paypal-corp.com/news-releases/news-releasedetails/paypal-reports-second-quarter-2018-results?ReleaseID=1072972, Accessed 10 Jan 2019

5. Lite coin. https://litecoin.com, Accessed 10 Jan 2019
6. Visa Net—Electronic payments network. https://usa.visa.com/aboutvisa/visanet.html, Accessed 10 Jan 2019
7. Engelhardt MA (2017) Hitching healthcare to the chain: an introduction to blockchain technology in the healthcare sector. Technol Innov Manag Rev 7:22–34
8. Aoki Y, Otsuki K, Kaneko T, Banno R, Shudo K (2019) Sim block: a blockchain network simulator. In: Proceedings of the cry block 2019 (in conj. With IEEE INFOCOM 2019)
9. Nakamoto, S (2008) Bit coin: a peer-to-peer electronic cash system
10. Falcon—A fast bit coin backbone. https://www.falcon-net.org/, Accessed 10 Jan 2019
11. Qtum (2018). https://qtum.org/en, Accessed 12 Mar 2019
12. Burniske C, Vaughn E, Cahana A, Shelton J (2016) How blockchain technology can enhance electronic health record operability. Ark Invest, New York, NY, USA
13. Jovanovich B, Rousseau PL (2005) General purpose technologies. in handbook of economic growth. Elsevier, New York, NY, USA
14. Androulaki E, Barger A, Bortnikov V, Cachin C, Christidis K, De Caro A, Enyeart D, Ferris C, Laventman G, Manevich Y et al (2018) Hyper ledger Fabric: a distributed operating system for permissioned blockchains. In: Proceedings of the thirteenth euro sys conference on European System '18. In: Association for computing machinery, New York, NY, USA, pp 1–30
15. Angraal S, Krumholz HM, Schulz WL (2017) Blockchain technology applications in health care. Circ Cardio vasc Qual Outcomes 10:e003800
16. Decker C, Wattenhofer R (2013) Information propagation in the bit coin net work. In: Proceedings of the IEEE P2P'13
17. Klarman U, Basu S, Kuzmanovic A, Sirer EG, blo Xroute: A scalable trustless blockchain distribution network whitepaper

Ensemble Learning-Based EEG Feature Vector Analysis for Brain Computer Interface

Md. Sadiq Iqbal, Md. Nasim Akhtar, A. H. M. Shahariar Parvez, Subrato Bharati, and Prajoy Podder

Abstract Brain Computer Interface (BCI) can be normally defined as the process of controlling the environment with the EEG signals. It is a real-type computer-based process which translates the human brain signals into necessary commands. Because there are many strokes attacked or neurologically affected patients in the world and they are not able to communicate effectively or share their emotions, thoughts with the outside world. Considering some extreme cases, tetraplegic, paraplegic (due to spinal cord injury) or post-stroke patients are factually 'locked in' their bodies, incompetent to exert strive motor nerves control after the stroke, paralyses or neurodegenerative diseases, requiring alternative techniques of interactive communication and control of organs. So, BCI is one of the best solutions for this purpose. This paper mainly discusses ensemble learning approaches for EEG signal classification and feature extraction. Bagging, Adaptive boosting, and Gradient boosting are quite popular ensemble learning methods, which are very effective for elucidation and explication of many practical classification problems. EEG signals may be classified as adopting

Md. Sadiq Iqbal · Md. Nasim Akhtar · A. H. M. Shahariar Parvez
Department of Computer Science and Engineering, Dhaka University of Engineering & Technology (DUET), Gazipur 1700, Bangladesh
e-mail: sadiq.iqbal@bu.edu.bd

Md. Nasim Akhtar
e-mail: drnasim@duet.ac.bd

A. H. M. Shahariar Parvez
e-mail: sha0131@gmail.com

Md. Sadiq Iqbal · A. H. M. Shahariar Parvez
Department of Computer Science and Engineering, Bangladesh University, Dhaka 1207, Bangladesh

S. Bharati
Department of EEE, Ranada Prasad Shaha University, Narayanganj 1400, Bangladesh
e-mail: subratobharati1@gmail.com

S. Bharati · P. Podder (✉)
Institute of Information and Communication Technology, Bangladesh University of Engineering and Technology (BUET), Dhaka 1000, Bangladesh
e-mail: prajoypodder@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_88

a set of features like autoregression, power spectrum density, energy entropy, and linear complexity. In this paper, we have used three illustrative procedures Adaptive Boosting, Bagging, and Gradient Boosting ensemble learning methods and extracted features using discrete wavelet transform (DWT), autoregression (AR), power spectrum density (PSD), and common spatial pattern (CSP).

Keywords AdaBoost · Gradient boosting · Bagging · Discrete wavelet transform (DWT) · Autoregression (AR) · Power spectrum density (PSD) · Common spatial pattern (CSP)

1 Introduction

The electric brain activities are recorded by using electrodes and that will be placed in the scalp and that signal is called EEG signals. To monitor brain activities, many techniques are introduced such as magnetoencephalography (MEG), MRI, imaging by using optical technique, and electroencephalography. From all these techniques electroencephalography is more inexpensive and it is more compatible with continuous tracking of brain activities [1–3]. Forthcoming samples can predict the hazardous characteristics of the trained classifier. EEG feature extraction is also an important task. Recently some researchers classified EEG signals by ensemble learning system with the fusion of linear discriminant analysis which can be explained as a hybrid feature extraction technique in order to recede generalization errors [4]. Two dipoles or two channels such as C3 and C4 exploration were offered to feature extraction of motor imagery tasks for the applications of brain–computer interface. By explaining the electroencephalography problem of some data, it is established that the basic methodology can adopt feature extraction techniques such as DWT, autoregression (AR), power spectrum density (PSD), and common spatial pattern (CSP) in order to extract some key features of the EEG data with the help of ensemble learning methods for the classification of signals. The rest of this EEG-based paper is organized as follows. In Sect. 3, the ensemble classification methods have been illustrated briefly which are used for classifying EEG signals. In the fourth section feature extraction methods and procedures have been discussed. Then in Sect. 5 experimental results of EEG signal classification with the training error calculation for a number of iterations has been discussed and tried to provide a comparative explanation.

2 Related Work

Several single classification algorithms have been evaluated already for the classification of EEG signals and they are neural networks, Fisher discriminant analysis, SVM classifier, hidden Markov models, Bayesian classifiers, and source analysis. Two nonlinear and linear classifiers [5] were used to classify the natural EEG signals for the period of five mental responsibilities, presenting that nonlinear classifiers provide only a little improvement results for classification. A methodology to genetic algorithm-based feature selection also offered initial results of EEG [6]. Classification of EEG and offline BCI were described by asymmetric hemisphere model with right hand and left hand performed when using HMMs [7]. The interface mechanisms asynchronously and practices 8 apparent electrodes to observe electroencephalogram signals and identifies statistical classifier 3 dissimilar mental states [8]. The experiment of EEG signal classification rising from BCI applications and recommend an on-line classifier considered through the unassociated LMS algorithm. Constructed on Gaussian models with a Bayesian classifier, that develop the common invention of gradient descent algorithms below the principle of LMS [9]. Sun et al. [2] classified EEG signal using adaptive feature extraction through the update of weighed signal covariance, the peak discriminative features associated with the present brain states are removed by the technique of multi-class CSP. Another research illustrated the classification of motor imagery tasks for BCI and that paper visualized movement of right or left hand without training its dataset [10]. However, as one of the principal current directions in ML, ensemble learning plays a significant role in order to solve problems faced in the above classification algorithm.

Ren [11] and his team adopted hybrid feature extraction method for exploiting the necessary information or data from the EEG signals combining three techniques such as Wavelet packet transform (WPT), autoregressive model (AR) and DWT.

Hongwei Mo et al. proposed a magnetic bacteria optimization algorithm method based on a support vector machine (SVM) classifier in his research paper for the purpose of constructing a proficient classifier for motor imagery EEG-based brain–computer interface (BCI) [12].

Robert Jenke reviewed different sets of features such as event-related potentials, fractal dimension, higher-order crossings, frequency domain features such as band power, time-frequency domain features such as DWT, and electrodes for emotion recognition (happy, curious, angry, sad, and quiet) from EEG data with over 30 studies in his paper. They described that the enactment of a multivariate feature selection system is comparatively better than univariate methods, generally requiring less than 100 features on average [13].

3 Ensemble Learning

In machine learning, ensemble learning is a model where several learners are trained to describe the similar problem [14]. In contrast to normal machine learning methods which try to study one theory from training data, ensemble procedures try to build a set of association and hypotheses of them for usage.

3.1 Creating Ensembles

Usually, an ensemble is created in two steps. In the first step, a quantity of base learners is created, which can be created in a parallel style or in a sequence model where a base learner has an impact on the peers of the following learners. Formerly, the base learners are united to use, wherever the furthermost popular arrangement systems are popular passed by vote for weighted averaging and ensemble model can also be performed classification and regression [15]. There are several effective ensemble approaches. Three ensemble learning-based approaches are illustrated in this paper and they are Adaptive Boosting, Bagging, and Gradient Boosting.

3.2 Adaptive Boosting

AdaBoost also known as Adaptive Boosting, is an ML meta-algorithm. It can be conducted in combination with various kinds of learning processes to develop performance. AdaBoost denote to a specific system of training a boosted algorithm [6]. A boosting algorithm is in the form

$$F_t(a) = \sum_{t=1}^{T} f_t(a) \tag{1}$$

where f_t is a weak learner, a is input and returns a value of the object represented for classes. Suppose, a dataset $\{(a_1, b_1),...., (a_{N,}\, b_n)\}$ is considered where separated item a_i has a related class $b_i \,\epsilon\, \{-1, 1\}$ for respective item and $\{k_1,\dots, k_\mathrm{L}\}$ is a weak classifier set for respective item which classification output is $k_\mathrm{j}(a_i) \,\epsilon\, \{-1, 1\}$ for respective item.

$$C_{(m-1)}(a_i) = x_i k_1(x_i) + \ldots. + x_{m-1} k_{m-1}(a_i) \tag{2}$$

3.3 Bagging

Bagging is also called bootstrap aggregating. It is a ML ensemble meta-algorithm. The algorithm can be designed in order to improve the accuracy of ML algorithms used in regression and classification. Variance is also reduced in bagging. Bootstrap aggregating or bagging helps to avoid overfitting [5].

4 Feature Extraction

Extraction of features is an exceptional procedure of dimensionality decrease. A feature is a distinctive measurement or a characteristic measurement, transform and structural component. This component must be extracted from a segment of a pattern. Changing the input data based on the creation of features is known as feature extraction. Feature extraction method is defined to choose the desired features or the most important information for the classification exercise. EEG signal frequency content delivers suitable data than representing the time domain. The wavelet transform provides us a description of the multi-resolution of a non-stationary signal. EEG is non-stationary signal therefore wavelet is suitable for EEG signals [2].

4.1 Extract Discrete Wavelet Transform (DWT) Feature

Wavelet transform is a non-stationary time-scale exploration technique appropriate to be recycled with EEG signals. DWT is a suitable tool used in order to dispersed and sort non-stationary signals into several frequency features in various time-scales [16]. It is defined in Eq. (5). There are many discrete wavelets such as Haar, Daubechies, and bi-orthogonal wavelet. In this paper, Daubechies 4 wavelet (db4) is applied.

$$\mathrm{DWT}(i, j) = \frac{1}{\sqrt{|2^i|}} \int_{-\propto}^{\propto} x(t)\psi\left(\frac{t - 2^i k}{2^i}\right)\mathrm{dt} \tag{3}$$

4.2 Extract Autoregression (AR) Feature

AR model is mainly applied to define the EEG in the static case. On the other hand, in this paper the AR parameters are permitted to differ from time where the EEG signal is a non-stationary signal [11]. AR is defined by this equation:

$$y_i = a_{1,i} * y_i + \ldots. + a_{p,i} * y_{i-p} + x_{i,} x_i = N\{0, \sigma_x^2(i)\} \tag{4}$$

where zero-mean-Gaussian-noise x_i develop with variance $\sigma_x^2(i)$; an integer i which refer to discrete intermediate time points.

4.3 Extract Power Spectrum Density (PSD) Feature

Power Spectral Density (PSD) is the response of frequency of a periodic or non-periodic signal. It expresses to us that somewhere the power is average which is distributed frequency of a function. The PSD of a non-periodic signal $x(t)$ can be communicated in one of two techniques that are equal individually. In order to achieve the PSD from the FFT values, each FFT value must be squared and divided by 2 times the frequency spacing on the x axis.

$$\mathrm{S_x}(\omega) \lim_{T\to\infty} E\{\frac{1}{2T}|\int_{-T}^{T} x(t)e^{-j2\pi\omega t}dt|^2\} \tag{5}$$

The power can be designed from a non-periodic signal in excess of a particular frequency band as given:

$$\mathrm{P} = \int_{-\propto}^{\propto} S_x(\omega)d\omega = R_x(0) \tag{6}$$

When frequency range is $\mathrm{f_1\_f_2}$, then

$$\mathrm{P_{12}} = \int_{-f1}^{f2} S_x(\omega)d\omega = R_x(0) \tag{7}$$

4.4 Extract Common Spatial Pattern (CSP) Feature

CSP feature and its algorithm is usually castoff in the BCI area and is recycled to discover spatial filters that can exploit the difference between programs on which it is trained [2]. Unmixing matrix is normally computed to the features whose variances are optimal in order to discriminate the two classes of EEG measurements. Normally CSP algorithm provides comparatively better results at the time of its execution on the optimal time windows and the frequency bands [17].

Assumed an EEG signal is extracted from the particular trail, E using dimension N × T since ith class, wherever N is the channels number which is in recording besides T is the time samples number that in use for respective channel. The equivalent covariance matrix can be termed as

$$C_i = \frac{EE'}{trace(EE')} \tag{8}$$

where *trace(EE')* is a function that determines a diagonal of a matrix with its sum and C_i is the average spatial covariance matrix for an individual class.

5 Simulation Results

5.1 Ensemble Learning

Figures 1 and 2 indicate iteration versus training error curve for AdaBoost ensemble learning method where the iteration number is 10, 50, and 100. In Fig. 1 where the iteration number is 2, training error is 0.4. Then the training error is decreased

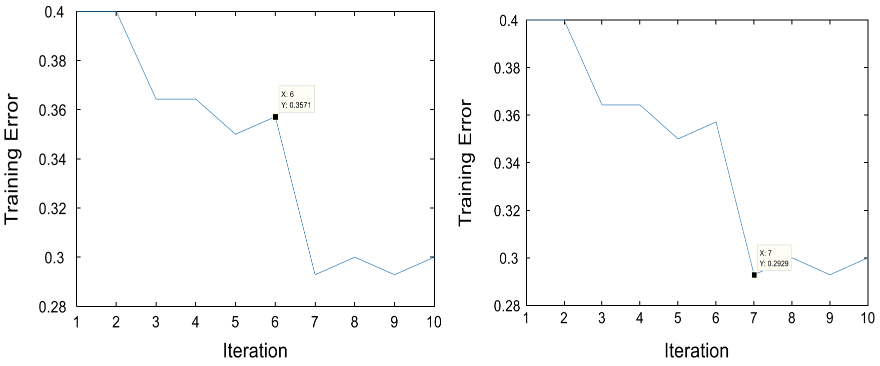
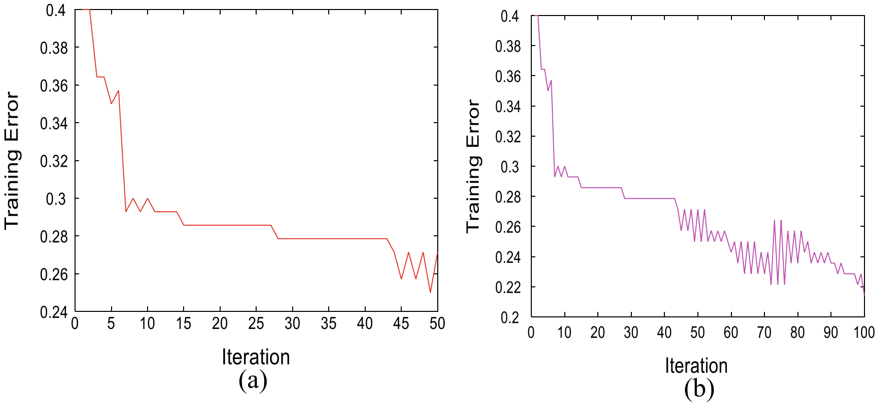


Fig. 1 Iteration versus training error curve for AdaBoost ensemble learning method where iteration number is 10



(a)

(b)

Fig. 2 Iteration versus training error curve for AdaBoost ensemble learning method where **a** iteration number is 50, **b** iteration number is 100

Table 1 Training error calculation for AdaBoost ensemble learning method

Iteration	Training error
1	0.400000
2	0.400000
3	0.364286
4	0.364286
5	0.350000
6	0.357143
7	0.292857
8	0.300000
9	0.292857
10	0.300000

Table 2 Training error calculation for bagging ensemble learning method

Number of iterations	Training error
1	0.3893
2	0.3875
3	0.3732
4	0.3839
5	0.3393
6	0.3393
7	0.3357
8	0.3357
9	0.3286
10	0.3232

for the next iteration number. On the other hand, when the iteration number is 8, training error increases in a small amount.Table 1, 2 and 3 illustrate the training error calculation for AdaBoost, Bagging and Gradient Boosting ensemble learning method respectively.

Figures 3 and 4 show iteration versus training error curve for bagging ensemble learning method where the iteration number is 10,50, and 100. In this simulation process, we have assumed 40 minimum parents where the base algorithm is classification tree.

Figure 5 illustrates iteration versus training error curve for gradient boosting learning method where the iteration number is 10, 50, and 100. Figure 5 also displays the peak value of error number with respect to the iteration number. Here, the base algorithm is regression tree. From the table, it can be said that when that number of iterations is increased, training error is decreased.

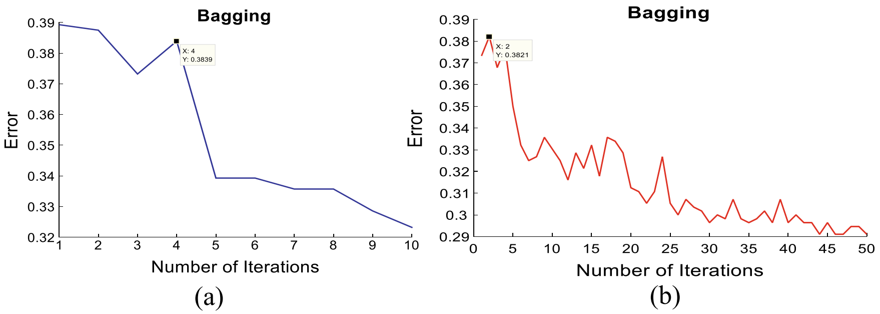


Fig. 3 Iteration versus training error curve for bagging ensemble learning method where **a** iteration number is 10, **b** iteration number is 50

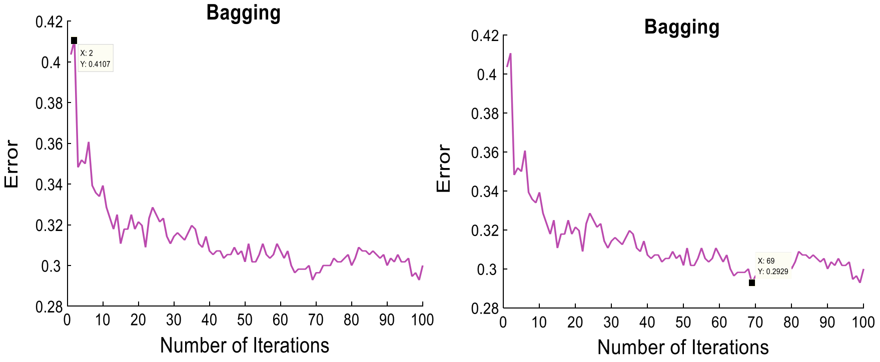


Fig. 4 Iteration versus training error curve for bagging ensemble learning method where iteration number is 100

Table 3 Training error calculation for Gradient Boosting ensemble learning method

Number of iterations	Training error
1	0.3964
2	0.3839
3	0.3375
4	0.3134
5	0.2839
6	0.2839
7	0.275
8	0.2679
9	0.2634
10	0.2571

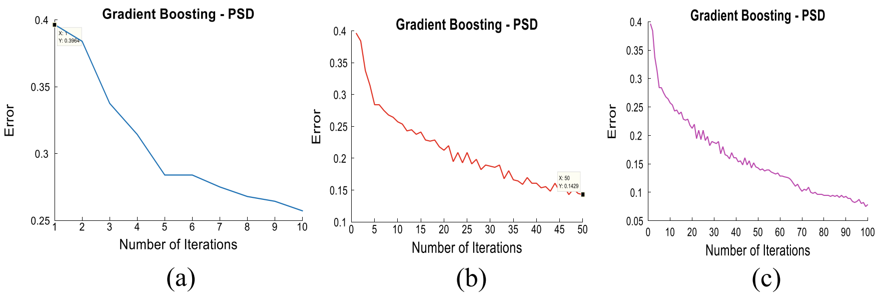


Fig. 5 Iteration versus training error curve for gradient boosting ensemble learning method where **a** iteration number is 10, **b** iteration number is 50, **c** iteration number is 100

5.2 *Feature Extraction*

Figure 6 visualizes 18 channels of electrodes such as A1, A2, C3, C4, Cz, and so on. We get two electrodes of channel C3 and C4 then extracted feature according to DWT, autoregression, power spectrum density (PSD), and common spatial pattern (CSP). For the purpose of extracting the features from the EEG dataset C3 and C4 channels of left hand and right hand movement are considered. Recording of right and left movement was arbitrary besides the recycled three EEG channels in excess of C3, Cz and C4.

Figures 7 and 8 visualize DWT, Autoregression, PSD, CSP features for C3 and C4 channels where blue color indicates left hand and red color indicates right hand correspondingly. The feature is extracted from an EEG signal. Sampling frequency value is 128. The autoregression order used in this experiment is 5.

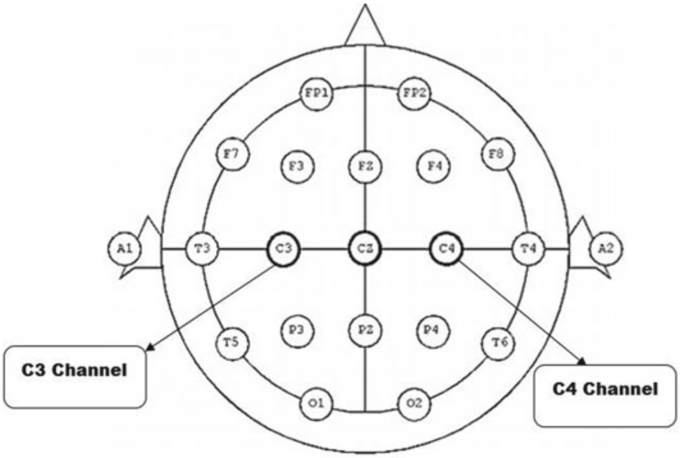


Fig. 6 Visualized distance of all electrodes with C3 and C4 channels

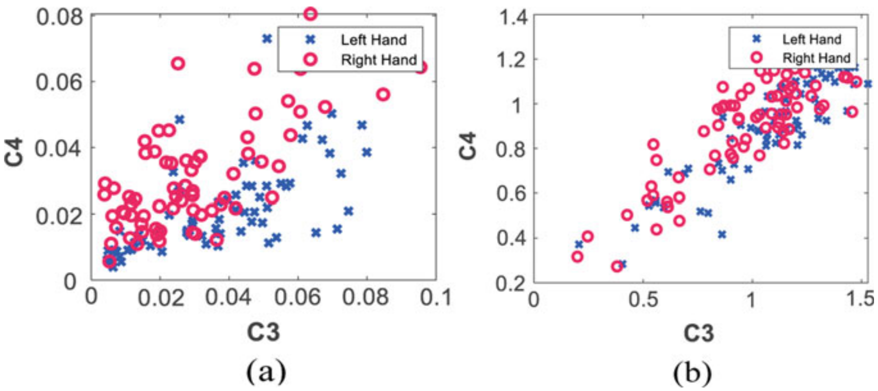


Fig. 7 Extract (a) DWT (b) Autoregression feature for the purpose of C3 and C4 channels

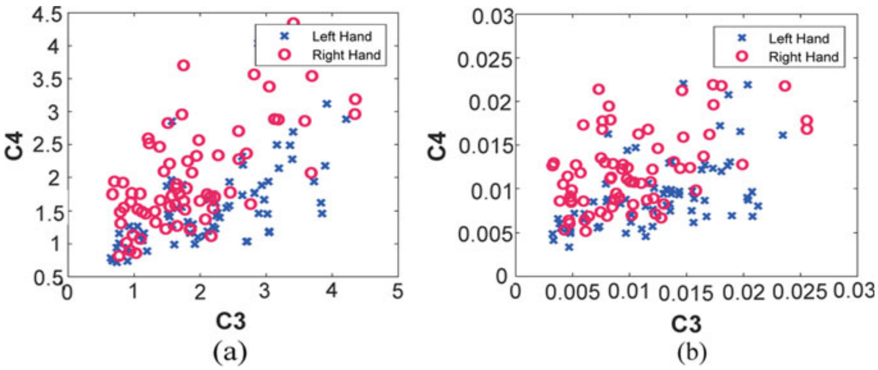


Fig. 8 Extract **a** Power spectrum density (PSD). **b** Common Spatial Pattern (CSP) feature for the purpose of C3 and C4 channels

Out of 280 trials in this dataset, 140 trials were specified as training data as well as enduring 140 were deliberated as testing data. Separately trial has 9 s length using sample frequency 128 Hz and it was filtered among 0.5 and 30 Hz. The CSP, AR, PSD, and DWT outputs were recycled as feature vector and were categorized into two classes right/left-hand movement exhausting ensemble learning which deliberated at the ensemble learning section. So to train these features, we take the help of ensemble learning. DWT, AR, and PSD have to be considered by mean of the total values of the changed signals, such as C3 and C4, coefficients separately for sub-band, and the rate of channel C3 is compared to channel C4 and output of this method is exposed in Figs. 7 and 8.

6 Conclusion

EEG signal is normally a quantification of potentials which reflect the electrical active motion of the human brain. So, EEG signal analysis, classification, and feature extraction are very significant for fruitful diagnosing of many neurological diseases such as brain stroke, brain tumor, epilepsy, head injury, and dementia. This paper mainly represents ensemble learning for the purpose of EEG signal classification applying AdaBoost, gradient boosting, bagging techniques, and visualized training error in order to increase the number of iterations. Where iteration number is increased, training error is decreased than the fluctuated training error. In simulation, it has been observed that for the 10th iteration, the value of training error of bagging, AdaBoost, and gradient boosting is respectively 0.3232, 0.3000, and 0.2571. In this paper, the feature extraction processes of EEG such as DWT, autoregression, PSD, and CSP feature have also been illustrated with respect to channels C3 and C4 where blue color indicates left hand and red color indicates right hand. Daubechies wavelet is used in the DWT feature extraction stage.

Ethical approval Authors do not violate the ethical statement.

References

1. Bruce H. Dobkin: The Clinical Science of Neurologic Rehabilitation. Oxford University Press, 147–160 (2003)
2. Sun S, Zhang C (2006) Adaptive Feature Extraction for EEG Signal Classification. Med Biol Eng Compu 44:931–935
3. Bharati S, Podder P, Al-Masud MR (2018) Brain Magnetic Resonance Imaging Compression Using Daubechies & Biorthogonal Wavelet with the Fusion of STW and SPIHT, 2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), Gazipur, Bangladesh, pp. 1–4. https://doi.org/10.1109/icaeee.2018.8643004
4. Ren W, Han M (2018) Classification of EEG Signals Using Hybrid Feature Extraction and Ensemble Extreme Learning Machine. Neural Process Lett. https://doi.org/10.1007/s11063-018-9919-0
5. Garrett D et al (2003) Comparion of Linear, Nonlinear, and Feature Selection Methods for EEG Signal Classification. IEEE Trans Neural Syst Rehabil Eng 11:141–144
6. Müller K-R, Anderson CW, Birch GE (2003) Linear and Nonlinear Methods for Brain-Computer Interfaces. IEEE Trans Neural Syst Rehabil Eng 11:165–169
7. Obermaier B, et al (2001) Asymmetric Hemisphere Modeling in an Offline Brain-Computer Interface. IEEE Transactions on Systems, Man, and Cybernetics: systems, Part C-Appl. Rev. 31, 536–540.
8. Millán JR et al (2004) Brain-Actuated Interaction. Artif Intell 159:241–259
9. Sun S, Zhang C (2005) Learning On-Line Classification via Decorrelated LMS Algorithm: Application to Brain-Computer Interfaces. In: Hoffmann A, Motoda H, Scheffer T (eds) DS 2005, vol 3735. LNCS (LNAI). Springer, Heidelberg, pp 215–226
10. Kamousi B, Liu Z, He B (2005) Classification of Motor Imagery Tasks for Brain-Computor Interface Applications by Means of Two Equivalent Dipoles Analysis. IEEE Trans Neural Syst Rehabil Eng 13:166–171

11. Ren W, Han M, Wang J, Wang D, Li T (2017) Efficient feature extraction framework for EEG signals classification, Seventh International Conference on Intelligent Control and Information Processing (ICICIP), IEEE
12. Mo H, Zhao Y (2016) Motor Imagery Electroencephalograph Classification Based on Optimized Support Vector Machine by Magnetic Bacteria Optimization Algorithm. Neural Process Lett 44(1):185–197. https://doi.org/10.1007/s11063-015-9469-7
13. Jenke R, Peer A, Buss M (2014) Feature Extraction and Selection for Emotion Recognition from EEG. IEEE Trans Affect Comput 5(3):327–339. https://doi.org/10.1109/taffc.2014.2339834
14. Bharati S, Podder P, Paul P (2019) Lung cancer recognition and prediction according to random forest ensemble and RUSBoost algorithm using LIDC data. Int J Hybrid Intell Syst 15(2):91–100. https://doi.org/10.3233/HIS-190263
15. Zhou ZH (2012) Ensemble methods: foundations and algorithms. CRC Press
16. Bharati S, Rahman MA, Mandal S, Podder P (2018) Analysis of DWT, DCT, BFO & PBFO algorithm for the purpose of medical image watermarking. In: 2018 International Conference on Innovation in Engineering and Technology (ICIET), Dhaka, Bangladesh, pp 1–6 (2018). https://doi.org/10.1109/ciet.2018.8660796
17. Lisi G, Morimoto J (2017) Noninvasive brain machine interfaces for assistive and rehabilitation robotics: a review. In: Ueda J, Kurita Y (eds) Human modelling for bio-inspired robotics. Academic Press, pp 187–216. https://doi.org/10.1016/B978-0-12-803137-7.00006-9
18. Ahnaf Rashik Hassan (2016) Siuly Siuly, Yanchun Zhang, Epileptic seizure detection in EEG signals using Tunable-Q factor wavelet transform and bootstrap aggregating. Comput Methods Prog Biomed 137:247–259. https://doi.org/10.1016/j.cmpb.2016.09.008
19. Bharati S, Podder P, Raihan-Al-Masud M (2018) EEG Eye State Prediction and Classification in order to Investigate Human Cognitive State, 2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), Bangladesh, pp 1–4. https://doi.org/10.1109/icaeee.2018.8643015

Cluster-Based Data Aggregation in Wireless Sensor Networks: A Bayesian Classifier Approach

Lokesh B. Bhajantri and Basavaraj G. Kumbar

Abstract The network is composed of wireless distributed sensor nodes with data computational capabilities. In each cluster, the cluster member nodes are used to send the sensed data to their respective cluster head to aggregate and classify the data effectively. In this work, the algorithm for cluster-based aggregation of data using the Naive Bayesian Classifier is proposed. The proposed scheme provides better performance rather than existing algorithms with accuracy, efficient energy utilization, and computation overhead.

Keywords Wireless sensor networks · Data aggregation · Cluster · Cluster head · Naive Bayesian classifier

1 Introduction

A Wireless Sensor Network (WSN) contains the number of distributed sensor nodes with high computation network. The sensors are scattered over the network, each of these nodes has the capability to process or route the data to a sink node or a base station through cluster head (CH) nodes. A sink node collects the data from its CHs from each cluster of the WSN for enhancing and decision-making of data. The WSN has been used for various real-time applications such as military applications, habitat monitoring, environmental monitoring, and so on. In WSN cluster member nodes in each cluster periodically send the data to their respective CHs, then CHs transmit the data over the network. The frequency of reporting sensed data usually depends on the particular applications to the sink node.

L. B. Bhajantri (✉)
Department of Information Science and Engineering, Basaveshwar Engineering College, Bagalkot, Karnataka, India
e-mail: lokeshcse@yahoo.co.in

B. G. Kumbar
Department of Computer Science and Engineering, BGMIT, Mudhol, Karnataka, India
e-mail: gkbasavaraj@gmail.com

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_89

In WSN, there is a grouping of sensor nodes into clusters to satisfy the load balancing, scalability, efficiency of energy, and better network lifetime. There are several problems that impact the design and performance of the WSNs as follows [1, 2]: (1) quality of Service-based communication and computing, (2) fault tolerance management issues, (3) routing issues, (4) context-aware issues in network, (5) scalability issues, (6) authentication issues, (7) key management issues in WSNs, (8) sensors are very limited in power, computational capacities, and memory, (9) management of resources, (10) dynamic topologies, (11) security, (12) aggregation and classification of data, and (13) data accuracy.

The energy is a limited resource for the computation of tasks in WSNs. When the number of nodes increases, the network complexity increases. In this regard, energy consumption is required for computing the tasks. So energy-efficient aggregation and classification of data approach are required in sensor networks. The data aggregation is a mechanism or a method for removing redundancy information or data from the sensor node data to increase the lifetime of the network and the accuracy of sensor data. This work focuses on data aggregation issues such as delay, data accuracy, data redundancy, and traffic load [3, 4]. Therefore, we get the accuracy of aggregated and classified data of sensor nodes in the WSN.

The following proposed work is structured as follows: Sect. 2 describes the literature survey toward the proposed work. Section 3 depicts the proposed work for cluster-based data aggregation using Naive Bayesian Classifier. The simulation work of the proposed work is presented in Sect. 4, and finally, it concludes the proposed work in Sect. 5.

2 Literature Survey

The following are related works described toward the proposed work: the work given in [5] depicts an efficient aggregation of sensor data using clusters in terms of energy and tree-based approaches. This paper addresses the problems in clusters of WSNs for data aggregation and transmission over the network. Also, it address the issue of unbalanced energy dissipation. The energy-efficient data aggregation based on clusters in WSNs is presented in [6]. In this work, the proposed scheme for cluster-based data aggregation comprises the four stages like cluster definition, CH election, data aggregation, and maintenance. The work given in [7] describes the secure cluster-based data aggregation in WSN. They have addressed the data aggregation and security issues together for data transmission over the network. This work identifies the untrustworthy nodes effectively with less energy utilization. A survey on energy-efficient hierarchical clustering in WSNs is discussed in [8]. This work addresses the issues and challenges for designing cluster schemes in terms of parameters and protocols. Also discussed and evalauted is the cluster approach for existing techniques in WSNs. Finally, they discussed the advantages, disadvantages, and applicabilities for clustering.

The work given in [9] describes the cluster-based compression data aggregation and routing in WSNs. In this work, tree-based data aggregation and routing of data are performed. The performance of the aggregation of data using clusters is evaluated in [10]. The entropy-based aggregation of data is discussed. The proposed method is simulated by the NS2 simulator with packet drops, network lifetime, aggregation, convergence rate, and transmission cost parameters. The dynamic clustering-based data aggregation is presented in [11]. They have proposed the scheme for dynamic fuzzy clustering data aggregation for the selection of clusters and data aggregation in WSNs. The proposed method is compared with existing algorithms in terms of energy and network lifetime. The cluster-based secure routing is depicted in [12]. In this paper, secure transmission of data over the network using the cryptography technique is discussed. Also, this scheme works on MAC operations for data encryption and transmission. The current status and future directions of research on data aggregation in WSNs are presented in [13]. In this work, they have discussed the challenges and limitations of the aggregation of data. Also presented are the techniques, tools, and methodologies for aggregation. The routing protocols for clusters is depicted in [14]. Also, some of the approaches are discussed toward cluster-based routing protocols. The work given in [15] presents the techniques for the aggregation of data in sensor networks. This paper also discusses data gathering and routing protocols using clusters. Some of the related works are given in [16–20].

The earlier works on WSNs are carried out to maximize the network lifetime by using the clustering technique based on LEACH, PEGASIS protocols, and other protocols or algorithms. These protocols are used to consume more amount of energy and time required for data communication over the network. As per the comparison between various data aggregation approaches (like centralized, cluster, tree, and in network-based approaches) in WSNs, cluster-based data aggregation approach is a more suitable approach because of load balancing, less energy consumption for data computation, increase in the lifetime of the network, and scalability. In this work, we have considered the Naive Bayesian Classifier-based aggregation and classification of data over the CHs in the network.

3 Proposed Work

In this work, the objective of the work is to remove the redundant data and perform efficient classification of data that increase the lifetime of the network and accuracy otherwise prolong or conserves the energy of the sensors during data aggregation in WSN. Therefore, which required to aggregate and classify the data in terms of syntactic and semantic information or structured data that is accomplished by the proposed Naive Bayesian Classifier algorithm. Cluster-based data aggregation and classification is the best approach among all the approaches in WSNs, which provide the following benefits like maximizing energy-efficient data computation, better network lifetime, accuracy, scalability, load balancing, and less computational overhead. In this section, the proposed model and algorithms for the proposed work are described below.

3.1 The Proposed Model

Figure 1 describes the data aggregation using the cluster-based approach in WSNs. A randomly or statically deployed WSN accepts the protocols for cluster formation in which the group of nodes in clusters randomly selects the highest energy of sensor node as a CH across various regions of the network. The randomly selected CH nodes are used to broadcast the message to their nodes. Sensor nodes receive and send the confirmation message to randomly selected CHs within the communication range. The cluster formation and CH selection mechanism is defined in the first proposed algorithm. After the selection of CHs in each cluster, CH starts the communication between cluster member nodes and sink node or base station through intra-cluster communication and inter-cluster communication in the proposed network. Periodically, cluster member nodes of a cluster sense and transmit the data to the CH node through single- or multi-hop communication, i.e., inter-cluster communication for aggregation of data. The sink node receives the aggregated data from its multiple CH nodes. Finally, the sink node performs the actions which are based on the aggregated and classified data from sensor nodes. The below algorithm shows the cluster formation scheme in the proposed work.

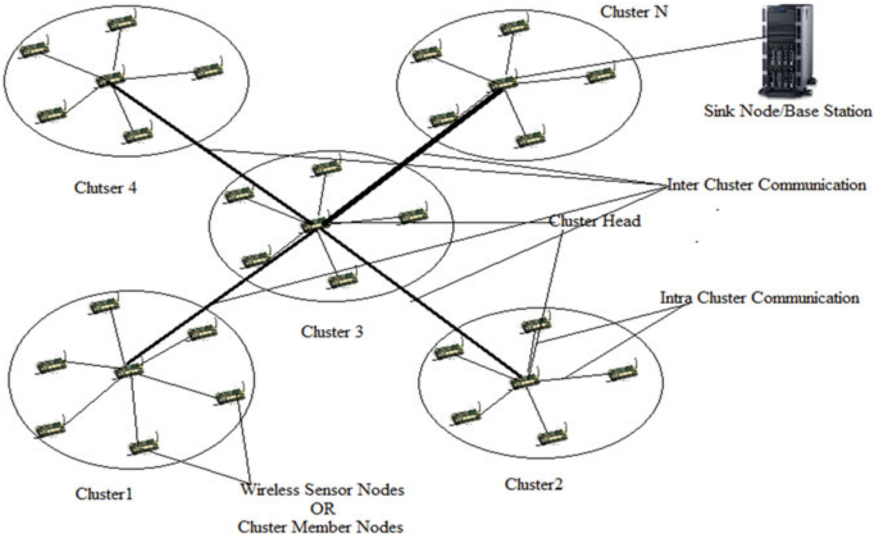


Fig. 1 The proposed model of data aggregation system

3.2 Algorithm for Cluster Formation

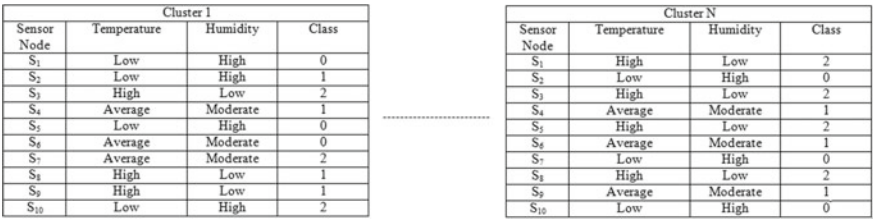
The cluster formation algorithm proposed is as follows:

Step 1: Place the number of sensor nodes in the environment randomly
Step 2: Randomly identify the active nodes as a CH with the highest amount of energy in various regions of the network
Step 3: Each elected CH node broadcasts the heartbeat message among the sensor nodes in the network
Step 4: Each CH receives the heartbeat messages from its sensor nodes to form the clusters within the communication range
Step 5: Sink node performs the computations in the WSNs.

3.3 Algorithm for Data Aggregation Using Naive Bayesian Classifier

In this work, we have used Naive Bayesian Classifier model for data classification, which is one of the widely used classification models in WSN. The Naive classification model has many advantages such as efficient computation, simplification, and better performance on classification on aggregated sensor data in the network. This classifier model works on the basis of Bayesian theory, which is a classical statistical classification algorithm [21]. The algorithm is on the basis of the independence of each attribute of the WSN. Consider the following various cluster information in the proposed work. The data set for each cluster is shown in Fig. 2. The data sets correspond to different WSNs that are used in various types of applications. The proposed algorithm for data aggregation and classification are described below.

In this work, we have considered the three ranges for temperature:
[0 °C–25 °C] = Low, [26 °C–35 °C] = Average, and [36 °C–50 °C] = High.
Ranges for humidity are as follows:
[10–40%] = Low, [41–60%] = Moderate, and [61–100%] = High.



Cluster 1			
Sensor Node	Temperature	Humidity	Class
S_1	Low	High	0
S_2	Low	High	1
S_3	High	Low	2
S_4	Average	Moderate	1
S_5	Low	High	0
S_6	Average	Moderate	0
S_7	Average	Moderate	2
S_8	High	Low	1
S_9	High	Low	1
S_{10}	Low	High	2

Cluster N			
Sensor Node	Temperature	Humidity	Class
S_1	High	Low	2
S_2	Low	High	0
S_3	High	Low	2
S_4	Average	Moderate	1
S_5	High	Low	2
S_6	Average	Moderate	1
S_7	Low	High	0
S_8	High	Low	2
S_9	Average	Moderate	1
S_{10}	Low	High	0

Fig. 2 Data set for clusters of WSNs

Begin

Step 1: Each sensor node/cluster member node measures the temperature and humidity of the environment in each cluster of the network periodically;

Step 2: Each CH collects the data set from its cluster member sensors in the network;

Step 3: Each Cluster Head aggregates and classifies the data using the proposed scheme as follows:

(i) Cluster Head reads the training data set;
(ii) Cluster Head converts the data set into the frequency table;
(iii) Cluster Head creates the Likelihood table by finding the probabilities for temperature and humidity;
(iv) Apply the following Bayesian theorem for data classification in WSNs:

$$P(Data1, Date2, \ldots DataN\ ClassN) = \prod_{i=1}^{m} P(DataN\ ClassN)$$

$$P(DataN\ ClassN) = \frac{P(ClassN\ DataN) \times P(DataN)}{P(ClassN)}$$

where N = 1, 2, 3,… i. Data could be temperature and humidity.

(V) Calculate the Posterior Probability of each class. The result of the highest probability of a class is the temperature and humidity of the environment.

End

4 Simulation

The proposed work is carried out using the NS2 simulator tool. The proposed work is carried out under the WSN environment. The work has been simulated with 100 iterations randomly. The proposed network model composed of the set of nodes, which are deployed randomly in a network environment. The following performance parameters of the proposed work measured are efficient energy consumption, accuracy, and computation overhead.

We have considered the following variables of the proposed work as $N = 100$–500 sensor nodes, Node energy (N_E) = 5Joules, Sink node (S_N) = 1, Data communication range between the nodes (R_N) = 300 m, Amount of energy for sensing of data by each node (A_S) = 60 nJ/Bit, Amount of energy for data transmission (A_T*)* = 60 nJ/Bit, and Value of Threshold with respect to energy (T_{HE}) = 0.05 J. The procedure for the simulation is given below:

Begin

(1) The set of sensor nodes are deployed randomly;

(2) Construct the clusters;

(3) Elect the CH with the highest energy among the sensor nodes randomly;
(4) Apply the proposed work for aggregation and classification of data in WSNs;
(7) Performs the system parameters of the scheme.
End

In this paper, we have evaluated the following parameters of our work:

1. *Computation Overhead (%)*: As the number of nodes increases in the cluster, the computation overhead gradually decreases over the network. The computation overhead is expressed as a percentage.
2. *Energy Consumption (mJoules)*: As the number of nodes increases in the network, it gradually decrements the energy of sensor nodes. The consumption of energy is evaluated in the form of a percentage.
3. *Accuracy (%)*: As the number of nodes increases, the accuracy of data aggregation algorithms gradually decreases for providing quality of data.

Figure 3 shows the comparative analysis of Naive Bayesian Classifier, KNN, and Decision Tree with respect to computation overhead on each node in the cluster. It clearly shows that the Bayesian Classifier is faster than the other algorithms used or computation overhead. As the Bayesian Classifier uses the probabilistic method for computation, which takes fewer iterations even for a large number of data sets, it is more efficient. Other algorithms like KNN and Decision Tree are less efficient for large data sets in the WSNs.

Figure 4 shows a clear analysis of the energy consumption of each node with respect to Naive Bayesian Classifier, KNN, and Decision Tree. Energy consumption in the Naive Bayesian Classifier algorithm is lesser than other algorithms, which is more efficient than KNN and Decision Tree algorithms. These algorithms are slower

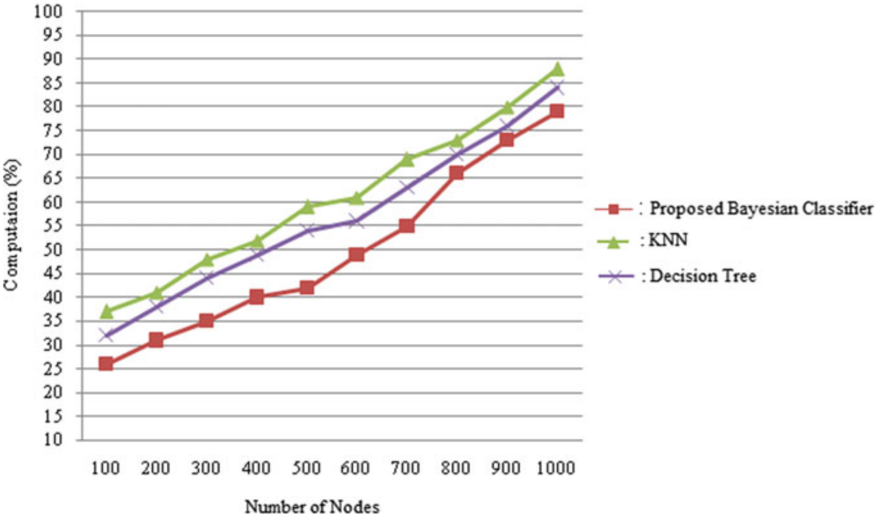


Fig. 3 Comparative analysis of computation overhead on each node for Bayesian Classifier, KNN, and Decision Tree

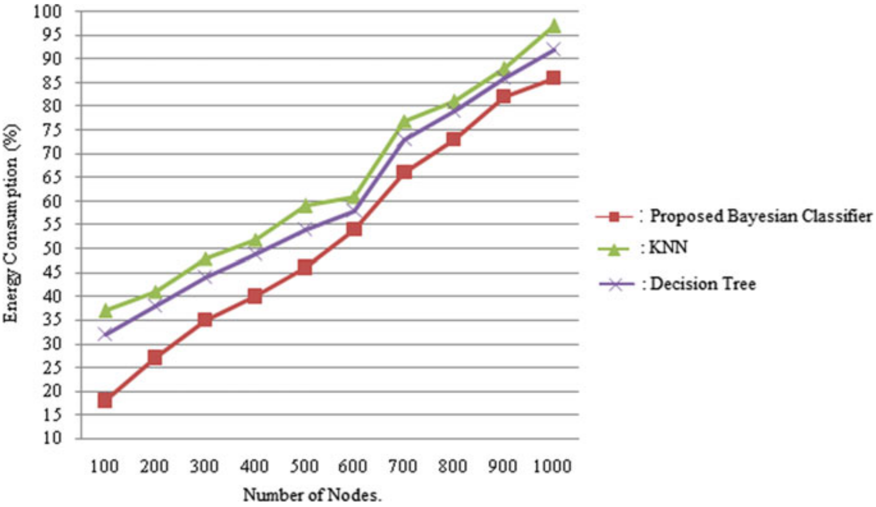


Fig. 4 Comparative analysis of energy consumption for the Naive Bayesian Classifier, KNN, and Decision Tree

for computing data whereas the Bayesian algorithm is probabilistic, which remains the same even for more nodes in the cluster.

The analysis of the accuracy of Naive Bayesian Classifier, KNN, and Decision Tree is shown in Fig. 5. It clearly depicts that the number of nodes increases in the cluster, the performance of Bayesian Classifier remains high at any instant of time comparing with the other two algorithms (KNN and Decision Tree).

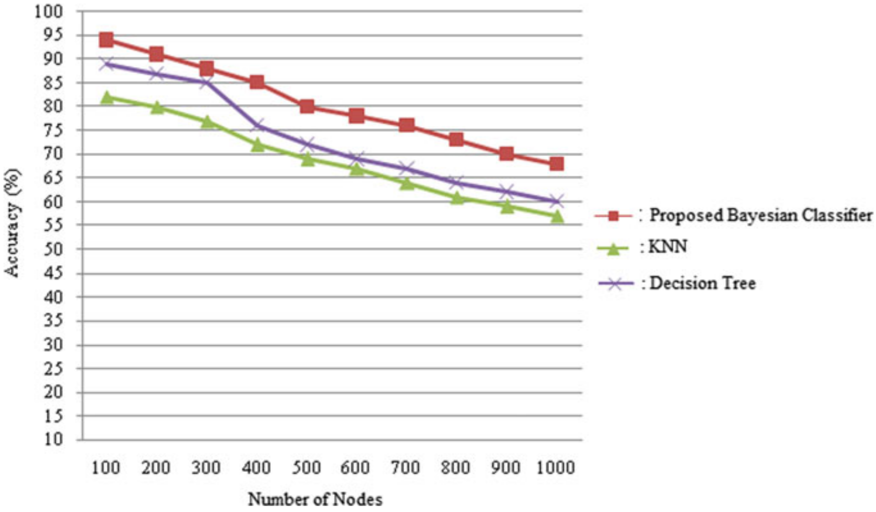


Fig. 5 Comparative analysis of accuracy for Bayesian Classifier and KNN

5 Conclusion

The proposed method for data aggregation and classification in clusters using Naive Bayesian Classifier is efficient with respect to computation overhead, energy consumption, and accuracy compared with the other two algorithms namely KNN and Decision Tree. The Naive Bayesian Classifier is a probabilistic method and time complexity for the training set is O(np) and for prediction, it is O(p) whereas for the other two it is KNN(O(np)) and Decision Tree(O(n^2p) and O(p)) where 'n' is the number of training samples and 'p' is the number of features. Hence Bayesian Classifier is faster and efficient than the KNN and Decision Tree. The simulation results in this paper clearly show that Naive Bayesian Classifier is more appropriate to classify the heterogeneous data produced in the cluster.

References

1. Sharma S, Bansal RK, Bansal S (2013) Issues and challenges in wireless sensor networks. In: International conference on Machine Intelligence and Research Advancement, pp 58–62
2. Suo H, Wan J, Huang L, Zou C (2012) Issues and challenges of wireless sensor networks localization in emerging applications. In: International conference on computer science and electronics engineering, vol 3:447–451
3. Ihsan A, Rahim K, Mehdi H, Mustaq, A (2013) Energy aware routing and data aggregation in wireless sensor Networks. In: 2nd International Conference on Machine Learning and Computer Science, pp 27–30
4. Bhaskar K, Deborah E, Stephen W (2002) The Impact of Data Aggregation in Wireless Sensor Networks. In: International conference on Distributed Computing Systems, pp 575–578
5. Yuea, J., Zhang, W., Xiao, W., Tang, D., Tang, J.: Energy Efficient and Balanced Cluster-Based Data Aggregation Algorithm for Wireless Sensor Networks. In: 2012 International Workshop on Information and Electronics Engineering, vol.29, pp. 2009–2015, (2012)
6. Ranjani, S. S., Radhakrishnan, S., Thangaraj, C.: Energy-efficient Cluster Based Data Aggregation for Wireless Sensor Networks. In: 2012 International Conference on Recent Advances in Computing and Software Systems, pp. 174–179, (2012)
7. Ranjani, S. S., Radhakrishnan, S., Thangaraj, C.: Secure Cluster Based Data Aggregation in Wireless Sensor Networks. In: IEEE International Conference on Science Engineering and Management Research, pp. 11–6, (2014)
8. Jan B, Farman H, Javed H, Montrucchio B, Khan M, Ali S (2017) Energy Efficient Hierarchical Clustering Approaches in Wireless Sensor Networks: A Survey. Journal on Wireless Communications and Mobile Computing 2017(6457942):1–14
9. Jothi S, Chandrasekaran M (2016) Cluster Based Compressed Data Aggregation and Routing in WSN. International Journal of Intelligent Engineering and Systems 9(4):69–78
10. Sinha A, Lobiyal DK (2013) Performance Evaluation of Data Aggregation for Cluster-Based Wireless Sensor Networks. Journal on Human-centric Computing and Information Sciences 3(13):1–17
11. Abasi, A., Sajedi, H.: Data Aggregation in Wireless Sensor Network Based On Dynamic Fuzzy Clustering. In: Fourth International Conference on Soft Computing, Artificial Intelligence and Applications pp. 125–138, (2015)
12. Mezrag, F., Bitam, S., Mellouk, A.: Secure Routing in Cluster-Based Wireless Sensor Networks, In: IEEE International Conference on Global Telecommunications, pp. 1–6, (2017)

13. Randhawa S, Jain S (2017) Data Aggregation in Wireless Sensor Networks: Previous Research, Current Status and Future Directions. Journal on Wireless Personal Communications 97(3):3355–3425
14. Fanian F, Rafsanjani MK (2019) Cluster-Based Routing Protocols in Wireless Sensor Networks: A survey. Journal of Network and Computer Applications 142:111–142
15. Dhand, G., Tyag, S. S.: Data Aggregation Techniques in WSN: Survey. In: 2nd International Conference on Intelligent Computing, Communication & Convergence, vol. 92, pp. 378–384, (2016)
16. Razaque A, Rizvi SS (2017) Secure Data Aggregation using Access Control and Authentication for Wireless Sensor Networks. Journal on Computers and Security 70:532–545
17. Nalini, S., Valarmathi, A.: Fuzzy Association Rule Based Cluster Head Selection in Wireless Sensor Networks. In: 2nd International Conference on Green High Performance Computing", pp. 1–5, (2016)
18. Harb H, Makhoul A, Tawbi S, Couturier R (2017) Comparison of Different Data Aggregation Techniques in Distributed Sensor Networks. Journal of IEEE Access 5:4250–4263
19. Haseeb K, Bakar KA, Abdullah AH, Darwish T (2017) Adaptive Energy Aware Cluster Based Routing Protocol for Wireless Sensor Networks. Journal on Wireless Networks 23(6):1953–1966
20. Khorasani F, Naji HR (2017) Energy Efficient Data Aggregation in Wireless Sensor Networks Using Neural Networks". International Journal of Sensor Networks 24(2):26–42
21. Mohammad M, Subhash C, Rami A (2010) Bayesian fusion algorithm for inferring trust in wireless sensor networks. J Netw 5(7):815–822

An Optimized Hardware Neural Network Design Using Dynamic Analytic Regulated Configuration

V. Parthasarathy, B. Muralidhara, Bhagwan ShreeRam, and M. J. Nagaraj

Abstract Due to the inherent parallelism offered by the Artificial Neural Networks (ANNs) and the rapid growth of Field Programmable Gate Array (FPGA) technology, the implementation of ANNs in hardware (known as Hardware Neural Networks, HNN) for complex control problems have become a promising trend. The basic design challenge in such a realization is the effective utilization of FPGA resources and the high-speed restructuring of the ANN circuits. These two factors are having a direct impact on the size, response time and cost of any HNN system. But these two aspects are competing variables, and controlling one factor may result in the violation of the other. In this paper, a simplified optimization technique known as Dynamic Analytic Regulated Configuration (DARC) is proposed. It has been used as a basic design methodology to increase the performance of an ANN architecture by applying a regulated restructuring on FPGA. A single layer feed-forward ANN system has been considered for this work. A brief explanation has been given about the workflow, the DARC structure and the design challenges. A comparison is provided between the static and dynamic restructuring outcomes. The result shows that optimizing both the above constraints is not effective for larger systems at present due to the design limitations and it can be an effective approach for small-scale implementation.

Keywords Hardware neural networks · ANN reconfiguration · Weights updation · Field programmable gate array

V. Parthasarathy (✉) · M. J. Nagaraj
Nitte Meenakshi Institute of Technology, Bengaluru, India
e-mail: parthasarathy.v@nmit.ac.in

B. Muralidhara
Higher College of Technology, Al-Khuwair, Muscat, Sultanate of Oman

B. ShreeRam
Lovely Professional University, Phagwara, Punjab, India

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_90

1 Introduction

The exponential growth of machine learning algorithms and hardware realization techniques has provided with a solid platform for the digital logic implementation of ANNs for various control system applications [1]. Since HNN design problems are getting momentum only since the last decade, there is very limited literature available that matches with the requirement of an HNN researcher. Some publications related to HNN design using deep learning algorithms that has their own limitations are available [3]. At present, the HNN designs continue to suffer some of the most crucial limitations including the device-dependent, insensitive to rapid modifications and board memory limitations. In some applications, due to the involvement of complex behaviors of the system parameters, the realization becomes a failure.

For example, in the clectric power system restoration process, the reactive power injection of the critically stable region is to be evaluated from the complex Fourier approximation [4]. If an ANN controller has been proposed for such an application, it is near impossible to realize the nonlinear nature of the problem and the controller parameters may not be as accurate as of the solution obtained from the conventional curve fitting or Lagrangian techniques [5].

When the basic building blocks are considered, some of the conventional computational units not only provide sufficient flexibility in the design but they also increase the range of the possible solutions for the given problem. This results in the generation of "system-dependent structures" [6]. Hence a system-independent system design will be the ultimate outcome of any such problem formulation. In this work, the DARC approach has been permitted the hardware to act independently for at least small-size ANN-realized models.

Also because of simplicity, many researchers have opted for the Modeling Period Restructuring (MPR) only. In image identification, robot wheel control, obstacle avoidance and data segregation problems, many authors have proved that the MPR is ineffective for dynamically varying inputs. For reducing the effective resource usage, some researchers [6, 7] had not used the beat possible number of Adjacent Connection Patterns (ACP) through which a neuron communicates with the other.

For example, in a real-time ANN structure, if there are 5 neurons, a whopping 120 possible patterns can be generated using which any given neuron may communicate with its adjacent element. But this will result in the large silicon space requirement. Hence in the practical implementations, very few (even less than 10 for some simple structures) connections were considered. This type of modeling may not be considered as an absolute realization because of the "logical constraint".

In some cases [8, 9] again to make things simple, an alternate approach is proposed. Once the ANN is generated in an FPGA, the actual plant, which is also controlled, will be simulated in a computer and the overall performance will be studied by the I/O measurement from the board [10]. If this process provides a satisfactory output, then the FPGA will be installed in the actual plant and real-time validation will be conducted. The present realization of work using DARC may try to minimize these design limitations.

2 The Hardware Neural Network Redefinition

Redefinition of ANN during the hardware implementation is a complex and rarely addressed issue [11, 12]. The updating of weights, changing or modifying the training patterns, transforming the activation function and structure revision are some of the cumbersome works in the HNN design and are generally termed as the "redefinition" process of ANNs.The redefinition process can happen during the modeling phase or the execution phase. In the modeling phase, well-defined, fixed weights and connections are decided at the beginning itself and the same has been transformed to equivalent hardware on FPGA. Any sort of modification required in the processing period cannot be done unless otherwise the procedure has been started from the beginning [13].

The advantage of such an approach is that there is no need for an auxiliary logical algorithm to permit external reprogramming of connection weights [14]. In such a case, a device resource is not utilized if it is not having any definite role in the operation of ANN during the design time. That is the connections between neurons are active only if the corresponding weights are nonzero. Due to this simple architecture, the design phase has opted for larger and complex ANN structures. But it is having its own limitations.

Because of the frequent repetition of the algorithm from the beginning, the restructuring is a time-consuming process [15]. The high-level coding describes ANN to be regenerated, resynthesized and placed and routed again [15]. Hence to develop a higher order ANN model, the design phase required more processing time. Hence an alternative approach is considered in this work termed as "Dynamic Analytic Regulated Configuration" (DARC) procedure, in which all the possible network connections are physically created when implementing an ANN network. The generated logic for a particular connection which may not be used in the beginning is to be preserved for future reference.

For example, if an ANN requires 20 neurons for its implementation, then in DARC, ANN must contain 400 synaptic weights and their corresponding connections. For this dynamic configuration, the necessary and sufficient condition is that the lookup tables must be addressable externally and rewritable. This step requires some complex coding work for communication and control strategies which have not been included in the conventional design algorithm. Once DARC is completed, the ANN is capable of realizing any architecture with improved speed. Obviously, the time required for the generation of training patterns will be minimal. But it was observed that the additional logics considered for DARC have an adverse impact on the size of the ANN to be implemented. It is to be accepted that selecting an approach between design-time reconfiguration and DARC is a very complicated analysis.

3 Block Diagram Representation of the Proposed System

There are four integrated design blocks considered for the proposed DARC architecture. A block diagram is shown in Fig. 1.

(i) Flow Authority Block (FAB)
(ii) Bit-flow Generator and Converter Block (GCB)
(iii) Hardware Neuron Block (HNB)
(iv) Signal Transferring Block (STB)

The proposed model nearly identical to the conventional HNN design modules but the introduction of the STB which takes care of the dynamic weight correction and updation process makes this system more flexible and faster than the existing systems. Figure 1 shows the functional block diagram of a DARC system. The primary function of the Flow Authority Block is to reset and initialize all the other blocks in case of power-up and to manage the timing cycle of the entire process. A simple 40 MHz onboard crystal oscillator is used for the timing process. If the timing needs of ANN are to be relaxed, the existing clock mechanism can be scaled down by 32 to create a 1.25 MHz base clock. The function of the base clock is to ignite an 8-bit counter which will, in turn, feed three comparators used for initialization work. After a certain number of pulses, the first comparator enables the system pulse generator. On further application of pulses, the bit flow generator is enabled.

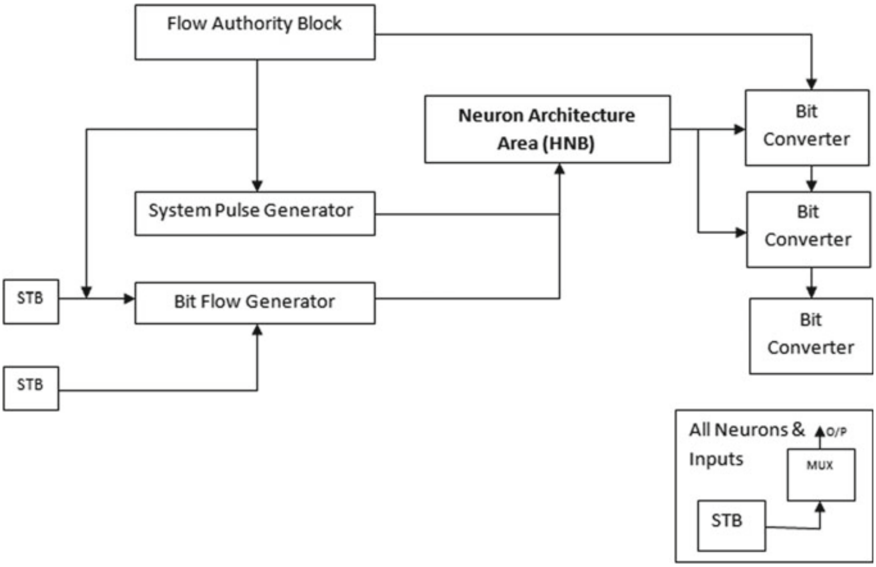


Fig. 1 Block diagram representation of the DARC system

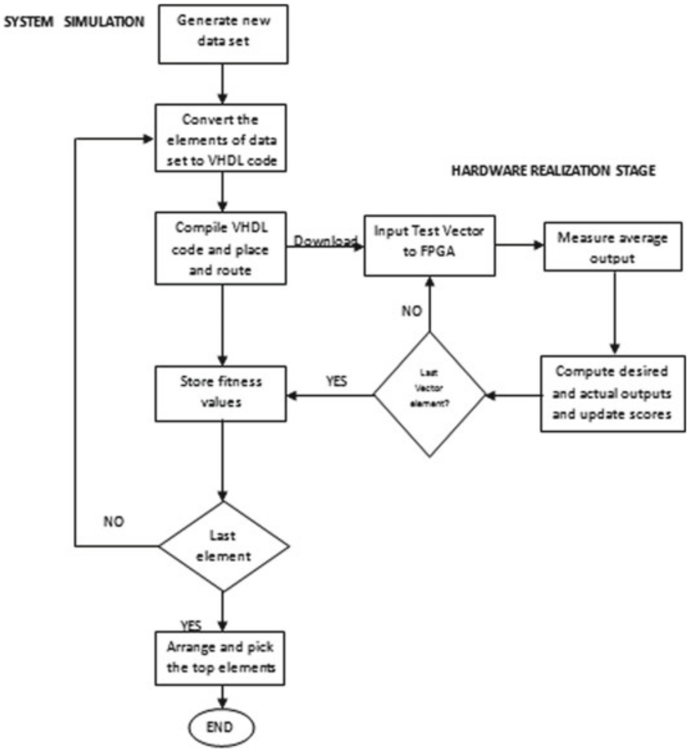

SYSTEM SIMULATION
Generate new data set
Convert the elements of data set to VHDL code
Compile VHDL code and place and route
Download
HARDWARE REALIZATION STAGE
Input Test Vector to FPGA
Measure average output
NO
Last Vector element?
Compute desired and actual outputs and update scores
YES
Store fitness values
NO
Last element
YES
Arrange and pick the top elements
END

Fig. 2 Flowchart for the DARC algorithm

4 The DARC Algorithm

As shown in the flowchart given in Fig. 2, first the arbitrary database is generated, next the DARC algorithm is used to find the better subset input variables for importing into ANN so that the prediction accuracy will be higher. The algorithm is transformed for each training data set into its equivalent program file such that it can be dumped on FPGA. Then the network is analyzed with the set of known input patterns. The response of the system for these known inputs has been compared with the predefined response and the closeness of the identification have been found out.

5 Experimentation Procedure

The DARC System experimental setup consists of a computer system equipped with a suitable multifunction Data Acquisition Board (DAB). The digital input–output pins of the DAB are connected to the Altera APEX DSP board which allows the hardware validation. The training board was tuned such that whatever preprogrammed structure available in an in-built memory will be dumped automatically to the FPGA. The design was developed with 28 redefinable network inputs, a single output and 3 single-layer neurons. The DARC works with the dynamic weight adjustment for every iteration and the system moving toward the convergence. Also, the reconfiguration time taken for the entire DARC process was relatively lesser than the conventional techniques having the basic hardware.

6 Results and Analysis

To evaluate the effectiveness of the sample identification ability of the developed system, 40 subsets of controller setting parameters have been generated using the direct mathematical relation. Apart from that, for knowing the accuracy standards, about 30 very close but still erratic sets were prepared (with an error range of 2% for each parameter). It is expected to keep the HNN to identify the accurate controller parameter sets and neglect the near accurate results when they are given as input to the system. After all the weights and the inputs were fed to the system, the output from the FPGA has been examined. The 8-bit output from the FPGA has been monitored by the data acquisition board. The output was monitored for 40 samples at random intervals for nearly 50 ms.

Then the average value of the output was calculated to conclude about the efficiency of the system. The process was repeated five times and every time the behavior of the developed model has been obtained and compared to finalize the deserved model.

The approximate formula for the evaluation of Realization Level Factor (RLF) is given by the following simple relations given in Eqs. 1, 2, and 3.

The Realization Level Factor (RLF) for controller gain

$$\begin{aligned}\mathrm{RLF}\big(K_{pss}(i)\big) &= 1 - \frac{x_m}{47.2} if x_m < 47.2\\ &= 0 if x_m > 47.2\end{aligned} \tag{1}$$

The RLF for Time constant of Lead network

$$\begin{aligned}RLF(T_1(i)) &= \frac{x_m}{0.95} - 1 if x_m < 0.95\\ &= 0 if x_m > 0.95\end{aligned} \tag{2}$$

The RLF for Time constant of Lag network

$$\begin{aligned}\mathrm{RLF}(T_2(i)) &= 1 - \frac{x_{m-0.2}}{0.08 + K_i} if x_m < 0.08\\ &= 0 if x_m > 0.08\end{aligned} \tag{3}$$

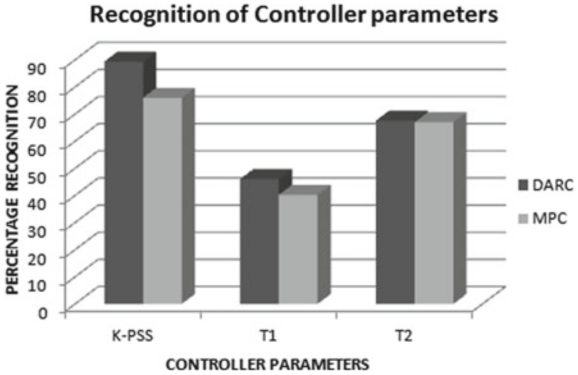


Fig. 3 Realization levels comparison

And the overall Fitness coefficient of the overall system

$$\sum_{i=1}^{S} \frac{Fitness_i}{S} (\mathrm{F}_0) \tag{4}$$

where S = total number of test vectors, x_m = digital circuit output, K_i = correction factor for lag network, $\mathrm{K}_{\mathrm{pss}}$, T_1 and T_2 are the controller (which is to be developed on FPGA) constants. Based on the fitness levels, the number of "fittest neurons" were identified by considering the realization level of 0.3 as threshold and it was found that the number of neurons with higher RL has been obtained from the DARC approach compared to the conventional modeling period configuration. It is an indication of the understanding level of DARC. The following graph shows the relationship between the number of fittest neurons and the various realization ranges.

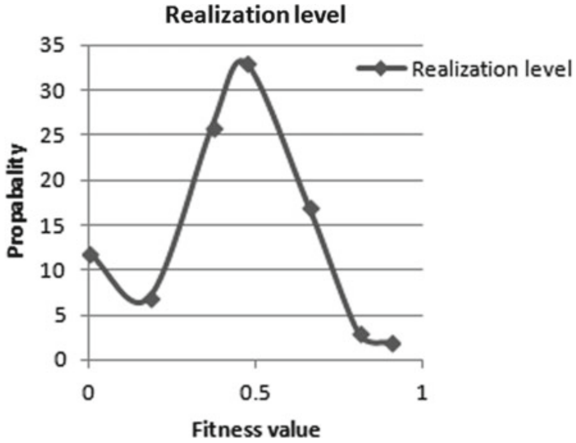
Out of the testing samples considered, to know which region or the range within which more number of fittest neurons were identified by arranging the realization level values, a graph was plotted between fitness value and the probability of occurrence.

It was observed that for the realization level 0.48, about 33 neurons were available. From the design perspective, this value is satisfactory and the structure has been finalized for future study. There is always a scope for increasing the number of fittest neurons by varying the DARC algorithm. Figure 3 gives the probability curve of the fittest neurons.

7 Conclusion

In this paper, a simplified technique termed "Dynamic Analytic Regulated Configuration" (DARC) is proposed mainly to minimize the precious time taken by the processor in the restructuring process of an ANN. A Signal Transfer Block is newly

Fig. 4 Propabality curve for realization levels



included in the conventional modeling period reconfiguration setup and its performance was validated. This permits the system to work even with more flexibility and effective resource utilization. It is also observed that the number of interneural communications has been very slightly improved because the number of connecting patterns used in this work is more than the conventional design time fixation works. The limitation of this work is that it is at present implemented for a 3-input neuron system. When the number of neurons was increased to the higher numbers (tried up to 30 neurons), it was observed that the redefinition is not effectively implemented. Hence it is concluded that there is a need for a detailed analysis before selecting a particular restructuring process. If the given problem involves very less number of neurons (about 10–20) and the response time is the major constraint, the DARC will be a better option wherein when is the number of neurons is very large (in terms of several hundreds), then due to unavailability of FPGA boards, the MPR approach is the only option. It is planned to investigate further to modify the existing DARC algorithm so that this concept can be implemented for larger ANN structures also. There is also scope for redesigning the Signal Transfer Block with some alternate components and structures. Hence, it is concluded that this work is having a lot of future scopes to produce more outcomes in the HNN design problems (Fig. 4).

References

1. Kizheppatt V, Suhaib AFahmy, “FPGA Dynamic and Partial reconfiguration: A Survey of Architectures, Methods, and Applications” ACM Computing Surveys, Volume 51, issue 4, September, 2018, Article No: 2
2. Florian Kastner, Benedikt Janßen, Frederik Kautz, Michael Hubner and Giulio Corradi “Hardware/Software Codesign for Convolutional Neural Networks exploiting Dynamic Partial Reconfiguration on PYNQ”, 2018 IEEE International Parallel and Distributed Processing Symposium Workshop, Proceedings of IEEE Computer society, 2018

3. Janardan Misra, Indranil Saha, “Artificial neural networks in hardware: A survey of two decades of progress”, Elsevier Journal Proceedings of Neuro computing 74 (2010) pp:239–255
4. P. Lysaght J. Stockwood, J. Law, D. Girma, “Artificial neural network implementation on a fine-grained FPGA. In: Proceedings of International Conference on Field Programmable Logic and Applications, Germeny
5. Smon Haykin, Neural Network:A ComprehensiveFoundation, 2nd Ed, Pearson Publication, 2005
6. Beatriz Prieto, Javier de Lope, Darío Maravall, “Reconfigurable Hardware Implementation of Neural Networks for Humanoid Locomotion” proceedings of International Work-Conference on the Interplay Between Natural and Artificial Computation, Almería, Spain, June,2019
7. B. Noory, Voicu Groza, “A reconfigurable approach to hardware implementation of neural networks” proceedings IEEE Canadian Conference on Electrical and Computer Engineering, pp 1861–1864 vol. 3, June 2003
8. Basu Abhirup, Bisaws Pinaki, Ghosh Sarmi, Datta Debarshi (December 2017) Reconfigurable Artificial Neural Networks. International Journal of Computer Applications 179(6):5–8
9. S. Moukhlis, A. Elrharras and A. Hamdoun, “FPGA Implementation of Artificial Neural Networks,” International Journal of Computer Science Issues, Vol. 11, Issue 2, No 1, March 2014
10. Skrbek M (1999) Fast Neural Network Implementation. Neural Network World 9(5):375–391
11. J. G. Eldredge, “FPGA Density Enhancement of a Neural Network Through Run-Time Reconfiguration”,Master’s thesis, Department of Electrical and Computer Engineering, Brigham Young University, May 1994
12. K.P. Sridhar, B. Vignesh, S. Saravanan, M. Lavanya and V. Vaithiyanathan, “Design and Implementation of Neural Network Based circuits for VLSI testing” World Applied Sciences Journal 29 (Data Mining and Soft Computing Techniques): 113–117, 2014
13. Carlos Alberto de Albuquerque Silva, Anthony Andrey Ramalho Diniz, Adrião Duarte Dória Neto and José Alberto Nicolau de Oliveira, “Use of Partial Reconfiguration for the Implementation and Embedding of the Artificial Neural Network in FPGA”,Proceedings of the 4th International Conference on Pervasive and Embedded Computing and Communication Systems (PECCS-2014), pages 142–150 ISBN: 978-989-758-000-0 Copyright c 2014 SCITEPRESS (Science and Technology Publications, Lda.)
14. Esraa Zeki Mohammed, Haitham Kareem Ali, Hardware Implementation of Artificial Neural Network Using Field Programmable Gate Array, International Journal of Computer Theory and Engineering, Vol. 5, No. 5, October 2013
15. Shastri Lokendra (2003) The role of temporal coding in the processing of relational information in the mindbrain. Proceedings of the International Joint Conference on Neural Networks 2:1379–1384

V. Parthasarathy is working as Associate Professor in Electrical and Electronics Engineering Department of Nitte Meenakshi institute of Technology, Bangalore. He is presently perusing his Ph.D. work under VTU, Belgaum. His research interests including Power system Stability, Intelligent prediction systems and Hardware Neural Networks.

Dr. B. Muralidhara is the Professor in the Electrical and Electronics Engineering department of Higher College of technology, (Ministry of man power), Al-Khuwair, Muscat, Sultanate of Oman.He is working extensively on Neural Hardware Realizations , Optimization techniques, Power electronic systems, Drives, ASDs, Signal processing and voltage stability studies

Dr. Bhagwan Shreeram is working as the Professor of Electrical and Electronics Engineering Department of Lovely Professional University, Punjab. His research interests including Renewable energy systems, Modeling, of Non linear systems and Artificial Neural networks.

M. J. Nagaraj is working as Assistant Professor in Electrical and Electronics Engineering Department of Nitte Meenakshi institute of Technology, Bangalore. His research interests are VLSI design, Embedded processors and Renewable energy systems.

Conceptual Online Education Using E-Learning Platform of Cloud Computing

Vivek Sharma, Akhilesh Kumar Singh, and Manish Raj

Abstract Education is the process of procurement of knowledge; accelerating learning, adroitness, worth, credence, and custom. An e-learning platform is built with a goal to provide knowledge to everyone without making them physically attain it and also at a minimum cost without any hassle. e-Learning presents solutions that include Learning Management Systems(LMS) integration, online courses, evaluation engines, and test prep systems on diverse distribution channels such as the Internet and mobile apps with rapidly evolving digital revolution and smartphone growth. Today's e-learning platform requires a huge cost to set up their required resources and various applications related to software. The acceptance of cloud architecture and its computing can help them to reduce expenditure on resources required for the setup of infrastructure, software, and human assets to an extensive level. The paper highlights important aspects of the present e-learning framework and its related models along with issues in present applications of e-learning. The paper also throws light on cloud computing principles and analyzes benefits for adopting them. A solution for applications related to e-learning using cloud computing is proposed herewith.

Keywords Online learning · Online education · E-Learning · Cloud architecture

1 Introduction

Education plays out an imperative situation in our achievement in non-open development. The more we learn, the more we develop. Education causes us numerous things like making viable a character to think, question, and see away from the self-evident.

V. Sharma (✉) · A. K. Singh · M. Raj
GLAU Mathura (uP), Mathura, India
e-mail: viveksharma.cea@gla.ac.in

A. K. Singh
e-mail: akhileshkr.singh@gla.ac.in

M. Raj
e-mail: manish.raj@gla.ac.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_91

Education is the quality method to delight our interest and want to break down more. Instructive innovation, every once in a while named EdTech, is the assessment and good demonstration of empowering e-realizing, which is the acing and improving execution by making, utilizing, and administering impeccable creative procedures and sources.

By advanced innovation, we mean the utilization of PC and innovation helped systems to help to learn inside schools. Approaches around there fluctuate broadly, however, for the most part include the following.

Technology for understudies, where students use projects or applications intended for critical thinking or open-finished learning; or technology for instructors, for example, intuitive whiteboards or learning stages.

Making an advanced learning condition isn't just about accommodation for understudies, it's tied in with setting them up for the future, says Renee Patton.

Advanced learning conditions not only include reading, listening, and writing. But, visualization effects in form of animations and videos makes a lot of difference.

Proof recommends that innovation approaches ought to be utilized to enhance other educating instead of supplanting progressively customary methodologies. It is improbable that specific advances realize changes in adapting straightforwardly. However, some can possibly empower changes in educating and learning associations. For instance, they can bolster instructors to give progressively successful input or utilize increasingly accommodating portrayals, or they can persuade understudies to rehearse more.

2 Concepts and usage of E-Resources in Online Learning

2.1 A. Transformation from Physical Learning to E-Learning

Online learning or e-learning exploits advances and correspondence frameworks to improve the learning experience. It can possibly change the manner in which we instruct and learn no matter how you look at it. It can increase expectations, and extend support in long-lasting learning. It can't supplant instructors and speakers, yet close by existing strategies it can improve the quality and reach of their educating, and decrease the time spent on organization.

Online learning or e-learning is particularly important because people find that it can make a significant difference in so many areas, e.g., on how fast and efficiently they learn a skill and how simple it is to study and enjoy learning. It has the capacity to make a contribution to learning: to improve its standards; to improve its quality; casting off boundaries to studying and involvement in learning; planning and preparation for employment; uplifting at job site; and ensuring that each and every learner achieves the height of their potential.

2.2 Online Learning or E-Learning Framework Model

The conventional model of instruction is homeroom based on the teacher drove preparing. The new worldview is online separation instruction. Web 2.0 [1, 2] advances make the conveyance of instruction substance increasingly intelligent and urge understudies to learn. The online learning frameworks alter the content of course in light of the client's capacity. The courseware personalization makes it simpler and allows clients to learn at their own speed, hence giving greater adaptability in the learning process. Online learning can be conveyed by various frameworks depending on the data transmission and the gadgets used to retrieve and access the understudies. Telesubmersion condition framework uses the video symbol and virtual board which gives understudies the sentiment in a classroom condition, animating up close, and personal study hall experience. With the 3-Dimensional empowered video multicasting and broadcasting, Tele-immersion [3] will be generally acknowledged with the help of understudies. The downside is the underlying expense of the venture for high goals video recorder gadgets. Huge bandwidth is likewise essential for moving information, and clients getting to gadgets must have a better video card and framework arrangement. Before the origin of Web 2.0 innovation, course structures were intended to the clients for accessing with relatively low data transfer capacity systems.

Clients didn't require very good quality PCs to get to the substance. Despite the fact that the personalization choice was accessible, the courseware didn't comprise high goals illustrations and video substance. Crossover Instructional framework is considered as a mix of the customary study hall and online learning concepts. Clients still are supposed to be present in the class, and what's more, they have the option to get to the content of the course product through online learning. This mix facilitates both and allows them in helping the understudies to move from study hall preparing to online learning mode. Understudies can adjust to the first half and the rest half framework as there is a smooth change. The courseware may be of intense impact point class introduction, referred and referenced books, understudy online journals, 3-Dimensional based, symbols, and so forth.

2.3 E-Learning Methodology

The way of teaching has been the same from time immemorial. There have been significant advancements in every field of life but not much in the field of education. The framework of lecture room type education is primarily a trainer based training. The latest change in context is online education. Web-based technologies and the growing advancements in the IT industry make the delivery of learning ingredients more experientially active and interactive and encourages students to understand and gain knowledge. The online learning platforms personalize the course content with respect to the user's ability to learn. e-Learning [4] can be represented by different frameworks and methods mostly emphasized on the network connection and the various equipment used in order to be accessible by the students. Self-learning and study is the most commonly accepted method that makes use of Wikipedia, reading materials like text files, PowerPoint files, and blogs to provide knowledge to the users. Videotape and audio podcast are the next commonly used methodology in creating demonstration-based graphics animations and videos to teach their users. They are useful in inventing a way of learning which allows the users to know and gain knowledge by viewing. Blended e-learning mixes both the Asynchronous and the Synchronous methods of learning. It combines online learning with classroom learning, where students can partially control the pace, time, and place of their learning. Mobile learning provides users an interactive environment where they can learn without actually requiring additional cost computers/laptops to get the knowledge. Other e-learning methods can be Social learning, Game-based learning, Computer-based training.

3 Problem Statement

The main focus of this exploration was to look into the best assets available that can ease the difficulties numerous e-learning framework engineers face specifically for stages that deliver a generous volume of sight and sound substance. The problems experienced in e-learning frameworks as a medium used to improve learning and spread of data have specially added to the planned journey toward an answer that can best be actualized in e-learning frameworks. The quest for such an answer was considering tending to the difficulties at the grass-roots level that is at the development phase of e-learning frameworks. Having viewed that numerous corporations can dispatch and host e-learning primarily based on frameworks to improve their gaining knowledge of criteria, there is a confinement of greater room.

4 Cloud Concepts

Putting in simple words, Cloud Computing [5] is the way of storing and accessing data over the Internet where cloud is just a metaphor for the Internet. It can provide a natural platform to furnish a guide to e-learning systems. It is also a fascinating technology for the instructing institutes with its dynamic scalability and utilization of virtualized sources, as a service via the Internet Cloud computing [6] is the use of empty assets of PC to expand productivity through improving use rate and lessening vitality utilization, one of the answers for decrease greenhouse impact.

4.1 Cloud Advantages

Digital content costs drastically less than printed material. These enable the university and it's students to have a cost effective solution for their studies. Cloud-based [7] materials can also be without difficulty updated so students continually have got the right of entry to the most current learning resources.

The cloud [8] also gives consistent incorporation among resources and college offices by means of record sharing, all through any device, so the workforce never again wants a specific medium where to team up or speak with others. For understudies, the quality of joint effort is likewise key with regards to becoming more acquainted with and preparing for the workforce. Understudies can work inside computerized homerooms, talking with college students far away, in real time, and mix the schooling of human beings from definitely exclusive backgrounds.

For example, open-access online learning apps such as Moodle is widely accepted by educational institutions.

4.2 Cloud Delivery Model

Online learning applications can be implemented using several services of cloud providers. Google, AWS, and Microsoft [4] provide cloud services as an on-demand or at various subscription-level plans. The architecture of the cloud can be considered consisting mainly of four layers [9]:

(1) The Physical Layer: The Physical Layer contains physical servers, network, and different components that can be physically managed and controlled.
(2) The Infrastructure Layer: Second layer in a cloud architecture that permits the Infrastructure as a Service (IaaS) customers to assemble and dissemble virtual machines along with their network as per their own business requirements. It includes storage facilities like virtual servers and networking.
(3) Platform Layer: Platform Layer serves as a platform for development and deployment. It presents the right platform for improvement and deployment of applications.
(4) Application Layer: The Application Layer is the one where end users have interactions in a direct manner. It mainly consists of software systems delivered as service. Examples are Gmail and Dropbox.

These four layers enable the user to use cloud computing services efficiently and acquire the results they are looking for from the system.

5 Proposed Solution

Based on the above results, it can be considered that cloud computing could provide a far better solution to the growing e-learning platforms. The e-learning platforms dealing with the hassle is typically concerning the lack of storage capabilities, and the updates and the hardware.

For the better utilization of cloud computing services provided by the different cloud providers, we can implement the services provided by AWS [10]. It provides a far better option to interact with the services and carry out the work seamlessly without any hassle. It provides several ways, but the two most suited to get the job done could be the following.

The first is manually running the instances and managing the scale in and scale out with the help of auto-scaling, connecting with the RDS instance to save the details of every new and existing user, and installing the lamp or wamp server to host the application on the Internet. And if we want to provide the application with its domain name, then contacting the domain name provider.

The other solution could be using the other services provided by AWS, such as AWS Beanstalk [11] and AWS Lightsail [12], where we need not worry about the availability and storage, it's all provided by the provider, and we need to carry our code and deploy it using AWS Beanstalk [13]. Our application is ready for hosting around the Internet. Lightsail is used to create the e-learning platform from scratch.

Lightsail provides ease of use of a cloud platform which offers everything required for building a website or an application, along with an affordable cost-effective, monthly basis plans. We can use it without worrying about the instances running or not and also about the instances being available or the load to the website or the application. It manages everything; we need to create a Lightsail instance and choose the WordPress that has its database running in its background, manage the users, and authore them against the details in the database.

We can host the application or website with the help of Lightsail at just monthly plans on on-demand basis.

We can also use other cloud providers such as Google Cloud Platform (GCP) [14], Microsoft Azure [15], or the IBM Watson [2]. They also provide a number of services the same as AWS in order to obtain the desired outcome.

6 Conclusion

Cloud-based learning systems are emerging as a good option for online learning and online education purposes. They help in decreasing the cost and provide reliable storage of data and sharing. They also allow countless and limitless opportunities for end users to use the Internet. However, it is mandatory for the users to have an Internet connection and the low speed can decrease the efficiency of provisioning of e-learning.

Innovation is being utilized progressively by foundations to give e-learning administrations. These establishments face a wide range of difficulties in actualizing these frameworks, for example, costs, a need for specialized assets, and obstruction by key partners to the execution of frameworks. Cloud-based learning frameworks are rising as an alluring strategy for giving e-learning administrations. They can diminish costs because of lower prerequisites of equipment and programming, and less requirement for on-location support. They are too simple to convey over various areas as they are midway regulated. They additionally offer advantages to end clients as far as openness, security, and similarity are concerned.

References

1. Bento Al (2011) Cloud computing: a new phase in information technology management. J Informat Technol Manag 22(1):39–46
2. https://lightsail.aws.amazon.com/ls/docs/all
3. Satyanarayana S (2012) Cloud computing: SAAS. J Comput Sci Telecommun 4(4):76–79
4. Gartner (2008c) Gartner says worldwide SaaS revenue in the enterprise application markets will grow 27 Per Cent in 2008. Gartner Press Release, 22 October 2008
5. Miller M (2008) Cloud computing: web-based applications that change the way you work and collaborate online. Que Publishing, Indianapolis
6. Reese G (2009) Cloud application architectures. O'Reilly Media, Sebastopol, CA
7. Sun (2009a) A guide to getting started with cloud computing. Sun White Paper. https://www.sun.com/offers/docs/cloud_computing_primer.pdf. Accessed 10 Jun 2009
8. Foster I, Zhao Y, Raicu I, Lu S (2008) Cloud computing and grid computing 360-degree compared
9. National Institute of Standards and Technology, The NIST Definition of Cloud Computing, Information Technology Laboratory (2009)
10. http://aws.amazon.com/what-is-aws/
11. https://aws.amazon.com/elasticbeanstalk/
12. https://docs.microsoft.com/en-us/azure/
13. https://docs.aws.amazon.com/elasticbeanstalk/latest/dg/awseb-dg.pdf
14. Sridhar S, Smys S (2016) A survey on cloud security issues and challenges with possible measures. In: International conference on inventive research in engineering and technology, vol 4
15. Wireless sensor network: a survey. Int J Eng Innovat Technol (IJEIT), May 2015

Efficient Deployment of a Web Application in Serverless Environment

Vivek Sharma, Akhilesh Kumar Singh, and Manish Raj

Abstract In the recent era to gain speed up, scale-up/down, and for cost-effective features, people are moving toward services provided by AWS or other computing paradigms, (salesforce, etc.) i.e., AWS Lambda which is applicable to build a serverless environment. Serverless environment doesn't mean there is no server but tells AWS to not only give all services which are provided by the server but also give metering on a pay per second basis. This project will entertain users with a full stack-based interface where they can request a unicorn ride from any place they want.

Keywords Amazon Web Service Lambda · Amazon Application Programming Interface · Gateway · Amazon S3 · Amazon DynamoDB · Amazon Cognito and serverless architecture

1 Introduction

Serveless in Amazon Web Service simply means serverless user interface. Serverless user interface is that which doesn't require to supervise servers. By working in a serverless environment, we are free from these inerrancies, i.e., compartmentalization of resources, scalability, and standardization. To be a serverless platform, it is necessary for it to provide several functionalities. There is no server supervisioning, adjustable scalability, and no idle capacity.

The AWS cloud consists of many services which directly or indirectly come in the scenario of a serverless web application. These are Amazon Web Service Lambda for

V. Sharma (✉) · A. K. Singh · M. Raj
GLAU Mathura (UP), Mathura, India
e-mail: viveksharma.cea@gla.ac.in

A. K. Singh
e-mail: akhileshkr.singh@gla.ac.in

M. Raj
e-mail: manish.raj@gla.ac.in

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8_92

computation [1], Amazon Application Programming Interface gateway for APIs [2], Amazon Simple Warehouse Service (Amazon S3) to provide storage [3], Amazon DynamoDB for databases [4], Simple Notification and Queue Service by Amazon (ASNQS, a combination of SNS and SQS service by Amazon) [5] and for inter-process managing [6], Amazon Web Server Step Functions [7] and Amazon Cloud Watch Events [8] for orchestration.

2 Literature Review

The first public cloud infrastructure vendor to offer abstract serverless computing AWS Lambda in 2014 was Amazon. Back in the year 2006, a cloud warehouse and storage service were launched by Amazon: Amazon Web Service S3 which presents a warehouse service that performs functions without the requirement of handling the maintenance of the servers. Later in late 2014, a serverless platform was launched by Amazon Web Services in which computation service became serverless. And it was from this point onwards that the new paradigm of cloud revolution began not so long after, that in 2016, other major cloud vendors released serverless platforms to compete with already existing Amazon Web Services, such as Google, Microsoft and IBM. Almost instantly, serverless computing started gaining popularity with organizations opting for the new technology to transform their businesses.

3 Preliminaries

The main emphasis of the paper presents an approach to provide high availability, flexible scaling, and reducing the overhead involved in server management. To achieve this, certain AWS services are used that are the key components in obtaining a serverless environment. Basically, the 4-step scheme is followed for developing a serverless environment. The four steps are—Hosting a non-dynamic web HTML, Management of end users, Building a serverless back end, and Restful Application Programming Interface.

4 Proposed Serverless Web Application

4.1 Description

In the proposed scenario for serverless web services-based application, we have developed a simple user interface which enables users to have a unicorn ride through the serverless platform from anywhere. In this, the user will request a unicorn from the front-end and their request will be accepted through restful web services (back end) and the unicorn will be dispatched to the nearby place. This project has been done on different levels. These levels are

1. **Computation level**:

The computation level manages the queries coming via external systems or end users. Authorization of the requests is also done here. It has the runtime platform where you can deploy your business logic.

- Amazon Web Services Lambda [1] runs serverless applications on managed platform which enables microservices, deployment, etc.
- Using Amazon Web Services Application Programming Interface gateway [2], we can deploy our business logic on fully managed rest Application Programming Interface integrated with Amazon Web Services Lambda.

The data level gives secure storage for states which our business logic has.

2. **Data level**:

- Amazon DynamoDB [4] provides us with managed no Structured Query Language database for storage purposes.
- Amazon Simple Storage Service [3] provides object storage for any range of use cases.

2. **User Management and Identity Level**:

User Management and Identity Level provides authentication, authorization, and identity to the customers (internal or external both).

- Amazon Cognito [9] provides simple, secured end-user registration, sign up, log in and log out, and user control.

4.2 Working of the Proposed Serverless Web Application

In our proposed scenario, Amazon Simple Storage Service helps in hosting the non-dynamic resources feasible for the web application. The simple logic behind this architecture is that you have to store all your static resources which include HyperText Markup Language, Cascading Style Sheet, Images, Multimedia, and other files. The user can directly view the content through the URL address spread by Amazon S3. You need not have to provision or supervise any server for it. Firstly, in Amazon Management console, select S3 under storage in service [10]. Create an S3 bucket for storing your static resources using a globally unique name [11], upload your content into it by selecting or dragging files, add the bucket policy to it by using the permission tab which allows the public access, and then save the bucket policy code. Select a static hosting card where the first field can be filled with index.html and other can be made blank but note that before saving this, keep in mind the endpoint URL address because after this you are going to use your web application using this URL address only, and then save it [12]. In Amazon Cognito you are going to register yourself in the web application which requires your email address and password and then it sends a verification code to your email address for authentication, and authorization. Then at login, a javascript function will communicate with Amazon Cognito and get back a set of keys for JSON Web Token which claims the identity of the user. Go to Cognito under services and select the manage your user pool and then click create user pool; create your pool with Wild Rydes and select review default and then create pool (write down pool id ap-south-1_qMZk3inhv), add an application to the user pool; from general settings click app clients and from there create app client and name this from Wild Rydes web application and uncheck the generate secret option and then save it and write down the application client id 386914j6o6upaa25ab42b6p5vo. Then update the config.js from pool id, application client id, invoke URL address, etc., and then resubmit it to S3 bucket which was made. Then on the web application register yourself. If registration becomes successful, you should get the notification that the application programming interface has not been configured [13–16].

AWS Lambda and Amazon DynamoDB are for building the serverless back ends for the request handling. You will implement the lambda function, each time user requests a unicorn. The module will choose a unicorn amongst a fleet and save the query in Amazon DynamoDB; once done reply to the front-end app with information of theunicorn going to be dispatched. Go to dynamo under services, select create table and name your table with Rides and the key with Rideid having the type string and by checking default checking box, make your table. (Note the ARN arn:aws:dynamodb:ap-south-1:741534142387: table/Rides). Then, to know what Amazon services can interact with the need to have an IAM role for which IAM under services should be selected, create your role by going to the left navigation bar 'roles'. Select lambda for the role type and then click next, give your role name with Wild Rydes Lambda and then create a role. Then on clicking on the role, select add inline policy, choose the service DynamoDB. Once done select the action put_item, add the ARN of the table in resources. Choose review policy and click creates then goes the lambda menu under service tab and creates a function over there. Write RequestUnicorn in name field and choose node.js 6.10 at execution time, write the existing role created and click to create function and update the index.js code with requestunicorn.js. Once done then click save. Then configure test event and add your event name TestRequestEvent and update the code. When the tested log appears, then this phase is completed [17–20].

Through restful APIs, you are going to invoke your lambda function and this static hosted website will convert into a dynamically hosted one by bringing the source side javascript which allows Asynchronous JavaScript and XML calls the shown Application Programming Interface. In Amazon Management Console, go to Application Programming Interface gateway under services and then click on create Application Programming Interface and then on new Application Programming Interface, name your Application Programming Interface (Wild Rydes) and in endpoint type select edge optimized. Then click on the create Application Programming Interface button. We are going to configure our Application Programming Interface authorizer which we have created to consume the user pool we made in Amazon Cognito. Under your newly created Application Programming Interface, choose authorizer, choose to create authorizer Application Programming Interface, name your authorizer, select Cognito, and fill the region where you created your user pool in Amazon Cognito. Then, enter your user pool name and then write Authorization in token source and then click create Go to resources under your Application Programming Interface and in action click on create resource, enter your resource name (Ride) and ensure that your resource path is ride, choose Enable Application Programming Interface Gateway CORS and then choose create resource button for the recently created resource, under action click create method. Select post on the new drop-down that appears, select lambda function for integration type and then check on the checkbox of use lambda proxy integration, select the region in which we have created our lambda function, and enter the name of our previously created lambda function and click save.

Choose deploy Application Programming Interface, and select new stage in the development stage drop-down menu. Enter stage name (prod) and choose deploy Application Programming Interface. Note the invoke Application Programming Interface URL. Update your config.js file and upload the changed file to Amazon Storage. Visit ride.html under your website domain; if redirected to the sign-in page, then sign in.

References

1. https://aws.amazon.com/lambda/
2. https://aws.amazon.com/api-gateway/
3. https://aws.amazon.com/s3/
4. https://aws.amazon.com/dynamodb/
5. https://ap-south1.console.aws.amazon.com/sns/v3/home?region=ap-south-1#/dashboard
6. https://ap-south-1.console.aws.amazon.com/sqs/home?region=ap-south-1
7. https://ap-south-1.console.aws.amazon.com/states/home?region=ap-south-1#/homepage
8. https://ap-south-1.console.aws.amazon.com/cloudwatch/home?region=ap-south-1#
9. https://aws.amazon.com/cognito/
10. https://s3.console.aws.amazon.com/s3/buckets/wildrydes-dolly-sen/?region=ap-south-1&tab=properties
11. https://s3.console.aws.amazon.com/s3/buckets/wildrydes-dolly-sen/?region=ap-south-1&tab=permissions
12. https://s3.console.aws.amazon.com/s3/buckets/wildrydes-dolly-sen/?region=ap-south-1&tab=overview
13. https://ap-south-1.console.aws.amazon.com/cognito/home?region=ap-south-1?
14. Miller M (2008) Cloud computing: web-based applications that change the way you work and collaborate online. Que Publishing, Indianapolis
15. Reese G (2009) Cloud application architectures. O'Reilly Media, Sebastopol, CA
16. Sun (2009a) A Guide to Getting Started with Cloud Computing. Sun white paper. https://www.sun.com/offers/docs/cloud_computing_primer.pdf. Accessed: 10 June 2009
17. Foster I, Zhao Y, Raicu I, Lu S (2008) Cloud computing and grid computing 360-Degree Compared

18. RightScale (2008) Define Cloud Computing. RightScale Blog, 26 May 2008. http://blog.rightscale.com/2008/05/26/define-cloud-computing/. Accessed 9 Jun 2009
19. Gartner (2008c) Gartner Says Worldwide SaaS Revenue in the Enterprise Application Markets Will Grow 27 Per Cent in 2008. Gartner press release, 22 October 2008
20. Singh AK, Singh P An approach for web based GIS Route Finder System, Int J Advanc Res Comput Sci Softw Eng (IJARCSSE)

Author Index

V. Suma et al. (eds.), *Evolutionary Computing and Mobile Sustainable Networks*,
Lecture Notes on Data Engineering and Communications Technologies 53,
https://doi.org/10.1007/978-981-15-5258-8