Lecture Notes in Electrical Engineering 475

Siddhartha Bhattacharyya
Tapan Gandhi · Kalpana Sharma
Paramartha Dutta
Editors

Advanced Computational and Communication Paradigms

Proceedings of International Conference on ICACCP 2017, Volume 1

Springer

Lecture Notes in Electrical Engineering

Volume 475

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, SCOPUS, MetaPress, Springerlink ****

Lecture Notes in Electrical Engineering (LNEE) is a book series which reports the latest research and developments in Electrical Engineering, namely:

- Communication, Networks, and Information Theory
- Computer Engineering
- Signal, Image, Speech and Information Processing
- Circuits and Systems
- Bioengineering
- Engineering

The audience for the books in LNEE consists of advanced level students, researchers, and industry professionals working at the forefront of their fields. Much like Springer's other Lecture Notes series, LNEE will be distributed through Springer's print and electronic publishing channels.

For general information about this series, comments or suggestions, please use the contact address under "service for this series".

To submit a proposal or request further information, please contact the appropriate Springer Publishing Editors:

Asia:

China, *Jessie Guo, Assistant Editor* (jessie.guo@springer.com) (Engineering)

India, *Swati Meherishi, Senior Editor* (swati.meherishi@springer.com) (Engineering)

Japan, *Takeyuki Yonezawa, Editorial Director* (takeyuki.yonezawa@springer.com)
(Physical Sciences & Engineering)

South Korea, *Smith (Ahram) Chae, Associate Editor* (smith.chae@springer.com)
(Physical Sciences & Engineering)

Southeast Asia, *Ramesh Premnath, Editor* (ramesh.premnath@springer.com)
(Electrical Engineering)

South Asia, *Aninda Bose, Editor* (aninda.bose@springer.com) (Electrical Engineering)

Europe:

Leontina Di Cecco, Editor (Leontina.dicecco@springer.com)
(Applied Sciences and Engineering; Bio-Inspired Robotics, Medical Robotics, Bioengineering; Computational Methods & Models in Science, Medicine and Technology; Soft Computing; Philosophy of Modern Science and Technologies; Mechanical Engineering; Ocean and Naval Engineering; Water Management & Technology)

(christoph.baumann@springer.com)
(Heat and Mass Transfer, Signal Processing and Telecommunications, and Solid and Fluid Mechanics, and Engineering Materials)

North America:

Michael Luby, Editor (michael.luby@springer.com) (Mechanics; Materials)

More information about this series at http://www.springer.com/series/7818

Siddhartha Bhattacharyya
Tapan Gandhi · Kalpana Sharma
Paramartha Dutta
Editors

Advanced Computational and Communication Paradigms

Proceedings of International Conference on ICACCP 2017, Volume 1

Springer

Editors
Siddhartha Bhattacharyya
Department of Computer Application
RCC Institute of Information Technology
Kolkata, West Bengal
India

Tapan Gandhi
Department of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi, Delhi
India

Kalpana Sharma
Department of Computer Science
and Engineering
Sikkim Manipal Institute
of Technology
Rangpo, Sikkim
India

Paramartha Dutta
Department of Computer
and System Sciences
Visva-Bharati University
Santiniketan, West Bengal
India

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-10-8239-9 ISBN 978-981-10-8240-5 (eBook)
https://doi.org/10.1007/978-981-10-8240-5

Library of Congress Control Number: 2018931454

Printed on acid-free paper

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

Siddhartha Bhattacharyya would like to dedicate this volume to his beloved wife Rashni, his cousin sisters Lakshmi, Rama, Soma, Saraswati, Mita, Dipa, Ruma and Runa

Paramartha Dutta would like to dedicate this volume to his father Late Arun Kanti Dutta and mother Bandana Dutta

Preface

With the progress of technology, the need for advanced computational techniques is always called for addressing the complex real-life problems. The objective of such computational paradigm is to give rise to fail-safe and robust solutions to the emerging problems faced by mankind. Typical applications encompass the fields of computer networking and data communication, cyber security, signal processing, computer vision and image processing, computational perception and cognition, human–computer interaction, adaptive computation and machine learning to name a few.

This volume of Lecture Notes in Electrical Engineering contains accepted papers presented at ICACCP 2017, the First International Conference on Advanced Computational and Communication Paradigms 2017. The volume is aimed to introduce to the prospective readers the latest trends in advanced computing technologies and communication paradigms. This conference was the first of its kind which was hosted by the Department of Computer Science and Engineering, Sikkim Manipal Institute of Technology, Majitar, Rangpo, East Sikkim, Sikkim, and technically collaborated with Computer Society of India (CSI), Kolkata.

The aim of ICACCP 2017 was to address significant areas of research and development in advanced computational and communication paradigms and cyber security, thereby providing due coverage of the advanced computational paradigms and communication techniques which give rise to fail-safe and robust solutions to the emerging problems faced by mankind and immense potential for future innovations and applications.

Sikkim is a northeastern state of India and is often preferred as a tourist spot due to the scenic beauty of nature amidst mountains and hills.

ICACCP 2017 received a tremendous response from the academic fraternity. There were more than 550 odd technical paper submissions. Finally after peer-reviewing, 185 high-quality papers were accepted and 142 papers were registered for oral presentation and possible publications in Springer (AISC and LNEE) series. This volume of LNEE series comprises 65 papers. In addition, 9 general sessions and 5 special sessions were scheduled in ICACCP 2017.

The organization of the ICACCP 2017 conference was entirely voluntary. The review process required an enormous effort from the members of the International Technical Program Committee, and we would therefore like to thank all its members for their contribution to the success of this conference. We would like to express our sincere thanks to the host of ICACCP 2017, Sikkim Manipal Institute of Technology, and to the publisher, Springer, for their hard work and support in organizing the conference.

Finally, we would like to thank all the authors for their high-quality contributions. The friendly and welcoming attitude of conference supporters and contributors made this event a success!

Kolkata, India — Siddhartha Bhattacharyya
New Delhi, India — Tapan Gandhi
Rangpo, India — Kalpana Sharma
Santiniketan, India — Paramartha Dutta
September 2017

Contents

About the Editors

Dr. Siddhartha Bhattacharyya [FIETE, FIEI, SMIEEE, SMACM, SMIETI, LMCSI, LFOSI, LMISTE, MIET (UK), MIAENG, MIRSS, MIAASSE, MCSTA, MIDES, MIISIP] did his bachelor's degree in Physics, bachelor's degree in Optics and Optoelectronics, and master's in Optics and Optoelectronics from University of Calcutta, India, in 1995, 1998, and 2000, respectively. He completed his Ph.D. in Computer Science and Engineering from Jadavpur University, India, in 2008. He is the recipient of the University Gold Medal from the University of Calcutta for his Master's. He is the recipient of the coveted Adarsh Vidya Saraswati Rashtriya Puraskar in 2016; Distinguished HoD and Distinguished Professor Awards conferred by Computer Society of India, Mumbai Chapter, India, in 2017; and Bhartiya Shiksha Ratan Award conferred by Economic Growth Foundation, New Delhi, in 2017. He received the NACF-SCRA, India Award for Best Faculty for Research in 2017. He received Honorary Doctorate Award (D. Litt.) from the University of South America in 2017. He is also the recipient of the South East Asia Regional Computing Confederation (SEARCC) International Digital Award ICT Educator of the Year at Colombo, Sri Lanka, in 2017. He is currently the Principal of RCC Institute of Information Technology, Kolkata, India. In addition, he is also serving as the dean (Research and Development and Academic Affairs) of the institute. He is a co-author of 4 books and the co-editor of 8 books and has more than 190 research publications in international journals and conference proceedings to his credit. He has got a patent on intelligent colorimeter technology.

Dr. Tapan Gandhi is currently working as Assistant Professor in the Department of Electrical Engineering, IIT Delhi, and also research affiliate to MIT, USA. He received his Ph.D. fellowship from PROJECT Prakash (MIT, USA) and obtained his Ph.D. from IIT Delhi in Biomedical Engineering. Following his Ph.D., he has spent 3+ years as Postdoctoral Research Scientist at MIT, USA. He was also awarded an INSPIRE Faculty in the Engineering and Technology Category of the Department of Science and Technology, Government of India. His research expertise spans from computational neuroscience, brain imaging, assistive technology, biomedical instrumentation, machine learning, cognitive computing to

artificial intelligence. He has published papers in top ranking journals like *Nature*, *PNAS, Current Biology, PloS Biology*. He has more than 60 publications in international journals and conference proceedings. He is PI & Co-PI of projects from industry as well as Government of India organizations.

Dr. Kalpana Sharma is currently Professor and Head of Computer Science and Engineering Department at Sikkim Manipal Institute of Technology, Sikkim Manipal University. She has pursued her B.E. from National Institute of Technology, Silchar, M.Tech. from Indian Institute of Technology, Kharagpur, and Ph.D. from Sikkim Manipal University. She has over 19 years of experience in academics. She is currently guiding two research scholars in the allied fields. She has 29 research publications in various national and international journals and conferences. Her areas of interest include security in wireless sensor networks (WSNs) and real-time systems. She has authored a book in WSN. She has executed many funded projects in the fields related to security in WSN and real-time task scheduling.

Dr. Paramartha Dutta was born in 1966, and is B. Stat. (Hons.) (ISI, Calcutta), M. Stat. (ISI, Calcutta), M. Tech. (Computer Science; ISI, Calcutta), and Ph.D. (Engineering; IIEST, Shibpur), SMIEEE, SMACM, SMIACSIT, FIETE, FIE, FOSI, LMISTE, LMIUPRAI, LMISCA having academic experience of about 23 years. He is presently a Professor in the Department of Computer and System Sciences, Visva-Bharati University. He has co-authored eight books and has also seven edited book to his credit. He has published more than 200 papers in various journals and conference proceedings, both international and national as well as several book chapters in edited volumes of reputed international publishing house like Elsevier, Springer-Verlag, CRC Press, John Wiley. He has guided four scholars and is guiding six for their Ph.D. He has served as guest editor of special volumes of international journals published by publishers such as Elsevier, Springer.

Design of Low-Noise Amplifier with High CMRR for Sensor Application

Debashis Jana(✉), Sumalya Ghosh, R S S M R Krishna, Sanjukta Mandal, and Ashis Kumar Mal

Department of Electronics and Communication Engineering, National Institute of Technology Durgapur, Durgapur 713209, India
{debashisjana,akmal}@ece.nitdgp.ac.in
{sg.12ece1106,rssmrk.16ec1101,sm.16ec1102b}@phd.nitdgp.ac.in

1 Introduction

In last few decades, there has been a growing interest on design of low-frequency signal conditioning system for sensor application specially for biomedical electronics and seismic detection [1–3]. In such cases, the differential amplifier is the key element which must be able to amplify the low-frequency signal detected by the sensors and reject the undesired noise. So, an amplifier with high gain and high CMRR is the basic necessary block for Analog Front-End (AFE) design. Apart from these issues, we are now in deep submicron region in CMOS technology where matching is a major problem due to scaling of the device. Therefore, mismatch is a critical factor in CMRR of differential amplifier for any kind of analog design. Other important factors are dopant variation and lithographic errors with the technology variation. So in this scenario, Differential difference Amplifier (DDA) is very suitable to realize high value of CMRR for its special architecture. Here, DDA has been designed with high-gain, high-CMRR and low-input referred noise using 0.18 m CMOS Technology.

2 Available Topologies for Low-Noise Amplifier

There are different topologies to achieve the challenging noise–area–power tradeoff in AFE design required for low-frequency signal conditioning application. The most common circuit topologies for implementing low-noise operational transconductance amplifiers (OTAs) are Open Loop Network (OLN), capacitive feedback network (CFN), Miller integrator Network (MIFN), Capacitive feedback Network (CAFN), etc.

Based on their architecture, different conclusions can be derived regarding the performance of the different LNA topologies as follows.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_1

2.1 OLN Topology

In this topology, the amplifier amplifies the input signal with the noise contributed by the input resistor. So, it becomes a dominant factor in the total input-referred noise. Generally, to achieve the lower value of the NEF, the bigger size of decoupling capacitors has been required. To avoid Miller effect in input side which would enhance the parasitic capacitance, the cascode connection is used to trade off the input signal attenuation and signal swing (Figs. 1 and 2).

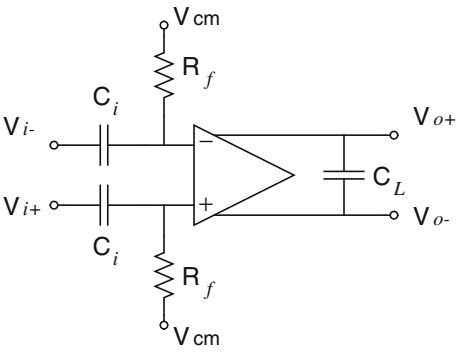


Fig. 1. OLN topology

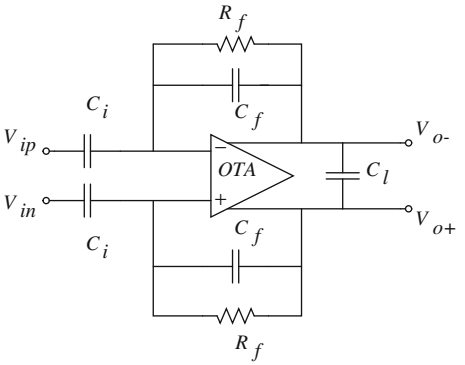


Fig. 2. CFN topology

2.2 CFN Topology

In CFN topology, the feedback resistor and capacitor determine the high-pass pole frequency, and the low-pass pole frequency obtained by operational

transconductance amplifier response. As the requirement of mid-band gain of LNA for low-frequency application is relatively high, cascode OTA is suitable to provide DC gains above 60 dB. Under low voltage supply, the use of telescopic OTA is not appropriate due to degradation of the output swing. It has been proposed in several literatures that the NEF value can be reduced in folded-cascode OTA using current scaling and current splitting techniques for CFN topologies.

2.3 MIFN Topology

Here, the frequency response of the first OTA controlled the low-pass corner frequency. The mid-frequency band gain is controlled by the DC gain of amplifier. This characteristic allows to realize high gain without resorting to large capacitor ratios, however, there may be strong variations in mid-frequency band gain because of technology process deviations. The noise performance of this topology is not better than CFN. To achieve low input referred noise, the large size of input capacitors is also a concerned factor in design (Figs. 3 and 4).

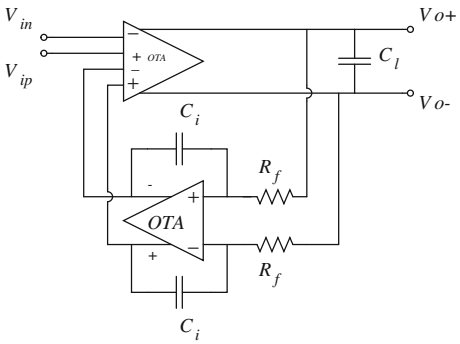


Fig. 3. MIFN topology

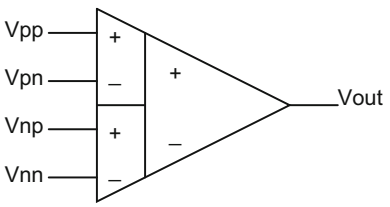


Fig. 4. Symbol of DDA

3 Noise Efficiency Factor (NEF)

For a particular application, it is extremely hard to compare and analyze the design trade-offs in various topologies. One figure of merit that has been used widely as a part of research is Noise Efficiency Factor (NEF).

The mathematical expression of NEF is defined as

$$NEF = \sqrt{2I_{total}/(\Pi \cdot U_t \cdot 4KT \cdot BW)},$$

where U_t or KT/q is the thermal voltage, q is the electron charge, K is the Boltzmann's constant, T is the absolute temperature, I_{total} is the total current consumption of the LNA, and BW stands for its 3 dB-bandwidth. Note that this paper focuses exclusively on low-noise amplifier design. So, flicker noise is also important factor in the noise characteristics of the amplifier, but it can be substantially reduced by using large transistor dimensions or chopper or auto-zero techniques.

4 Differential Difference Amplifier (DDA)

The Differential Difference Amplifier (DDA) is an emerging CMOS analog design building block. It is basically an extension of the conventional operational amplifier. An operational amplifier employs only one differential input, whereas DDA has two differential inputs. Two voltage-to-current converters of DDA convert the differential voltage into the current, later these currents are subtracted and converted into voltage by current-to-voltage converter and amplified (Fig. 5).

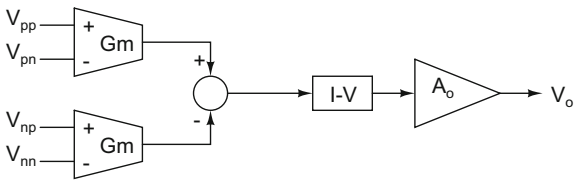


Fig. 5. Schematic of DDA

So, the mathematical expression of the V_{out} of above symbol is expressed as

$$V_{out} = A_o[(V_{pp} - V_{pn}) - (V_{np} - V_{nn})],$$

where, Ao is the open loop gain of DDA. Similar to traditional operational amplifier operation, as A_o infinite, $(V_{pp} - V_{pn}) = (V_{np} - V_{nn})$. Now, for finite open loop gain, the difference between the two differential voltage increases. Therefore, the open loop gain is required to be as large as possible to achieve better performance.

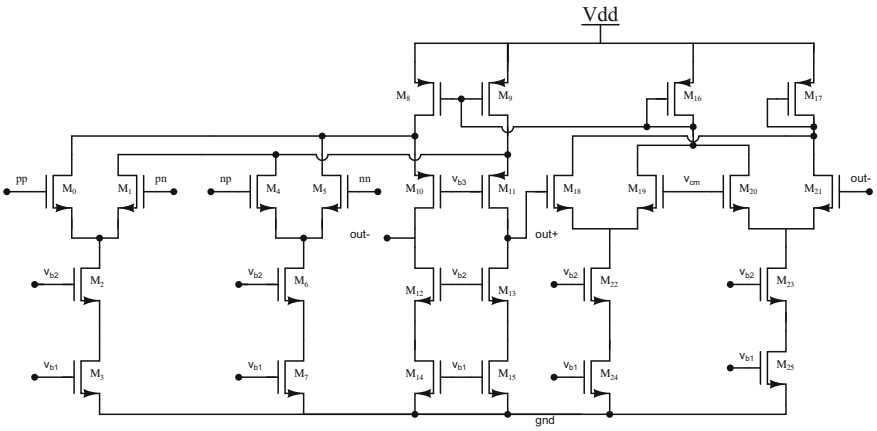


Fig. 6. Schematic of DDA

4.1 Reported Works on DDA

J. Huijsing first introduced the DDA by using CMOS technology. Many basic circuits such as comparator with floating inputs, level shifter, instrumentation amplifier, and resistor less unity gain inverting amplifier using DDA have been realized by Sackinger and Guggebuhl [4]. The DDA attracted the researchers due to its inherent high input impedances and requirement of less passive components for realizations of various circuits. Ismail et al. realized DDA-based adder, subtractor, multiplier, integrator, filters, and phase lead-lag compensator [5]. Soliman et al. proposed current feedback differential difference amplifier with constant bandwidth, independent of the closed loop gain and with higher slew rate [6].

Most of the high-performance analog integrated circuits incorporate fully differential signal paths. In 2001, Alzaher et al. [7] proposed the fully differential difference amplifier. Fully differential architectures have several advantages over the single ended outputs. They provide a larger output voltage swing and are less susceptible to common-mode noise. Also, even-order nonlinearities are not present in the differential output of a balanced circuit, which is symmetric with perfectly matched elements on either side of an axis of symmetry.

However, the DDA has been discussed in very few literature, particularly for fully-differential applications. Here, FDDA schematic and its transient response, AC response, and noise response are shown (Fig. 6).

5 Simulation Result

The DDA architecture has been designed by Potential Division Methodology (PDM) using the model parameters of SCL CMOS 0.18 m technology. Simulation has been done using Cadence design tools. A Common Source (CS) Amplifier has been incorporated as second stage of the amplifier to achieve high gain. The first stage of the DDA provides the open loop gain of 74 dB. The second stage CS Amplifier offers the open loop gain of 25 dB. The gain and the phase response of the designed circuit have been depicted in Fig. 7 and Fig. 8 respectively in typical-typical design corners.

The amplifier has also been simulated for different process variations considering five different design corners and the results are tabulated in Table 1. Then the Common Mode Rejection Ratio (CMRR) of amplifier has been depicted in Fig. 9. It is clearly indicated in Table 1 that the CMRR of the proposed amplifier is near to 130 dB considering the variation of all process corners. The input-referred noise of DDA has been shown in Fig. 10. Eventually, the proposed architecture with simulated results strongly implies that it can be used as high-CMRR and low-noise CMOS amplifier.

5.1 AC Response

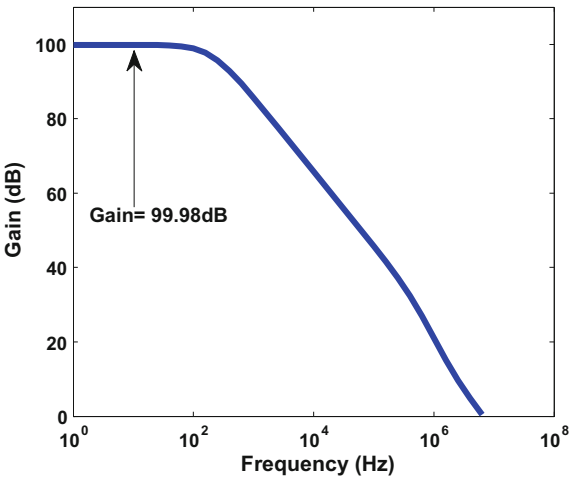


Fig. 7. Gain of DDA

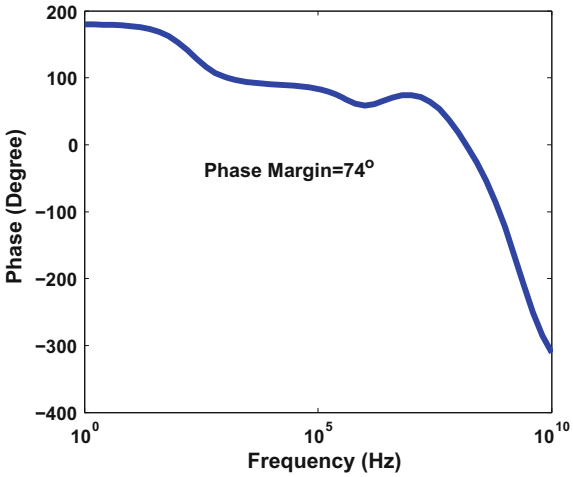


Fig. 8. Phase of DDA

5.2 CMRR and Noise Analysis

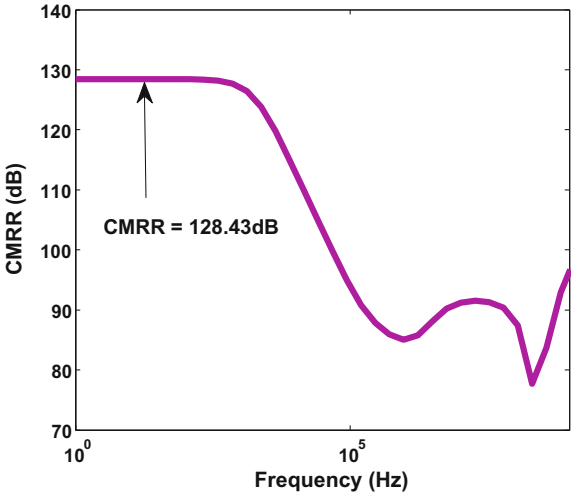


Fig. 9. CMRR of DDA

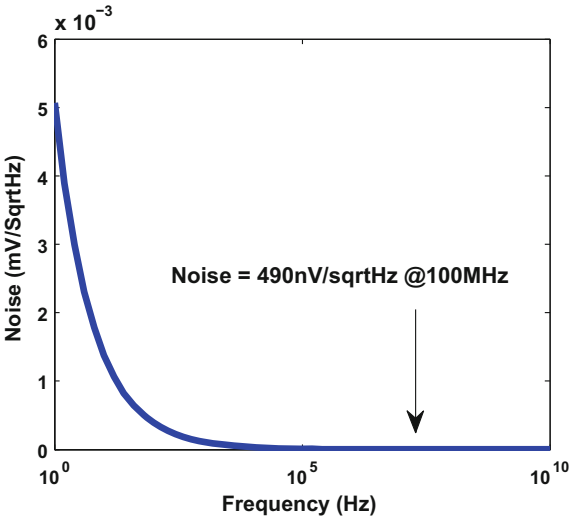


Fig. 10. Input referred noise of DDA

Table 1. Variation of the simulated results for five different design corners

Parameters	Unit	SS	SF	TT	FS	FF
Gain	dB	103.4	74	99.98	83.2	73.45
Phase margin	(°)	64	57	74	74	72
3-dB BW	Hz	85	1.7 K	200	1.2 K	3 K
UGB	MHz	3.5	2.9	6.69	6.8	5.6
CMRR	dB	134.22	124.05	128.43	122.67	120.03
Power dissipation	μW	131.2	168.3	211.4	226.3	251.5

6 Conclusion

The challenge of processing the very small sensor signal in the presence of noise is met by a specially designed very low-noise analog front-end amplifier. The simulation results indicate that this proposed high-gain and high-CMRR amplifier could be used for low-frequency signal amplification such as seismic detection system, etc. The designed DDA has high CMRR considering all design corners. However, the input referred noise can be minimized by using techniques such as chopper stabilization, autozeroing, and correlated double sampling, etc.

Acknowledgements. The authors gracefully acknowledge Dr. Debashis Dutta, Ministry of Electronics and Information Technology, Govt. of India for leading the Chip to System Design activities (SMDP-C2SD) in India. We are deeply indebted to Dr. Dutta and his team for their constant encouragement and special care for NIT Durgapur. The

authors also acknowledge Prof. Pradip Mandal, Prof. Anindya Sundar Dhar of Department of Electronics and EC Engineering, IIT Kharagpur and Prof. Rajat Mahapatra of Department of Electronics and Communication Engineering of NIT Durgapur for their kind advice throughout the project. The authors also acknowledge the people of SMDP-VLSI Lab for their enthusiastic participation in reviewing the work and critical assertions.

References

1. Gosselin B (2011) Recent advances in neural recording microsystems. Sensors 11:4572–4597
2. Harrison RR, Charles C (2003) A low-power low-noise CMOS amplifier for neural recording applications. IEEE J Solid-State Circuits 38(3):958–965
3. Perelman Y, Ginosar R (2007) An integrated system for multichannel neuronal recording with spike/LFP separation, integrated A/D conversion and threshold detection. IEEE Trans Biomed Eng 54:130–137
4. Sackinger E, Guggenbuhl W (1987) A versatile building block: the CMOS differential difference amplifier. Solid-State Circuits 22:287–294
5. Huang S-C, Ismail M, Zarabadi SR (1993) A wide range differential difference amplifier: a basic block for analog signal processing in MOS technology
6. Soliman AM, Soliman MA (1997) A new CMOS realization of the differential difference amplifier and its application to a MOS-C oscillator. Int J Electron 83(4):455–466
7. Alzaher HA, Elwan H, Ismail M (2003) A CMOS fully balanced second-generation current conveyor. IEEE Trans Circuits Syst II: Analog Digit Signal Proc 50(6):278–287
8. Zhao W, Li H, Zhang Y (2009) A low-noise integrated bioamplifier with active DC offset suppression. IEEE Trans Biomed Circuits Syst 5–8
9. Saberhosseini SS, Zabihian A, Sodagar AM (2012) Low-noise OTA for neural amplifying applications. In: Proceedings of the 8th international caribbean conference on devices, circuits and systems (ICCDCS12), Playa del Carmen, Mexico, 14 March 2012. pp 14
10. Zhang F, Holleman J, Otis BP (2012) Design of ultra-low power biopotential amplifiers for biosignal acquisition applications. IEEE Trans Biomed Circuits Syst 6:344–355
11. Chae M, Kim J, Liu W (2008) Fully-differential self-biased bio-potential amplifier. Electron Lett 44:1390–1391
12. Chae M, Liu W, Zhi Y, Chen T, Kim J, Sivaprakasam M, Yuce M (2008) A 128-channel 6 mW wireless neural recording IC with on-the-fly spike sorting and UWB transmitter. In: Proceedings of the IEEE international solid-state circuits conference (ISSCC08), San Francisco, CA, USA, 37 February 2008. pp 146–603
13. Gosselin B, Ayoub AE, Roy JF, Sawan M, Lepore F, Chaudhuri A, Guitton D (2009) A mixed-signal multichip neural recording interface with bandwidth reduction. IEEE Trans Biomed Circuits Syst 3:129–141
14. Majidzadeh V, Schmid A, Leblebici Y (2011) Energy efficient low-noise neural recording amplifier with enhanced noise efficiency factor. IEEE Trans Biomed Circuits Syst 5:262–271
15. Zou X, Liu L, Cheong JH, Yao L, Li P, Cheng MY, Goh WL, Rajkumar R, Dawe G, Cheng KW Je M (2013) A 100-channel 1-mW implantable neural recording IC. IEEE Trans Circuits Syst I Regul Pap 60:2584–2596

16. Chang CW, Chiou JC (2013) A wireless and batteryless microsystem with implantable grid electrode/3-dimensional probe array for ECoG and extracellular neural recording in rats. Sensors 13:4624–4639
17. Sepehrian H, Mirbozorgi A, Gosselin B (2014) A low-power current-reuse analog front-end for multi-channel neural signal recording. In: Proceedings of the 12th IEEE international NEWCAS conference (NEWCAS14), Trois-Rivieres, Quebec, Canada, 22 June 2014

Performance Improvement of WiMAX (IEEE 802.16d) Using Combined RS–CC Forward Error Channel Coding Technique

Arun Agarwal[1,2(✉)] and Saurabh N. Mehta[3]

[1] Department of Information Technology, AMET University, Chennai 603112, Tamil Nadu, India
arunagrawal@soa.ac.in

[2] Department of EIE, ITER, Siksha 'O' Anusandhan Deemed to be University, Khandagiri Square, Bhubaneswar 751030, Odisha, India

[3] Department of Electronics and Telecommunication Engineering, Vidyalankar Institute of Technology, Mumbai 400037, Maharashtra, India
saurabh.mehta@vit.edu.in

1 Introduction

WiMAX (Worldwide Interoperability for Microwave Access) belongs to wireless communication IEEE 802.16 set of standards which provide multiple physical layers (PHY) and Media Access Control (MAC) options. It is expected to serve about 30–40 Mbps data rates for mobile stations and around 1 Gbps for fixed stations as per the current update [1, 2]. Typical coverage range for Broadband Access (BWA) using WiMAX for fixed receivers is 50 km, and between 5 and 15 km for mobile stations, whereas on the other hand, the wireless local area network standard, i.e., the Wi-Fi/802.11, has limited coverage range up to 30–100 m. Considering performance, coverage, and quality of service (QOS) of cellular networks, WiMAX is actually the same as 802.11/Wi-Fi IP-based networks that provide wireless broadband access technology. WiMax is an updation to wireless version of Ethernet to serve as an alternative to wire technologies (such as cable modems, DSL, and T1/E1 links) while providing high-speed data access to customer premises. WiMAX can be compared with LTE and HSPA+ as energy-efficient pre-4G technique.

Both the MAC and PHY layers constitute the WiMAX air interface standard. For efficient bandwidth utilization, the MAC layer usually uses TDM/TDMA scheduled uplink/downlink frames. Due to the inclusion of ATM, IPv4, IPv6, Ethernet, and Virtual Local Area Network (VLAN) services in IEEE 802.16, WiMAX is made applicable to a number of high bandwidth broadband wireless applications. There are two types of WiMAX versions as per IEEE 802.16 working group. One is a fixed and nomadic access model (IEEE 802.16-2004) providing connection in Line of Sight (LOS) and Non-Line of Sight (NLOS) environment using OFDM and the other is a portable/mobile access model (IEEE 802.16e) having scalable OFDM access for handoffs and roaming. A brief overview of the 802.16 standards was presented in [4, 5] by us.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_2

PHY of WiMAX OFDM—Fixed: A 256-point Fast Fourier Transform (FFT) is used in fixed WiMAX physical layer. Out of 256 subcarriers, 192 is used for data transmission, 8 are reserved as pilot for synchronization and channel estimation purpose, and the last 56 subcarriers act as guard band. This PHY operates in the 2–11 GHz frequency band using TDD technique. With 20 MHz maximum, the bandwidth of the channel is chosen to be multiple of 1.25, 1.5, 1.75, 2, and 2.75 MHz [3]. This physical layer has inner Convolutional Code (CC) and outer Reed–Solomon (RS) code as channel coding technique to combat errors during transmission. This PHY performs well under Rayleigh multipath propagation.

PHY of WiMAX OFDM—Mobile: This PHY incorporates 2048-point FFT OFDMA system with a variable FFT size from 128 to 2,048. The FFT size increases with bandwidth, to maintain 10.94 kHz subcarrier spacing. This PHY operates below 11 GHz frequency. Important parameters for both fixed and mobile WiMAX physical layers were discussed in [4, 5].

This research paper deals with the fixed WiMAX physical layer implementation. The analyses were carried out using different digital modulation schemes (BPSK, QPSK, and QAM) and coding rate in diverse transmission channels. Bit Error Rate (BER) curve was used as the performance metrics. After the introduction, the remaining part of the paper is arranged as below. The detailed description of the WiMAX physical layer is discussed in Sect. 1. Section 2 will present the details about modeling and simulation of proposed system using MATLAB. Section 3 presents the simulation results. Section 4 finally concludes the paper.

2 Simulation Model Implemented

Here, we present the implemented simulation model for evaluating the IEEE 802.16 WiMAX systems BER performance. Figure 4 indicates the functional blocks of baseband WiMAX transceiver. After randomization at the transmitter, the data is subjected to convolutional encoder with variable code rates. Then, we perform mapping into QPSK/QAM symbols. The FFT will be used to implement the subcarriers, for maintaining the orthogonality. Then, the Inverse Fast Fourier Transform (IFFT) modulates the symbol onto subcarriers. After conversion to serial data, we add a cyclic prefix to avoid Inter-symbol Interference (ISI). Finally, the output is transmitted to Rayleigh fading channel whereby we add additive white Gaussian noise characteristics. At the receiver, we perform the reverse operations.

3 Results and Discussion

Here, we discuss and present the Matlab simulation results for Rayleigh fading (Channel-1) and Rician (Channel-2) under three path fading channels with BER curve analysis. We have taken different values of rms delay spread for these fading channels

to evaluate the performance along with effect of path loss (including shadowing), Doppler spread, multipath delay, spread, co-channel, and adjacent channel interference.

Parameters as discussed in [4, 5] were taken for simulation of WiMAX system. The effect of Reed–Solomon encoder on the performance of the system is also shown along with how much there is performance degradation without RS encoding. Figures 1, 2, 3, 4, 5, 6, 7, and 8 present the effect on noise performance due to RS encoder on various digital modulations in terms of the BER versus SNR plot for Channel-1 and Channel-2. It can be seen from Figs. 1, 2, 3, 4, 5, 6, 7, and 8 that addition of RS improves the system BER performance. From Fig. 1 for a BER of 10^{-4}, it can be readily evaluated that for BPSK, the SNR value is 18.5 dB for only CC coding whereas it reduces drastically to 10 dB for RS–CC concatenated coding. This shows an SNR improvement or coding gain of about 8.5 dB. Table 1 indicates the summary of noise performance regards improvement in SNR with use of RS codec for various digital modulation schemes.

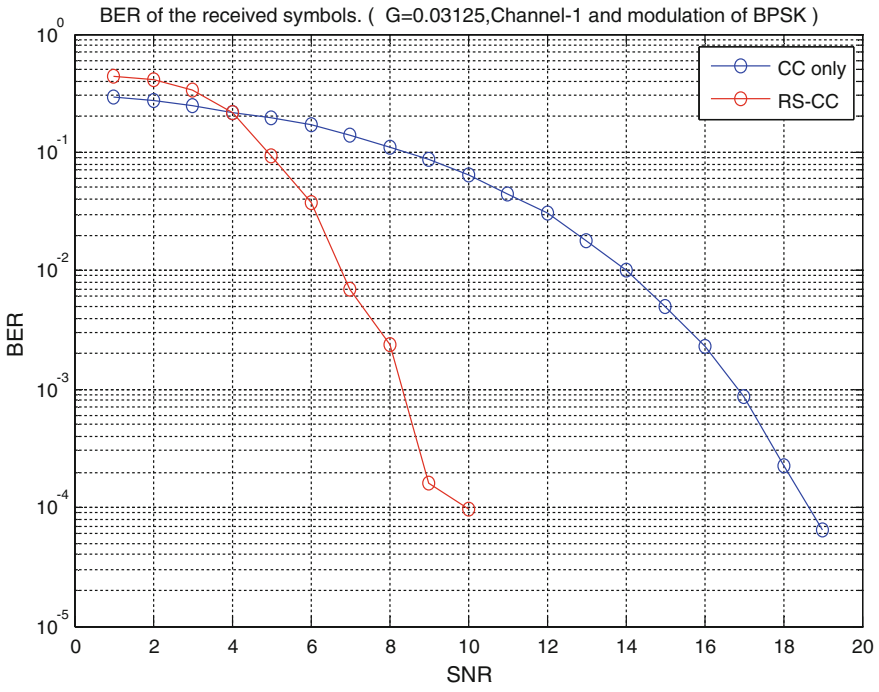


Fig. 1. Channel-1 BER curve using BPSK modulation

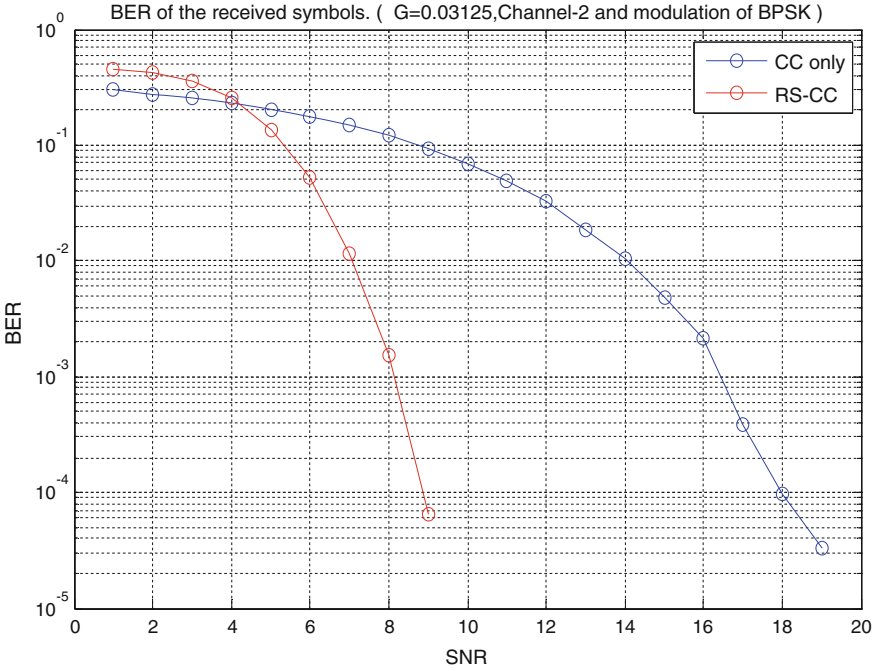


Fig. 2. Channel-2 BER curve using BPSK modulation

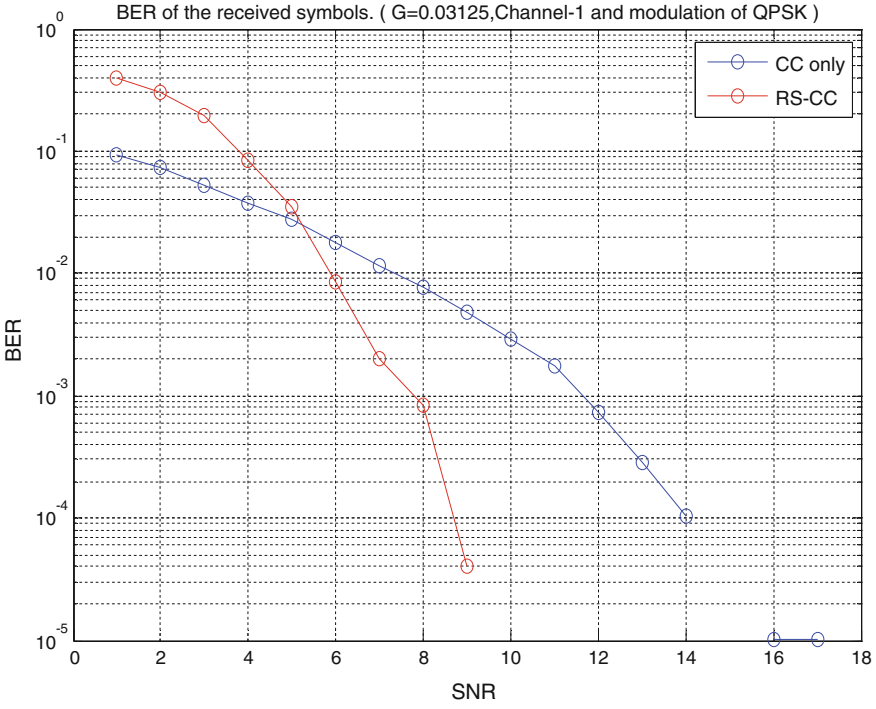


Fig. 3. Channel-1 BER curve using QPSK modulation

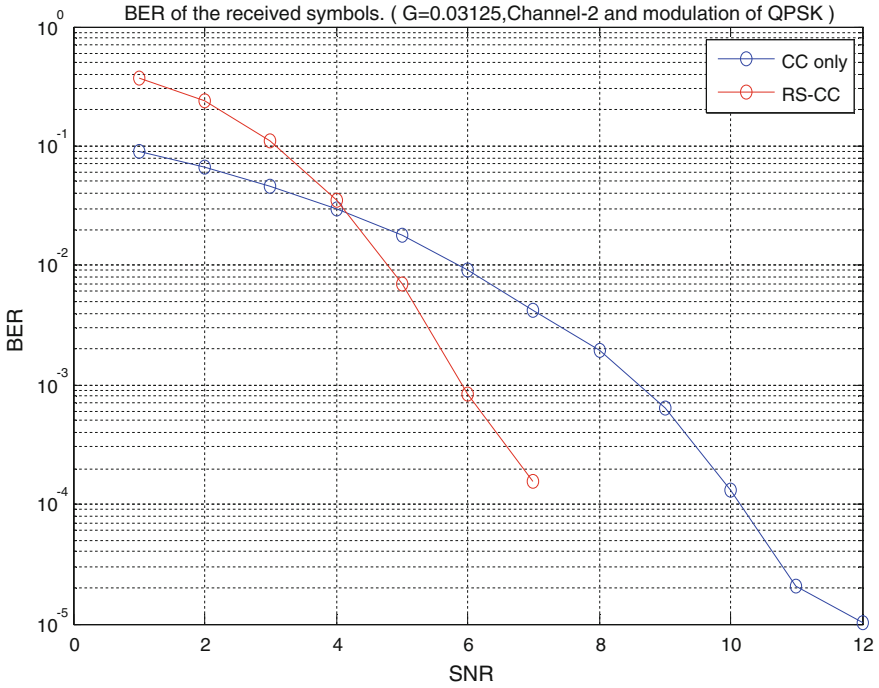


Fig. 4. Channel-2 BER curve using QPSK modulation

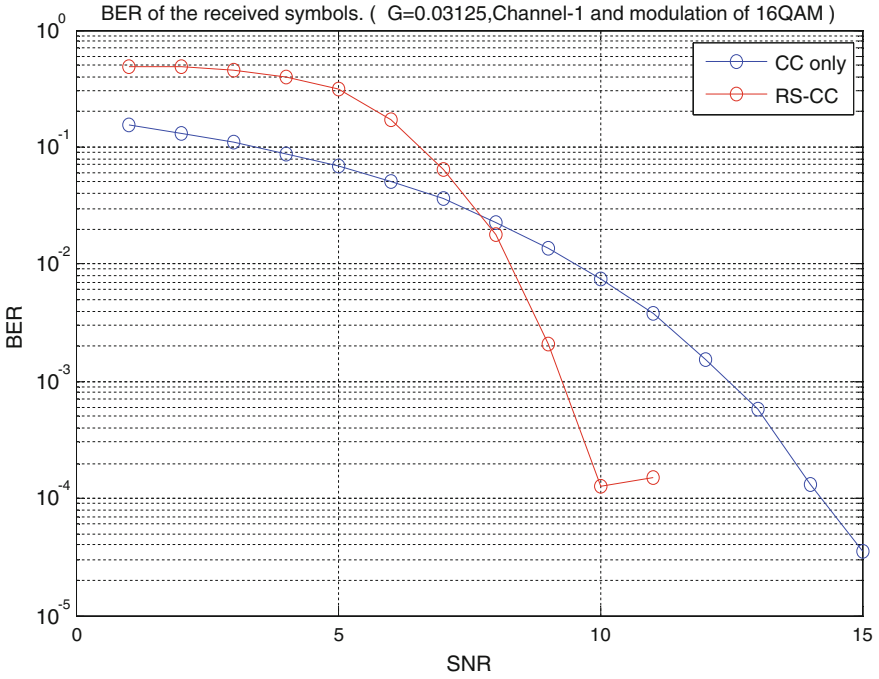


Fig. 5. Channel-1 BER curve using 16-QAM modulation

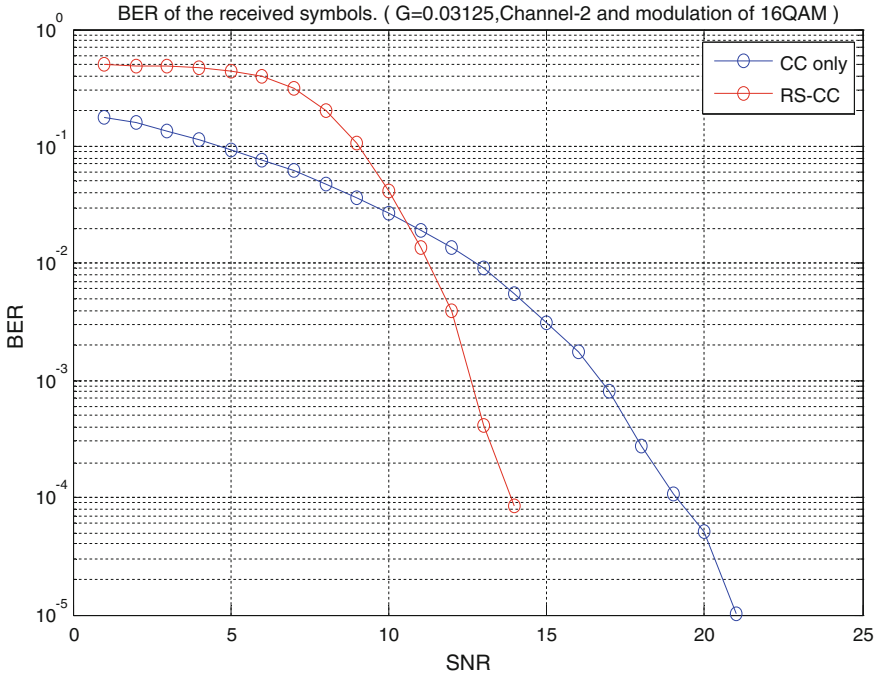


Fig. 6. Channel-2 BER curve using 16-QAM modulation

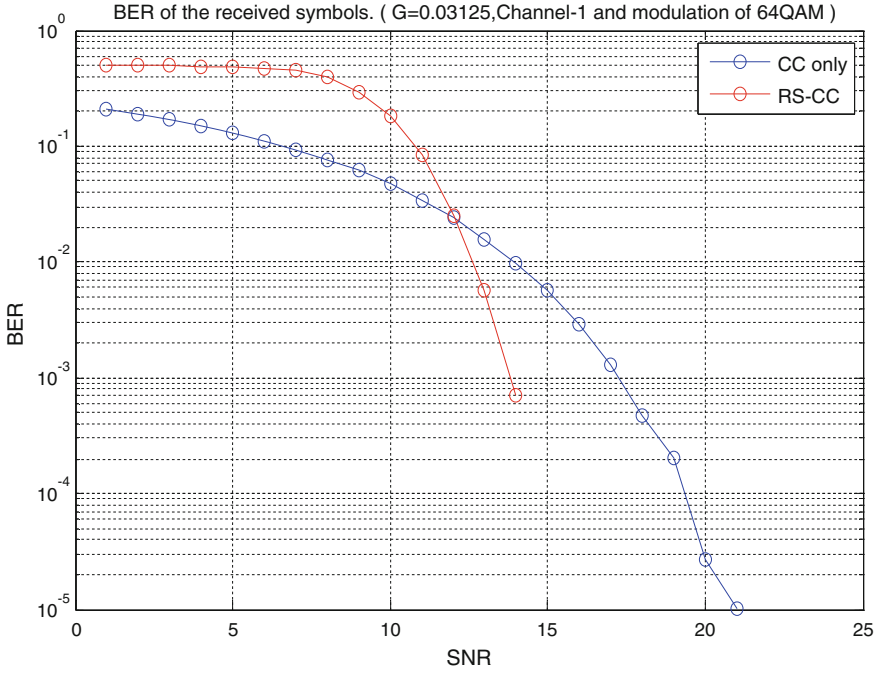


Fig. 7. Channel-1 BER curve using 64-QAM modulation

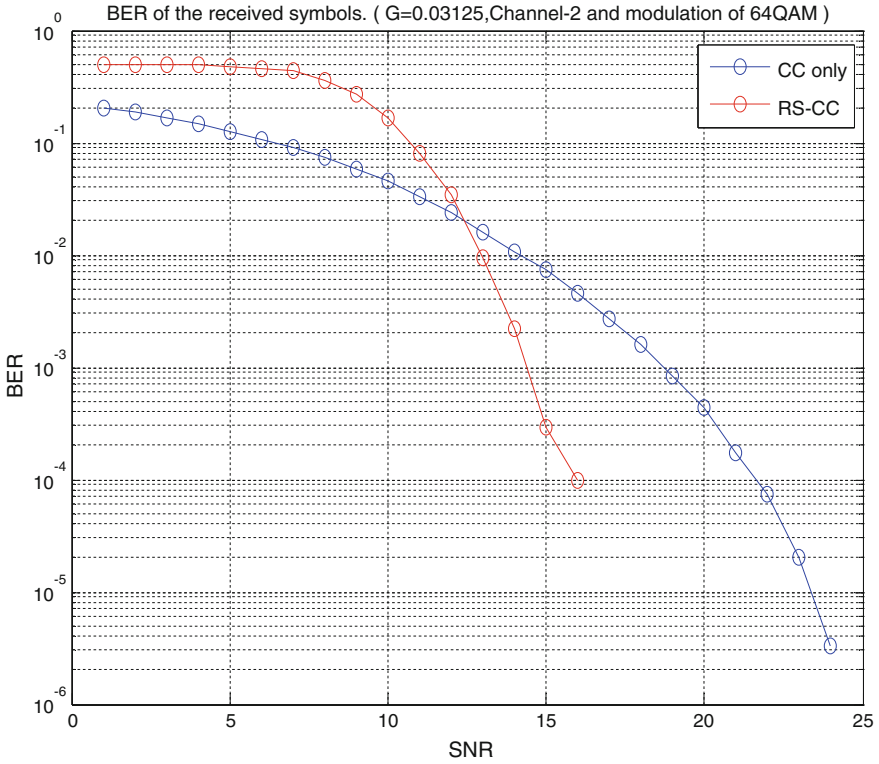


Fig. 8. Channel-2 BER curve using 64-QAM modulation

Table 1. Performance improvement due to concatenated coding

Modulation	BPSK		QPSK		16-QAM		64-QAM	
	Channel 1	Channel 2	Channel 1	Channel 2	Channel 1	Channel 2	Channel 1	Channel 2
Coding gain (dB) at BER 10^{-4}	8.5	9.1	5.2	3.1	4	5.5	4	6

4 Conclusion and Future Directions

The design and simulation of PHY layer of IEEE 802.16 OFDM-based WiMAX system was the important contribution of this work. This work analyzed the system BER performance under RS–CC channel coding techniques showing good impact on transmission reliability as well. The comparative analysis shows improvement in the noise performance using concatenated channel coding with minimum SNR, thereby decreasing the transmitter power. This has a powerful impact of saving of batteries of mobile-driven devices.

In addition, turbo coding, LDPC coding, and polar coding are some future directions for extension of the proposed work. The performance can be evaluated in worst channels like Stanford University Interim (SUI) channel models. Smart antenna technology like MIMO can be added to test the system performance.

References

1. WiMAX Forum Mobile WiMAX—Part I: a technical overview and performance evaluation, Aug 2006
2. WiMAX Forum Mobile WiMAX. Part II: A Comparative analysis, May 2006
3. IEEE 802.16-2004 IEEE standard for local and metropolitan area networks-Part 16: air interface for fixed broadband wireless access systems, 1 Oct 2004
4. Agarwal A, Mehta SN (2015) Combined effect of block interleaving and FEC on BER performance of OFDM based WiMAX (IEEE 802.16d) system. Am J Electr Electron Eng 3 (1):4–12. https://doi.org/10.12691/ajeee-3-1-2
5. Agarwal A, Agarwal K (2015) Performance prediction of WiMAX (IEEE 802.16d) using different modulation and coding profiles in different channels. Int J Appl Eng Res (IJAER) 10 (2):2221–2234

A Hybrid Control Algorithm for Extraction of Voltage Harmonics to Mitigate Power Quality Problems Using UPQC

Mashhood Hasan[1], Jai Prakash Pandey[2], Imran Ahmed Khan[1(✉)], and Rajeev Kumar Chauhan[3]

[1] Galgotias College of Engineering and Technology, Greater Noida, India
{mirmashhood2010, imran.vlsi}@gmail.com
[2] KNIT, Sultanpur, India
tojppandey@rediffmail.com
[3] AKTU, Lucknow, India
mmmec.rkc@gmail.com

1 Introduction

Normally, power electronics based apparatuses have drawbacks. It creates multiple power quality (MPQ) problems in AC network. These apparatuses generate harmonics and raise the extra burden of reactive power to the utility side. The techniques are required to solve the new problems imposed on the AC network. The idea of custom power devices has been proposed in the late 1980s. In this concept, the controller is incorporated to improve power quality and controllability of the power system [1, 2]. Among the controllers based on power electronics, the first one is conventional thyristor switched capacitors (TSC) or thyristor switched reactors (TSR) and the second is self-commutated switching converters (SSC). The response time and compensation flexibility in the SSC have been investigated better than conventional TSC. Because the size of the passive elements limits the application of SSC, the most promising custom devices are dynamic voltage restore (DVR), static compensator in distribution system (D-STATCOM), and unified power quality conditioners (UPQCs). These custom devices are used in the ACDS to improve voltages and currents quality. The UPQC is most effective to improve the power quality in DS which offers advantages over DVR and D-STATCOM [3, 4].

It is very crucial to control the IGBT-based shunt and series converters of UPQC [5]. The PAC and symmetrical component theory based techniques are used for controlling the series [6] and shunt converters, respectively [7]. The voltage harmonics can be abolished if sources itself produces voltage harmonics [8]. Moreover, it is incapable to share reactive load of the shunt convert under steady state (S.S). This shortcoming can be overcome using PAC. The PAC improves the utilization factor of series part and shares the reactive loads of shunt part of the UPQC. In [5], mathematically, it has been seen that PAC is incapable to eliminate voltage harmonics of the utility. The *d-q* transformation based control technique [9] is unified with PAC to mitigate the multiple voltage qualities of the utility of low-voltage AC network. Moreover, IGBT-based shunt converter of

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_3

UPQC is controlled by the gate pulse using mathematical concept of instantaneous symmetrical voltage component. A schematic diagram of UPQC based on proposed three-phase three-wire AC network is shown in Fig. 1. This configuration has been taken as a MATLAB model to analyze the performance of the control techniques. The MPQ of low-voltage AC network are controlled using series and shunt control algorithm based on simple mathematical concepts which are described in Sect. 2. The results based on proposed control techniques are given in Sects. 3 and 4 that represents the conclusion.

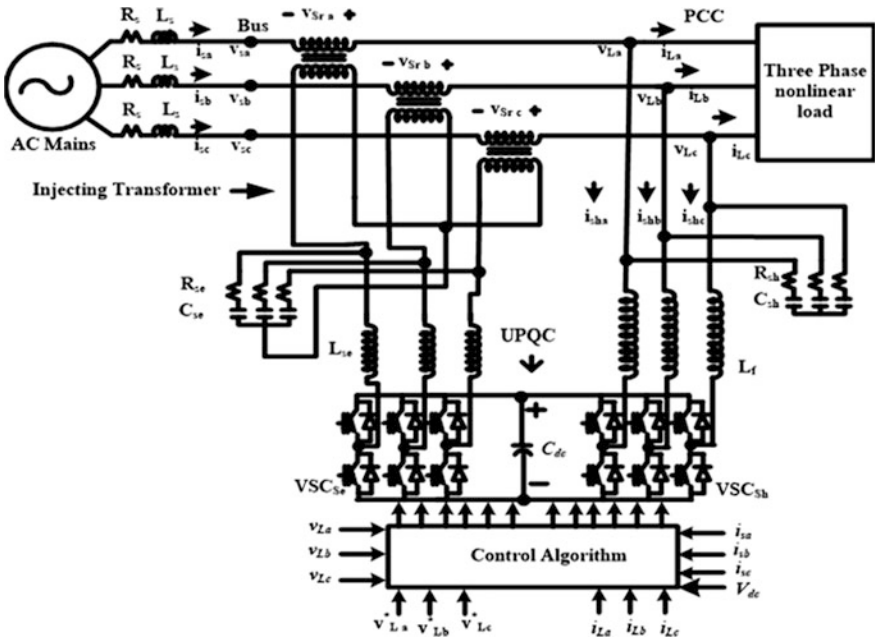


Fig. 1. Three-phase AC network configuration with UPQC

2 Mathematical Model of Control Techniques for UPQC

The mathematical model represents the design of control algorithm of hybrid or unified controller for IGBT-based series converter. In addition, it represents control concepts based on symmetrical voltage component theory to design firing a gate pulse of IGBT-based shunt converter. The control techniques are as follows.

2.1 An Algorithm to Generate Pulses of Series Converter

PAC can compensate reactive power of the voltage harmonics. Therefore, a hybrid controller is used to overcome this issue of PAC. The mathematical expression is developed under various conditions of AC mains.

2.1.1 Under Normal Steady State Condition

Under the normal operating condition, the voltage at PCC and AC mains voltage are same. The series converter injects few volts to mitigate the wire losses. The PAC is unified to d-q based algorithm to mitigate harmonics. To extract the harmonics of the voltage of the three-phase polluted voltages are transformed to two-phase voltages. It can be of direct axis (v_{ds}^{s}) and quadrature axis (v_{qs}^{s}) of utility voltage which can be estimated as

$$\begin{pmatrix} v_{ds}^{s} \\ v_{qs}^{s} \end{pmatrix} = \sqrt{3/2}\begin{pmatrix} 1 & -1/2 & 1/2 \\ 0 & \sqrt{3/2} & -\sqrt{3/2} \end{pmatrix}\begin{pmatrix} v_{as} \\ v_{bs} \\ v_{cs} \end{pmatrix} \tag{1}$$

It is further converted into synchronous rotating frame (SRF) direct axis (v_{ds}^{e}) and quadrature axis (v_{qs}^{e}) voltages which can be expressed as

$$\begin{pmatrix} v_{sd}^{e} \\ v_{sq}^{e} \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} v_{sd}^{s} \\ v_{sq}^{s} \end{pmatrix} \tag{2}$$

θ represents angular position of the voltage signals. The rotating reference voltage has harmonic voltage ($\tilde{v}_{sd,hr}^{e}$) and fundamental voltage ($\bar{v}_{sd,f}^{e}$) of utility; it can be evaluated by (3). The polluted voltage is to be filtered by high-pass filter (HPF). The HPF separates harmonics of voltage and then it can convert two-phase to three-phase voltage to get reference harmonics voltage for three-phase. Mathematical evolution is given as

$$\begin{pmatrix} v_{sd}^{e} \\ v_{sq}^{e} \end{pmatrix} = \begin{pmatrix} v_{sd,f}^{e} + v_{sd,hr}^{e} \\ v_{sq,f}^{e} + v_{sq,hr}^{e} \end{pmatrix}\begin{pmatrix} v_{sd}^{s} \\ v_{sq}^{s} \end{pmatrix} \tag{3}$$

$$\begin{pmatrix} v_{sd,hr}^{s} \\ v_{sq,hr}^{s} \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} v_{sd,hr}^{e} \\ v_{sq,hr}^{e} \end{pmatrix} \tag{4}$$

$$\begin{pmatrix} v_{sa,hr}^{*} \\ v_{sb,hr}^{*} \\ v_{sc,hr}^{*} \end{pmatrix} = \sqrt{3/2}\begin{pmatrix} 1 & 0 \\ -1/2 & \sqrt{3/2} \\ 1/2 & -\sqrt{3/2} \end{pmatrix}\begin{pmatrix} v_{sd,hr}^{s} \\ v_{sq,hr}^{s} \end{pmatrix} \tag{5}$$

The voltage harmonics of the three-phase polluted voltage is calculated with the help of Eq. (5) and the values are added with series injection voltage to abolish the voltage harmonics. The designing of PAC is now the part of the voltage harmonics too. The three-phase series injection voltages require series voltage (v_{inj}) and power angle (δ) which are represented as

$$v_{inj} = 2 \times v_{rms} \times \sin\frac{\delta}{2} = k(say), \tag{6}$$

where δ represents the angle between utility voltage and voltage at PCC (Fig. 2), while δ can be evaluated by the ratio of reactive power Q_{sr} to the active power of the load P_L which is given as

$$\delta = \sin^{-1}\frac{Q_{sr}}{P_L}, \tag{7}$$

where Q_{sr} is shared by series converter. The injection voltage and the extracted voltage harmonics are added to find actual series injection voltage, under voltage harmonics generation of the AC mains. The magnitude of the actual reference voltage can be calculated as follows:

$$v_{inj.a} = k\ \sin(\omega t + \phi_{inj}) + v^*_{sa,hr} \tag{8}$$

$$v_{inj.b} = k\ \sin(\omega t + \phi_{inj} - 120^\circ) + v^*_{sb,hr} \tag{9}$$

$$v_{inj.c} = k\ \sin(\omega t + \phi_{inj} + 120^\circ) + v^*_{sc,hr} \tag{10}$$

where Φ_{inj} is the angle of series voltage injection, which is calculated as

$$\phi_{inj} = 90^\circ + \frac{\delta}{2} \tag{11}$$

In Fig. 3a, the reference voltage signals for nominal condition ($v_{inj,a}v_{inj,b}$, and $v_{inj,c}$) are directly injected with a certain angle to share reactive loads of shunt converter. For nominal condition, the reference injection voltages ($v^*_{inj,a}$ $v^*_{inj,b}$, and $v^*_{inj,c}$) depend only on power angle; that is why it is compared with the output of the series voltage converter to produce the actual series injection voltage.

2.1.2 Under Voltage Sag Condition of the AC Mains

In Fig. 2, the operating voltage of the AC mains is reduced to v_{sabc1} under voltage sag condition. Thus, series voltage injections ($v_{inj1,a}$, $v_{inj1,b}$, and $v_{inj1,c}$) under voltage sag condition are evaluated as follows:

$$v_{inj1,a} = v_{inj,a} + (v^*_{La} - v_{sa1}) \tag{12}$$

$$v_{inj1,b} = v_{inj,b} + (v^*_{Lb} - v_{sb1}) \tag{13}$$

$$v_{inj1,c} = v_{inj,c} + (v^*_{Lc} - v_{sc1}), \tag{14}$$

where the actual source voltage ($v_{sa1,}$ $v_{sb1,}$ v_{sc1}) under voltage sag condition is subtracted with the reference load voltage v^*_{Labc} to get the actual pulse generation through series converter.

2.1.3 Under Voltage Swell Condition of the AC Mains

In Fig. 2, the operating voltage of the AC mains is increased to v_{sabc2} under voltage swell condition. Thus, series voltage injections ($v_{inj2,a}$, $v_{inj2,b}$, and $v_{inj2,c}$) under voltage sag condition are evaluated as follows:

$$v_{inj2,a} = v_{inj,a} + (v^*_{La} - v_{sa2}) \tag{15}$$

$$v_{inj2,b} = v_{inj,b} + (v^*_{Lb} - v_{sb2}) \tag{16}$$

$$v_{inj2,c} = v_{inj,c} + (v^*_{Lc} - v_{sc2}), \tag{17}$$

where the actual source voltages ($v_{sa1,}$ $v_{sb1,}$ v_{sc1}) under voltage swell condition are subtracted with the reference load voltage v^*_{Labc} to get the actual pulse generation through series converter.

In Fig. 3a, the MATLAB model of the series converter to generate reference series voltage ($v^*_{sabc,hr}$) and vector angle of projection (Φ_{Sr}) is needed, which is given by Eq. (11). The angle ωt is produced with the help of phase locked loop (PLL) and get it added with (Φ_{Sr}). The angle sum is acquired through the sine function and get it multiplied with sum of harmonics $v^*_{sabc,hr}$ and reference series injection voltage ($v^*_{inj.abc}$). The final value is compared with the sensed voltage of series transformer output voltage to generate pulse signals for series converter.

2.2 An Algorithm to Generate Pulses of Shunt Converter

The object of shunt converter is to maintain harmonic-free currents at AC mains and desired DC link voltage. In Fig. 3b, shunt controller is needed for shunt currents (i^*_{Ca}, i^*_{Cb}, and i^*_{Cc}) injection for getting the pulse of three-phase shunt converter, which is given by

$$i^*_{ca} = i_{la} - i^*_{sa} = i_{la} - \frac{v^+_{la1} + \gamma(v^+_{lb1} - v^+_{lc1})}{\Delta^+_1}(P_{lavg} + P_{loss}) \tag{18}$$

$$i^*_{cb} = i_{lb} - i^*_{sb} = i_{lb} - \frac{v^+_{lb1} + \gamma(v^+_{lc1} - v^+_{la1})}{\Delta^+_1}(P_{lavg} + P_{loss}) \tag{19}$$

$$i^*_{cc} = i_{lc} - i^*_{sc} = i_{lc} - \frac{v^+_{lc1} + \gamma(v^+_{la1} - v^+_{lb1})}{\Delta^+_1}(P_{lavg} + P_{loss}), \tag{20}$$

where Δ and γ can be expressed as (21) and (22). The term Φ is input power factor angle or the angle between AC mains voltage and AC mains current.

$$\Delta = \sum\nolimits_{j=a,b,c} (v^+_{lj1})^2 \Delta = \sum\nolimits_{j=a,b,c} (v^+_{lj1})^2 \tag{21}$$

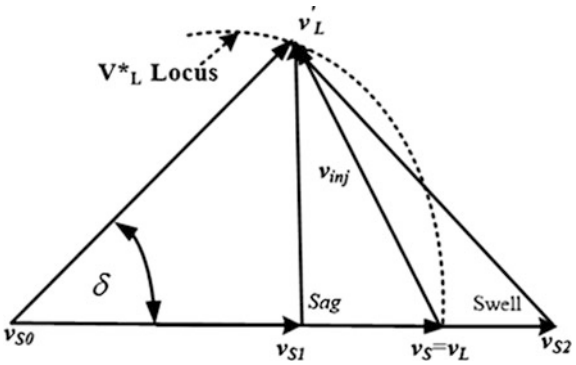


Fig. 2. Phaser control of proposed controller

$$\gamma = \tan\frac{\phi}{\sqrt{3}} \tag{22}$$

The highly polluted load demands the desire voltage and current. The load demands affect the AC mains voltage and current, and it also affects voltage drop across the line impedance and generates polluted voltage at AC mains (Microgrid). Fundamental positive sequence voltages v_{la1}^{+},v_{lb1}^{+}, and v_{lc1}^{+} at PCC are sensed to develop the control algorithm of the shunt converter. The average power P_{lavg} of the load is also required to produce reference currents. These are generated using Eqs. (18), (19), and (20). The converter losses P_{loss} of the UPQC are the output of a PI voltage controller at DC bus. For fast control realization, reference DC voltage V_{dc}^{*} must be kept equal to actual DC voltage V_{DC}, in order that power loss P_{Loss} through PI voltage controller is zero. The magnitude of P_{Loss} is given by the following equation:

$$P_{Loss} = K_P(V_{DC}^{*} - V_{DC}) + K_i \int (V_{DC}^{*} - V_{DC})^2 \tag{23}$$

In that manner, shunt current controller system supplies the needed power to make fast control of UPQC. In Fig. 3b, the IGBT-based shunt converter is mitigated reactive demand to inject currents and it keeps supply current to maintain DC voltage.

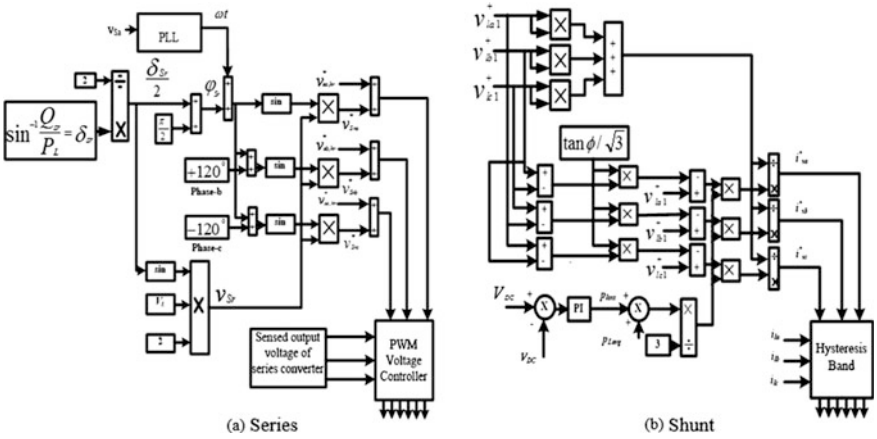


Fig. 3. Pulse techniques generation for IGBT-based series and shunt converters

3 Real Results and Discussion

Consider the hybrid control techniques using PAC and d-q theory are implemented to analyze the performance of series converter and instantaneous symmetrical theory based controller to analyze the performance of shunt converter of three-phase three-wire AC network connecting nonlinear loads with UPQC. All the parameters for the simulation are represented in Table 1.

3.1 Performance of Hybrid Control Algorithm for Series Converter of UPQC Under Various Conditions of AC Mains

In Fig. 4, the performance of hybrid control algorithm has been tested for series converter of UPQC under the various conditions of the AC mains voltage. The simulation results have given for t = 0.4–0.7 s. For, t = 0.4–0.45 s, only the power angle controller of series converter works, so the voltage at PCC and voltage at AC mains are same. The series converter shares only reactive power without the voltage harmonics elimination. The hybrid control algorithm works for t = 0.45–0.5 s under normal operating condition of AC mains. It injects voltage $v_{inj.hr}(V)$ to share reactive power and eliminates harmonics voltage at PCC. For the voltage sag condition of AC mains for t = 0.5–t = 0.55 s, series converter adds extra voltage to compensate the voltage sag at PCC voltage $v_{Labc}(V)$ and it is also used for the elimination of AC mains voltage harmonics using hybrid control algorithm. During t = 0.6–0.65 s, the AC mains under voltage swell condition, the hybrid control algorithm adds extra voltage to maintain the constant voltage at PCC.

3.2 Performance of Shunt Control Algorithm Under Various Conditions of AC Mains

In Fig. 5, the performance of shunt control algorithm has been tested for UPQC under the various conditions of the AC mains voltage for t = 0.4–0.7 s. During t = 0.4–0.5 s, t = 0.55–0.6 s, and t = 0.65–0.7 s, the shunt converter runs under normal operating condition and it injects harmonics to eliminate harmonics of the AC mains currents $i_{sabc}(A)$. Moreover, it maintains the DC link $v_{DC}(V)$ at required value *725(V)*. The shunt currents are $i_{sha}(A)$, $i_{shb}(A)$, *and* $i_{shc}(A)$ for voltage sag (t = 0.5–0.55 s) and voltage swell condition (t = 0.6–0.65 s); it eliminates the harmonics and maintains its DC link voltage. Throughout the simulation, the load currents $i_{La}(A)$, $i_{Lb}(A)$, *and* $i_{Lc}(A)$ are constant under various conditions of the AC mains. Thus, it shows that the shunt control algorithm works satisfactorily.

3.3 Performance Analysis of Harmonics Spectrum

In Fig. 6, harmonics of voltage and current are shown, whereas at t = 0.4–0.45 s, the AC mains current contains current harmonics 2.67 while load current contains 21.29% which is shown in Fig. 6a and b. The utility (AC mains) voltage and PCC voltage harmonic contains above the IEEE slandered which are shown in Fig. 6c and d.

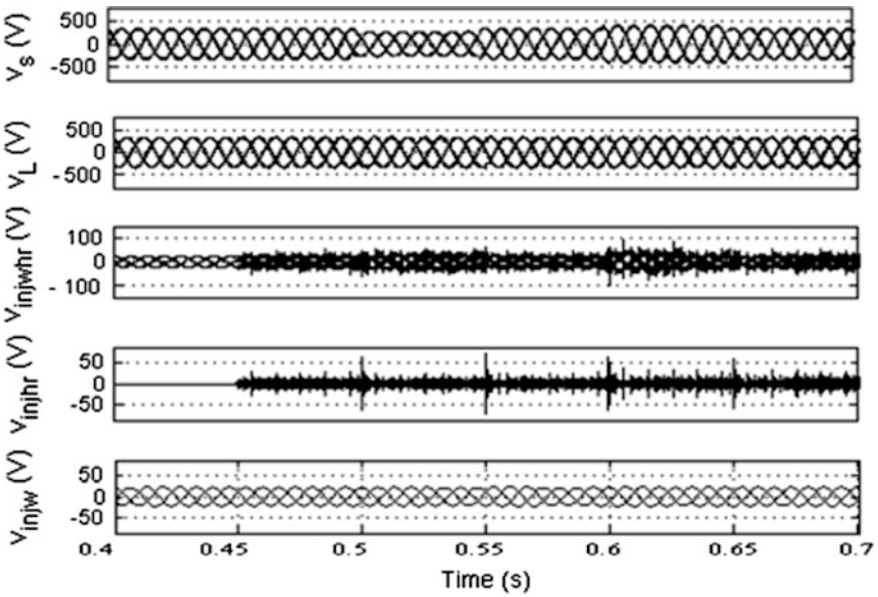


Fig. 4. Shows the performance of hybrid control approaches under various voltages conditions of utility, utility voltages ($v_{abc}(V)$), PCC voltage ($vL_{abc}(V)$), series transformer output voltages with harmonic pollution ($v_{inj.hr}(V)$), and series transformer voltage without harmonic pollution ($v_{inj}(V)$)

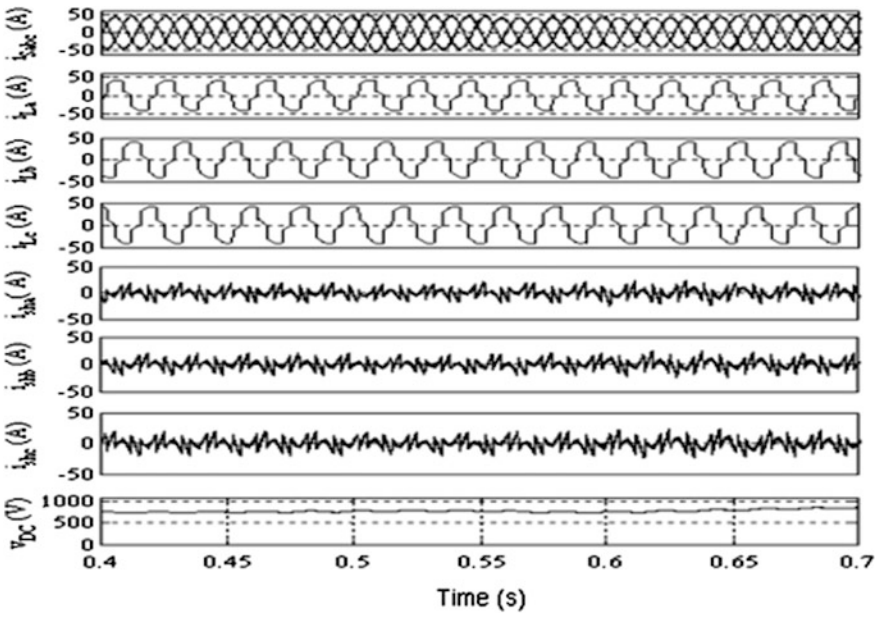


Fig. 5. Shows the performance of shunt control approaches under various voltage conditions of utility, utility currents ($i_{abc}(A)$), load currents ($iL_a(A)$), ($iL_b(A)$), and ($iL_c(A)$), shunt currents ($i_{sha}(A)$), ($i_{shb}(A)$), and ($i_{shc}(A)$), DC voltage (V_{DC} (V))

It happened because of the absence of voltage harmonics injection through series controller. Moreover, when the series hybrid controller injects voltage harmonic, during t = 0.45–0.7 s, the AC mains voltage and PCC voltage contain 2.75 and 2.9% voltage harmonics which are shown in Fig. 6e and f. Thus, the hybrid series controller can decrease voltage harmonics. While shunt converter is used to abolish current harmonic and reduce the burden of reactive power of the utility, during t = 0.5–0.55 s, the AC mains voltage is reduced by 20%. The active and reactive powers of the AC mains are compensated by extra series voltage injection using series converter. The extra power is seeking through shunt converter under voltage sag condition, thus the AC mains currents are raised to maintain the DC link voltage which is shown in Fig. 5. During t = 0.6–0.65 s, the AC mains is under voltage swell condition. The series voltage is injected 180° out of phase to compensate the swell voltage of the loads. Performance of proposed UPQC is represented in Fig. 7.

3.4 Active and Reactive Power Sharing

Figure 7 represents the condition of active and reactive powers of the AC mains. For t = 0.4–0.5 s, the AC mains supplies only active power of 12.5 kW and load demand of reactive 1.9 kVAR are supplied through UPQC. For t = 0.5–55 s, the AC mains is under voltage sag condition, and additional active and reactive powers are supplied through series converter with DC link; under this condition, the shunt converter maintains output of the DC link to supply the extra active and reactive power to the

load. The angle δ is fixed at 5.29°. The AC mains supplies voltage and current at unity power factor during t = 0.6–0.65 s (under voltage swell). The series converter consumes extra active power and back to the shunt converter to maintain the output current, and thus load active power is always constant throughout the simulation work.

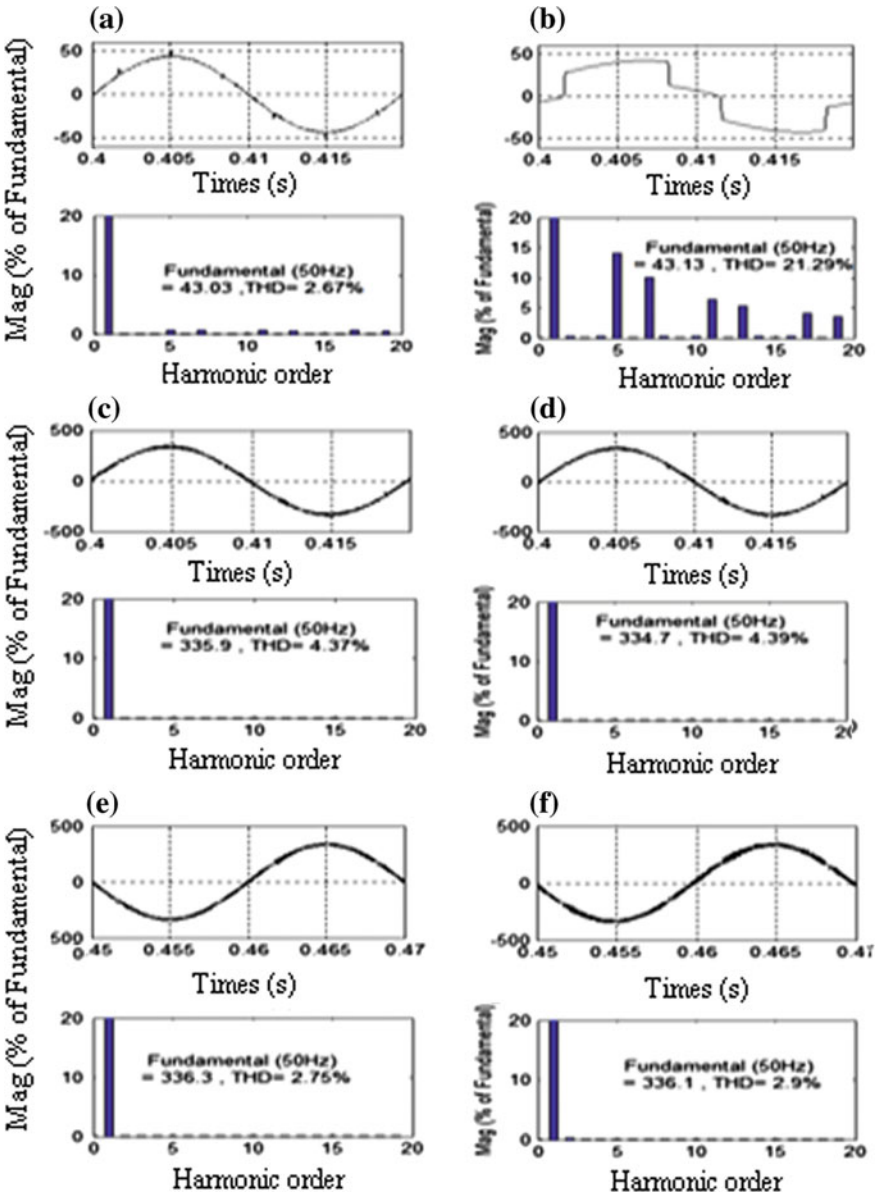


Fig. 6. Harmonics of the voltage and current. **a** utility current. **b** Load current. **c** Utility voltage without hybrid controller. **d** Load voltage without hybrid controller. **e** Utility voltage with hybrid controller. **f** Load voltage with hybrid controller

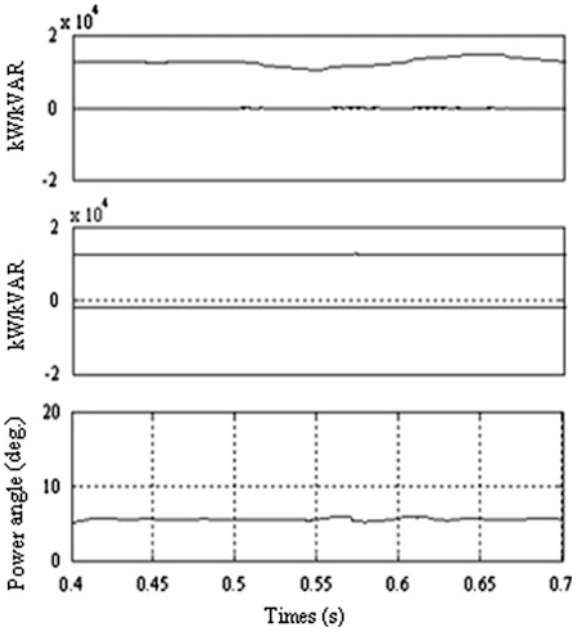


Fig. 7. Active and reactive power flow under constant power angle, utility active power *(P_s (kW))*, utility reactive power *(Q_s (kVAR))*, Load active power *(P_L (kW))*, load reactive power *(Q_L (kVAR))*, and power angle between utility and PCC voltage

Table 1. Summary of 3P3W system

Parameters	Magnitude
Three-phase line voltage	415 (L-L)
Line impedance	0.001 + j0.8 Ω
Nonlinear load and linear load passive filter parameters	Three-phase DBR load feeding a R-L load of 20 Ω–200 mH
	C_{se} = 10 μF, Rse = 10 Ω
	C_{sh} = 10 μF, R_{sh} = 10 Ω
Series inductance for series converter	L_{se} = l mH,
Series inductance shunt converter	L_f = 2 mH
DC capacitor voltage	725 V
Transformer	1:2, 5 kVA
Pulse width modulation	10 kHz

4 Conclusion

The power angle concept based controller is incapable to enhance the MPQ of voltage in AC distribution network. Thus, a hybrid controller is proposed in this work to generate gate pulses of IGBT-based series converter. A hybrid control approach is used

to enhance the multiple voltage-related issues. The performance of the hybrid control algorithm has been worked satisfactorily on MATLAB platform. The hybrid control algorithm can eliminate the voltage harmonics generated by source. These approaches can also be used in solar-based microgrid where source generates voltage harmonics. The IGBT-based shunt converter is used to enhance the multiple current qualities. Moreover, the performance of shunt control algorithm performs satisfactorily under various conditions of AC mains.

References

1. Hingorani NG (1995) Introducing custom power. IEEE Spectrum V 32(6):41–48
2. Fujita H, Akagi H (1998) The unified power quality conditioner. The integration of series- and shunt-active filters. IEEE Trans Power Electron 12(2):315–322
3. Zhilei Y, Lan X (2013) Control of single-phase grid-connected inverters with nonlinear loads. IEEE Trans Ind Electron 60(4):1384–1389
4. Ragavan S, Subramanian M (2016) Intelligent controller for UPQC using combined neural network. Circuits Syst 7(6):680
5. Abdul QA, Bhim S, Mashhood H (2015) Algorithm for power angle control to improve power quality in distribution system using unified power quality conditioner. IET GTD 9(9):1439–1447
6. Koteswara RU, Mahesh KM (2009) A new control strategy for load compensation in power distribution system. In: International conference on power systems, Kharagpur, pp 1–6
7. Alireza J, Abdelhamid H, Lyne W, Kamal A (2016) Experimental investigation on a hybrid series active power compensator to improve power quality of typical households. IEEE Trans Ind Electron 63(8):4449–4459
8. Subhashish B, Deepak D (1995) Design and implementation of a hybrid series active filter system. In: 26th Annual IEEE, power electronics specialists conference, Atlanta, GA, vol 1, pp 189–195
9. Mashhood H, Abdul QA, Bhim S (2015) Parameters estimation of a series VSC and shunt VSC to design a unified power quality conditioner (UPQC). In: 39th national system conference, Noda, pp 1–6

Comparative Performance of the Various Control Techniques to Mitigate the Power Quality Events Using UPQC

Rajeev Kumar Chauhan[1(✉)], Jai Prakash Pandey[2], and Mashhood Hasan[3]

[1] IMS Engineering College (affiliated to AKTU, Lucknow), Ghaziabad, Uttar Pradesh, India
mmmec.rkc@gmail.com
[2] Department of EE, Kamla Nehru Institute of Technology, Sultanpur, India
tojppandey@rediffmail.com
[3] Department of EE, Galgotia College of Engineering and Technology, Gr. Noida, India
mirmashhood2010@gmail.com

1 Introduction

The power qualities [1] in AC distribution system are voltage flickering, voltage fluctuation, unbalanced voltage, voltage sag/swell, current harmonics, and reactive power, and when PQ comprise with harmonics is called multiple power quality (MPQ) [2]. The PQ and MPQ are the common exiting problems in AC distribution network. This problem ascended because of highly necessitated electronics based device, like computer, mobile, laptop, power electronics oven, washing machine, etc. Electronics or power electronics devices generate pollution in the AC distribution network. The multiple mitigation devices are available to enhance PQ and MPQ. Earlier the passive elements like capacitor and inductor are grouped to mitigate the PQ and MPQ, as the passive elements have limitation. Therefore, active filters are trendy to enhance the PQ of AC networks. The most popular active power filter is unified power quality conditioners (UPQC) which is the combination of dynamic voltage restorer (DVR) and the D-STATCM. It enhances unified power quality condition of the AC distribution network, so it is named as UPQC. The DVR is series part of the UPQC, whereas D-STATCOM is shunt part of the UPQC. The DVR is used to enhance the voltage profile of AC distribution network and D-STATCOM enhances the current profile of the network. The DVR, D_STATCOM, and UPQC are insulated gate bipolar transistor (IGBT)-based converter, which requires gate drive circuit for proper operation of the converter-based active filters like DVR, D-STATCOM, and UPQC. The UPQC is universal choice to enhance the PQ and MPQ in low-level three-phase AC voltage circuit or three-phase four-wire circuits. The UPQC is categorized on the basis of sharing of active and reactive power through series converter/DVR of the UPQC. The active and reactive power sharing depends on the angle of voltage injection with the utility (AC mains) current. If the voltage injection is in series with utility currents, to improve the voltage profile [3], it means UPQC shares active power so it is

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_4

named as UPQC-P (Where P stands for active power). The second mode of series voltage injection in 90° with utility currents is to improve the voltage quality of the point of common coupling (PCC); under this case, the UPQC uses reactive power so it is named as UPQC-Q [4–7] (Where Q stands for reactive power). And the third mode of voltage injection creates an angle with utility currents which makes two components as active and passive components. These components share complex power (Active and reactive power), so UPQC is named as UPQC-S (Where S stands for complex power). Moreover, the injection of voltage at a certain angle has more advantageous than that of series injection and quadrature injection. As the series and quadrature injection of voltage have minimum utilization factor of series converter of the UPQC than that of the UPQC-S [8–13], on the other hand, the UPQC-P requires active power to mitigate voltage sag/swell, whereas the UPQC-Q is unable to compensate voltage swell. Therefore, UPQC-S is advantageous to opt as an active filter in AC distribution network.

2 IGBT Gate Firing Techniques of UPQC

The output of the UPQC depends on the customer demands. The customer demands are fulfilled by its control techniques; therefore, control techniques to fire the gate of insulated gate bipolar transistors (IGBTs) are the heart of the UPQC. The concepts of the multiple control techniques for series converter and single control technique for shunt converter of the UPQC are presented in this section.

2.1 Techniques to Fire the Gate of IGBT-Based Series Converters of UPQC

The objective of IGBT-based series converter of the UPQC is to control the voltage quality (voltage sag/swell and reactive power compensation). Here, the three techniques are evaluated using simple addition/subtraction mathematical concepts for IGBT-based series converter. These concepts are based on how to inject the voltage with respect to utility current in AC distribution network. The three control techniques are as follows.

2.1.1 Power Angle Control Technique

The vector diagram representation of power angle control technique is presented in Fig. 1a. The concepts of injection voltages (v_{injS}) under various conditions of voltage are given. In power angle control technique, a power angle or load angle (δ_S) is established in between utility voltage (V_S) and voltage at PCC (V_L) under steady state (S.S) condition of AC distribution network. In S.S condition, the IGBT-based series converter injects voltage (v_{injS}) without disturbing in AC network system. An angle of v_{injS} leads the phasor of load current I_L to I_L^I and consequently phase angle (Φ_L) of the load currents leads to an angle ($\Phi_L - \delta_S$). An angle of v_{injS} results in compensating of reactive power under S.S condition of AC distribution network. Therefore, both the converters of UPQC-S compensate reactive under S.S condition.

In Fig. 1a, given physical parameters are estimated using triangle law to design a control algorithm for firing the gate pulses of IGBT-based series converter. The physical parameters are load angle (δ_S), magnitude of injection voltages (v_{injS}), and an angle of voltage injection with respect to utility current ($\Phi_{inj.S}$) which can be estimated as

$$\delta_S = \sin^{-1} \frac{Q_{sr}}{P_S = P_L} \tag{1}$$

$$V_{injS} = V_L \times 2 \sin(\frac{\delta_S}{2}) \tag{2}$$

$$\phi_{injS} = (90 + \frac{\delta_S}{2}) \tag{3}$$

Q_{Sr} represents sharing compensated reactive of IGBT-based shunt converter by IGBT-based series converter of the UPQC. And P_S represents active power of the utility. Equations (1–3) require an instantaneous value to develop power angle concept based controller. An instantaneous voltage injection angle ($\Phi_{injS} = 90° + \delta_S/2$) get plus with angular angle ωt. The ωt is extracted using phase lock loop (PLL). The sum of (Φ_{injS} + ωt) is passed by trigonometric sine function and gets it multiply by magnitude of instantaneous v_{injS} (t). It generates reference injection voltage ($v^*_{injSabc}$) for three-phase AC distribution network. The reference voltage is subtracted by sensed output of series transformer voltage and passes this signal through pulse width modulation (PWM) to generate the required injection voltage in AC distribution network.

2.1.2 Quadrature Angle Technique

The penetration voltage v_{injQ} in quadrature to the utility current under voltage sag condition is presented in Fig. 1b. In this technique, no active power is consumed to mitigate the voltage sag; however, this technique is unable to mitigate voltage swell condition. The instantaneous voltage sensor is required to design such a controller. Instantaneous voltage gets divided by the constant reference voltage magnitude. The output term gives $cos\delta$, and thus power angle δ can be calculated. Using control mathematics [5], the required v_{injQ} is to be mitigated by the voltage sag condition. A PLL is required to generate the terms sinωt and cosωt which generates reference signals at 90° that lead to the utility current signals. These signals pass through PWM to penetrate output voltage by IGBT-based series converter.

2.1.3 Zero and 180° Angle Techniques

Figure 1c and d represents the phaser diagrams of voltage sag and swell condition techniques, respectively. In this concept, voltage is injected in phase or zero degree angle with utility current under voltage sag condition, whereas at 180° or out of phase voltage is penetrated under voltage swell condition and it remains inactive under normal condition. A simple algorithm unit vector template generation (UVGT) is useful for generation of reference signals. The UVTG is used to sense magnitude of utility voltages and divide the sensed magnitude by V_m, where V_m is the maximum peak magnitude of utility voltage; accordingly, outputs are unit magnitude under all condition of utility. The unit magnitude gets pass through PLL output terms sine or cosine

and multiplied by desire load voltage magnitude V_{Lm}. These combinations of mathematics give the reference voltage signals. The difference of actual signals and reference signals pass through PWM to get feedback through IGBT-based series converter.

2.2 Technique to Fire the Gate of IGBT-Based Shunt Converters of UPQC

The purpose of IGBT-based shunt converter is to improve current quality of utility and maintain the DC link voltage under all condition of the utility. The injection of shunt current needs a feedback controller. A feedback controller is designed using the concept of instantaneous symmetrical component theory which is given as below.

2.2.1 Control Theory for IGBT Shunt Converter

The block diagram of instantaneous symmetrical component theory based controller is shown in Fig. 2. This theory is extracted zero sequence (v_{Sa}^{0}), positive sequence (v_{Sa}^{+}), and negative sequence voltages (v_{Sa}^{-}) component which are calculated as

$$\begin{pmatrix} v_{Sa}^{0} \\ v_{Sb}^{+} \\ v_{Sb}^{-} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & a & a^2 \\ 1 & a^2 & a \end{pmatrix} \begin{pmatrix} v_{Sa} \\ v_{Sb} \\ v_{Sc} \end{pmatrix}, \tag{4}$$

where a is equal to operator $e^{j2\pi/3}$. The utility currents are given as $i_{sa,}$ $i_{sb,}$ and i_{sc} to gather to evaluate the satisfactory condition to compensate the load reactive power. Three necessary conditions are needed to improve the current quality of the utility. These are as follows:

$$i_{Sa} + i_{Sb} + i_{Sc} = 0 \tag{5}$$

The second is the reactive power compensation through the phase angle controlled by positive sequence voltage and current, which are given as

$$\angle v_{Sa1}^{+} = \angle v_{Sa1}^{+} + \phi^{+} \tag{6}$$

This implies that

$$\angle(v_{Sa1} + a v_{Sa1} + a^2 v_{Sa1}) = (v_{Sa1} + a v_{Sa1} + a^2 v_{Sa1}) + \phi^{+} \tag{7}$$

Further simplifying equations,

$$\angle(v_{Sb} - v_{Sc} + 3\gamma v_{Sa}) i_{Sa} + (v_{Sc} - v_{Sa} + 3\gamma v_{Sb}) i_{Sb} + (v_{Sa} - v_{Sb} + 3\gamma v_{Sc}) i_{Sc} = 0 \tag{8}$$

where the $\gamma = \tan\phi^{+}/\sqrt{3}$ and the Φ^{+} is the angle between instantaneous value of v_{Sa1}^{+} and i_{Sa1}^{+}. Similarly, negative component of the source voltage and the current equations to compensate the reactive power are given as

$$v_{Sa1}^{-} = \angle v_{Sa1}^{-} + \phi^{-} \tag{9}$$

$$\angle(v_{Sc} - v_{Sb} + 3\gamma v_{Sa})i_{Sa} + (v_{Sa} - v_{Sc} + 3\gamma v_{Sb})i_{Sb} + (v_{Sb} - v_{Sa} + 3\gamma v_{Sc})i_{Sc} = 0, \tag{10}$$

where β is equal to $\tan\phi^{-}/\sqrt{3}$ and ϕ^{-} is the angle between v_{Sa1}^{-} and i_{Sa1}^{-}.

When phasor sum of currents and voltage is zero for both positive and negative sequences, both the angle γ and β are zero. The calculation is given as follows:

$$(v_{Sb} - v_{Sc})i_{Sa} + (v_{Sc} - v_{Sa})i_{Sb} + (v_{Sa} - v_{Sb})i_{Sc} = 0 \tag{11}$$

And the third condition is that utility should supply the average loads power (P_{Lavg}) as

$$v_{Sa}i_{Sa} + v_{Sb}i_{Sb} + v_{Sc}i_{Sc} = P_{Lavg} \tag{12}$$

The extraction of three-phase reference currents can be evaluated using Eqs. (5), (11), and (12), which are given as

$$i_{Sa}^{*} = \frac{v_{Sa} - v_{Sa}^{0}}{\Delta}$$

$$i_{Sb}^{*} = \frac{v_{Sb} - v_{Sb}^{0}}{\Delta}$$

$$i_{Sc}^{*} = \frac{v_{Sc} - v_{Sa}^{0}}{\Delta}, \tag{13}$$

where v_{Sa}^{0} is equal to 1/3 $\sum_{j=a.b,c} v_{Sj}$ and Δ is equal to $\sum_{j=a.b,c} v_{Sj}^{2} - 3(v_{Sa}^{0})^{2}$. Under balanced condition, zero sequence voltage v_{Sa}^{0} is zero, and under unbalance condition it is nonzero. The required reference source currents are subtracted from the load currents to generate the gating pulse to the IGBT-based shunt converter, which can be evaluated as

$$i_{sha}^{*} = i_{la} - i_{sa}^{*} = i_{la} - \frac{v_{Sa} - v_{Sa}^{0}}{\Delta}(P_{Lavg} + P_{loss})$$

$$i_{shb}^{*} = i_{lb} - i_{sb}^{*} = i_{lb} - \frac{v_{Sb} - v_{Sb}^{0}}{\Delta}(P_{Lavg} + P_{loss})$$

$$i_{shc}^{*} = i_{lc} - i_{sc}^{*} = i_{lc} - \frac{v_{Sc} - v_{Sc}^{0}}{\Delta}(P_{Lavg} + P_{loss}) \tag{14}$$

The switching loss and average load power are presented as P_{lavg} and P_{loss}, respectively. The ohmic loss and switching losses are compensated by the difference of reference DC voltage (v_{DC}^{*}) and the actual sensed DC voltage $v_{DC}(n)$ for nth sampling

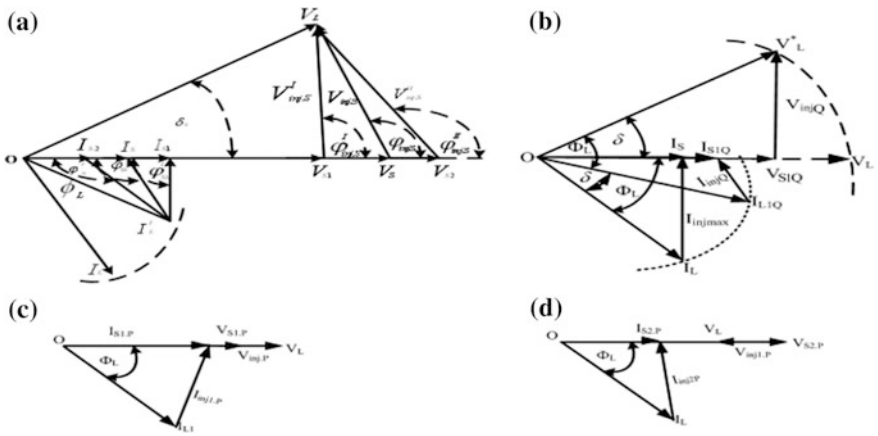


Fig. 1. **a–d** Phaser representation of proposed control techniques under various conditions of voltage

time. An error $\varepsilon(n)$ of DC voltage is passed through a PI controller to maintain the DC link voltage of the UPQC. The error and losses can be calculated as

$$\varepsilon(n) = v_{DC}^{*}(n) - v_{DC}(n) \tag{15}$$

$$p_{loss}(n) = p_{loss}(n-1) + k_p\{\varepsilon(n) - \varepsilon(n-1)\} + k_i(n), \tag{16}$$

where k_p and k_i are the PI gain constants.

3 Results and Discussions

The single-line diagram of UPQC is represented as in Fig. 2. The same model has been developed on MATLAB–Simulink to analyze the graphical approaches of comparative performance of UPQC under S.S and voltage sag conditions. The IGBT-based series converter is performed under three control techniques, while IGBT-based shunt converter is performed under instantaneous symmetrical component theory based controller. All the MATLAB-based Simulink parameters are presented in Table 1.

The notation of results based on 415 V (L-L), three-phase, system parameters are as follows—utility voltage v_s (V), utility current i_s (A), load current i_L (A), load voltage v_L (V), and DC link voltage v_{DC} (V). The notation $\{v_s$ (V) & i_s (A)$\}$ shows the utility voltage and utility current phase. The performances of series injection voltages and shunt penetration currents are observed under 20% voltage sag for three types of UPQC (UPQC-S, UPQC-Q, and UPQC-P). The optimized power angle has been taken as 0.1 rad. ($\approx$5.7°) of UPQC-S to share the reactive shunt converter. In Fig. 3a, complete wave signals are given from 0.1 to 0.25 s, while between 0.1–0.15 s and 0.2–0.25 s, the UPQC-S works under S.S conditions. Moreover, the IGBT-based series converter generates voltage ($v_{inj.S}$) 23.83 V per phase and shares 0.233 kVAR per phase and the

IGBT-based shunt converter generates current ($i_{inj.S}$) 3.2 A to share 0.768 kVAR per phase. Both the IGBT-based converters shared total reactive powers of the loads; thus, wave signals of {v_s (V) & i_s (A)} show the utility works under unity power factor (upf). On other hand, utility runs from 0.15 to 0.2 s, under 20% voltage sag. Then, IGBT-based series converter generates 49.17 V to compensate the voltage sag at PCC, while IGBT-based shunt converter generates 2.5 A to maintain the v_{DC} at 700 V and compensate the reactive power. Thus, wave signals of {v_s (V) & i_s (A)} show utility works under upf. The performance of controller based on proposed techniques works satisfactorily.

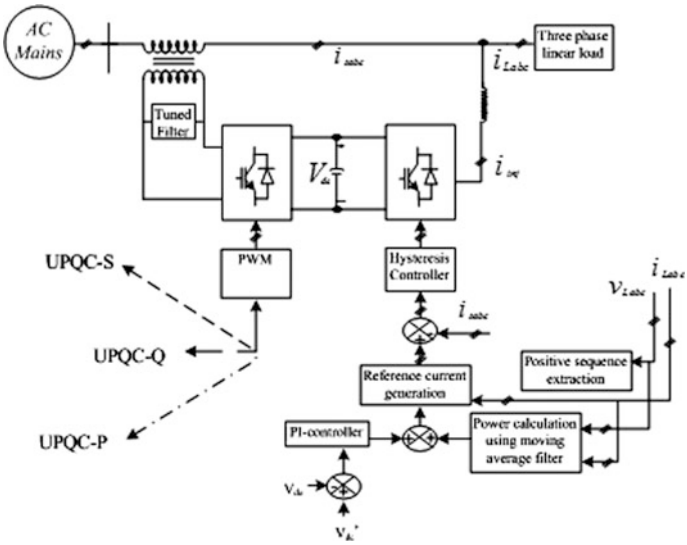


Fig. 2. Single-line diagram of AC network with UPQC

In Fig. 3b, complete wave signals are given from 0.25 to 0.4 s, while between 0.25–0.3 s and 0.2–0.25 s, the UPQC-Q works under steady state conditions. Moreover, the IGBT-based series converter is inactive under S.S condition. Only IGBT-based shunt converter generates shunt currents 4.2 A to compensate total reactive power demand by load. The wave signals of {v_s (V) & i_s (A)} show that utility voltage and current are in same phase. Moreover, for 0.3–0.35 s, utility runs under 20% voltage sag. The IGBT-based series converter generates 144 V to maintain the desired voltage at PCC. And at same time IGBT-based shunt converter of UPQC-Q generates 0.9 A to share the load reactive power and maintains DC link voltage. It has been seen that series converter generates very high voltage under voltage sag to maintain the load voltage. Therefore, VA rating of the series transformer and VA rating of IGBT-based series converter are high, and the overall rating of the UPQC-Q is also very high. Hence, this kind of control approach is unacceptable as a universal choice. In Fig. 3c, complete wave signals are given from 0.4 to 0.55 s, while between 0.4–0.45 s and 0.5–0.55 s, the UPQC-P works under S.S conditions. In UPQC-P types of control approach, the

IGBT-based series converter is inactive mode and incapable to share reactive power of load. The IGBT-based shunt converters compensate total reactive demand of the load and it maintains DC link voltage at desire value 780 V. It generates shunt currents 4.2 A under S.S condition to compensate 1 kVR reactive power per phase and the wave signals of {v_s (V) & i_s (A)} show that utility voltage and current are in same phase. When utility runs during 0.45–0.5 s under 20% voltage sag, the IGBT-based series converter generates 48 V to maintain the desired voltage at PCC. The IGBT-based series

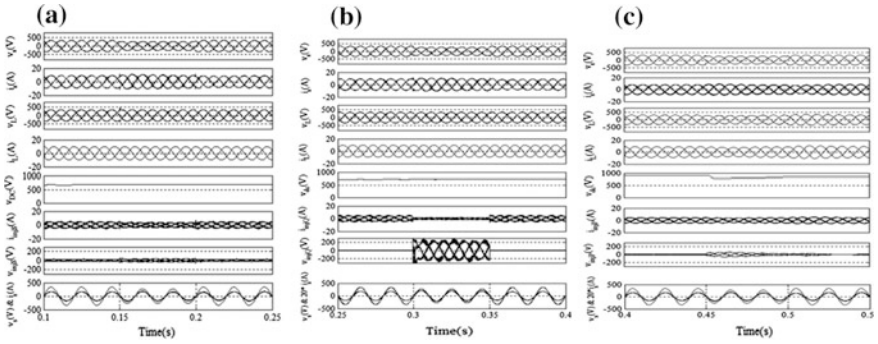


Fig. 3. a–c Performance of controller of UPQC-S, UPQC-Q, and UPQC-P represent, respectively, utility voltage *($v_s(V)$)*, utility current *($i_s(A)$)*, voltage at point of common coupling *($v_L(V)$)*, current at point of common coupling *($i_L(A)$)*, capacitor DC voltage ($V_{DC}(V)$), shunt converter current at point of common coupling *($i_{inj.S}(A)$)*, series converter voltage *($v_{inj.S}$ (V))* for UPQC-S, shunt converter current at point of common coupling *($i_{inj.Q}(A)$)*, series converter voltage *($v_{inj.Q}$ (V)) for UPQC-Q*, shunt converter current at point of common coupling *($i_{inj.P}(A)$)*, series converter voltage *($v_{inj.P}$ (V)) for UPQC-P, respectively*

Table 1. AC distribution network parameters

Parameters	UPQC-S	UPQC-P	UPQC-Q
DC link voltage (V_{dc}) (V)	680	780	700
DC link capacitance (C_{dc}) (μF)	2500	2500	2500
AC inductance for shunt converter (L_f) (mH)	2	4	2
AC inductance for series converter (L_{se}) (mH)	2.5	2.5	2.5
Voltage rating of switch $V_{Sw} = V_{dc}$ + 10% of V_{dc} (V)	1200	1200	1200
Current rating of switch (I_{Sw}) (A)	50	50	50
Ripple filters of series converter	C_{se} = 10 μF	C_{se} = 10 μF	C_{se} = 10 μF
	R_{se} = 10 Ω	R_{se} = 10 Ω	R_{se} = 10 Ω
Ripple filters of series converter	C_{se} = 10 μF	C_{se} = 10 μF	C_{se} = 10 μF
	R_{se} = 10 Ω	R_{se} = 10 Ω	R_{se} = 10 Ω
Injection transformer	5 kVA, 120 V/240 V	5 kVA, 120 V/240 V	10 kVA, 120 V/240 V
Linear load	5 kVA at 0.8 pf	5 kVA at 0.8 pf	5 kVA at 0.8 pf
PWM (kHz)	10	10	10

voltage generation needs active power which shares from DC link. Thus, amount of DC link voltage decreases from 0.45 to 0.5 s. And the IGBT-based shunt converter generates 2.5 A to maintain the DC link voltage and compensate the reactive power of the load. Thus, wave signals of $\{v_s$ (V) & i_s (A)$\}$ is shown at the same phase under all conditions of utility. It has been noticed that the utilization factor of IGBT switches for UPQC-P and also for UPQC-Q less than that of UPQC-S. Thus, UPQC-S has good choice for power quality enhancement in AC distribution networks.

4 Conclusion

As UPQC–Q is unable to work under swell condition, the performance of the minimized VA size is compared under steady state and voltage sag conditions. It has been observed that the utilization factor of the IGBT switches of series converter is maximum for UPQC-S whereas minimum for UPQC-Q and UPQC-P. The IGBT switches of series converter work only in voltage sag condition for UPQC-Q and UPQC-P, whereas IGBT switches work under all conditions of utility. Moreover, the magnitude of voltage under voltage sag condition requires nearly same for UPQC-S and for UPQC-P, while UPQC-S requires less active power than that of UPQC-P. As per MATLAB results, UPQC-S, UPQC-P, and UPQC-Q require 49.7 V, 48 V, and 144 V magnitudes of voltage correspondingly to compensate 20% voltage sag. Again it is seen that magnitude of UPQC-Q is nearly three times more than other types of UPQC. The factors VA size, utilization factor of switches, minimum active power utilization, and effective cost which is proportional to the VA size of UPQC prefer the UPQC-S over the UPQC-Q and UPQC-P.

References

1. Mohammed S, Ibrahim R, Rao SR, Perumal N (2013) Performance evaluation of R-UPQC and L-UPQC based on a novel voltage perfections detection algorithm. Int Rev Electr Eng 8(4)
2. Padiyar KR (2007) FACTS controllers in power transmission and distribution. New Age International Publisher
3. Ameli A, Bahrami S, Khazaeli F, Haghifam MR (2014) A multi objective particle swarm optimization for sizing and placement of DGs from DG owner's and distribution company's viewpoints. IEEE Trans Power Delivery 29(4):1831–1840
4. Malabika B, Das SP, Dubey GK (2002) Performance study of UPQC-Q for load compensation and voltage sag mitigation. In: IEEE 28th annual conference. Industrial Electronics Society, pp 698–703
5. Khadkikar V, Chandra A (2006) A novel control approach for unified power quality conditioner Q without Active power injection for voltage sag compensation. In: Proceeding IEEE international conference industrial technology, pp 779–784
6. Basu M, Das SP, Dubey GK (2008) Investigation on the performance of UPQC-Q for voltage sag mitigation and power quality improvement at a critical load point. IET Gener Transm Distrib 2(3):414–423

7. Lee WC, Lee DM, Lee TK (2010) New control scheme for a unified power-quality compensator-Q with minimum Active power injection. IEEE Trans Power Deliv 25(2):1068–1076
8. Kolhatkar Y, Das S (2007) Experimental investigation of a single-phase UPQC with minimum VA loading. IEEE Trans Power Deliv 22(1):371–380
9. Khadkikar V, Chandra A (2011) UPQC-S: A novel concept of simultaneous voltage sag/swell and load reactive power compensations utilizing series inverter of UPQC. IEEE Trans Power Electron 26(9):2414–2425
10. Ansari AQ, Singh B, Hasan M (2015) Algorithm for power angle control to improve power quality in distribution system using unified power quality conditioner. IET Gener Transm Distrib 1:39–47
11. Hasan M, Ansari AQ, Singh B (2015) Parameters estimation of a series VSC and shunt VSC to design a unified power quality conditioner (UPQC). In: 39th national system conference, Noida, pp. 1–6
12. Basu M, Das SP, Dubey GK (2007) Comparative evaluation of two models of UPQC for suitable interface to enhance power quality. Electr Power Syst Res 77:821–830
13. Pal Y, Swarup A (2014) A comparative analysis of UPQC-P, UPQC-Q, and UPQC-VA_min: a simulation study. In: Proceedings of the international conference on power systems, energy, environment, Interlaken, Switzerland

A Mobile Health Intervention to Support TB Eradication Programme for Adherence to Treatment and a Novel QR Code Based Technique to Monitor Patient–DOTS Provider Interaction

K. Navin[1(✉)], G. Vadivu[1], Avinash Maharaj[1], Tom Thomas[2], and S. Lavanya[1]

[1] SRM University, Chennai, India
navin.k@ktr.srmuniv.ac.im
[2] Olive Touch Healthcare Services Limited (MARG), Chennai, India

1 Introduction

Smartphone app-based mobile health intervention is becoming a viable solution to support TB eradication programme. It could bring unique solutions by adapting newer technologies available through the technological growth. But it could be effectively utilized only if proper human–computer interaction (HCI) is taken care of. The proposed mHealth solution to support TB programme is implemented through TB track mobile app and web application supported by database. The app provides support to the TB patients in taking right medicine for the day, motivating and educating the patients with regular messages, and helping to report symptoms of uneasiness to invite attention. Virtual medicine strips, symbolic representation, and localization features were employed in the designing of app with mobile HCI aspects in mind. A QR code technique based remote attendance monitoring system using smartphones to monitor patient–DOTS provider interaction is proposed to support Directly Observed Therapy (DOT) concept recommended by WHO. The proposed mHealth solution is designed to address the challenges mentioned in the literature survey. The methodology and results through pilot testing demonstrate the outcome of implementing the smartphone-based mHealth solution to support the TB programme.

2 Literature Survey

World Health Organization (WHO) TB statistics reports for the year 2015 point out that estimated incidence of TB for India is around 2.2 million cases out of a global incidence of 9.6 million with more than 2 million new cases every year [1]. Monitoring treatment response, treatment adherence, maintaining records for all TB patients, public health responsibility, notification of TB cases, and health education are quite few

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_5

important guidelines where the National TB programme implements on TB patients in India as per the recommendations of WHO [2]. WHO insists the patients should be observed while taking their medicine which could be implemented through DOT as the TB control strategy [3]. The crucial element of the TB programme is to focus on strict adherence to TB treatment by patient until cure. It involves continual medicine intake and periodical TB test for patients until cure and avoidance of wrong dosage of medicine. Failing to do so will result in increase in cases of drug resistance TB which will complicate in further treatment. Treatment supervision by DOT in clinic will be an additional burden due to resource constraint and cost-intensive for health clinics. Patients might also have constraints in visiting clinic daily due to the reasons of travel and waiting time for treatment. The alternate choice could be a community-based DOT (CB-DOT) which can be from organized community groups, peer groups, chosen members of the community, and family members if trained could act as potential supervisors to ensure completion of treatment and hence cure [4]. Patient should not be left as mere passive recipient of services; they should be motivated and educated and should be an active partner in TB programme. They should be made to have responsibility in sharing information to the public health, adhering treatment properly, and come forward in sharing their expertise experienced during treatment [5]. Increasing penetration of smartphones among common people because of its lowering cost and the proliferation of high-speed mobile networks and services available at cheaper prices, the smartphone-based mHealth solution will play a far greater role in health care in both developed and developing countries in the future [6–8]. Success of mHealth solution will be based on good HCI design which should be user-centric; in fact, technological developments and its impact could be effectively utilized only if proper focus is given to the usability of the technology fitting into daily life of common people. The design approach should be based on intellectual traditions of several disciplines including human factors, computer science, information systems, psychology, sociology, and visual design [9–11]. Ergonomics and psychology are also needed to be considered in designing the TB track mobile app [12–14]. With technology becoming viable, mHealth solution will have to become an inevitable part of strategic design for setting future TB programme standards where mobile phone will the easiest mode to reach outpatients who are located in places where difficulty in approaching them is high. Proof of concept and studies have shown that mHealth has been influential in TB treatment through medication monitoring, record keeping, and possibly help in motivating and educating patients. This could be useful when scarcity of human observers is an issue [15].

3 Methodology

The architecture diagram of TB track application framework is shown in Fig. 1. Smart mobile phones are provided to identify tuberculosis patients with mobile application TB track loaded to it. The app enables to monitor them throughout the course of treatment. The system also has web-based application intended for TB units where patient details and schedules for treatment can be entered. The web application can also

generate reports on patients record like daily medicine intake, their reported symptoms, and more. The system ties up with third-party SMS service provider which enables web application to implement message communication to be sent to respective patients, DOTS providers, and medical supervisors. The web app can compute reminders separately for each patient if the patient had skipped medicine. It can send alert messages to intended DOTS provider if the patient has reported symptoms of uneasiness. It also sends motivational and educational messages at regular interval. The DOTS provider is also provided with mobile app to report their attendance. The collected data such as patient medical intake, patient symptoms occurring during course of treatment, report of sputum test, and DOTS provider attendance are stored in database; these data can be analyzed and report can be generated for perusal.

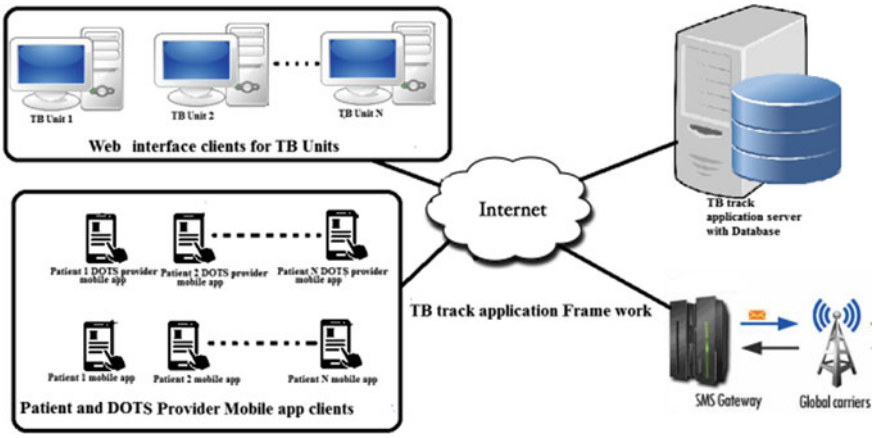


Fig. 1. Architecture diagram of TB track application for TB patients

4 TB Track Application

4.1 TB Patient Mobile App

The internet-enabled mobile application is used by patients to report their daily intake of medicine during the different phases of treatment namely intensive phase, prolongation phase, and continuous phase. The data will be collected by the web application and stored in database. Figures 2 and 3 show the TB track mobile app screenshot of a patient in "intensive phase" who is presented with virtual tablet strip and voice support. The patient needs to report by clicking the medicine images to TB track database to confirm that correct medicine intake is taken. Figures 4 and 5 show the screenshot of patient in "continuous phase" who is presented with different tablet dosages for intake.

Figures 6 and 7 show the screenshot of the app through which patient can report any symptoms occurring during the course of medication immediately through SMS to DOTS provider which is also recorded in database. The patient will receive educational, motivational messages, and also reminders or alert messages for intake of

Fig. 2. Virtual medicine strip of patient in intensive phase

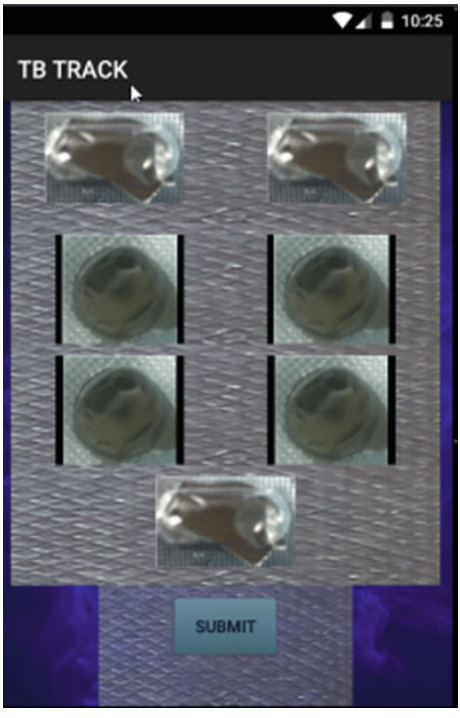


Fig. 3. Patient has submitted medicine intake

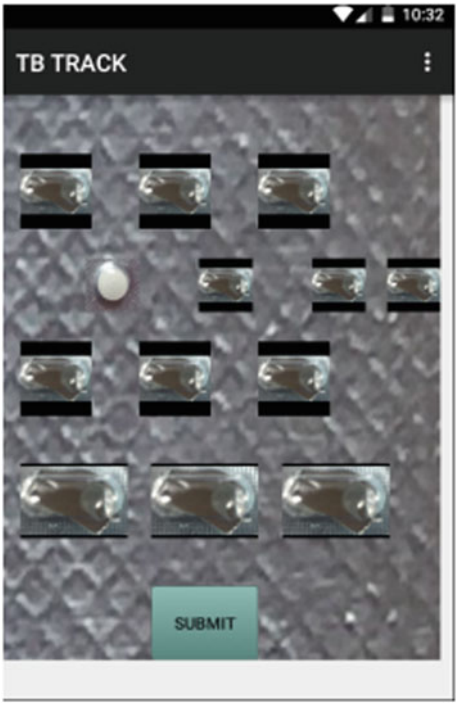


Fig. 4. Virtual medicine strip of patient in continuous phase

Fig. 5. Virtual medicine strip of patient in continuous phase

Fig. 6. Symptoms reporting screen

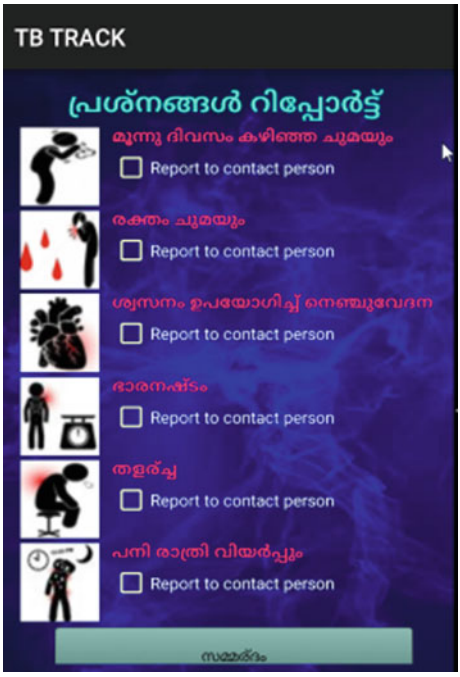


Fig. 7. Symptoms reporting screen in Malayalam language

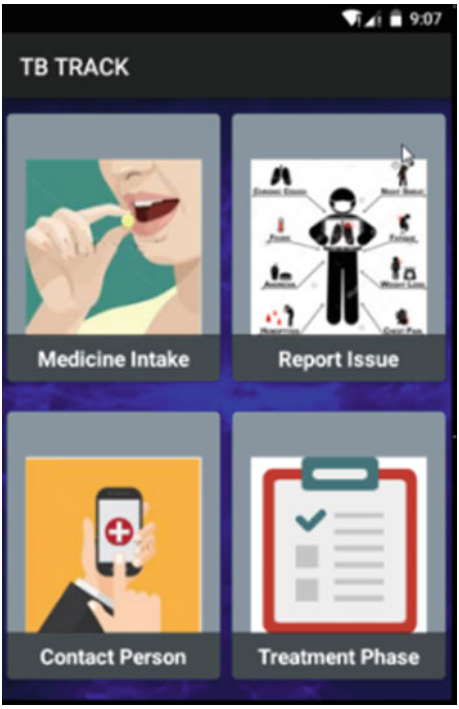


Fig. 8. Main menu of app

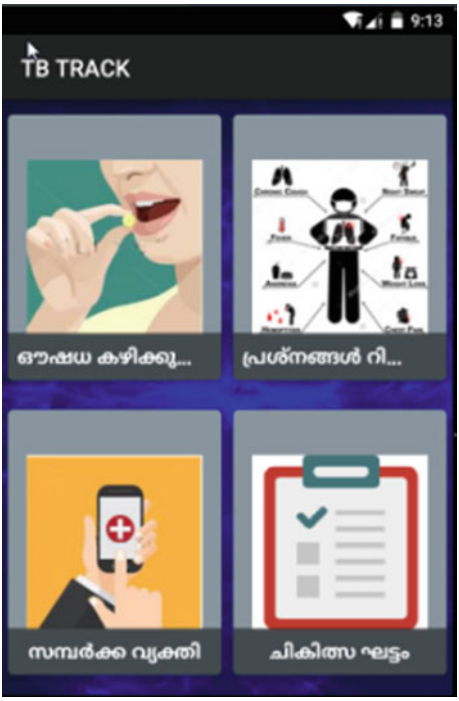


Fig. 9. Main menu of app in Malayalam language

medicines or alerting patient when medical test is due. The app provides feature for one-touch calling and messaging to connect to contact person or DOTS provider in case of emergency as shown in Fig. 10. It also provides details like course of treatment and schedule for next sputum test of the patient by fetching the values from database as shown in Fig. 11 (Figs. 8 and 9).

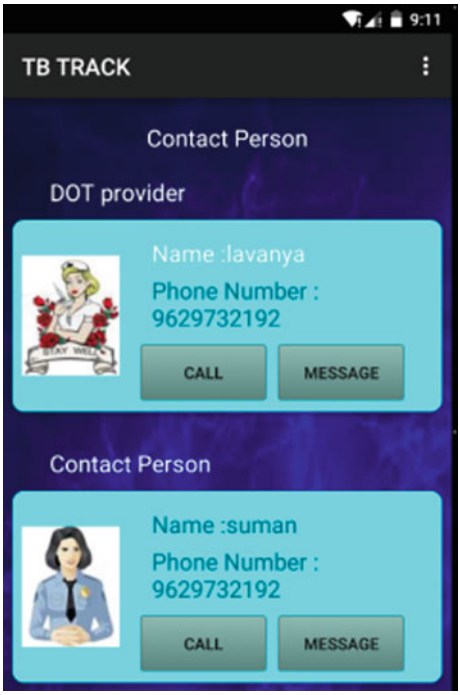


Fig. 10. One-touch emergency call screen

4.2 TB Track DOTS Provider Mobile App

The DOTS provider mobile app requires internet enabled which works on Quick Response (QR) code based technique as a mechanism for registering attendance remotely into the web portal for field personals interacting with respective recipient client. This is a foolproof mechanism where only registered smartphones are used as attendance registering device. The attendance protocol insists DOTS provider while the schedules to meet the patient needs to ask the patient's mobile app to invoke attendance module. It connects to web attendance service module which generates a 6-digit random number which is converted into QR code image and gets downloaded to the mobile app. The DOTS provider has to invoke his attendance module which has QR code scanner which needs to scan the QR code image downloaded in patients' mobile app. The scanned code is uploaded to the web attendance service module with secured code in the registered app. The web attendance service module compares the code

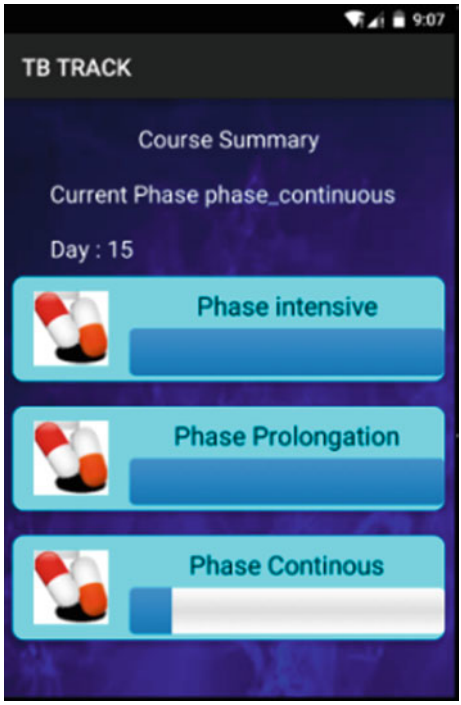


Fig. 11. Patient treatment stage summary

Fig. 12. Patient mobile app attendance module displaying QR code

Fig. 13. DOTS provider mobile app module displaying QR code scanner scanning QR code generated in patient mobile app

generated for patient mobile app and uploaded code from DOTS provider mobile app and registers timestamp against the patient entry in database. Provision to send SMS to TB unit officer on attendance registration can help to verify interaction between patient and DOTS provider. This brings accountability in DOTS provider. Figure 13 shows the screenshot of QR code scanner component used to scan the generated QR code in patient's TB track app Fig. 12 for registering attendance which will be recorded in TB track database with timestamp. We can also record time spent by DOT provider with patient with entry and exit times.

4.3 TB Track Web Application

The web application part of TB track can be used by medical staffs, doctors, and lab technician of TB units with respective access level. When a patient is identified with TB, a TB sheet is generated with the entire dataset being entered into the database through web interface. A unique ID and password will be generated for the registered patient, which will be given to the patient along with smartphone with TB track app

loaded. Similarly, a unique ID and password will be generated for the DOTS providers to enable them to log in into their app. The patient will be using the user ID and password every time to log in to the mobile app for daily reporting medicine intake and report symptoms if he had any during the day. The web application runs a scheduled job routine twice a day at stipulated time to identify that each registered patient has taken medicine properly or not. It escalates any issue of nonadherence of treatment by a patient through reporting to respective DOTS providers for the patient. The web application further escalates and reports to the treatment supervisor if DOTS provider has not responded with in time. The web application also sends separate educational and motivational messages for each patient. A set of messages will be sent in sequence on daily basis at different time intervals based on which phase of treatment course the patient is in. The web application also allows entering diagnostic test reports and treatment schedule which could be viewed by patient and helps in alerting patient through SMS for going to medical tests when it is due.

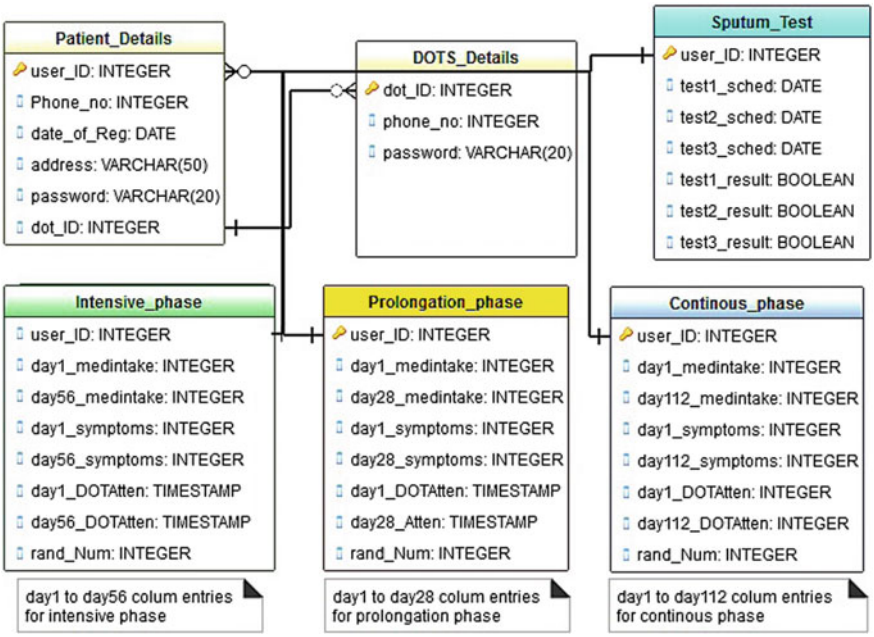


Fig. 14. TB track application database schema

4.4 TB Track Database

TB track application records all the activities like patient interacting with the mobile app to report medicine intake and symptoms, TB unit entering patient data such as their details, sputum test reports, etc. through web interface, and DOTS provider registering attendance through DOTS provider mobile app recorded in TB track database. Figure 14 shows the partly database schema used by TB track application. Patient

details while registering and the patient sputum test schedule and reports are entered through web interface in TB unit in "Patient_Details" table and "Sputum_Test" table, respectively. Patient's medicine intake details and symptoms details are recorded through their mobile app into the "Intensive_phase", "Prolongation + phase", or "Continous_phase" table depending on which phase they are in. Similarly, DOTS provider attendance is recorded as timestamp against the patient entry in the respective table. Consolidated report can be generated for patients registered in TB unit for getting details like the medicine intake during the phase of treatment, symptoms recorded by patients during the phase of treatment, and DOTS providers' attendance recorded during the patient interaction.

4.5 TB Track SMS Communication

The TB track application framework uses third-party bulk SMS service for sending periodical alerting, educational, and motivational messages to the TB patients. The SMS application is leveraged with the TB track web application for sending messages to patients, DOT providers, and medical treatment supervisors in their local native language whose phone numbers are stored in TB track database. Messages were sent twice a day to DOTS providers and medical officers on their respective patients' status of intake of medicine so that they can act upon if they have not submitted medicine intake status through app. Educational and motivational messages were sent to patients on day-to-day basis depending on the particular day of treatment. Computed alert messages will be sent to patients and DOTS provider to remind them if they had not taken medicine, remind the patients for taking regular checkup when it is due, and alert DOTS provider and medical supervisors if patient has reported symptoms through app.

5 Pilot Testing and Discussion

We selected 20 patients of the TB care unit mostly from rural background, and they were provided with smartphones loaded with TB track app and were trained to use it. The patient needs to log into the app to view medicine intake for the day in the form of virtual medicine strip if they had any for the day. The patient needs to press each tablet image and press submit button to report the TB track system that they had taken right medicine for the day. Patients were asked to report symptoms through app if they had experienced during the course of treatment. Mock tests were conducted by asking patients to report symptoms through app. TB patients were sent with motivational and educational messages through SMS besides all important reminders to remind them for taking periodical sputum test.

5.1 Outcome

Pilot testing showed promising results. 17 out of 20 patients (85%) found overall usability of app was good. They gave feedback that virtual medicine strip provided

support for correct medicine intake. Patients were asked to report requested correct symptoms from app. Out of 200 mock responses collected from patients for symptoms during the phase of treatment where patients were asked to submit mock symptoms through app, we were able to collect 81.5% of responses being reported correctly (5 cases during the mock test had real symptoms and they reported them correctly). DOT providers were asked to attend their respective patients for monitoring their patients' medicine intake as well as the time when they had reported symptoms. Attendance was registered through app brought accountability in them.

6 Conclusion

The proposed work presents a smartphone-based mHealth solution for supporting TB treatment. The educational background, age, and family background would still pose major challenge and limit the prospect of inculcating mHealth solution into TB treatment programme. It is also observed that success of smart app-based mHealth solution is largely based on good HCI design which is instrumental in acceptance of app by users. The results of pilot testing of project on usability showed virtual medicine strip and symbolic presentation to report symptoms, localization, and voice support that substantially aided the effort needed to train the patients for using the app. The TB track application infused responsibility and accountability to DOTS providers ensured better patient monitoring through the QR code based attendance system. With cheaper smartphones and availability of internet services due to improved information and communication technology (ICT), smartphone-based mHealth solution for TB programme is a viable strategy to be adopted along with the existing practice at least at selected areas where technology is available.

Acknowledgements. Pilot testing of the TB track app framework was done in SRM University Medical Hospital located in Kattankulathur, Chennai with the support of school of public health department at SRM University Medical College. The app was also tested with eight patients in Thrissur District TB Hospital, Kerala. We thank Mr. Tom Thomas Founder Chairman Olive Touch Healthcare Services Limited & Mutual Assistance Resource Group (MARG) P.O., Ayyanthole, Near State Bank of Travancore, Ayyanthole Thrissur, Kerala 680003—India for providing support.

References

1. TB Statistics for India. http://www.tbfacts.org/tb-statistics-india/. Accessed 21 Jan 2017
2. WHO Reference. Standards for TB care in India. www.tbcindia.nic.in/showfile.php?lid=3061. Accessed 21 Jan 2017
3. Wright CM, Westerkamp L (2015) Community-based directly observed therapy (DOT) versus clinic DOT for tuberculosis: a systematic review and meta-analysis of comparative effectiveness. NCBI Resources. https://doi.org/10.1186/s12879-015-0945-5
4. Article. Centre for Disease control and prevention. Drug-Resistant TB. U.S. Department of Health & Human Service. Accessed 21 Jan 2017

5. Treatment of Tuberculosis (2010) Guidelines: Supervision and patient support. NCBI resources, 4th edn. https://www.ncbi.nlm.nih.gov/books/NBK138737/
6. Denkinger CM, Grenier J (2013) Mobile health to improve tuberculosis care and control: a call worth making [Review article]. Int J Tuberc Lung Dis 17(1):719–727
7. Chaiyachati KH, Loveday M, Lorenz S (2013) A pilot study of an mHealth application for healthcare workers: poor uptake despite high reported acceptability at a rural South African community-based MDR-TB treatment program. PLoS ONE
8. Lee J-H (2016) Future of the smartphone for patients and healthcare providers. NCBI resources
9. Poole ES (2013) HCI and mobile health interventions. NCBI resources. https://doi.org/10.1007/s13142-013-0214-3
10. Shim K, Yim J (2016) Design and implementation of a mobile shopping app specializing in regional product. Int J Multimedia Ubiquit Comput 11(1):43–54
11. Gündüz F, Pathan AK (2012) Usability improvements for touch-screen mobile flight booking application: a case study. In: 2012 International conference on advanced computer science applications and technologies. ACM Digital Library, pp 49–54. https://doi.org/10.1109/acsat.2012.9
12. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. NCBI resources. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3080339. Accessed 21 Nov 2016
13. Jeon JH (2016) Development of mobile app for self-management performance of patients with CHB. Adv Sci Technol Lett Mech Eng 129:229–233
14. Nazir M, Iqbal I, Shakir H (2014) Title future of mobile human computer interaction research—a review. In: 2014 IEEE 17th International topic conference (INMIC). Electronic ISBN: 978-1-4799-5755-2. Print ISBN: 978-1-4799-5754-5
15. Narasimhan P, Bakshi A, Kittusami S (2014) A customized mHealth system for improving Tuberculosis treatment adherence and follow-up in south India. Health Technol 4(1):1–5

Modeling and Behavioral Simulation of PID-Controlled Phase-Locked Loop

Geetamoni Konwar(✉), Munmee Borah, and Tulshi Bezboruah

Department of Electronics & Communication Technology, Gauhati University, Guwahati, India
{geetak222, munmeeborah01}@gmail.com,
zbt_gu@yahoo.co.in

1 Introduction

The earliest research work towards the phase-locked loop (PLL) goes back to 1932. Because of the number of tuned stages in a superheterodyne receiver, a new type of receiver was developed, called the homodyne. Later on, it was renamed as synchrodyne receiver. In synchrodyne system, a local oscillator tuned the desired input frequency and multiplied it with the input signal. An automatic correction signal is applied to the oscillator, maintaining it in the same phase and frequency as desired. This type of feedback circuit initiated the evolution of the PLL [1, 2]. The basic PLL concepts were published by de Bellescize in 1932 [2–4]. The techniques were mainly used for synchronous reception of radio signals [3]. The PLL is widely used in communication systems. Some of its uses include frequency synthesizer, clock data recovery, FM demodulation, mobile phones, microprocessors, and satellite communications.

In certain applications, such as, in frequency hopping spread spectrum (FHSS) and wireless local area networks (WLANs) which require fast frequency switching methods, the settling time of the PLL is a very important parameter [5, 6]. In this paper, we propose a proportional–integral–derivative (PID)-controlled PLL model by placing a PID control block into the loop for decreasing the settling time.

1.1 Brief Overview of PLL

The PLL is a negative feedback control system which has the ability to align output phase and frequency to the phase and frequency of the reference signal at locked condition. It includes a phase frequency detector (PFD), a loop filter (LF), and a voltage-controlled oscillator (VCO) in the forward path and a frequency divider (FD) in the feedback path as shown in Fig. 1. The PFD is basically a comparator which compares the phase and frequency error between the reference signal and VCO signal. The PFD output is followed by LF to remove any unwanted high-frequency components and generate a dc signal for controlling the VCO. The VCO produces a high-frequency signal which is feedback to the input of the PFD. The FD in the feedback path is usually used to generate a low-noise, high-frequency digitally programmable signal from a low-frequency low-noise crystal oscillator [7]. In free running state of the PLL, there is no input voltage applied to the PFD. In capture range, VCO

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_6

begins to change and produce an output frequency. When output frequency and phase is adjusted to become equal to the frequency and phase of the input signal, then this state is called locked state of the PLL.

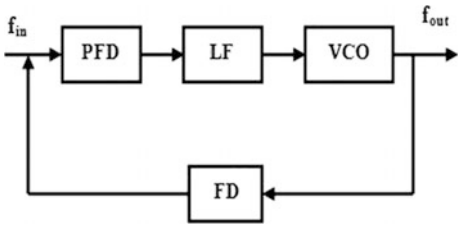


Fig. 1. Different blocks of PLL

2 Related Works

H. U. Uyanik and N. Tarrim in 2007 reported a novel aided-acquisition technique based on PID-controlled PLL where a significant reduction in the settling time is observed from 1.114 μs to 293 ns with an improvement of around 75% [5]. In 2009, V. Minambres, M. I. Milanes, B. Vinagre, and E. Romero designed and analyzed a proportional–integral (PI), a PID, and a novel fractional–proportional–integral (FPI) controller in continuous time domain to solve the phase tracking by improving time response, phase error, and overshoot. Phase margin is obtained to be 32.8°, 26.1°, and 42.9°, and settling time is obtained to be 0.22 s, 0.16 s, and 2.5 s for PI, PID, and FPI controller, respectively [8]. In 2011, P. Karuppanan and K. K. Mahapatra reported a PID controller with PLL synchronization controller based shunt active power filter for compensating harmonic currents and reactive power under both balanced and unbalanced conditions [9]. L. A. dos Santos, M. dos Santos Kaster, and S. A. O. da Silva in 2012 developed an adaptive nonlinear PID controller applied in a single-phase PLL algorithm which provides a transient response of 2.3 times faster than the conventional PI [10]. K. Bora and T. Bezboruah developed and implemented a PID-controlled higher order PLL model with low settling time in the year 2013. They obtained the settling time for second-, third-, and fourth-order PLL with LPF to be 1.49×10^{-5} s, 1.08×10^{-8} s, and 5.4×10^{-9} s, respectively, and for PID-controlled PLL, it is obtained to be 4.4×10^{-9} s [6]. In 2014, S. Golestan, M. Monfared, Francisco D. Freijedo, and Josep M. Guerrero developed a systematic and efficient approach to design the control parameters of the synchronous reference frame PLL (SRF-PLL) with pre-filtering stage [11]. In 2015, S. Golestan, Francisco D. Freijedo, and Josep M. Guerrero presented a systematic approach to design high-order PLLs control parameters for first-, second-, third-, and fourth-order LPF, respectively, and they observed that the reduced-order model provides good accuracy in all the cases [12].

3 Objective and Methodology

The main objective of the proposed work is to design and simulate a PID-controlled PLL. A PID controller is inserted in place of LPF in an attempt to reduce the settling time of the system. The simulation of the proposed model is performed in MATLAB platform to study (i) the stability, (ii) settling time, (iii) phase margin (PM), and (iv) bandwidth (BW) of the system.

The methodologies toward implementation of the proposed work are: (a) derivation of TF in s-domain for each block of the model, (b) derivation of the system transfer function by integrating individual TF of each block of the model, and (c) simulation of the model in MATLAB platform to study the various aspects of the system.

4 Theoretical Estimation for the PLL Components

The block diagram of the proposed model is given in Fig. 2.

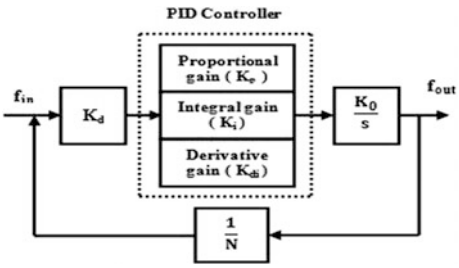


Fig. 2. Functional block diagram of the model

4.1 Estimation for PFD

For linear analysis of the model [2], it is assumed that the loop is locked and the PFD is linear [13]. Then, the PFD output voltage can be written as:

$$V_{PD} = K_d(\phi_{ref} - \phi_{out})$$

$$K_d = \frac{V_{PD}}{(\phi_{ref} - \phi_{out})}, \tag{1}$$

where ϕ_{ref} is the phase of the reference signal, ϕ_{out} is the phase of the VCO output signal, K_d is the PFD gain constant measured in volts/radian, and V_{PD} is the PFD output in volt.

4.2 Estimation for VCO

For linear analysis of the loop [14], the VCO output can be expressed as:

$$\phi_{out} = K_0 \int V_{tune} dt,$$

where K_0 is the VCO gain factor and V_{tune} is the VCO control voltage.

The TF of the VCO can be written as

$$\frac{\phi_{out}}{V_{tune}} = \frac{K_0}{s}, \tag{2}$$

where $K_0 = \frac{1}{CR}$; C and R are the capacitance and resistance used in VCO design.

4.3 Estimation for LF

We have used second- and third-order passive LPF in the loop as shown in Fig. 3a and b.

The TF of the second-order passive LF can be derived as:

$$F_{2nd}(s) = \frac{1 + sT_2}{sA_0(1 + sT_1)}, \tag{3}$$

where $A_0 = C_1 + C_2$, $T_1 = \frac{C_1 C_2 R_2}{C_1 + C_2}$, and $T_2 = C_2 R_2$.

Thus, putting these values in Eq. (3), the TF for second-order LF can be simplified as:

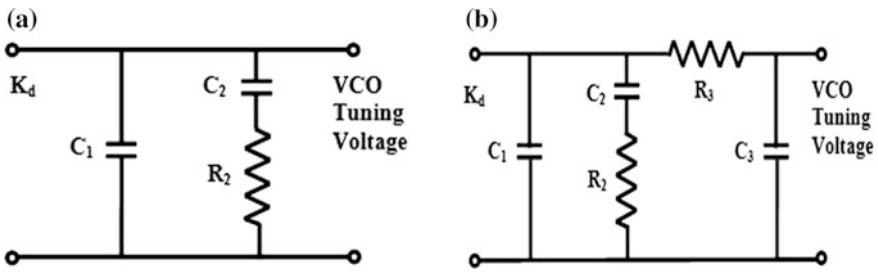


Fig. 3. **a** Second-order passive LPF. **b** Third-order passive LPF

$$F_{2nd}(s) = \frac{1 + sC_2R_2}{s^2C_1C_2R_2 + s(C_1 + C_2)} \tag{4}$$

The TF of the third-order passive LF can be derived as:

$$F_{3rd}(s) = \frac{1 + sT_2}{sA_0(1 + sT_1)(1 + sT_3)}, \tag{5}$$

where $A_0 = C_1 + C_2 + C_3$, $T_1 = \frac{C_1C_2R_2}{C_1 + C_2 + C_3}$, $T_2 = C_2R_2$, and $T_3 = C_3R_3$.

Putting these values in Eq. (5), the TF for third-order LF can be simplified as:

$$F_{3rd}(s) = \frac{1 + sC_2R_2}{s^3C_1C_2C_3R_2R_3 + s^2\{C_1C_2R_2 + C_3C_3R_2 + C_3R_3(C_1 + C_2)\} + s(C_1 + C_2 + C_3)} \tag{6}$$

4.4 Estimation for FD

An FD takes an input signal of a frequency f_{in} and produces an output signal of frequency f_{out}, where $f_{out} = f_{in}/n$; and n is an integer.

The TF for the FD can be derived as:

$$F_{div} = \frac{1}{N}, \tag{7}$$

where N is the division ratio.

4.5 Estimation for PID Controller

The TF of PID controller can be expressed as

$$C(s) = K_p + \frac{K_i}{s} + sK_{di} \tag{8}$$

where K_P is the proportional gain, K_i is the integral gain, and K_{di} is the derivative gain.

4.6 The System TF of the Model

The system TF of the proposed model can be derived as:

$$H(s) = \frac{Forward\ gain}{1 + Loop\ gain} \tag{9}$$

4.6.1 System TF with LF

The forward gain of the model can be expressed as:

$$Forward\ Gain = K_dF(s)\frac{K_0}{s} \tag{10}$$

The loop gain is the product of the individual TF of each block used in the model and it can be derived as:

$$Loop\ Gain = \frac{K_d K_0 F(s)}{Ns} \tag{11}$$

Combining Eqs. (4), (9), (10), and (11), the system TF with second-order LF can be derived as:

$$H_{2nd}(s) = \frac{K_d K_0 (1 + sC_2R_2)}{s^3 C_1 C_2 R_2 + s^2 (C_1 + C_2) + sC_2 R_2 K + K}, \tag{12}$$

where $K = \frac{K_d K_0}{N}$ = loop gain constant [15].

Combining Eqs. (6), (9), (10), and (11), the system TF with third-order LF can be derived as:

$$H_{3rd}(s) = \frac{K_d K_0 (1 + sC_2R_2)}{s^4 C_1 C_2 C_3 R_2 R_3 + s^3 \{C_1 C_2 R_2 + C_3 C_3 R_3 + C_3 R_3 (C_1 + C_2)\} + s^2 (C_1 + C_2 + C_3) + sC_2 R_2 K + K} \tag{13}$$

4.6.2 System TF with PID Controller

Combining Eqs. (8), (9), (10), and (11), the system TF with PID controller can be derived as:

$$H_{pid}(s) = \frac{K_d K_0 \left(s^2 K_{di} + sK_p + K_i\right)}{s^2 \left(\frac{K_d K_0 k_{di}}{N} + 1\right) + s\left(\frac{K_d K_0 K_{di}}{N}\right) + \frac{K_d K_0 K_{di}}{N}} \tag{14}$$

5 Simulations

We simulate the system TF of the model given in Eqs. (12), (13), and (14) in s-domain by using MATLAB to study various aspects of the system. We analyzed the behavior and performance of the model by considering different parameters, namely: (i) K_d, (ii) K_o, (iii) N, (iv) second- and third-order LF components R_2, R_3, C_1, C_2, and C_3, and (v) PID controller gain K_p, K_i, and K_{di}. Various test cases for experiments with second- and third-order LF and also with PID controller are given in Table 1 and Table 2, respectively. The value of capacitor C_2 for second-order LF is chosen at least three times of that of the VCO input capacitance. The capacitor C_3 for third-order LF is chosen at least four times of that of the VCO input capacitance and R_3 is chosen as small as possible.

5.1 Simulation for Settling Time

The step response is used to determine the settling time of the system and provides information about stability. For PLL, settling time is the time for the system output response to reach and stay within 2–5% of its steady-state value. The settling time is set

by the LF components, K_d, K_o, and N. It is directly related to loop BW. The LF BW may be widened to speed up settling time and is also used to optimize PM [15, 16]. We simulate the system TF of the model with second- and third-order LF and with PID controller to evaluate the settling time. Four sample step responses of simulations are given in Fig. 4a, b (Sl No. 2, 4, 12, 15 as highlighted in Table 1) and Fig. 4c (Sl No. 4, 12, 14, 16 as highlighted in Table 2).

Table 1. Second- and third-order LF parameters of the model

Sl No.	Second-order LF parameters			Third-order LF parameters					K_d (V/rad)	K_0 (MHz/V)	N
	C1 (pF)	C2 (pF)	R2 (Ω)	C1 (pF)	C2 (nF)	C3 (pF)	R2 (Ω)	R3 (Ω)			
1	1.000	0.6	0.100	0.100	2.2	0.6	1.00	1.0	0.7	1.00	1.0
2	**0.100**	**0.6**	**0.200**	**0.100**	**1.2**	**1.2**	**1.50**	**1.0**			
3	0.100	2.2	0.200	0.010	2.2	1.5	1.00	1.0			
4	**0.100**	**0.1**	**0.500**	**0.002**	**4.7**	**1.5**	**1.20**	**1.0**			
5	0.500	1.0	0.100	0.001	1.0	2.2	1.20	1.0			
6	0.400	1.2	0.100	0.010	1.0	2.2	1.00	1.0			
7	1.500	1.5	0.100	0.050	6.0	1.2	1.50	1.0			
8	4.700	2.2	0.047	0.010	3.3	1.0	1.00	1.0			
9	3.300	1.5	0.047	0.002	1.0	1.5	1.00	1.0			
10	2.200	1.0	0.100	0.050	2.2	2.2	0.47	1.0			
11	6.000	3.3	0.050	0.100	1.2	1.2	1.00	1.0			
12	**0.600**	**1.0**	**0.150**	**0.002**	**9.0**	**0.9**	**1.00**	**1.0**			
13	1.200	1.5	0.075	1.000	4.7	1.2	1.00	1.0			
14	0.001	4.4	4.700	1.000	6.0	2.2	0.68	1.0			
15	**0.005**	**1.5**	**1.200**	**1.000**	**1.0**	**0.6**	**0.68**	**1.0**			
16	0.001	4.7	0.075	1.000	2.2	0.1	1.00	1.0			
17	0.010	3.3	0.075	1.200	1.0	2.2	0.68	1.0			
18	0.010	1.5	0.050	0.900	1.2	4.7	0.68	1.0			
19	1.200	2.2	0.100	0.900	6.0	1.2	1.00	1.0			
20	1.000	1.2	0.047	0.200	1.2	0.9	1.00	1.0			

Table 2. PID-controlled PLL parameters of the model

Sl No.	K_d (V/rad)	K_o (kHz/V)	N	K_p	K_i	K_{di}
1	0.1	1.00	1.00	0.1×10^{9}	0.1×10^{-9}	0.1×10^{-9}
2	0.1	1.00	2.00	1.5×10^{9}	1.5×10^{-9}	1.5×10^{-9}
3	0.5	1.00	1.00	0.2×10^{9}	1.5×10^{-9}	1.5×10^{-9}
4	**0.7**	**1.00**	**1.50**	$\mathbf{0.3 \times 10^{9}}$	$\mathbf{0.1 \times 10^{-9}}$	$\mathbf{0.5 \times 10^{-9}}$
5	1.0	1.00	5.00	1.0×10^{9}	0.5×10^{-9}	0.5×10^{-9}
6	0.1	0.10	2.00	2.5×10^{9}	0.1×10^{-9}	0.6×10^{-9}
7	0.1	0.50	4.00	0.1×10^{9}	0.1×10^{-10}	0.9×10^{-9}

(*continued*)

Table 2. (*continued*)

Sl No.	K_d (V/rad)	K_o (kHz/V)	N	K_p	K_i	K_{di}
8	1.0	0.50	1.00	0.6×10^{9}	2.2×10^{-9}	1×10^{-10}
9	0.1	1.00	10.00	3.3×10^{9}	2.2×10^{-9}	1.5×10^{-9}
10	0.1	1.00	5.00	6.6×10^{9}	1.5×10^{-10}	1×10^{-10}
11	0.9	0.10	2.00	1.5×10^{10}	3.3×10^{-10}	2.5×10^{-10}
12	**0.7**	**1.00**	**5.00**	$\mathbf{1.5 \times 10^{9}}$	$\mathbf{6.6 \times 10^{-10}}$	$\mathbf{3.3 \times 10^{-10}}$
13	0.7	0.40	5.00	2.2×10^{9}	0.12×10^{-9}	1.2×10^{-9}
14	**0.1**	**1.00**	**1.00**	$\mathbf{5.0 \times 10^{9}}$	$\mathbf{0.09 \times 10^{-9}}$	$\mathbf{1.0 \times 10^{-9}}$
15	0.2	2.50	1.00	1.2×10^{9}	0.7×10^{-9}	0.8×10^{-9}
16	**0.1**	**1.00**	**4.00**	$\mathbf{1.1 \times 10^{9}}$	$\mathbf{1.5 \times 10^{-10}}$	$\mathbf{1.5 \times 10^{-10}}$
17	0.2	10.00	9.0	0.5×10^{9}	1.5×10^{-10}	1×10^{-10}
18	0.1	1.00	5.00	6.0×10^{9}	0.09×10^{-9}	9×10^{-9}
19	0.1	0.20	1.00	1.2×10^{10}	0.1×10^{-10}	1×10^{-10}
20	0.1	1.25	1.00	1.5×10^{9}	0.6×10^{-9}	0.8×10^{-9}

5.2 Simulation for PM and System BW

Bode plot is a useful tool that directly gives the relative stability in terms of PM and gain margin. The PM is equal to 180° plus the phase shift of the TF in an open-loop condition [2, 16]. System stability is related to PM and a system is considered to be stable if the PM value is positive [13, 15]. In PLL system, the loop BW is the frequency at which the magnitude of the open-loop TF is equal to 1. The choice of loop BW typically involves a trade-off between spur level and lock time [15]. We simulate the TF of the model for PM and loop BW by using Bode function. Few sample bode responses are shown in Fig. 5a, b (Sl No. 2, 4, 12, 15 as highlighted in Table 1) and Fig. 5c (Sl No. 4, 12, 14, 16 as highlighted in Table 2).

6 Results and Discussion

The results obtained from the simulated responses for various test cases of the model with second- and third-order LF and PID-controlled PLL for different response parameters are given in Table 3. It shows that the settling time for PLL with second- and third-order LF is in the range from 0.103 ns to 31.8 ns and 0.109 ns to 7.68 ns, respectively, whereas the PLL with the PID controller in the loop shows a settling time ranging from 0.062 ps to 31.3 ps. However, the minimum PM for PLL with second- and third-order LF is 40.9° and 42.6°, where as the maximum comes out to be 76.9° and 71.7°, respectively. For PID-controlled PLL, the minimum PM is observed to be 90.1° and the maximum to be 90.6°. With the change in the settling time, the three systems show a different BW which varies within the range of 119.87 MHz to 6141.4 MHz for the PLL with second-order LF. The PLL with third-order LF and the PLL with PID controller shows a variation from 56.441 MHz to 192.75 MHz and 19.846 GHz to 9923.62 GHz, respectively.

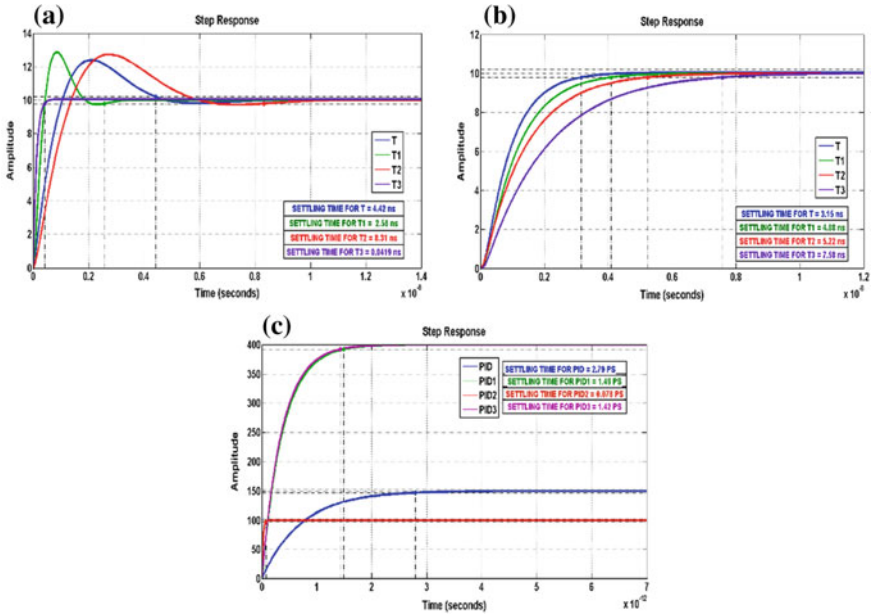


Fig. 4. **a** Step responses of the simulated second-order LF model. **b** Step responses of the simulated third-order LF model. **c** Step responses of the simulated PID controller model

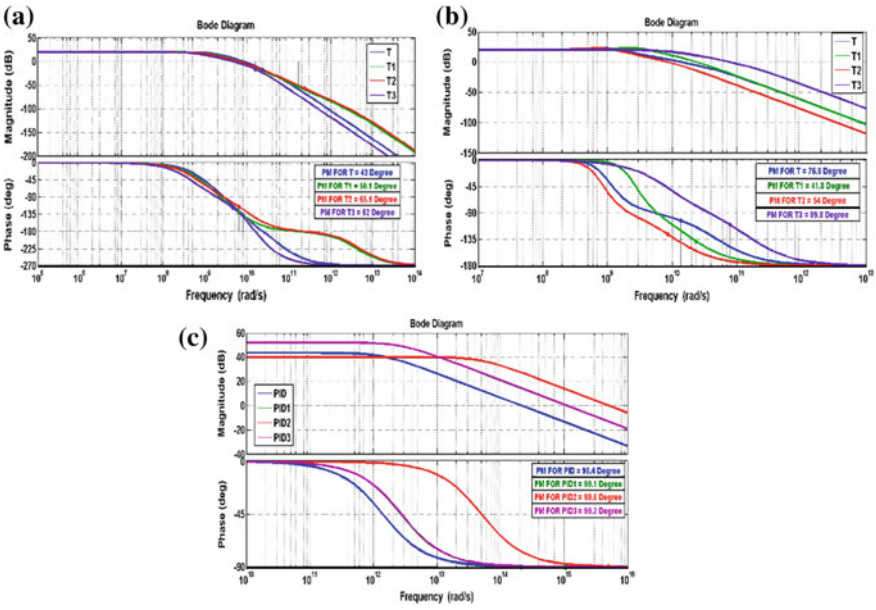


Fig. 5. **a** Bode responses of the simulated second-order LF model. **b** Bode responses of the simulated third-order LF model. **c** Bode responses of the simulated PID controller model

Table 3. Simulation results for second- and third-order LF and PID-controlled PLL model

Sl No.	PM (Degree)			Settling time			BW		
	Second-order LF	Third-order LF	PID-controlled model	Second-order LF (ns)	Third-order LF (ns)	PID-controlled model (ps)	Second-order LF (MHz)	Third-order LF (MHz)	PID-controlled model (GHz)
1	51.6	69.3	90.6	13.3	5.05	3.910	272.71	117.31	158.77
2	**76.9**	0.43	90.3	4.49	3.15	5.220	345.30	**192.75**	119.08
3	71.2	53.8	90.6	5.86	4.82	0.391	485.54	123.94	1587.7
4	41.8	62.6	90.4	2.58	5.06	2.790	887.51	119.68	222.29
5	0.69	**42.6**	90.1	12.2	3.64	0.196	231.14	159.10	3175.6
6	73.6	45.7	90.3	0.10	4.77	3.130	216.41	130.53	198.46
7	50.5	46.4	90.1	11.2	3.27	**31.30**	200.88	191.84	**19.846**
8	0.48	50.1	90.6	**31.8**	4.08	1.560	137.32	151.93	4763.34
9	47.8	0.54	90.1	31.6	5.04	0.130	161.68	123.97	52.39
10	**40.9**	65.1	**90.1**	15.9	5.22	2.960	220.71	118.66	209.59
11	46.9	55.2	90.3	30.1	5.12	5.800	**119.87**	122.25	107.17
12	0.54	61.8	90.1	8.31	**0.109**	1.860	269.71	**56.441**	333.42
13	62.5	0.62	90.1	15.7	7.58	3.180	183.85	80.742	195.61
14	45.9	**71.7**	**90.6**	**0.103**	7.54	0.078	**6141.4**	79.348	7938.80
15	69.8	58.3	90.6	0.419	5.08	0.652	1413.0	120.92	952.66
16	0.70	59.1	90.2	13.7	4.92	1.420	311.43	121.07	436.64
17	62.6	68.7	90.6	13.6	7.65	0.352	367.77	79.742	1764.23
18	0.76	66.5	90.2	0.20	**7.68**	0.326	176.66	79.897	1905.40
19	56.4	60.5	90.6	12.4	5.14	0.163	180.78	119.93	3810.64
20	59.3	57.5	90.1	26.5	5.18	**0.062**	188.78	120.47	**9923.62**

7 Conclusion

From the analysis of the proposed PID-controlled PLL model and PLL with second- and third-order LF, we can conclude that the settling time of PID-controlled PLL is too much faster (0.062 ps) than that of the PLL with LF (0.103 ns for second-order LF and 0.109 ns for third-order LF). Bode responses for PM of the model show that the PM values for both the cases of PLL with LF and PID-controlled PLL are well within the limit of stability. Also, it is observed that the PID-controlled PLL system is highly stable. So, the model may be applied as a trade-off for research and industrial applications for acquiring better settling time in communication system.

Acknowledgements. The authors are thankful to the Head, Department of Electronics & Communication Technology (ECT), Gauhati University, for providing necessary infrastructural facility toward the present work.

References

1. Bellescize H (1932) La reception synchrone. Onde Electrique 11:230–240
2. Kalita K, Maitra A, Bezboruah T (2014) Modeling and behavioral simulation of a wide band phase-locked loop for wireless communication receivers. Int J Control Theory Appl 7(1): 27–39
3. Lai MF, Nakano M (1996) Special section on phase-locked loop techniques, guest editorial. IEEE Trans Ind Electron 43(6):607–608
4. Prasad V, Sharma C (2012) A review of phase locked loop. Int J Emerg Technol Adv Eng 2 (6):98–104
5. Uyanik HU, Tarim N (2007) PID-controlled PLL for fast frequency-hopped systems. In: 6th IEEE Dallas circuits and systems workshop on system on-chip, pp 1–3
6. Bora K, Bezboruah T (2013) Modeling and software implementation of PID controlled higher order PLL. In: Proceedings of the world congress on engineering, vol 2
7. Hajimiri A (2001) Noise in phase–locked loops (Invited). In: Southwest symposium on mixed-signal design, pp 1–6
8. Minambres V, Milanes MI, Vinagre B, Romero E (2009) Comparison of controllers for a three-phase phase locked loop system under distorted conditions. In: Compatibility and power electronics, pp 79–85
9. Karuppanan P, Mahapatra KK (2011) PLL synchronization with PID controller based shunt active power line conditioners. Int J Comput Electr Eng 3(1):42–47
10. dos Santos LA, dos Kaster M, da Silva SAO (2012) Applying a nonlinear PID in a single phase PLL control. In: IEEE International conference on power electronics, drives and energy systems, pp 1–4
11. Golestan S, Monfared M, Francisco DF, Guerrero MJ (2014) Performance improvement of a prefiltered synchronous-reference-frame PLL by using a PID-type loop filter. IEEE Trans Ind Electron 61(7):3469–3479
12. Golestan S, Francisco DF, Guerrero MP (2015) A systematic approach to design high-order phase-locked loops. IEEE Trans Power Electron 30(6):2885–2890
13. Gardner FM (1979) Phase lock techniques, 2nd edn. Wiley, NY

14. Goldman S (2007) Phase locked loop engineering handbook for integrated circuits. Artech House
15. Banerjee D (2006) PLL performance simulation and design, 4th edn. Dog Ear Publisher
16. Kalita K, Handique J, Bezboruah T (2012) Modeling and Behavioral simulation of a high-speed phase-locked loop for frequency synthesis, published in IET. Signal Process 6(3):195–204

Deviant Calcium Channels: Role in Ventricular Arrhythmias—A Computational Study on Human Ventricular Tissue

G. Gulothungan(✉) and R. Malathi

Department of EIE, Annamalai University, Annamalai Nagar, India
{g.gulothungan, vsmalu}@gmail.com

1 Introduction

Amendment of the electric properties of single myocyte may favor the occurrence of ventricular arrhythmias in these patients by exerting influence on early or delayed after depolarizations. Mathematical models of the cellular action potential and its underlying ionic currents could help to elucidate possible arrhythmogenic mechanism on a cellular level [1]. Cardiac arrhythmias mechanisms are well understood by means of theoretical studies to improve and form a well-established area of research. Computational approach is one of the most important applications to studies of human cardiac myocytes, because the chance for doing fact-finding and clinical studies in human cardiac myocytes is very limited. Then, animal cardiac myocytes used for fact-finding studies may vary notably from human cardiac myocytes [heart size, heart rate, AP shape, duration, restitution, vulnerability to arrhythmias, etc.]. At last, major cardiac arrhythmias are taken place in ventricular region, which are three-dimensional phenomena, whereas fact-finding observations are still mostly driven to exterior recordings. Computational approach to simulations of arrhythmias in the human cardiac myocytes can get the better of some of these problems.

To perform simulation studies of different arrhythmias in human ventricles, we need a mathematical model that on the one hand replicates biophysically full properties of single human ventricular cells, such as the major ionic currents, including the roles of concentration changes, ion channels, calcium transients, AP duration (APD) restitution (APDR), and important properties of wave propagation in human ventricular tissue, such as conduction velocity (CV) restitution (CVR). On the other hand, it should be computationally efficient enough to be applied in the large-scale spatial simulations needed to study different types of arrhythmias. A mathematical model of the AP of human ventricular cells that include a high level of electrophysiological detail is developed by ten Tusscher et al. in 2004 [2].

The major ionic currents in ten Tusscher model are based on the current investigational data and it is most computationally cost-effective sufficient to be applied in large-scale spatial simulations for the study of reentrant arrhythmias. This model also predicts current experimentally observed data on APDR, which is a key property for reentrant arrhythmias. The CVR of this model is broader than in other human cardiac models and agrees better with obtainable data. Model has lack of calcium transient

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_7

process information. Ten Tusscher models [2] are updated to include a fuller description of calcium dynamics by Tusscher and Panfilov et al. in 2006 [3]. The models mainly focused on dynamical instabilities in APD, conditions for electrical instability in single cells, for reentrant waves in a ring of cells, and for reentry in two-dimensional (2D) sheets of ventricular tissue. This model has fast sodium current recovery dynamics, which shows a key factor for the onset of instability. Both APDR and CVR are very key properties for the event and balance of cardiac arrhythmias [4–7]. APDR was measured in cardiac myocytes for three reasons: first, to match the fact-finding data, which was obtained in cardiac myocytes; then, to avoid sensitive dependence on the stimulus [4, 8, 9]; and last, to obtain the minimum diastolic interval (DI) for propagation. The restitution curves of all models were calculated in tissue using 1D cable 2 cm in length [4].

This aim is to develop a novel computational 1D model and also to predict the same restitution property in that model. To avoid the more complex cable equation model, in this work, an initiative taken to formulate a novel computational theoretical 1D and 2D multicellular fiber model for human ventricular cells. It is also methodical for wide-reaching spatial simulations of cardiac arrhythmias. The model includes a simple calcium dynamics that reproduces realistic calcium transients and a positive human contraction staircase and allows us to realistically model calcium-dominated I_{Cal} inactivation, while at the same time maintaining a low computational load. In our study of ventricular arrhythmias due to deviant calcium channel, we take ten Tusscher model [2] for my single cell and novel one-dimensional (1D) and 2D computational multicellular fiber model. In the cardiac myocytes, potential-dependent ion channels, such as sodium, calcium, and potassium, and respective ions are significant keys in giving rise to and forming the AP as well as exchangers and pumps that donate ion fluxes [10]. Na^+–Ca^{2+} exchanger in plasma membrane is a high priority component of messaging pathways in several tissues.

In that cardiac myocytes, Na^+–Ca^{2+} exchanger especially has high activity and also is an important regulator of contractility. Ca^{2+} is major intracellular agent, and nature has developed numerous mechanisms to balance Ca^{2+} levels. Some of these mechanisms are Na^+–Ca^{2+} exchanger, sarcoplasmic calcium pump, leakage channels, and voltage-dependent calcium channels. In a survey of 100 recent Na^+–Ca^{2+} exchange keyword listing in medline, 65 of the articles involved the use of myocardial tissue. Na^+–Ca^{2+} exchanger in plasma membrane is high priority component of messaging pathways in cardiomyocytes and plays an important role in contractile events [11]. The Ca^{2+} ions inducing contraction in cardiac muscle arise from two sources, namely voltage-dependent calcium channels and SR calcium channels. This Ca^{2+} binds to and activates Ca^{2+} release channels (ryanodine receptors) of the SR through a Ca^{2+}-induced Ca^{2+} release (CICR) process. Ca^{2+} from both sources can then induce contraction. In a survey, Ca^{2+} ions estimates indicate that 20% of the Ca^{2+} persuading contraction comes from extracellular fluid sources and 80% of the Ca^{2+} is released from the SR [12].

The ability of the Na^+–Ca^{2+} exchanger to rapidly remove Ca^{2+} from cardiac myocytes is readily demonstrated. Bridge et al. [13] showed that the SR function is disabled by the presence of caffeine and ventricular myocytes that would not relax in the absence of external Na^+. By raising the Na^+ during AP process, ventricular relaxation rapidly happened. If 80% of coupling Ca^{2+} originates from the SR, then that

same amount of Ca^{2+} must be returned to the SR. Thus, the SR Ca^{2+} pump competes very effectively with Na^{+}–Ca^{2+} exchange [14]. Cardiac contractility in ventricular tissue is prominent by calcium channels like voltage-dependent calcium channels, Na^{+}–Ca^{2+} exchanger, SR calcium pump, and leakage channels. Our aim of this article is to study the impact on AP generation and propagation in single ventricular myocyte, 1D ventricular myocyte, and 2D ventricular myocyte under different deviant calcium channel conditions.

2 Methods and Materials

2.1 General

In this paper, ten Tusscher model of human ventricular cell is used for simulation analysis [2]. The electrophysiological behavior of a single cell can, hence, be described with the following differential equation.

$$C_m(dV/dt) = -(I_{ion} + I_{stim}), \tag{1}$$

where

C_m Membrane capacitance in (μF)
V Membrane potential in (mV)
t Time in (ms)
I_{ion} is the sum of the transmembrane ionic currents (pA/pF)
I_{stim} is an externally applied stimulus current in (pA/pF)
I_{ion} is given by summation of the number of ionic currents.

$$\begin{aligned} I_{ion} = & I_{Kr} + I_{Ks} + I_{K1} + I_{to} + I_{Na} + I_{b,Na} \\ & + I_{Ca,L} + I_{b,Ca} + I_{NaK} + I_{NaCa} + I_{p,Ca} + I_{p,K} \end{aligned} \tag{2}$$

where L-type Ca^{2+} ($I_{Ca,l}$), transient outward K^{+} (I_{to}), rapid and slow delayed rectifier K^{+} (I_{kr} & I_{ks}), time-independent potassium current (I_{k1}), fast Na^{+} (I_{Na}), Na^{+}–Ca^{2+} exchanger (I_{NaCa}), Na^{+}–K^{+} pump (I_{NaK}), background Na^{+} ($I_{b,Na}$), sarcolemmal calcium pump ($I_{p,Ca}$), background Ca^{2+} ($I_{b,Ca}$), and potassium pump ($I_{p,K}$).

2.2 2D Array of Ventricular Cell with Gap Junction

A novel computational model is developed for 2D array of ventricular cells with gap junction 2.5 nS based on Henriquez [15, 16]. In that, each single-cell ventricular dynamics is considered as a single potential node. Using nodal voltage principle, the coupling current between two ventricular cells in 2D model is given by

$$I_{cup(2,2)} = (C_{22} - C_{21}) * RG_{21} + (C_{22} - C_{12}) * CG_{12}, \tag{3}$$

where

$I_{cup\ (2,\ 2)}$	Coupling current between cell1 and cell2
V1	Membrane potential at cell1
V2	Membrane potential at cell2
G1	Conductance between cell1 and cell2.

The gap junctions were distributed uniformly along the length of the fiber, and the gap junction was represented as a linear, constant conductance of 2.5 nS [17]. All 2D models are developed by cable theory (cell-to-cell connection by space step) and solved using partial differential equations. But here cell-to-cell connection by gap junction and 2D model is developed by ohms law principle and solved by general voltage nodal algebraic equation.

2.3 Preparation of Failure Calcium Channels

We modified the formulation of I_{NaCa}, $I_{Ca,L}$, I_{NaK}, I_{up}, and I_{leak} proposed by ten Tusscher et al. [2] for human ventricular cells. The scaling factor of I_{NaCa} is modified to reproduce enhance activity of Na^{+}–Ca^{2+} exchanger (i.e., the maximal I_{NaCa}, (K_{NaCa} = 1650 (pA/pF)) 65% increase from its normal level in endocardial cell). The maximum current of I_{NaK} and $I_{Ca,L}$ is modified to reproduce hypertrophy condition in endocardial cell (i.e., (P_{Nak} = 0.78 (pA/pF)) 45% reduce from its normal level and ($I_{Ca,L}$ = ($I_{Ca,L}$/2) (pA/pF)). In same way, SR calcium pump I_{up} (V_{maxup} = 0.0015(mM/ms) and SR calcium leak channel I_{leak} (V_{leak} = 0.00055 (mM/ms) activation time are also modified to reproduce deviant CICR process. All the data are taken from human myocytes by Pribe and Beukelmann et al. [1]. 2D model simulation coded in MATLAB and membrane potential and gating differential equations are solved using Euler integration method (Fig. 1).

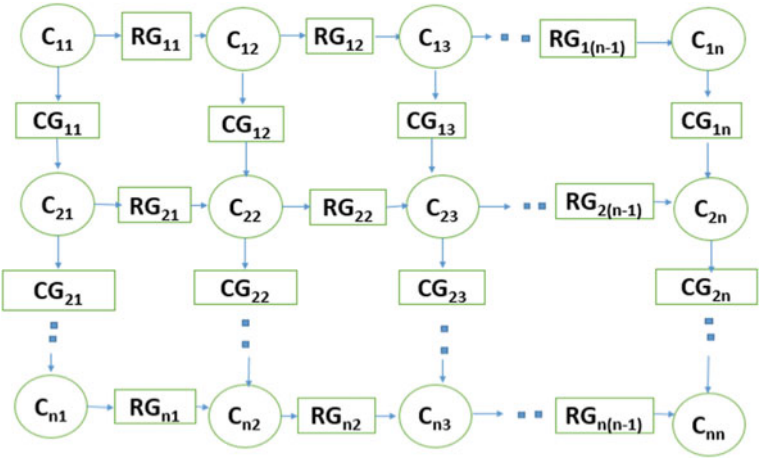


Fig. 1. Structure of 2D array of ventricular cell

3 Results and Discussions

Major cardiac arrhythmias are taken place in ventricular region, which are three-dimensional phenomena, whereas fact-finding observations still mostly drive to exterior recordings. Computer 2D simulations of arrhythmias in the human ventricular cell must be predicted the AP propagation and generation effects much more detailed information to help possible arrhythmogenic mechanism on a cellular level. In this article, 2D simulation study made on how deviant calcium channels ventricular portion AP affects the normal ventricular portion AP. The 2D simulation was performed on a 100 × 100 square lattice of nonfailing endocardial ventricular cell with gap junction conductance (gcon) = 2.5 nS. The results of these computations are shown in Fig. 2a–i (red region is the depolarised region and blue region is the resting region). Figure 2a and b shows membrane potential (V) at time (t) = 1 ms in 2D square simulation domain of side length (L) = 10 mm and L = 0.6 mm. At time 1 ms, total propagation of ventricular sheet covers the region of 0.6 mm. In 0.6 mm region, 28% is fully depolarized, 16% is half depolarized, 20% reaches above threshold potential, 24% starts its depolarization, and finally 12% is at resting state. At time 10 ms, total propagation of ventricular sheet covers the region of 4.6 mm. Similarly, at time 20 ms, total propagation of ventricular sheet covers the region of 9.1 mm. Figure 2c and d shows V at time 27 ms in 2D square simulation domain of side length 10 mm. In 10 mm region, 99.66% is fully depolarized, 0.13% is half depolarized, and finally 0.21% is at resting state. Figure 2e and f shows V at time 250 ms and t = 290 ms in 2D square simulation domain of side length 10 mm. In 10 mm region at t = 250 ms, the total ventricular sheet is in slow repolarization state and also at t = 290 ms, and the total ventricular sheet is in resting potential state. Plateau potential region of normal ventricular sheet is shown in Fig. 2g. In 5.1–10 mm region at t = 25–60 ms, the total ventricular sheet is in plateau potential state. The 2D simulation was performed on a 100 × 100 square lattice of nonfailing endocardial ventricular cell with gcon = 2.5 nS. The results of these resting potential computations are shown in Fig. 2h and i (red region is the deviant resting potential state and blue region is the normal resting potential state). Figure 2h shows V at time 290 ms in 2D square simulation domain of side length 10 mm. At time 290 ms, total ventricular sheet is in resting potential state in the range between −83 and −86 mV. Similarly, Fig. 2i shows V at time 295 ms in 2D square simulation domain of side length 10 mm. At time 295 ms, total ventricular sheet is in completely resting potential state (−86.2 mV). This 2D representation of electrical activities of human ventricular tissue is an important phenomenon to study and analyze the cardiac arrhythmias. In this article, we also study the characteristic of AP propagation effects due to deviant calcium channel AP.

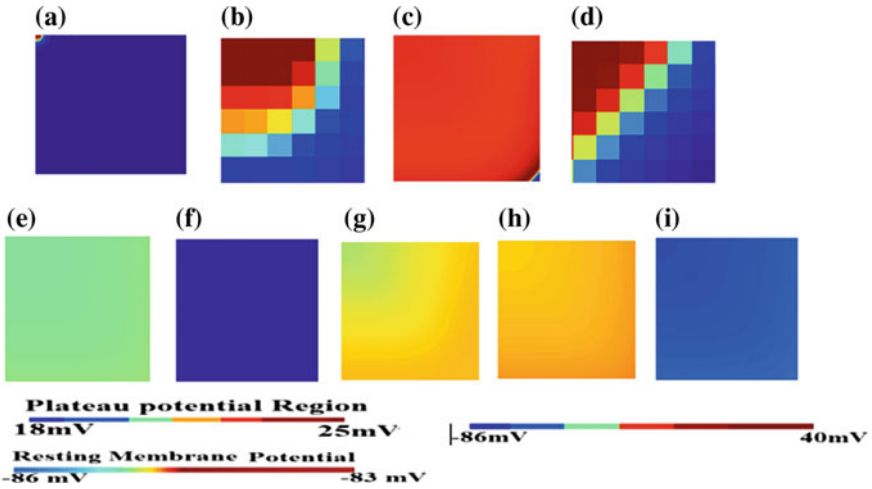


Fig. 2. AP generation and propagation in 2D simulations performed on a 100 × 100 square lattice of normal endocardial human ventricular cell with gap junction (gcon = 2.5 nS). **a** Plane waves shown via pseudocolor plots of membrane potential (V) at time (t) = 1 ms in 2D square simulation domain of side length (L) = 10 mm. **b** V at t = 1 ms of L = 0.6 mm. **c** V at t = 27 ms of L = 10 mm. **d** V at t = 27 ms of L = 9.4–10 mm. **e** V at t = 25–60 ms of L = 10 mm (Plateau potential region). **f** V at t = 250 ms of L = 10 mm. **g** V at t = 290 ms of L = 10 mm. **h** V at t = 290 ms of L = 10 mm (Resting membrane potential region). **i** V at t = 295 ms of L = 10 mm (Resting membrane potential region)

The 2D simulation was performed on a 100 × 100 square lattice of nonfailing endocardial ventricular cell with gcon = 0.5 nS. The results of these computations are shown in Fig. 3a–d (red region is the depolarised region and blue region is the resting region). Figure. 3a and b shows V at time 1 ms in 2D square simulation domain of side length 10 mm and L = 0.6 mm. In 0.6 mm region, 1% is fully depolarized, 8% is half depolarized, 12% starts its depolarization, and finally 79% is at resting state. Compared with Figs. 2b and 3b, 80% of ventricular sheet is at resting state because of gcon = 0.5 nS. At time 10 ms, total propagation of ventricular sheet covers the region of 1.6 mm, which is 63% less propagation compared to gcon = 2.5 nS. Similarly, at time 20 ms, total propagation of ventricular sheet covers the region of 2.7 mm which is shown in Fig. 3c. Figure 3d shows V at time 94 ms in 2D square simulation domain of side length 10 mm. With g_{con} = 2.5 nS, the ventricular sheet of length 10 mm propagation time is 28 ms, whereas when g_{con} = 0.5 nS, the ventricular sheet of length 10 mm propagation time is 94 ms. This result of AP propagation delay in multicellular fiber depends on different g_{con} = 0.5–2.5 nS, which is absolutely great evidence of AP propagation that depends on gap junction coupling current (I_{cup}) and not on the applied stimulus, for excitation.

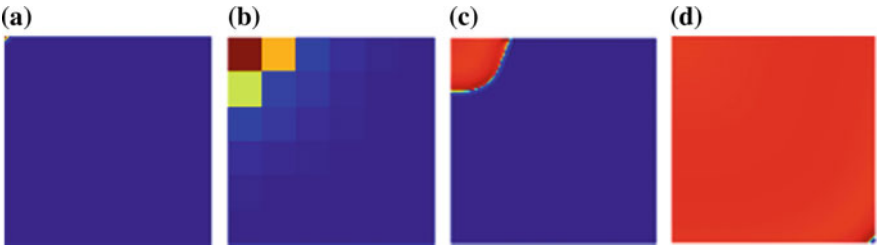


Fig. 3. AP generation and propagation in 2D simulations performed on a 100 × 100 square lattice of normal endocardial human ventricular cell with gcon = 0.5 nS. **a** V at t = 1 ms of L = 10 mm. **b** V at t = 1 ms of L = 0.6 mm. **c** V at t = 20 ms of L = 10 mm. **d** V at t = 94 ms of L = 10 mm

AP generation and propagation in 2D simulations are performed on the first half 50 × 50 square lattice of enhanced activity of Na^+–Ca^{2+} exchanger (failing myocyte) and the second half 50 × 50 square lattice of normal (nonfailing myocyte) endocardial human ventricular cell with gcon = 2.5 nS. The results of these computations are shown in Fig. 4a–d (red region is the depolarised region and blue region is the resting region). Figure 4a and b shows V at time 27 ms in 2D square simulation domain of side length 10 mm. In 10 mm region, 99.72% is fully depolarized, 0.13% is half depolarized, and finally 0.15% is at resting state. Compared with Figs. 2d and 4b, 10% of ventricular sheet resting state is reduced because of enhanced activity of Na^+–Ca^{2+} exchanger. Plateau potential region of ventricular sheet is shown in Fig. 4c. In 5.1–10 mm region at t = 25–60 ms, the total ventricular sheet is in reduced plateau potential state. In the second half 50 × 50 square lattice of normal (nonfailing myocyte), endocardial human ventricular cell region has plateau potential effect due to the first half 50 × 50 square lattice of enhanced activity of Na^+–Ca^{2+} exchanger (failing myocyte). Compared with Figs. 2f and 4d, the major effects of enhanced activity of Na^+–Ca^{2+} exchanger (failing myocyte) reduce the APD, which is clearly observed from Fig. 4d of ventricular sheet at time t = 250 ms that is going to resting state very early.

AP generation and propagation in 2D simulations performed on the first half 50 × 50 square lattice of hypertrophy condition (failing myocyte) and the second half 50 × 50 square lattice of normal (nonfailing myocyte) endocardial human ventricular cell with gcon = 2.5 nS. The results of these computations are shown in Fig. 5a–d (red region is the depolarised region and blue region is the resting region). Figure 5a and b shows V at time 27 ms in 2D square simulation domain of side length 10 mm that is same as enhanced activity of Na^+–Ca^{2+} exchanger (failing myocyte). But in plateau potential region (red region is the maximum plateau potential and blue region is the minimum plateau potential), ventricular sheet having tremendous changes compared to nonfailing myocyte is shown in Fig. 5c. In 5.1–10 mm region at t = 25–60 ms, 50% of ventricular sheet is reduced to minimum plateau potential state. In the second half 50 × 50 square lattice of normal (nonfailing myocyte), endocardial human ventricular cell region has plateau potential effects due to the first half 50 × 50 square lattice of

hypertrophy condition (failing myocyte). Compared with Figs. 2f and 5d, the major effects of hypertrophy (failing myocyte) reduce the APD, which is clearly observed from Fig. 5d that ventricular sheet at time t = 250 ms goes to completely resting state.

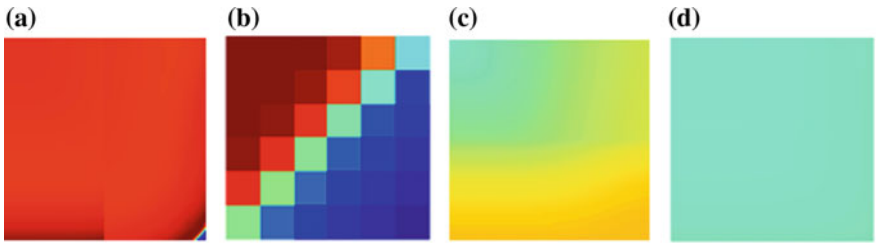


Fig. 4. AP generation and propagation in 2D simulations performed on the first half 50 × 50 square lattice of enhanced activity of sodium–calcium exchanger (failing myocyte) and the second half 50 × 50 square lattice of normal (nonfailing myocyte) endocardial human ventricular cell with gcon = 2.5 nS. **a** V at t = 27 ms of L = 10 mm. **b** V at t = 27 ms of L = 9.4–10 mm. **c** V at t = 25–60 ms of L = 10 mm (Plateau potential region). **d** V at t = 250 ms of L = 10 mm

AP generation and propagation in 2D simulations are performed on the first half 50 × 50 square lattice of SR calcium uptake channel and calcium leakage channel failure condition (failing myocyte) and the second half 50 × 50 square lattice of normal (nonfailing myocyte) endocardial human ventricular cell with gcon = 2.5 nS. The results of these computations are shown in Fig. 6a–d (red region is the depolarised region and blue region is the resting region). Figure 6a and b shows V at time 27 ms in 2D square simulation domain of side length 10 mm that is same as enhanced activity of Na^+–Ca^{2+} exchanger (failing myocyte). But in resting potential region (red region is the deviant resting potential state and blue region is the normal resting potential state), ventricular sheet has tremendous changes compared to nonfailing myocyte as shown in Fig. 6c and d. In the second half 50 × 50 square lattice of normal (nonfailing myocyte), endocardial human ventricular cell region has resting potential effects due to the first half 50 × 50 square lattice of SR calcium uptake channel and calcium leakage channel failure condition (failing myocyte). Figure 6c shows V at time 290 ms in 2D square simulation domain of side length 10 mm. At time 290 ms, 70% of ventricular sheet is in deviant resting potential state in the range −83 mV. Similarly, Fig. 6d shows V at time 295 ms in 2D square simulation domain of side length 10 mm. Compared with Figs. 2i and 6d, the major effects of SR calcium uptake channel and calcium leakage channel failure condition (failing myocyte) reduced the CICR and excitation–contraction coupling, which is clearly observed from Fig. 6d that ventricular sheet at time 295 ms goes to 65% deviant resting state.

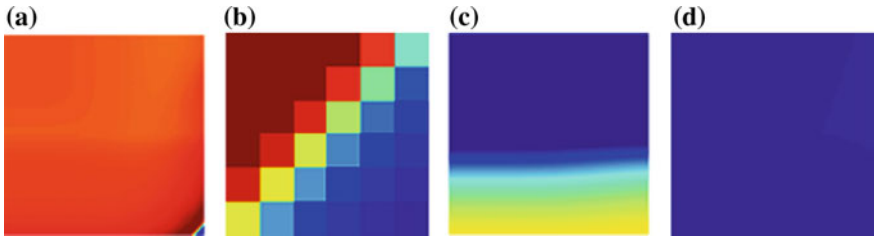


Fig. 5. AP generation and propagation in 2D simulations. **a** V at t = 27 ms of L = 10 mm. **b** V at t = 27 ms of L = 9.4–10 mm. **c** V at t = 25–60 ms of L = 10 mm (Plateau potential region). **d** V at t = 250 ms of L = 10 mm

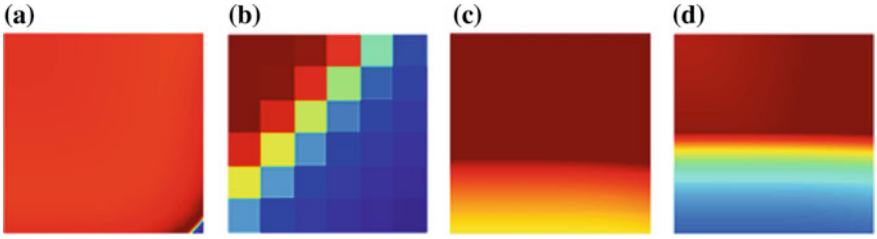


Fig. 6. AP generation and propagation in 2D simulations. **a** V at t = 27 ms of L = 10 mm. **b** V at t = 27 ms of L = 9.4–10 mm. **c** V at t = 290 ms of L = 10 mm (Resting membrane potential region). **d** V at t = 295 ms of L = 10 mm (Resting membrane potential region)

4 Conclusions

In summary, amendment of the electric properties of single myocytes may favor the occurrence of ventricular arrhythmias in patients with exert influence on early or delayed after depolarizations. Cardiac contractility in ventricular tissue is prominent by calcium channels like voltage-dependent calcium channels ($I_{Ca,L}$), sodium–calcium (Na^+–Ca^{2+}) exchanger, sarcoplasmic reticulum (SR), calcium pump ($I_{p,Ca}$), and leakage channels (I_{leak}). Theoretical ventricular sheet is framed by integrating two cells with linear constant gap junction conductance. The conduction velocity (CV) of the 1D model is found to be 50 cm/s (Normal value g_{con} = 2.5 μS) and agrees better with available data. The simulation results of this work illustrate that shorter APD, instability resting potential, and affected calcium-induced calcium release (CICR) due to deviant calcium channels potentially have a substantial effect on cardiac contractility and relaxation. Computational study of single ventricular cells and tissue action potential and its underlying ionic channel currents could aid to clarify possible arrhythmogenic mechanism on a cellular level.

References

1. Priebe L, Beuckelmann DJ (1998) Simulation study of cellular electric properties in heart failure. Circ Res 82:1206–1223
2. ten Tusscher KHWJ, Noble D, Noble PJ, Panfilov AV (2004) A model for human ventricular tissue. Am J Physiol Heart Circ Physiol 286:H1573–H1589
3. ten Tusscher KHWJ, Panfilov AV (2006) Alternans and spiral breakup in a human ventricular tissue model. Am J Physiol Heart Circ Physiol 291:H1088–H1100
4. Bueno-Orovio A, cherry EM, Fenton FH (2008) Minimal model for human ventricular action potentials in tissue. J Theor Biol 253:544–560
5. Chudin E, Goldhaber J, Garfinkel J, Weiss A, Kogan B (1999) Intracellular Ca^{2+} dynamics and the stability of ventricular tachycardia. Biophys J 77:2930–2941
6. Garfinkel A, Kim YH, Weiss JN, Chen PS (2000) Preventing ventricular fibrillation by flattening cardiac restitution. Proc Natl Acad Sci USA 97:6061–6066
7. Karma A (1994) Electrical alternans and spiral wave breakup in cardiac tissue. Chaos 4: 461–472
8. Syed Z, Vigmond E, Leon LJ (2005) A trial cell action potential parameter fitting using genetic algorithms. Med Biol Eng Comput 43:561–571
9. Henry H, Rappel WJ (2005) Dynamics of conduction blocks in a model of paced cardiac tissue. Phys Rev E 71:051911
10. Jespersen T (2012) The cardiac ion channels. ISBN: 978-953-51-0126-0
11. Philipson KD, Nicoll DA (2000) Sodium-calcium exchange: a molecular perspective. Annu Rev Physiol 62:111–133
12. Delbridge LM, Bers DM (1996) steady-state twitch Ca^{2+} fluxes and cytosolic Ca^{2+} buffering in rabbit ventricular myocytes. Am J Physiol 270:C192–C199
13. Bridge JHB, Smolley JR, Spitzer KW (1990) The relationship between charge movements associated with I_{Ca} and I_{NaCa} in cardiac myocytes. Science 248:376–378
14. Bers DM, Spitzer KW (1989) Intracellular Ca^{2+} transients during rapid cooling contractures in guinea-pig ventricular myocytes. J Physiol 417:537–553
15. Henriquez AP, Weingart R, Cascio WE (2001) Influence of dynamic gap junction resistance on impulse propagation in ventricular myocardium: a computer simulation study. Biophys J 81:2112–2121
16. Bondarenko VE (2009) Action potential propagation in inhomogeneous 2D mouse ventricular tissue model. Int J Med Heart Biomed Bioeng 3(6)
17. Nayak AR, Panfilov AV, Pandit R (2013) Spiral-wave dynamics in a mathematical model of human ventricular tissue with myocytes and fibroplasts. PLoS ONE 8:e72950

Modeling and Simulation of 1/f Noise During Threshold Switching for Phase Change Memory

Rutu Parekh[1(✉)], Maryam Shojaei Baghini[2], and Bipin Rajendran[3]

[1] DA-IICT, Gandhinagar, India
rutu_parekh@daiict.ac.in
[2] IIT-Bombay, Powai, Mumbai, India
mshojaei@ee.iitb.ac.in
[3] New Jersey Institute of Technology, Newark, NJ, USA
bipin@njit.edu

1 Introduction

The ITRS considers phase change memory (PCM) as an existing baseline and prototypical memory technology [1–3]. It is currently most mature of the new memory technologies under research. Development of PCM with high speed, density, non-volatility, decreased energy consumption, scalability, 3D integration, endurance, retention, and multi-bit operation would initiate a revolution in computer architecture, known as nanostores, which will be the future of information processing. PCM has potential to be a universal memory that works across multiple layers of memory hierarchy.

PCM relies on chalcogenide material typically $Ge_2Sb_2Te_5$ (GST) which features the capability to change from amorphous to polycrystalline phase and vice versa by giving electrical pulses and subsequent joule heating. Figure 1 shows the schematic of a PCM device. It comprises a thin GST film, heater (TiN), top and bottom electrode contacts, and an active GST that undergoes phase transition between high resistance amorphous state (RESET state) and low resistance polycrystalline GST (SET state) when the PCM cell is stimulated thermally. The active region thickness of u_a and the contact area, A, shall be used to study the effect of scaling on threshold switching. The polycrystalline GST surrounds the active GST and it does not change during program. With proper heating applied, Joule heating is induced resulting in current crowding near the heater-active area interface, and hence phase change of the active GST.

In the reset state, PCM exhibits a threshold switching behavior in the I–V characteristics which is a negative differential resistance (NDR) effect beyond threshold voltage, V_T. It defines a boundary between read and write current and voltages which determines the speed for programming the memory cell. It is therefore important to estimate threshold switching and speed for developing PCM memory. Threshold switching corresponds to two distinct electronic states of the amorphous material which are subthreshold state (OFF state) at low current and a high current ON state.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_8

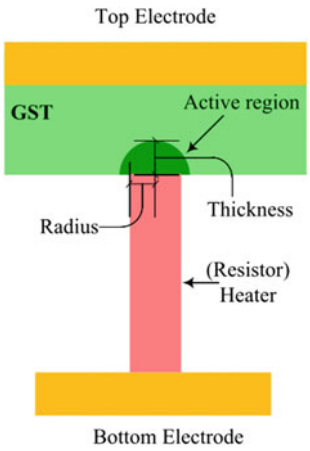


Fig. 1. Structure of a PCM cell

For modeling and design purpose, it is significant to predict the dependence of I–V characteristics on PCM device physical parameters, voltages, and temperature for amorphous chalcogenide during subthreshold conduction and threshold switching. In addition, low-frequency 1/f-like noise is a matter of concern for nanoscaled nonvolatile memory devices, as it produces fluctuations in current which can cause errors during memory readout. Research by Lavizzari et al. has addressed the impact of 1/f fluctuating current during threshold switching by measuring the effect on PCM cell and developed a physical model to account for the statistical delay in the device [4]. A compact model for 1/f noise during threshold switching that can be accounted for simulating PCM array is still missing. In this work, we have developed a compact model for PCM cell in VerilogA of Cadence environment [5] which accounts for subthreshold conduction and threshold switching considering 1/f noise.

2 Modeling and Simulation of 1/F Noise in Threshold Switching

PCM compact model developed in VerilogA is based on analytical model that can address subthreshold conduction, threshold switching, negative differential resistance region, and ON regime. The model constitutes Poole–Frenkel (PF) conduction mechanism in deep traps in the amorphous material [4, 6]. The model was further developed to account for 1/f current fluctuations which can predict delay statistics in PCM. The model developed in VerilogA is implemented in Spectre Circuit Simulator and Virtuoso Analog Design Environment of Cadence. This has a facility of schematic editor for design and simulation of hybrid PCM circuits with CMOS, active and passive devices. Figure 2 shows the schematic implemented in Cadence environment to study the threshold switching during transient simulation. Figure 3 shows I–V curves of a PCM cell in different programmed states benchmarked with that presented by Ielmini

et al. [6]. Reset pulses with different amplitudes equal to 3.2 V, 3.5 V, and 5 V are applied at source V_A for amorphous layer thickness of 18 nm, 27 nm, and 39 nm, respectively, and transient simulation carried out to get the I–V curves. The list of the model parameters used in the simulation is listed in Table 1.

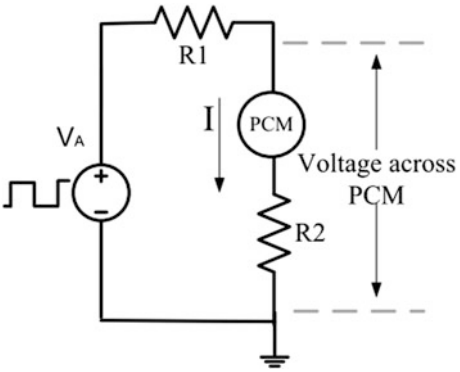


Fig. 2. Schematic for transient simulation of a PCM cell

There is just a little decrease in programmed resistance corresponding to different pulse amplitudes even though the pulse width is reduced to 10 ns. This is because of a delay to achieve a thermal steady state in the heated volume due to large thermal resistivity of the GST film. A RESET pulse of 10 ns and SET pulse of 20 ns are sufficient for a decade change in the resistance of a PCM cell resulting in a sufficient read margin [7]. In addition, reduction in reset pulse width improves PCM cell endurance.

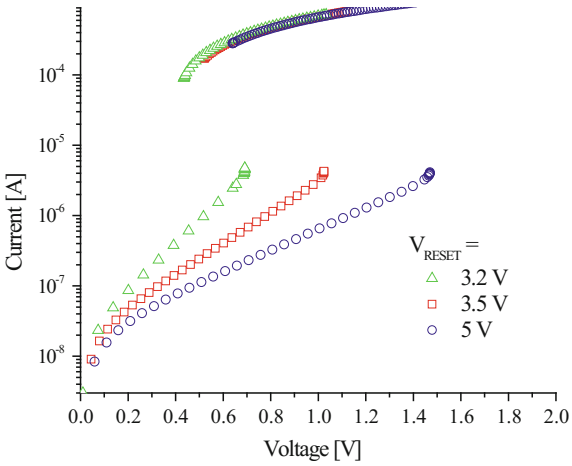


Fig. 3. Simulated I–V characteristics of a PCM cell using the compact model

Table 1. Amorphous chalcogenide physical parameters used in the circuit simulation [6]

PCM device parameters	Values
Elementary charge, q	1.6e–19 C
Average intertrap spacing, Δz	7 nm
Electronic vibration time, τ_0	1e–15 s
Energy distance between equilibrium Fermi level E_F and the conduction band-edge energy E_C, ($E_C - E_F$)	0.36 eV
Boltzmann constant, k_B	8.617e–5 eVK^{-1}
Sample temperature, T	300 K
Concentration of deep states, N_{T1}	1e19 cm^{-3}
Shallow-trap concentration, N_{T2}	1e19 cm^{-3}
Effective time for electron relaxation from shallow to deep traps, τ_n	1e–9 s
Fowler–Nordheim coefficient, B_{12}	1e9 Vm^{-1}
Energy distance between deep-trap energy E_{T1}, and the shallow-trap energy E_{T2}, E_{T2}–E_{T1}	0.29 eV

3 PCM Transient Simulation with 1/f Noise Fluctuations

Transient simulation of the PCM cell considering 1/f noise was carried out as shown in Fig. 4 with applied bias voltage V_A of 800 mV just below V_T for a PCM device thickness of 21 nm and radius of 13.8 nm. As seen in the diagram, 1/f noise can result in statistical variation of the delay time. When the applied voltage is close to V_T, noise fluctuations can lead in current to exceed the threshold voltage which triggers the switching event. This can result in large delay time as shown below. During read operation, threshold switching must not occur else it can disturb the state of the cell. Thus, the delay time can impact the device operation resulting in read disturb and programming errors.

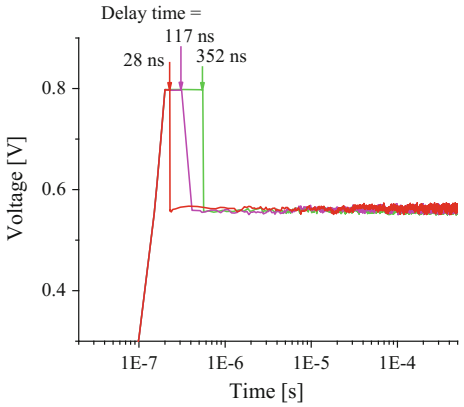


Fig. 4. Simulation result of transient behavior of a PCM device with 1/f noise fluctuations

4 Effect of PCM Device Scaling on Threshold Switching Considering 1/f Noise

The scaling capability is an important virtue of the PCM device as it allows the technology to be scaled for many years. As shown in Fig. 1, the size of the bottom electrode determines the writing current since the heat generated depends on the current density. With reduction in contact size, the current density increases and hence the Joule heating. Thus, a lower reset current can achieve the required temperature. The smaller the active GST material, the less the power will be needed to heat up and hence less programming current. The power dissipated (P) at the GST and heater interface can be given by [7],

$$\mathrm{P} = V_h \mathrm{I} + \eta R_{heater} I^2 \tag{1}$$

Here, V_h is the voltage drop inside the chalcogenide, I is the programming current, η is the heater efficiency, and R_{heater} is the heater resistance. Transient simulation

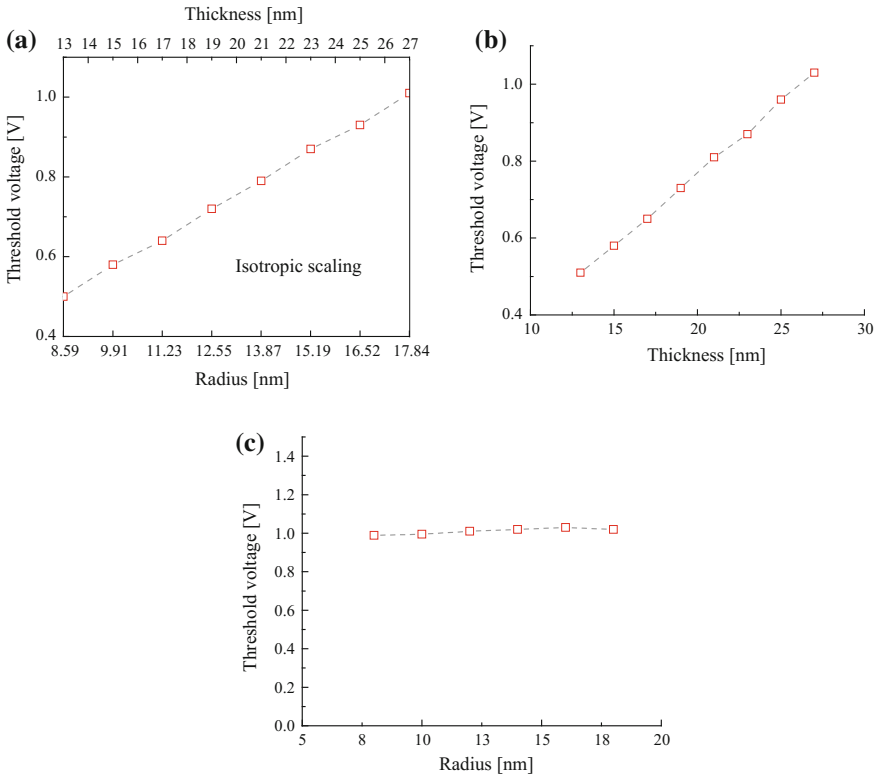


Fig. 5. PCM device simulation result. **a** Threshold voltage dependence on thickness and radius of the amorphous chalcogenide of a PCM device in case of isotropic scaling. **b** Threshold voltage dependence on amorphous chalcogenide layer thickness for a fixed area. **c** Threshold voltage dependence on amorphous chalcogenide layer radius for a fixed thickness

considering 1/f noise is used to study the effect of phase change device scaling on the threshold switching voltage. The PCM device shown in Fig. 1 can be scaled down by shrinking the amorphous chalcogenide layer thickness, its contact area, or both.

Figure 5a shows the plot of V_T versus amorphous chalcogenide layer radius and thickness for an isotropic scaling wherein a constant aspect ratio = thickness/radius = 1.5 initially obtained from thickness = 27 nm for A = 1000 nm^2 is maintained. The graph shows that V_T increases with PCM dimensions in case of isotropic scaling. Figure 5b shows the effect of varying amorphous layer thickness on V_T while keeping its area constant to 1000 nm^2. It can be seen that V_T increases with increase in amorphous chalcogenide layer thickness. Figure 5c shows that V_T remains almost constant in spite of changing contact radius while maintaining a fix thickness (27 nm). This is because, by decreasing the heater thickness, the overall thermal resistance, R_{th} and R_{heater}, increases. The programming current must decrease by the same factor to melt the GST by providing a temperature change of $\Delta T = R_{th} P$ of 630 °C [7]. As a consequence, the overall voltage drop V, across the cell that is equal to $V_h + R_{heater} I$, remains unchanged. Minor deviation in the graph is because of the 1/f noise.

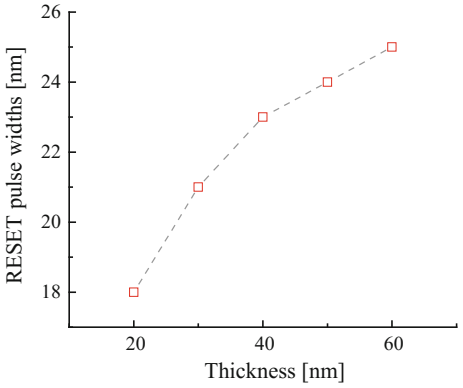


Fig. 6. Effect of amorphous chalcogenide layer thickness on minimum required RESET pulse width

Further, the thickness of the amorphous layer effects the programming speed. By observing minimum RESET pulse width that can induce cell amorphization, it is possible to measure the programming speed of a PCM cell. To study the effect by simulation, we fix the pulse amplitude to 3.5 V with rise and fall time as 1 ns. Figure 6 shows that with increase in thickness of the amorphous chalcogenide the minimum width of the reset pulse also increases.

5 Conclusion

We have simulated a compact VerilogA model of a PCM cell which can predict the subthreshold characteristics and threshold switching. The compact model is developed from a numerical model based on PF conduction and energy gain for the transient simulation of threshold switching. Then, 1/f current fluctuations are added to the PCM model in VerilogA which is able to account for delay statistics and contributes to an insight of the physics during the switching transients.

Simulation results for delay time for an applied voltage just below V_T show a large statistical variation that can result in read disturb and programming errors. In addition, the simulation results for the threshold voltage versus PCM scaling shows that the threshold voltage increases with thickness and remains almost constant with the increase in the area of the amorphous chalcogenide layer. Finally, the simulation result shows that PCM cell with thinner layer has a shorter pulse width for RESET operation and hence low-power consumption.

References

1. The International Technology Roadmap for Semiconductors, 2013 Edition. http://www.itrs.net/
2. Kim H, Seshadri S, Dickey CL, Chiu L (2014) Evaluating phase change memory for enterprise storage systems: a study of caching and tiering approaches, IBM Almaden Research. In: Proceedings, 12th USENIX conference on file and storage technologies (FAST '14), Santa Clara, CA USA, 17–20 Feb 2014
3. Burr GW, Breitwisch MJ, Franceschini M, Garetto D, Gopalakrishnan K, Jackson B, Kurdi B, Lam C, Lastras LA, Padilla A, Rajendran B, Raoux S, Shenoy RS (2010) Phase change memory technology. J Vac Sci Technol B 28(2):223–262
4. Lavizzari S, Ielmini D, Lacaita AL (2010) Transient simulation of delay and switching effects in phase-change memories. IEEE Trans Electron Devices 57(12):3257–3264
5. Cadence Design Systems. http://www.cadence.com
6. Ielmini D, Zhang Y (2007) Analytical model for subthreshold conduction and threshold switching in chalcogenide-based memory devices chalcogenide-based memory devices. J Appl Phys 102:054517-1–054517-13
7. Bez R, Gleixner RJ, Pellizzer F, Pirovano A, Atwood G (2009) Phase change memory cell concepts and designs. In: Raoux S, Wuttig M (eds) Phase change materials. Springer, Boston, MA, pp 355–380

Pilot Subcarrier Based Channel Estimation in OFDM System

Harjeet Singh[1(✉)] and Savina Bansal[2]

[1] IKG Punjab Technical University, Kapurthala, India
sharjeetl5@yahoo.co.in
[2] GZS Campus College of Engineering & Technology, Bathinda, Punjab, India
savina.bansal@gmail.com

1 Introduction

Channel estimation is the first and foremost issue for today's high data rate transmission for wireless communication system with multipath propagation. For high data rate transmission with minimum impairments, orthogonal frequency division multiplexing is a technique with long symbol duration, spectrally efficient and high bandwidth efficiency for frequency-selective multipath fading channels. In OFDM system, pilot subcarriers are scattered with data symbols in time-varying channel to track time- and frequency-selective channel characteristics. Studies on the design of pilot patterns and pilot density are taken in [1–3] for time- and frequency-selective channel to minimize the mean square error (MSE) assuming equi-spaced pilot subcarrier. Allocation of fixed number of pilot subcarrier in data stream for OFDM system, where the channel impulse response varies rapidly, degrades the system performance. Some methods for pilot pattern design were proposed in [1, 4, 5] where minimum delay spread and minimum Doppler frequency are considered. Exiting pilot patterns cannot provide optimal performance for exact distribution of power delay profile and Doppler frequency.

In this paper, first we present various pilot pattern design techniques for frequency-selective multipath fading channels. An algorithm is discussed for mapping of 4-bit pilot symbols using gray-coded 16-QAM constellation, equivalent to superimposing two ASK signals on quadrature carriers (I and Q components), and also suggest the technique to insert gray-coded pilot symbol constellation into OFDM data stream before transmission on channel. Here, we choose low Doppler frequency to observe the effect of vehicle moving in a heavy traffic passing.

2 Related Work

Many works are available in literature that is suggested to improve efficiency of estimating the channel characteristics for time-varying channels under multipath environment. Generation of random pilot pattern is more challenging in time-varying

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_9

systems due to its high complexity. Inserting pilot symbols before transmission for proper channel estimation is basically an overhead of the system, and it is desirable to keep the number of pilot symbols to a minimum. For time-varying system characterized by Doppler shift, pilot insertion was taken up in [6], and it was shown that the number of pilot symbols for a desired bit error rate and Doppler frequency is highly dependent on the pilot patterns used. So, by choosing a suitable pilot pattern, we can reduce the number of pilot symbols, and still retain the same performance. Commonly used pilot patterns in literature are block and comb pilot arrangements [7–9] with the latter performing much better than the former in fast varying environments [2]. A scheme has been proposed in [10] using the improved discrete stochastic approximation to optimize the pilot placement in OFDM systems. The channel data is employed to search the near-optimal pilot placement before the transmission for accurate channel estimation. Three pilot design schemes were proposed in [11] to obtain near-optimal pilot pattern for spare channel estimation. Techniques named stochastic sequential search (SSS) and stochastic parallel search (SPS) are based on the stochastic search, and iterative group shrinkage (IGS) technique employs a tree-based searching structure and removes rows in a group instead of removing a single row at each step. The SSS, SPS, and IGS schemes converge much faster than the cross-entropy optimization and the exhaustive search and are thus more efficient. Moreover, SSS and SPS outperform IGS in terms of channel estimation performance. An approach to channel estimation by singular value decomposition taken up in [12] and channel estimation characterized by Doppler shift over Rayleigh fading multipath channel is available in [8, 13–17].

3 Pilot Patterns

Pilot symbol patterns having regular structure with equi-powered and equi-spaced patterns are used for better performance in communication system. For block-type arrangement, subcarriers are transmitted periodically and the interval between two consecutive symbols must be shorter than the channel coherence time. This type is suitable for slow fading channels. In comb-type arrangement, subcarriers are reserved for pilot signals which are transmitted continuously and spacing among the pilot subcarriers must be less than the coherence bandwidth of the channel. This arrangement is used for accurate and effective channel estimation in fast fading channels.

Rectangular grid, parallelogram spaced grid, and hexagonal grid type pilot subcarriers provide subsampling of 2-D channel response. To avoid the aliasing effect in both dimensions, sampling theorem must be obeyed by the said arrangement. Scattered pilot arrangement of all these three arrangements improves spectral efficiency and reduces the pilot density.

4 Pilot Symbol M-ary Modulation

In this paper, an algorithm is proposed as outlined below for data-aided channel estimation in multipath fading channel using 16-QAM modulation technique for time-varying OFDM system.

The passband $M = 2^b$-ary QAM signal waveforms can be written as

$$\mathrm{S_m}(t) = A_m \sqrt{\frac{2}{T_s}\cos(w_c t)}$$

with $A_m = \sqrt{A_{mc}^2 + A_{ms}^2}$ and $\theta_m = \tan^{-1}\frac{A_{ms}}{A_{mc}}$.

Let the amplitudes or phases of QAM signal waveforms ($M = 2^b$-ary) be arranged in such a manner that they are represented by a rectangular constellation in the signal space having minimum distance of 2d among the signal points. Then, the average number of adjacent signal points (N_b) and the average of signal powers (E_s, a_v) for a signal point vary with the modulation order $2^M = b$ (bits per symbol) as $\mathrm{M} = 2^4 = 16$

$$\mathrm{E_{s,a_v}} = \frac{1}{M}\sum_{m-1}^{M-1} d_m^2 = \frac{4d^2 \times (1^2 + 3^2) \times 4}{16} = 10d^2 = \frac{2(M-1)}{3}d^2$$

$$N_b = 4 - \frac{4}{\sqrt{M}}$$

In case of circular constellation, we have same average energy ($\mathrm{E_{b,av}}$) with larger minimum distance but the difference is very small for $\mathrm{M} \geq 16$. In case of $\mathrm{M} = 2^{\mathrm{b}}$-ary signaling with rectangular constellation, signal can easily be implemented by two independent b-ary PAM signalings. Thus, QAM rectangular constellation signaling is widely used. Symbol probability error for LN-ary QAM signaling (combining an L-ary PAM signaling and an N-ary PAM signaling) can be found as

$$\begin{aligned}\mathrm{P_{e,s}(M = LN)} &= 1 - \mathrm{P\ (probability\ of\ correct\ detection)} \\ &= 1 - \big(1 - \mathrm{P_{e,s}(L)}\big)\big(1 - \mathrm{P_{e,s}(N)}\big)\end{aligned}$$

$$P_{e,s}(M = LN) \leq \frac{4(L-1)}{L} Q\left[\sqrt{\frac{3b/2}{M-1} SNR_{r,b}}\right] \text{ with } \mathrm{L} \geq \mathrm{N}.$$

Here, the information symbols are gray-coded so that the codes for adjacent signal points differ by one bit; the relationship between the symbol and bit errors, when the frequently occurring symbol errors contain just one of the b bits mistaken can be written as $P_{e,b} = \frac{1}{b} P_{e,s}$.

In the proposed algorithm, mapping of four bits pilot symbols is taken up based on Karnaugh maps equivalent to superimpose two ASK signals on quadrature carriers (I and Q components) using gray-coded 16-QAM constellation. Let PQRS be the 4-bit pilot symbol; here, first two bits (PQ) are considered as the 2-bit ASK modulated (also called pulse amplitude modulation) on the in-phase arm and next two bits (RS) are considered as the two bits of ASK modulated on the quadrature arm. For 16 QAM

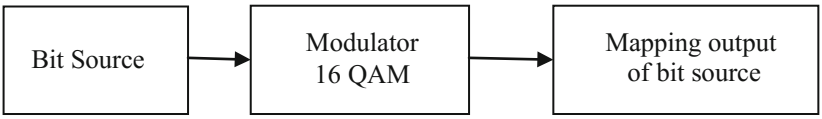
constellation, each point represents 4-bit pilot symbols named as ABCD, composed of two (AB) in-phase bits on I axis and two (CD) quadrature bits on Q axis. Here, bits ABCD are mapped to the levels (± 3, ± 1) using gray code. In this coding, when the received symbol is corrupted by noise as it is misinterpreted as an adjacent constellation point, it ensures that the demodulator will make a single-bit error. In OFDM system, the input data stream of random binary bits is mapped into I and Q QAM symbols.

4.1 Proposed Algorithm

The main steps involved are as follows:

1. Mapping of pilot bit into respective gray-coded constellations.
2. Inserting of gray-coded pilot symbols into OFDM data stream before transmission.

   - Set the sub-symbol number of OFDM (N) and symbol number of one simulation (N_L) from mapping output of bit source as follows.



   [N, NL] = size (Mapping output of bit source)
   Here, bit source is obtained from subcarriers index and OFDM symbols.

   - Mapping the gray-coded pilot constellation using 16 QAM corresponding to each binary input pilot symbol as follows.

(-3,3)	(-1,3)	(1,3)	(3,3)
(-3,1)	(-1,1)	(1,1)	(3,1)
(-3,-1)	(-1,-1)	(1,-1)	(3,-1)
(-3,-3)	(-1,-3)	(1,-3)	(3,-3)

3. Set the sequence length based on sub-symbol number of OFDM (N), symbol number of one simulation (N_L), and pilot interval.

$$\textbf{Sequence length} = \textbf{zeros}\,(\mathbf{N},\ (\mathbf{N_L} + \textbf{ceil}(\mathbf{NL}/\textbf{pilot interval}))).$$

4. Set the pilot sequence length based on the input pilot bit taken into consideration.

$$\textbf{Pilot sequence length} = \textbf{complex symbol of the pilot symbol} * \textbf{ones}\,(\mathbf{N},\ \mathbf{1})$$

5. Proposed gray coded pilot constellation insert with data stream in OFDM frame after equal interval as following proposed algorithm.

```
Count = 0;
 i=1;
While
  Count < (symbol number of one simulation (N_L)/pilot interval);
     Output sequence series = pilot sequence length;
     Count=Count+1;
  if
  Count * pilot interval <= symbol number of one simulation (N_L)
  Output sequence series =mapping (:, (( count-1)* pilot interval+1): count* pilot
interval);
  else
     Output sequence series = mapping (:,(( count-1)* pilot interval+1): N_L);
end
          i=i +pilot interval+1;
end
```

5 Simulation Results

Performance evaluations of the proposed OFDM system with multipath delay channel for gray-coded pilot symbol bits are carried out using simulations. Channel bandwidth of OFDM system and the sampling rate were chosen as 1 MHz. Total number of subcarriers was set to 128, which yielded the subcarrier spacing of 7.8125 kHz and the OFDM symbol time was kept as 128 μs. Here, we took guard interval ratio of, which yielded a guard interval of 32 μs. The total OFDM block time then becomes 160 μs. For OFDM multipath channel estimation over Rayleigh fading channel, channel coefficients and Doppler frequency are kept constant for lowering the system

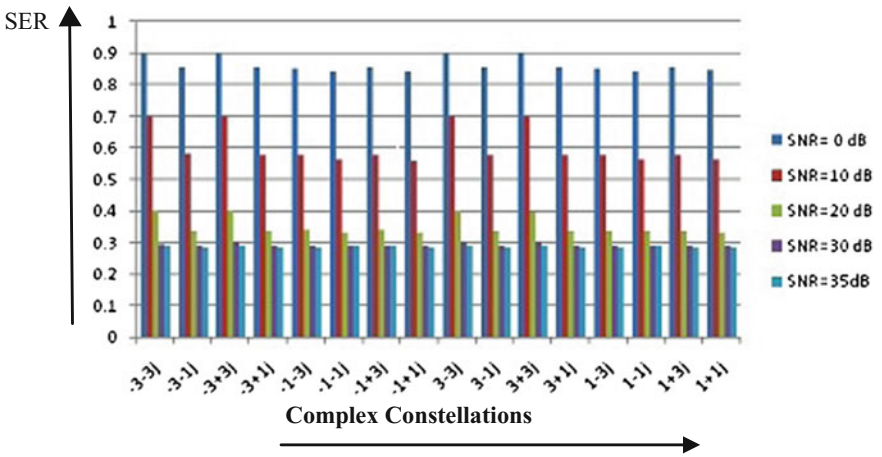


Fig. 1. Variation of SER versus pilot symbol complex constellations

complexity. Here, we consider five paths for transmission, delay is kept at [0, 2, 4, 8, 12 μs], and RMS delay spread of the channel is 4 μs.

Performance of OFDM system is observed for gray-coded pilot symbol complex constellation as shown in Fig. 1. It is seen that with input gray-coded complex constellation changed by one bit, the system performance changes at each constellation point due to variation of minimum distance among them.

The simulation is extended further over the time-varying channel having Doppler shift of 50, 100, and 200 Hz. Here, we consider gray-coded constellation training symbol 0010 (−3 + 3j) and 0111 (−1 + j) gives least and maximum SER, respectively,

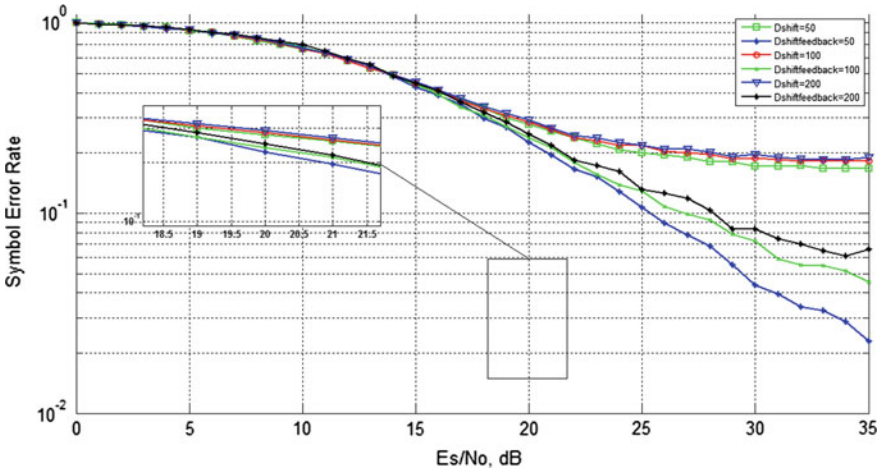


Fig. 2. Comparison of LS and feedback LS techniques for gray-coded complex pilot constellation (−3 + 3j)

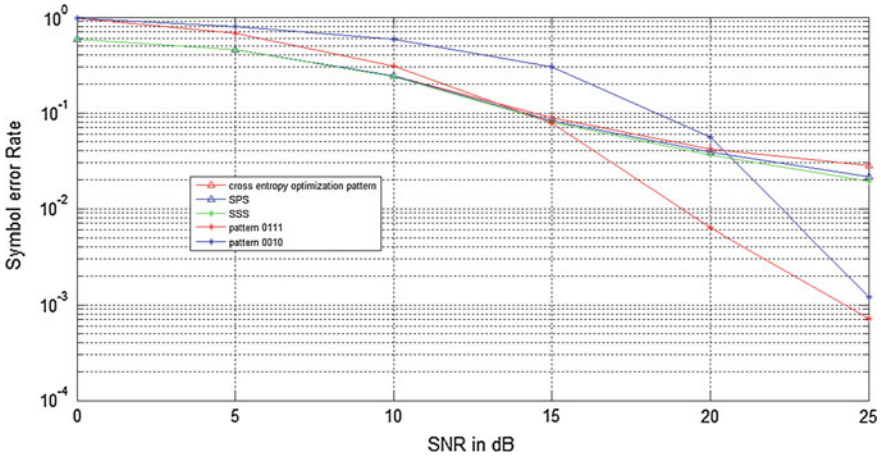


Fig. 3. Channel estimation comparison of different pilot patterns

as shown in Fig. 1. Simulation results for the time-varying channel are shown in Fig. 2 for the Doppler shift variation considering heavy traffic area in metro cities. It is seen that proposed system using gray-coded symbol (0010 (−3 + 3j)) performed better at 100 Hz Doppler shift for LS and feedback LS estimation technique as proposed in [18] by the authors. Also, feedback LS estimation technique improved the SER performance over LS channel estimation technique.

In Fig. 3, performance of different pilot patterns for cross-entropy optimization, stochastic sequential search (SSS), and stochastic parallel search (SPS) taken up in [11] are compared with the pattern scheme proposed in this work. Results show that proposed pilot pattern 0111(−1 + j) resulted in lowest SER beyond 14 dB among all the compared techniques. Similarly, proposed pilot pattern 0010 (−3 − 3j) also performed better though beyond 21 dB.

6 Conclusion

In this paper, we propose an algorithm for inserting a gray-coded pilot symbol into OFDM data stream. Simulation is carried out for all pilot symbols constellation to find the best possible constellation for DFT-based time-varying OFDM system. Our simulation results show that among all gray-coded pilot constellation for 16-QAM, the constellation (0111 (−1 + j)) performs much better as compared to other constellations. The proposed pilot patterns improved the system performance compared to cross-entropy optimization, stochastic sequential search (SSS), and stochastic parallel search (SPS) pilot patterns.

References

1. Ohno S et al (2011) Preamble and pilot symbol design for channel estimation in OFDM system with null subcarriers. EURASIP J Wirel Commun Netw 1–17
2. Coleri S et al (2002) Channel estimation techniques based on pilot arrangement in OFDM system. IEEE Trans Broadcast 48(3):223–229
3. Zhang W et al (2006) Optimal training and pilot pattern design for OFDM systems in Rayleigh fading. IEEE Trans Broadcast 52(4):50–57
4. Dong M, Tong L, Sadler BM (2002) Training placement for tracking fading channels. In: Proceedings of international conference acoustics, speech and signal processing (ICASSP), Orlando, FL, pp 2189–2192
5. Li Y, Cimini LJ, Sollenberger NR (1998) Robust channel estimation for OFDM systems with rapid dispersive fading channels. IEEE Trans Commun 46(7):902–915
6. Tufvesson F, Maseng T (1997) Pilot assisted channel estimation for OFDM in mobile cellular systems. In: Proceedings of IEEE Vehicular Technology Conference, Phoenix, USA, pp 1639–1643
7. Liu M (2012) A novel data-aided channel estimation with reduced complexity for TDS-OFDM systems. IEEE Trans Broadcast 58(2):247–260
8. Ohno S, Giannakis GB (2002) Optimal training and redundant precoding for block transmissions with applications to wireless OFDM. IEEE Trans Commun 50(12):2113–2123

9. Hsieh M-H, We C-H (2009) Channel Estimation for OFDM system based on comb type pilot arrangement in frequency selective fading channel. IEEE Trans Wirel Commun 44(1): 217–255
10. Qi C, Wu L (2011) Optimized pilot placement for sparse channel estimation in OFDM systems. IEEE Signal Process Lett 18(12)749–752
11. Qi C et al (2014) Pilot design schemes for sparse channel estimation in OFDM systems. IEEE Trans Veh Technol 64(4):1493–1505
12. Edfors O, Sandell M, van de Beek J-J, Wilson SK, Baorjesson PO (1998) OFDM channel estimation by singular value decomposition. IEEE Trans Commun 46(7):931–939
13. Senol H et al (2012) Non data-aided joint channel estimation and equalization for OFDM systems in very rapidly varying mobile channels. IEEE Trans Signal Process 60(8): 4236–4253
14. Tang X et al (1999) Effect of channel estimation error on M-QAM BER performance in Rayleigh fading. IEEE Trans Commun 47(12):1856–1864
15. Moon JK, Choi SI (2000) Performance of channel estimation methods for OFDM systems in a multipath fading channels. IEEE Trans Consum Electron 46(1):161–170
16. Lin DD, Lim TJ (2007) The variation inference approach to joint data detection and phase noise estimation in OFDM. IEEE Trans Signal Process 55(5):1862–1874
17. Cai X et al (2004) Error probability minimizing pilots for OFDM with M-PSK modulation over Rayleigh-fading channels. IEEE Trans Veh Technol 53(1):146–153
18. Singh H, Bansal S (2017) Improvement in time varying multipath channel estimation. In: 5th International Conference on Advancements in Engineering and Technology (ICAET-2017), pp 124–126. ISBN No. 978-81-924893-2-2

Spectrum Allocation in Cognitive Radio Networks—A Centralized Approach

Monisha Devi(✉), Nityananda Sarma, and Sanjib Kumar Deka

Department of Computer Science & Engineering,
Tezpur University, Tezpur 784028, Assam, India
mnshdevi@gmail.com, nitya@tezu.ernet.in, sdeka@tezu.ernet.in

1 Introduction

The growing demand for radio spectrum in telecommunication sector has resulted in inadequacy of this scarce resource. Measurements captured by Federal Communications Commission (FCC) reveal that most of the licensed spectrum, assigned using fixed spectrum assignment policy, remains unused by the primary users (PUs) that initiate formation of spectrum holes. Such inefficiency in spectrum utilization promotes a new communication standard, called as Cognitive Radio (CR) [2], that allows dynamic and opportunistic access of the frequency bands. CR enables the secondary users (SUs) to opportunistically exploit the spectrum opportunities (SOP) in such a way that there is no any disruptive interference to PUs. Idle bands are sensed using a sensing process and following that, spectrum sharing takes over the responsibility of distributing the spectrum among SUs. Radio spectrum allocation corresponds to the designing of allocation mechanisms by formulating different design models [1] so as to enhance the spectrum utility performance.

This paper proposes a spectrum allocation mechanism for CRN that relies on a central entity for adopting the channel assignment decisions. This centralized approach takes into account the dynamics of channel availability of SUs and assigns the channels in a sequential manner. Also, an allocation strategy gets applied in this model wherein one channel can be allotted to only one SU at a time, and one SU can exploit at most one channel at a time. A simulation study has been performed to evaluate the performance of the proposed method and results are compared to a random allocation approach.

The rest of the paper is organized as follows. Section 2 presents the related work. The proposed model is described in Sect. 3. Section 4 discusses the performance evaluation. And finally, Sect. 5 gives the conclusion and future work in this field.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_10

2 Related Works

Spectrum sharing in CRN provides a platform for designing allocation mechanisms that render a fair distribution of the spectrum holes. Niyato and Hossain [4] design a non-cooperative game model where Nash equilibrium defines a solution for sharing the bandwidth awarded by PUs. A conflict graph-based channel allocation model is discussed in [5]. In [6], spectrum allocation is carried out considering heterogeneous nature of available channels where randomized rounding algorithm is deployed with several design constraints. Another design model that works with the sharing process is auction theory. In [7], a truthful double-sided auction game is formulated that allows spectrum reusability among non-interfering SUs. Kash et al. [3] discussed a strategyproof auction mechanism for assigning the spectrum among shared and exclusive-use devices by carrying out bucketing and ironing techniques to achieve a monotone allocation rule for proper channel allocation. A study on the spectrum sharing models designed for CRN tells that most of the work does not take into concern the dynamics in SOP of the SUs. Moreover, deploying of distributed approach requires the SUs to communicate among them that may cost a higher overhead in the network. In the proposed scheme, the set of channels sensed by each SU varies with their position with PUs. And, a centralized method, which needs no information exchange among the SUs, is allowed for allocation of the idle channels. The following section provides a clear discussion about the proposed model.

3 Spectrum Sharing Mechanism for CRN

3.1 System Model

The designed model accommodates a cognitive radio network, consisting of SUs, along with primary network, having PUs. PUs are licensed with M idle channels that can be shared among N SUs. It is assumed that $N > M$. The basic module is a central entity, C_E, that manages dynamic allocation of the spectrum. All SUs are considered to be within the transmission range of C_E. Before the allocation process, SUs sense for vacant channels from PUs since channel availability may vary with the location of SUs with respect to PUs' location. The list of idle channels for SU i is maintained as $C_i = \{i_1, i_2, i_3, \ldots, i_{N_i}\}$, where i_j represents the jth channel in C_i of SU i. $N_i = |C_i|$ specifies the number of channels in set C_i. A channel availability matrix, C_{Avail}, organizes the SOP list for all N SUs at C_E. As such, $C_{Avail} = \{c_{ij} | c_{ij} \in \{0, 1\}\}_{N \times M}$, where $c_{ij} = 1$ if channel j is sensed as available at SU i, and $c_{ij} = 0$ otherwise. A graph $G=(V,E)$ can illustrate the CR network, where V represents the set of nodes corresponding to CR users and E represents the set of edges between nodes. An edge, e, exists between nodes p and q if they have at least one common channel in their SOP lists, i.e., $C_p \cap C_q \neq \emptyset$. A common control channel (CCC) is assumed for exchanging messages. SUs gaining access to a channel are maintained in a set W, such

that $W \subset \{1, 2, \ldots, N\}$. Also, on sharing a channel j, a vector $H_j = (h_{j1}, h_{j2}, \ldots, h_{jN})$ builds up during the sharing process that can be defined as

$$h_{ji} = \begin{cases} 0 & \text{if } j \notin C_i \\ 1 & \text{if } j \in C_i \wedge i \in W \\ -1 & \text{if } j \in C_i \wedge i \notin W \end{cases} \quad (1)$$
$$\text{where, } i \in \{1, 2, 3, \ldots, N\}$$

Similarly for M channels, H_1, $H_2, \ldots, H_M$. To represent the channel assignments, a channel allocation matrix, A, is designed such that $A = \{a_{ij} | a_{ij} \in \{0, 1\}\}_{N \times M}$, where $a_{ij} = 1$ if channel j is allocated to SU i, and $a_{ij} = 0$ otherwise. Following the allocation strategy, $\sum_{j=1}^{M} a_{ij} \leq 1$, $\forall i \in N$ and $\sum_{i=1}^{N} a_{ij} \leq 1$, $\forall j \in M$ get imposed in A.

3.2 Sharing Mechanism

During this sequential spectrum allocation, when a SU wins a channel, the SU-channel pair does not participate for the further allocation process. C_E has a knowledge on the channels that are not being used by PUs. Initially, to gather the SOP of N SUs, C_E sends out a *READY* message over CCC. A SU i, on receiving this message, sends back a *READY REPLY* that contains C_i and N_i of the user. Accordingly, if a channel $j \in C_i$, then C_{Avail} maintained at C_E shows $c_{ij} = 1$. Thereafter, C_E starts the sharing process and sends a $SHARE_j$ message, $\forall j \in \{1, 2, ..., M\}$, in a row. If a SU i has the channel j in its SOP list, i forwards a $SHARE_j$ *REPLY* to C_E that contains a reward value, b_{ij}, representing the maximum bandwidth that SU i can acquire by using channel j. As such, a reward vector, B_j, gets formed at C_E such that $B_j = \{b_{ij}\}_{N \times 1}$, where $b_{ij} = 0$ if $j \notin C_i$ or if $j \in C_i$ and $i \in W$. This gives B_1, $B_2, \ldots, B_M$ for M channels.

On receiving $SHARE_j$ *REPLY* from SUs, C_E initializes the H_j vector such that if a SU i sends the message, then $h_{ji} = 1$, otherwise, $h_{ji} = 0$. Also, for channel j, a SU list, $S_j \subseteq \{1, 2, \ldots, N\}$, builds up where $i \in S_j$ if $j \in C_i$ and $h_{ji} = 1$. Now, $\forall i$ in S_j, $Q_i^j = (C_i - \{j\})$ is framed wherein all channels in SOP list of i are included in Q_i^j except channel j. Subsequently, all the channels in Q_i^j, $\forall i \in S_j$, are collected in C_{Choose}^j defined as

$$C_{Choose}^j = \bigcup_{\forall i \in S_j} Q_i^j; \quad i \in \{1, 2, ..., N\},\ j \in \{1, 2, ..., M\} \quad (2)$$

Next, with every channel k in C_{Choose}^j, C_E looks into the channel availability matrix and computes $X_j = \{x_k\}_{1 \times |C_{choose}^j|}$ as follows:

$$x_k = \sum_{i=1}^{N} c_{ik}; \quad \forall k \in C_{choose}^j \quad (3)$$

The x_k values in X_j are then arranged in an ascending order which also organizes the channels in C^j_{choose} list according to their x_k values. However, with the ascending ordered X_j vector, two conditions may arise $\forall x_k \in X_j$ computed using (3).

Case I: If $x_k = 1$, s.t. $\exists k \in C^j_{choose}$, then the kth column of C_{Avail} is searched to get the mth row such that $c_{mk} = 1$. This implies that channel k is present only with SU m. So, if $m \in S_j$ and $h_{jm} = 1$, change h_{jm} value to -1 in H_j specifying that channel j cannot be given to SU m. Otherwise, if j gets allotted to m, channel k shall remain unassigned. However, if $h_{jm} = -1$, it remains unchanged. Also, when $m \notin S_j$, $h_{jm} = 0$ remains same.

Case II: If $x_k > 1$, s.t. $\exists k \in C^j_{choose}$, then H_j comprising of 0, 1, -1 remains unchanged.

After going through all the x_k values in X_j, the updated H_j vector is now taken into consideration for proceeding with the allocation process. A count of the number of 1's in H_j is to be examined for deciding the allocation. Three conditions that can appear are (1) $\sum_{i=1}^{N}[h_{ji} = 1] = 1$, (2) $\sum_{i=1}^{N}[h_{ji} = 1] > 1$, and (3) $\sum_{i=1}^{N}[h_{ji} = 1] = 0$. The channel allocation strategy for this approach is described in Algorithm 1 that takes as input the updated H_j vector to decide upon the winner for channel j. Once a winner SU i is decided for j, then $h_{ji} = 1$, rest H_j values with 1 are set to -1. Also, channel allocation matrix shows $a_{ij} = 1$ and i gets entered into the winner list W.

The designed channel allocation algorithm leases the idle channels in such a way that the link utilization gets maximized while trying to award a channel to an SU requesting for higher reward, in cases possible. This improves the spectrum utility. As an example, Fig. 1 shows a Cognitive Radio network with five SUs and four vacant channels. Initially, channel 1 is shared. Let $b_{31} > b_{21} > b_{11}$. Since, there appears a single 1 value in updated H_1, channel 1 gets allocated to SU 1. However, if channel 1 goes to SU 3, channel 4 shall remain unallocated. Similarly, channel 2 goes to SU 2, channel 3 goes to SU 4 considering $b_{43} > b_{53}$, and channel 4 goes to SU 3.

4 Performance Evaluation

This section explores the performance evaluation of the proposed spectrum sharing approach under two different network scenarios: number of SUs {20, 30, 40, 50, 60} when number of channels is kept fixed at 6 and number of channels {4, 6, 8, 10, 12, 14} when number of SUs is kept fixed at 20. A network environment is deployed where users are distributed in an area of 600×600. Simulations are accomplished using MATLAB where all the results are averaged over 500 rounds. Performance of the proposed model is evaluated by comparing it with a random

Algorithm 1 Channel Allocation Algorithm

Input: H_j={(h_{j1},h_{j2},...,h_{jN}) | h_{ji}∈{0,1,-1}} to share channel j among N SUs.
Output: Winner SU allotted to channel j.
Steps:

1: Counting the number of 1's and -1's in H_j i.e. count(1)=$\sum_{i=1}^{N}[h_{ji}=1]$, count(-1)=$\sum_{i=1}^{N}[h_{ji}=-1]$.
2: If count(1) = 1
  2.1: Get the i^{th} SU s.t. ∃! h_{ji}=1, $h_{ji} \in H_j$. Winner = i.
3: If count > 1
  3.1: $\forall h_{ji}$=1, get the corresponding b_{ij} values. Create SU list and Reward list s.t. i∈ SU list and b_{ij}∈ Reward list if h_{ji}=1
  3.2: Arrange Reward list in descending order of the values and accordingly arrange SU list.
  3.3: $\forall i$∈ SU list, compute the following until winner is determined:
    3.3.1: Get $Q_i^j = (C_i - \{j\})$.
    3.3.2: If $Q_i^j = 0$. Winner = i.
    3.3.3: If $Q_i^j \neq 0$, then take $L_j = Q_i^j$ and get length(L_j). Now, $\forall k \in Q_i^j$
      3.3.3.1: Look into the k^{th} column of C_{Avail} and get the rows with value 1. $R \subseteq \{1,2,....,N\}$ s.t. a SU $q \in R$ if c_{qk}=1.
    3.3.4: $\forall q \in R$, if (SU q) = (SU i), q is not considered. Otherwise, get Q_q^j=(C_q - {j})
    3.3.5: If $\forall q \in R$, $q \neq i$ and $Q_q^j = L_j$, go through steps 3.3.5.1 to 3.3.5.2. Else, if ∃q $\in R$, q$\neq i$ and $Q_q^j \neq L_j$, go through steps 3.3.5.3 to 3.3.5.4.
      3.3.5.1: If length(L_j)=1, and $\forall q$, $s \in Q_q^j$ gives $x_s \geq 2$, $\forall s$, then Winner = i.
      3.3.5.2: If length(L_j)>1 and $\forall q$, $s \in Q_q^j$ gives $x_s \geq 2$, $\forall s$, then if x_s>length(L_j), Winner = i. Else, get next i and go through steps 3.3.1 to 3.3.5.
      3.3.5.3: If length(L_j)>1 and $\forall q$, $s \in Q_q^j$ gives $x_s \geq 2$, $\forall s$, then if $\forall q$, $q \neq i$, length(Q_q^j)=1, Winner = i. Else, goto step 3.3.5.4.
      3.3.5.4: If length(L_j)≥1, $\forall q$, $s \in Q_q^j$, x_s is computed. Take all Q_q^j as 1st level.
        3.3.5.4.1: $\forall s \in Q_i^j$, $x_s \geq$length(L_j) with atmost one x_s=2.
        3.3.5.4.2: If $\exists s \notin Q_i^j$, go through steps 3.3.3.1 and 3.3.4 for next level. If $\exists r \in R$ in next level s.t. Q_r^j is equal to some Q_q^j of previous level, no more computation. Else, 3.3.5.4.2 is repeated.
        3.3.5.4.3: If 3.3.5.4.1 is satisfied and if $\forall s \notin Q_i^j$, x_s at all levels give x_s>length(L_j), Winner = i. Else, get next i and go through steps 3.3.1 to 3.3.5.
4: If count(1) = 0
  4.1: If count(-1) = 1
    4.1.1: Get the i^{th} SU s.t. ∃! h_{ji}=-1, $h_{ji} \in H_j$. Winner = i. Set h_{ji}=1.
  4.2: If count(-1) > 1
    4.2.1: Get the i^{th} SU s.t. h_{ji}=-1 and b_{ij} = max(B_j). Winner = i. Set h_{ji}=1.

end

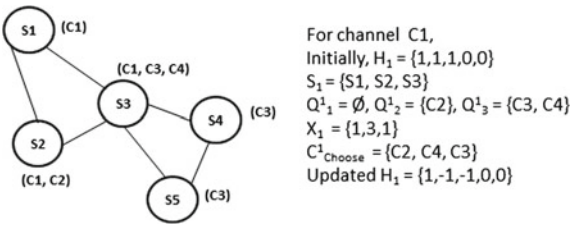


Fig. 1. An example of the proposed spectrum sharing approach

allocation where channels get randomly allocated among SUs. Performance metrics considered are spectrum utilization and throughput. Spectrum utilization is defined as the sum of allocated reward values, i.e., $S_u = \sum_{j=1}^{M} \sum_{i=1}^{N} a_{ij} b_{ij}$. And throughput earned by the allocation is given as $T_r = b_{ij} \log_2(1 + P_{T_j} \cdot \frac{P_{L_i}}{I_i + \sigma^2})$. This model assumes that the channels are homogeneous with their quality, and as such, all channels exhibit the same transmission power, $P_{T_j} = 0.01$. σ^2 is the noise variance and taken as 10^{-5} for all, PL_i is the path loss factor between SU i's transmitter and receiver, and I_i is the interference from PUs. b_{ij} is the reward value of winner i for channel j.

Figure 2 shows the spectrum utilization, S_u, earned for different sets of SUs while varying the number of channels. It is observed that on increasing the number of channels, S_u increases for all sets. With rising number of idle channels, SUs get more chance to acquire the spectrum. However, S_u for different SU sets is more or less the same with changing number of channels. Increasing number of SUs somewhat supplies a wider range of reward values, as to which a slight growth in S_u might be observed. In Fig. 3a, spectrum utilization, S_u, for

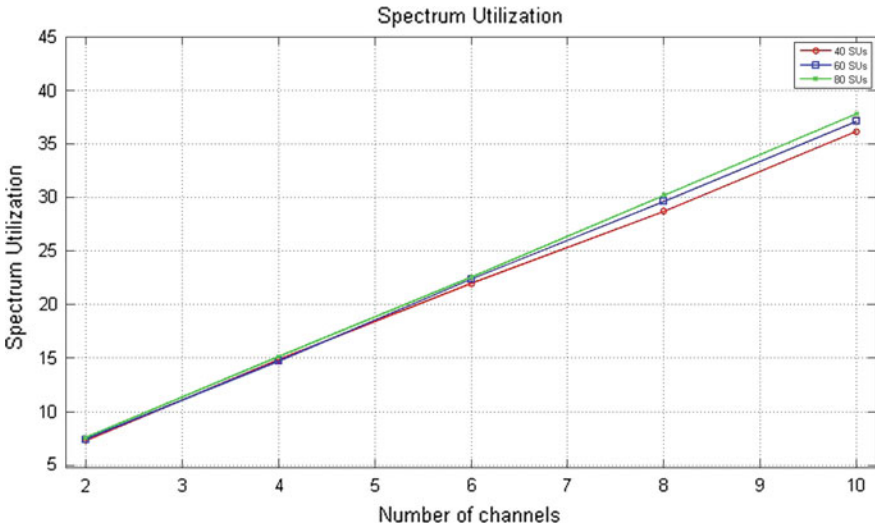


Fig. 2. Spectrum utilization for different sets of SUs with varying number of channels

proposed scheme is compared with random allocation while varying the number of SUs. S_u for the proposed model shows a better performance under the designed scenario. This is because random allocation picks any SU for channel allocation without consideration of reward values, as to which it may show a degraded result even on increasing number of SUs. However, the proposed model tries to allocate a channel to the SU who wishes to pay higher reward value. Also, there

may be situations where random allocation selects the SU with highest reward, whereas the proposed scheme does not go for that SU. This is when the proposed model finds that on assigning the channel to the SU having highest reward the remaining channels with the SU shall remain unallocated, thus reducing the spectrum usage. Figure 3b displays that the throughput, T_r, attained by the proposed sharing mechanism outperforms the random allocation with changing number of SUs. This accounts to the similar reason, wherein, during random allocation, low reward value SUs may be chosen for channel allocation, or in other case, channels may somehow remain unassigned due to improper allocation strategy. In Fig. 4a, the proposed model shows a comparison with random

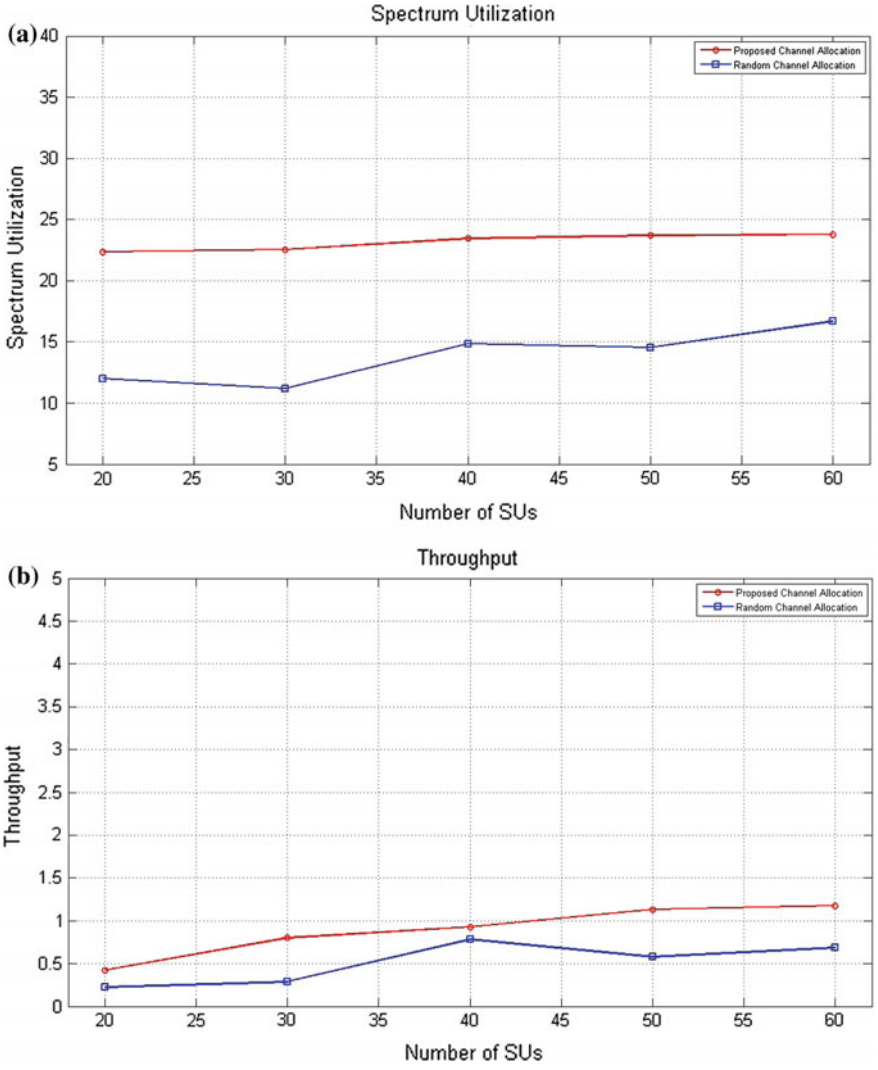


Fig. 3. **a**, **b** show spectrum utilization and throughput of proposed and random allocation with respect to number of SUs, respectively

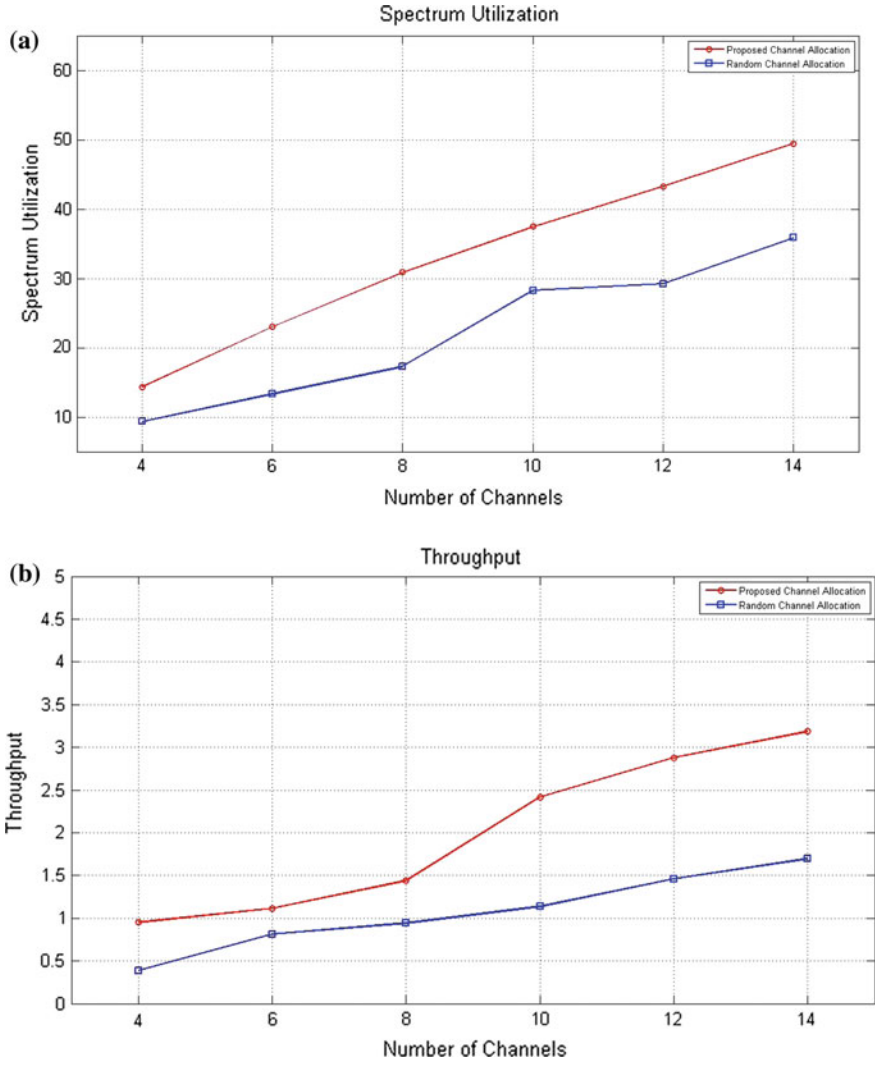


Fig. 4. a, **b** show spectrum utilization and throughput of proposed and random allocation with respect to number of channels, respectively

allocation for spectrum utilization, S_u, while altering the number of channels. It is found that the proposed scheme provides a more acceptable S_u than the random allocation. For the designed sharing approach, S_u improves with increasing number of channels, since resource availability grows among the SUs. Similarly, Fig. 4b presents a comparison for the throughput, T_r, where random allocation features a deteriorated performance on matching with the proposed approach. Such circumstances appear due to characteristics applied in the proposed model, wherein channels are allocated following some strategic policy so that link utilization can be made the most.

5 Conclusion

This paper explores a centralized spectrum sharing approach for CRN that goes along with a sequential policy and the allocation strategy to allocate the idle channels among rightful SUs so as to boost the spectrum usage. Simulation results have shown the improved spectrum utilization compared to the random strategy. Future research directions can include incorporation of heterogeneous channels in the designed model and develop concurrent policy-based spectrum sharing approach.

Acknowledgements. This research work has been supported by DST/INSPIRE Fellowship/REL1/2016/1, INSPIRE Reg No. IF150900, Govt. of India.

References

1. Ahmed E, Gani A, Abolfazli S, Yao LJ, Khan SU (2016) Channel assignment algorithms in cognitive radio networks: taxonomy, open issues and challenges. IEEE Commun Surv Tutor 18(1):795–820
2. Akyildiz IF, Lee W-Y, Vuran MC, Mohanty S (2006) NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey. Elsevier Comput Netw 50(13):2127–2159
3. Kash IA, Murthy R, Parkes DC (2014) Enabling spectrum sharing in secondary market auctions. IEEE Trans Mob Comput 13(3):556–568
4. Niyato D, Hossain E (2007) A game-theoretic approach to competitive spectrum sharing in cognitive radio networks. In: 2007 IEEE international conference on wireless communications and networking (WCNC). IEEE
5. Wu Y, Wang B, Liu KJR, Clancy TC (2008) A multi-winner cognitive spectrum auction framework with collusion-resistant mechanisms. In: 3rd IEEE symposium on new frontiers in dynamic spectrum access networks (DySPAN). IEEE
6. Zhang W, Deng L, Kiat YC (2016) Dynamic spectrum allocation for heterogeneous cognitive radio network. In: 2016 IEEE international conference on wireless communications and networking (WCNC). IEEE
7. Zhou X, Zheng H (2009) TRUST: a general framework for truthful double spectrum auctions. In: IEEE INFOCOM 2009. IEEE, pp 999–1007

Mixing Test Set Generation for Bridging and Stuck-at Faults in Reversible Circuit

Mousum Handique(✉) and Joinal Ahmed

TSSOT, Department of Computer Science and Engineering, Assam University, Silchar 788011, Assam, India
mousum78@yahoo.co.in, joinalahmed@gmail.com

1 Introduction

The reversible circuit technologies are more promising future alternative as compared to conventional circuit technologies in the scenario of high-performance computation. Rolf Landauer [1], 1961, showed that whenever using a logically irreversible operation, it dissipates energy into the environment. The reversible logic operations are those operations which can reuse a fraction of the signal energy that theoretically can approach near to 100%. Therefore, the reversible logic circuit is the most popular technology to achieve this performance. The properties of the reversible circuit are simpler than a conventional circuit. The basic properties of reversible circuits are (a) number of inputs are equal to the number of outputs, (b) only bijection operation is allowed, and (c) no fanout and feedback connection is allowed.

Due to the reversibility property, the efficient test pattern generation of a reversible circuit is relatively simpler than the traditional logic circuit. The other facet of the reversible circuit is that it has produced unique output vector from each corresponding input vector and vice versa, which gives high controllability and observability [2]. A test set is a collection of the input test vectors which are applied to a reversible circuit to detect and observe that the faults are occurred in the circuit. The test set is called as complete test set if it is capable of detecting all the faults in a given circuit. The relation between different fault models of the reversible circuit has been discussed in [3]. This paper has drawn parallels between the bridging and stuck-at faults for generating the test patterns. The test vectors that set the two lines by the opposite logic values "01" and "10" are used to detect the two lines of bridging faults at the same level [4]. In other way, for detecting the stuck-at faults at any level, the lines at every level are set by 0 and 1. Generated complete test set for individual fault model is already in the

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_11

literature but "generated test patterns for one particular fault model is derived such that it is capable of detecting another fault model" is our key concern. Based on this concept, we have generated the test patterns for the input bridging faults and further reconstructed the test vectors for the input stuck-at faults which are obtained from bridging fault model.

The rest of the paper is organized as follows: Sect. 2 provides some basic background on reversible logic circuits and an overview of stuck-at and bridging fault model in the reversible circuit. Sect. 3 describes our proposed method for generating the test set for detecting input bridging and stuck-at faults. The experimental result and conclusion are provided by Sect. 4 and Sect. 5, respectively.

2 Background

2.1 Reversible Logic Circuits

A reversible logic circuit is used to implement the reversible computation and it is formalized in terms of gate-level circuits. The reversible circuit structures are cascade structure [5]. All the operation of the reversible circuit has to be performed in a reverse form. The reversible circuit allows only bijective operations [2] and maintained the bijective operation; the circuit does not include any concept of fan-out and feedback connection [6]. In reversible circuit design, several gates have proposed over the past decades. They are the controlled NOT (CNOT) proposed by Feynman [7], Toffoli [8], Fredkin and Toffoli [9], etc. In this paper, we are using only NCT library that contains NOT, CNOT, Toffoli gate, and GT library containing generalized (n-bit) Toffoli gate. This NCT library was introduced by Toffoli [8] in 1980.

2.2 Fault Models in Reversible Circuit

A fault model has described the different levels of abstraction of physical faults in a system. These levels of abstraction can be defined as behavioral, functional, structural, and geometric [4]. A fault model is a mathematical model which represents various faults possibilities and it helps to generate the tests for detecting and reducing all the possible faults in a given circuit [3]. Numerous fault models have been introduced in the reversible circuit. In this work, we have considered stuck-at fault and bridging fault model in the reversible circuit.

1. **Bridging Faults in a Reversible Circuit**: The bridging faults occur when two signals are connected together but they should not be. If two or more lines

involved in bridging faults, then logical effects of this faults are categorized by wired-OR or wired-AND bridging faults [4]. The single-input bridging faults occur, if two input lines are operated by the opposite logic values [10]. Sarkar and Chakrabarti [11] proved that if any test set is complete for the single-input bridging faults, then that particular test is also completed for detecting the multiple-input bridging faults. The authors also showed that the generated test sets are capable of detecting all the possible single- and multiple-input bridging faults if generated test sets are eligible for detecting all the single-input stuck-at faults in a given reversible circuit.

2. **Stuck-at Faults in a Reversible Circuit**: The stuck-at faults occur in a circuit when one of its inputs and outputs to be fixed at either logic value 0 (stuck-at-0) or logic value 1 (stuck-at-1) without considering the input value. This is a very common fault model used for irreversible circuits. If the stuck-at fault is involved only for one line in the circuit, then it is called as the single stuck-at fault (SSF), and if stuck-at fault is involved more than one line, then it is called as multiple stuck-at faults (MSF). Patel et al. [2] stated that any test set is complete for detecting the stuck-at faults, if and only if each line at every level can be set to both 0 and 1 by the test sets. The paper [2] also showed that the test set is complete for all single stuck-at faults, and then it also becomes complete test set for all multiple stuck-at faults.

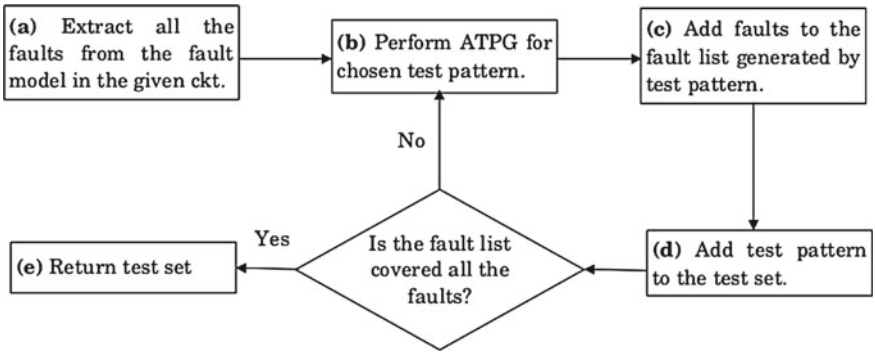


Fig. 1. ATPG and FDL flow

3 Proposed Method

We have divided into two modules in our method. The first module described the test pattern generation of single-input bridging faults. After generating the test patterns, we have derived these patterns for further use of single-input stuck-at faults. Also, we have introduced fault description list (FDL) in our work. The detailed discussion has been given below.

3.1 Fault Description List

Fault description list contains all the possible faults which are generated by the test vectors based on the position of the binary bit sequence present in that test vector. Using the fault description table, we have checked that which test vector is covered by a maximum number of faults and combined the minimum number of test vectors such that it covers all the possible faults in the given reversible circuit.

The construction of ATPG and fault description list (FDL) flow is shown in Fig. 1. In Fig. 1, step (a) extracts all the possible single-input bridging and stuck-at faults in the given reversible circuit. Step (b) generated all the possible test pattern from our proposed method (ATPG). All the faults are stored in the fault list which is generated by test pattern mentioned in step (c). In step (d), each individual test pattern of corresponding faults adds to the test set. Now, we have checked that fault list is covered all the faults, if not then take another combination of test patterns and continue the same process. If some selected test set is able to cover all the faults, then that test set is the final one for detecting all the faults; it has mentioned in step (e) and also this test set is minimized form because the combination of test vectors is growing in increasing order.

3.2 Test Set Generation Algorithm for Single-Input Bridging Fault

In this section, we have arranged our proposed method into two algorithms. Algorithm 1 has explained initially how to extract all the test vectors (TV) based on the bridging faults property. To obtain the binary bit sequences, we are starting from n = 3, where n is the number of input lines of the reversible circuit. In Algorithm 2, the FDL has ensured that which test vector(s) is capable of detecting all single-input bridging faults $(l_1, l_2), (l_1, l_3) \ldots (l_i, l_j)$, where $i < j \leq n$

and $(l_i, l_j) \in F_{B_i}$ in the given circuit. Finally, select the minimized fault detecting test set T_B of N number of input lines in a reversible circuit.

Algorithm 1: Extracting the TV from bridging fault property

Input: S is the set of all binary sequences of 3-input lines.
Output: T is minimized form of test vectors (TV).
Step 1: List out all the opposite logic value of two-bit binary sequence for n=3 from set S and P is the set of opposite logic values.
Step 2: Assign 'X' at missing bit position in P and replaced X by 0 and 1.
Step 3: P with duplicate terms then removes until no duplicate terms are exist and then remove complement terms at P.
Step 4: P is the set of final minimized binary sequences and T← P.
Step 5: return T.

Algorithm 2: Complete test set generation for detecting single input bridging faults

Input: A set of test vectors T from Algorithm 1.
Output: Final minimized fault detecting test set T_B using FDL.
Step 1: Compute T_B, initially empty.
Step 2: All the faults F_{B_1},F_{B_2}, F_{B_3} ..... F_{B_i} are assigned to F_B in a given n-input reversible circuit.
Step 3: Each test vector of T is identified single input bridging faults F_B using the binary bit position.
Step 4: FDL stores the information of each fault identified by TV in T.
Step 5: Choose the increasing order combination of test vectors in T.
Step 6: If T is cover all the faults F_B using FDL then $TS_B \leftarrow T$ and **return $\mathbf{T_B}$**, otherwise continue from Step 5 until covers all the faults.
Step 7: Update the value of n = 4, 5, 6, ...
Step 8: Appending the "0" or "1" at the LSB of each test vector in T_B.
Step 9: T ← current T_B for n-input lines.
Step 10: Continue the same process from Step 2.

Lemma 1. *Detecting single- and multiple-input bridging faults in a given n-input reversible circuit, the* $(\lceil n/2 \rceil)$ *number of test vectors are sufficient.*

Proof. Here, each of the input lines can be set to both opposite logic values 0 and 1. The test vector TV_i=$(\langle b_{0i}\ b_{1i}\ b_{2i}\ \ldots\ b_{(n-1)i} \rangle)$, where $1 \leq i \leq \lceil n/2 \rceil$, $b_{(n-1)i}$ is the $(n-1)$th bit of ith test vector and $b_{(n-1)i} \in \{0, 1\}$. As explained in Algorithm 1, the test set T=$\{TV_1, TV_2, TV_3\}$ is the test set for n = 3-input lines

(initially) in the given reversible circuit, where no duplicate and complement test vectors are present. The total single-input bridging faults are C(n, 2) = 3 in the three-input reversible circuit. In the binary bit position for three-input lines, any one of the test vectors is capable of detecting the number of faults $C(n, 2)/2$ or more than $C(n, 2)/2$ of the total faults but only one test vector is unable to detect all the faults. Because at least one logic value for each test vector is similar to the other logic value, to maintain the opposite logic value for each test vector, the test set T_B has exactly produced $\lceil n/2 \rceil$ test vectors. Therefore, according to Algorithm 2, we formed the test set $T_B = \{TV_1, TV_2\}$ that is capable of detecting all the faults in the given three-input reversible circuit. If we go for (n + 1), i.e., four-input lines in the reversible circuit, then adding extra input line is created new faults $(l_1, l_4), (l_2, l_4), (l_3, l_4)$. For detecting new faults, we are adding 0 or 1 to the least significant bit (LSB). We have only checked the new form of faults because existing faults are automatically covered by the existing test vector (for n = 3). Hence, $\lceil n/2 \rceil$ test vectors are required to detect all the single-input bridging faults.

3.3 Test Set Generation Algorithm for Single-Input Stuck-at Fault

In this section, we have introduced our algorithm for detecting single-input stuck-at faults. Algorithm 3 extracted the test set T_B which is derived from the single-input bridging fault model and generated the final test set T_S for detecting the single-input stuck-at faults.

Algorithm 3: Test set generation for detecting single input stuck-at faults

Input: T_B is the current test set obtained from Algorithm 2.
Output: Test set T_S for detecting single input stuck-at faults.
Step 1: Compute test set T_S, initially $T_S = T_B$.
Step 2: All faults F_{S_1}, F_{S_2}, $F_{S_3}, \ldots, F_{S_i}$ of the given circuit assign to F_S.
Step 3: FDL stores the information of each fault identified by the T_S.
Step 4: If T_S is covered all the faults of F_S from FDL then **return $\mathbf{T_S}$** else goto next Step.
Step 5: Complement of any one test vector of T_S and update T_S with adding new complement test vector and goto Step 3.

Lemma 2. *Detecting all the single- and multiple-input stuck-at faults in a given n-input reversible circuit, the* $(\lceil n/2 \rceil) + 1$ *number of test vectors are sufficient enough.*

Proof. In Lemma 1, it has proved that test set $T_B = \{TV_1, TV_2\}$ is capable of detecting all the single-input bridging faults in three-input reversible circuit. Any combination of test vectors (excluding complement of test vector), at least one-bit position, is occurred with same logic value. Due to this fact, it is unable to cover either one (or more) stuck-at-0 or stuck-at-1 fault(s). If we are introducing new test vector which is the complement of any test vector (e.g., TV_1 or TV_2),

then that bit position of complement test vectors must occur at least one opposite logic value. Therefore, extra one more test vector is needed for detecting the single-input stuck-at faults in our proposed method, i.e., the test has required exactly $(\lceil n/2 \rceil) + 1$ number of test vectors.

Example 1. Consider the reversible benchmark circuit *ham*3 consisting of three-input and three-output lines as shown in Fig. 2a. The number of single-input bridging faults is $C(3, 2) = 3$. Algorithm 1 generated the minimized test vectors in $T = \{001, 010, 011\}$. Then, using Algorithm 2, the final minimized test set $T_B = \{001, 010\}$ with the help of FDL, which is the complete test set in a *ham*3 benchmark circuit for detecting the input bridging faults. In this circuit, the total numbers of single-input stuck-at faults are $3 \times 2 = 6$. It has observed that the test set T_B is not capable of capturing all the single-input stuck-at faults. Hence, we apply Algorithm 3 and generated the complete test set $T_S = \{001, 010, 110\}$, where "110" is the complement form of "001". The test set T_S covered all the possible single-input faults in *ham*3 benchmark circuit.

Example 2. In this example, we are considering *mpsk*_4_49_13 reversible benchmark circuit consisting of four-input and four-output lines as shown in Fig. 2b. The number of input bridging faults is $C(4, 2) = 6$. The Step 9 in Algorithm 2 generated the minimized test set $T_B = \{0011, 0101\}$ with the help of FDL, which is the complete test set for detecting the single-input bridging faults in a *mpsk*_4_49_13 benchmark circuit. Now if we consider the single-input stuck-at faults of this circuit, then a total number of single-input stuck-at faults occur $4 \times 2 = 8$. Similarly, the test set T_B is not capable to detect all the input stuck-at faults. Therefore, Algorithm 3 is generated the test set $T_S = \{0011, 0101, 1100\}$, where "1100" is the complement form of "0011". The test set T_S is the complete test set for detecting single-input stuck-at faults in *mpsk*_4_49_13 benchmark circuit.

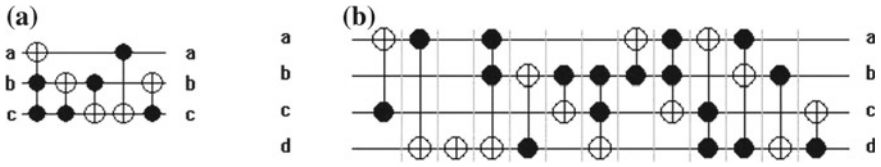


Fig. 2. **a** *ham*3 benchmark circuit **b** *mpsk*_4_49_13 benchmark circuit

4 Experimental Results

The proposed algorithms are generating the mixing test vectors applied to various reversible benchmark circuits [12] with NCT and GT models. The test sets are generated by our proposed method that is compared with the existing method [11] which is shown in Table 1. It has observed that the proposed method gives good performance in both the fault models and also covered 100% faults.

Table 1. Detection of input faults (BF = input bridging faults and SAF = input stuck-at faults)

Benchmark circuits	Gate model	No. of input/output	No. of input stuck- at faults	No. of input bridging faults	No. of test vectors [11]	No. of test vectors [Proposed]	% fault coverage [Proposed]
					BF+SAF	BF+SAF	BF+SAF
3_17	NCT	3/3	6	8	3	3	100
6sym	NCT	6/1	12	90	6	4	100
9sym	NCT	9/1	18	352	9	6	100
hwb7	GT	7/7	14	152	7	5	100
hwb8	GT	8/8	16	238	8	5	100
hwb6	GT	6/6	12	90	6	4	100
rd73	NCT	7/3	14	152	7	5	100
rd84	NCT	8/4	16	238	8	5	100
ham7	GT	7/7	14	152	7	5	100
ham15	GT	15/15	30	1848	15	9	100
mod1024 Adder	GT	20/20	40	4598	20	11	100
mod10485 76 adder	GT	40/40	80	21222	40	21	100

5 Conclusion

This paper observed that there is a close relation between stuck-at and bridging faults in the reversible circuit. This paper concludes that ($\lceil n/2 \rceil$) test vectors are generated at first for detecting the input bridging faults and reconstructed the test vectors from the bridging fault model such that adding only one test vector is sufficient to detect input stuck-at faults in the n-input reversible circuit. There will be a possibility to design a technique such that similar type of test vectors can be detected which may be some other fault models. This may lead to the future work.

References

1. Landauer R (1961) Irreversibility and heat generation in the computing process. IBM J Res Dev 5(8):183–191
2. Patel KN, Hayes JP, Markov IL (2004) Fault testing for reversible circuits. IEEE Trans Comput Aided Design Integ Circuits Syst 23(8):1220–1230
3. Rice J (2013) An overview of fault models and testing approaches for reversible logic. In: 2013 IEEE pacific rim conference on communications, computers and signal processing (PACRIM). IEEE, pp 125–130
4. Jha NK, Gupta S (2003) Testing of digital systems. Cambridge University Press
5. Maslov D (2003) Reversible logic synthesis. PhD Dissertation, University of New Brunswick
6. Nielson MA, Chuang IL (2000) Quantum computation and quantum information. Monograph Collection (Matt-Pseudo)
7. Feynman RP (1986) Quantum mechanical computers. Found Phys 16(6):507–531
8. Toffoli T (1980) Reversible computing. Springer
9. Fredkin E, Toffoli T (2002) Conservative logic. Springer
10. Rahaman H, Kole DK, Das DK, Bhattacharya BB (2007) Optimum test set for bridging fault detection in reversible circuits. In: 16th Asian test symposium, ATS07. IEEE, pp 125–128
11. Sarkar P, Chakrabarti S (2008) Universal test set for bridging fault detection in reversible circuit. In: 3rd international design and test workshop, IDT 2008. IEEE, pp 51–56
12. Maslov D, Dueck G, Scott N (2005) Reversible logic synthesis benchmarks page

A Comparative Study of Biopotentials Acquired from Left and Right Hands of Human Subjects

Arindam Sarkar(✉), Aditi Bhattacharya, Ratna Ghosh, and Bhaswati Goswami

Instrumentation and Electronics Engineering, Jadavpur University, Kolkata, West Bengal, India
arindamsarkar86@gmail.com, aditi.bhattacharya08@gmail.com, ratna.ghosh@jadavpuruniversity.in, bhaswati.goswami@gmail.com

1 Introduction

The human system is known to exhibit morphological as well as functional bilaterality with respect to the sagittal plane [1]. However, exact measurements taken from a great number of cadavers and living subjects have established that the two halves of the human body are not exactly symmetrical.

Researchers have studied the bilateral symmetry for health-related characteristics [2], specially to investigate human performance [3] and also human infirmities [4]. Indices like BSI [5] and BBSI [6] have also been proposed to study the bilateral symmetry of brain signals.

One characteristic feature of the brain cells, as well as all living cells, is the control of the flow of specific electrically charged ions across the cell membrane [7]. In this sense, it is well established that biological processes should be viewed as electrodynamic processes, as in the cases of ECG, EEG, and EMG, to name a few [7].

The active or passive electrodynamic phenomena that can be observed in or on the skin is termed as electrodermal activity (EDA) [8]. The endosomatic-type EDA measurement technique [8] has been adopted for the present study. In this technique, no external current is used for stimulation. Instead, only potentials originating from the skin surface are recorded. The signals have been acquired from identical locations on the left and right hands of the subjects. Thereafter, the qualitative as well as statistical characterization of these signals has been performed in order to compare and contrast these signals and to determine whether any bilateral relationship exists within them.

The paper has been organized as follows. In Sect. 2, the scheme of recording the biopotential signals has been described. In Sect. 3, the data analysis has been presented and a new parameter has been defined. The conclusions have been stated in Sect. 4.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_12

2 Acquisition of Biopotentials

The schematic of the instrumentation system used to acquire the biopotentials has been shown in Fig. 1.

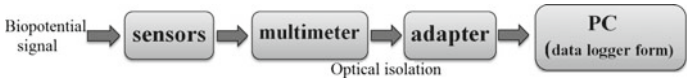


Fig. 1. Schematic diagram

The sensor used for acquiring the biopotential from each hand comprises a pair of silver (Ag) electrodes, which have been connected to the intermediate phalanges of the middle finger and the forefinger. So, the biopotential acquired from any one hand is the differential voltage recorded from this pair of electrodes. As the fingers are situated far away from the major biopotential generating organs like heart, brain, etc., it is expected that these acquired biopotential signals would not reflect the information of any specific organ.

The data was recorded in the form of dc voltage signals from the two hands simultaneously for a time span of 2 min. Two digital multimeters (Make: Rishabh Instruments, model-RISH Multi 18S) [9] were used for simultaneously recording the biopotential data from both hands. Two adapters (RISH Multi SI232) [10] were attached to the RISH Multi 18S multimeters for direct online transmission of the measured data to PC in data logger format using Rishcom 100 software.

These multimeters have a resolution of 10 μV, which is known to be the typical minimum value of biopotentials [11]. Since the typical bandwidth of the EDA signals is known to be 0–3 Hz [8]; hence, the sampling frequency has been set to 20 samples/s in this study. It is known that more than $5mA$ of current is harmful for human beings [12]. The inbuilt optical isolation between the multimeter and the adapter helps to avoid any such electrical hazard to the human body. These features justify the use of current acquisition system for the present study.

The recorded data is directly acquired into the PC without any filter or filtering technique. Systematic static calibrations have been performed with a digital millivolt calibrator (Make: Libratherm, Model-LC-05), followed by dynamic calibrations using a function generator (Make: Agilent, Model: 33210A LXI). These ensure that the recordings are noise free, while keeping all the biological information intact.

A total number of 426 sets of biopotentials were recorded from 23 human subjects, who were between 23 and 58 years and from both genders, over a period of 7 months. The recording was done as follows. Each set of data was recorded while the subjects were in lying condition. For each subject, one set of data was recorded in a day and this procedure was repeated at regular intervals over a number of days. It is to be noted that in all the cases, the subject was allowed to settle down for about a minute in calm, temperature-controlled environment

before recording the data. The lab conditions were maintained almost the same as far as possible for all the recordings. All the subjects chosen were generally healthy.

3 Analysis of Acquired Potentials

3.1 Raw Biopotential Signals

Biopotential data recorded from left hand and right hand are henceforth denoted as LH and RH biopotentials, respectively. It is observed that no two simultaneous recordings are identical or exactly symmetrical to each other, although they are quite similar in most cases. This characteristic has been observed for all subjects, and also for the data recorded from the same subject on different days.

In order to study the bilateral relationship of these dynamically changing human biopotentials, the simultaneous records of both hands have been considered as a pair of signals, as shown in Fig. 2. It is observed that all the patterns possible for such a pair of signals exist in the various recorded data, namely converging, diverging, crossover, and parallel. The basis for this classification is the difference, or gap, between the LH and RH data at the start and end points of the set. If the gap decreases over time, the signals are classified as converging, while the reverse scenario is classified as diverging. If the gap remains within ± 4 mV for the total duration, then these signals are considered as parallel. In the crossover pattern, as the name suggests, the gap between the LH and RH biopotentials initially reduces and then increases in the reverse sense.

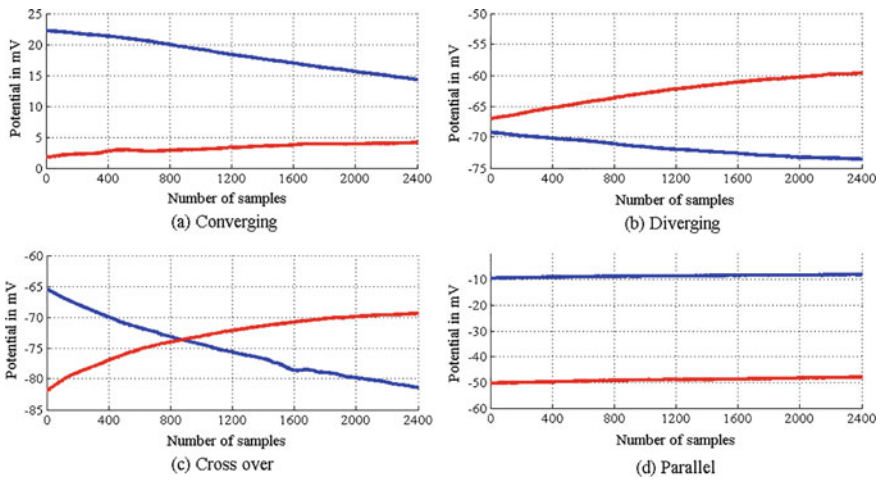


Fig. 2. Patterns observed for the pair of LH and RH biopotentials

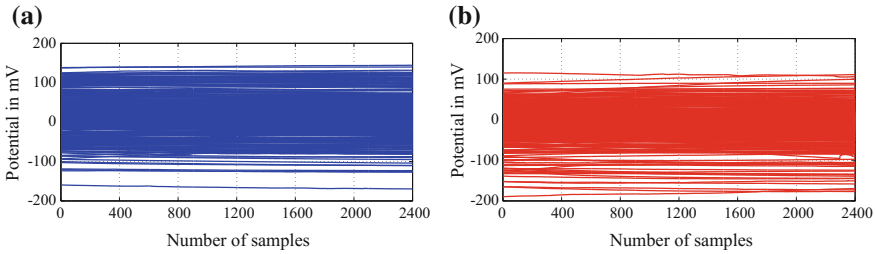


Fig. 3. Time plots of 426 sets of **a** LH and **b** RH biopotentials

It is observed that the converging patterns (38.26%) are the most prevalent, followed by the parallel (33.80%) and then the diverging (22.77%) patterns. Crossover is detected only in a few specific cases (5.16%). A further analysis shows that the individual signals may be increasing, decreasing, or constant. However, in the plot of all the 426 sets of data recorded for the LH and RH biopotentials shown in Fig. 3, the signals appear to be parallel. This is due to the fact that the setwise mean values of the individual signals vary over a much larger range, from −165.59 to 141.50 mV for LH biopotentials and from −180.58 mV to 111.06 mV for RH biopotentials, while the corresponding standard deviations of the individual signals are much smaller, typically within ±0.03 mV to ±9.06 mV for LH biopotentials and ±0.02 to ±13.64 mV for RH biopotentials.

Histograms of the overall LH and RH biopotentials for bin widths of 20 mV are shown in Fig. 4, while their standard statistical parameters have been tabulated in Table 1. It is observed that the distributions of both LH and RH biopotentials are unimodal, with slight left-skewed natures.

However, the ranges, mean (μ), and standard deviation (σ) are different for LH and RH biopotentials. For an almost normal distribution, the skewness lies within ±0.5 and the kurtosis value is very close to 3. It is observed that while the skewness of the LH biopotentials lies within the range, the corresponding value of the RH biopotentials lies beyond the range. Furthermore, the kurtosis of RH is higher than that of LH. Since these are unimodal distributions, it

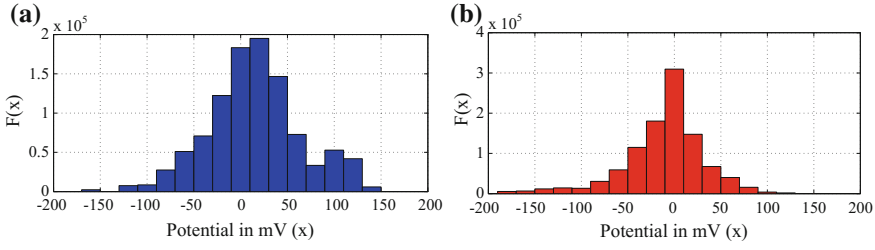


Fig. 4. Histograms of **a** LH and **b** RH biopotentials

Table 1. Standard statistical parameters of the acquired biopotentials

Parameter	Range in mV		μ in mV	σ in mV	Skewness	Kurtosis
	Min	Max				
LH	−169.68	144.29	14.89	50.01	−0.02	3.34
RH	−189.64	115.31	−10.87	43.11	−0.87	5.12

can be inferred that in comparison to the LH biopotential distribution, the RH biopotential distribution exhibits an increased peakedness with an associated long tail.

3.2 Bias of the Biopotential Signals

The previous observations indicate that the nominal gap between the LH and RH biopotentials can be used to characterize their bilateral relationship, since the signal deviations are significantly smaller. In order to do so, the bias, or central tendency, of an individual data set has been considered to be the mean, in this case, the time average, of the set.

Plots of the cumulative distribution function (cdf) of the mean values of the LH and RH biopotentials are shown in Fig. 5b. It is observed that these plots are almost identical to the cdf plots of the raw LH and RH biopotential signals, as shown in Fig. 5a. So, in this case, the bias of the signals, hereafter denoted as μ_{LH} and μ_{RH}, respectively, can be considered to be representative of the actual signals. Chi-square goodness-of-fit tests have been performed with 5 and 10% significance levels to check the normality of the signal mean values. The test results, stated in Table 2, show that neither μ_{LH} nor μ_{RH} is distributed normally.

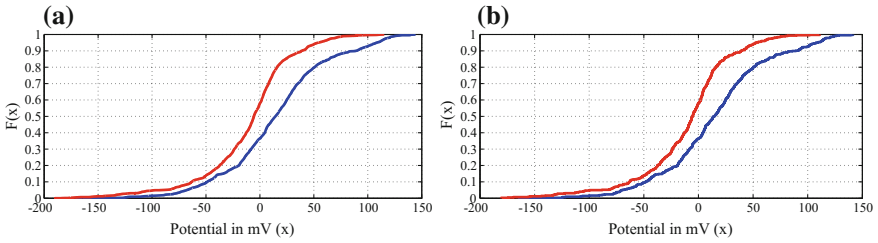


Fig. 5. Cumulative distribution function (cdf) plots of **a** LH and RH biopotentials and **b** mean of LH and RH biopotentials

Table 2. Chi-square goodness-of-fit tests for μ_{LH} and μ_{RH}

Hypothesis	H_0: μ_{LH} is normally distributed		H_0: μ_{RH} is normally distributed	
	H_A: μ_{LH} is not normally distributed		H_A: μ_{RH} is not normally distributed	
df = 15	$\chi^2_{0.10}$	$\chi^2_{0.05}$	$\chi^2_{0.10}$	$\chi^2_{0.05}$
Theoretical	22.307	24.996	22.307	24.996
Experimental	39.704	39.704	85.465	85.465
Conclusion	H_0 rejected	H_0 rejected	H_0 rejected	H_0 rejected

Table 3. Standard statistical parameters of μ_{diff}

Parameters	Range in mV		μ in mV	σ in mV	Skewness	Kurtosis
	Min	Max				
μ_{diff}	−189.05	284.37	25.76	54.58	0.16	4.15

3.3 Differential Bias (μ_{diff})

The nominal gap between the LH and RH biopotentials has been quantified using the difference between their mean values. This is being proposed as a new parameter, which is termed as differential bias (μ_{diff}). Thus,

$$\mu_{diff} = \mu_{LH} - \mu_{RH} \tag{1}$$

In order to characterize the differential bias, its statistical parameters have been tabulated in Table 3. The corresponding cdf plot and quantile–quantile plot (QQ plot) are shown in Fig. 6, while the chi-square test results for normality are tabulated in Table 4.

Table 4. Chi-square test for μ_{diff}

Hypothesis	H_0: μ_{diff} is normally distributed			
	H_A: μ_{diff} is not normally distributed			
df = 11	$\chi^2_{0.975}$	$\chi^2_{0.95}$	$\chi^2_{0.05}$	$\chi^2_{0.025}$
Theoretical	3.816	4.575	19.675	21.920
Experimental	3.639	3.639	3.639	3.639
Conclusion	H_A rejected	H_A rejected	H_A rejected	H_A rejected

As shown in Fig. 6a, the nature of cdf of μ_{diff} shows a normal-like distribution. From this figure and the chi-square test results, it can be seen that the data lies within the 97.5% confidence limit for the total range. It is to be noted that in comparison with the mean and median of the parameters μ_{LH} (14.89 mV and 18.29 mV, respectively) and μ_{RH} (−10.87 mV and −5.52 mV, respectively), the mean (25.76 mV) and the median (23.20 mV) for μ_{diff} are very close to each

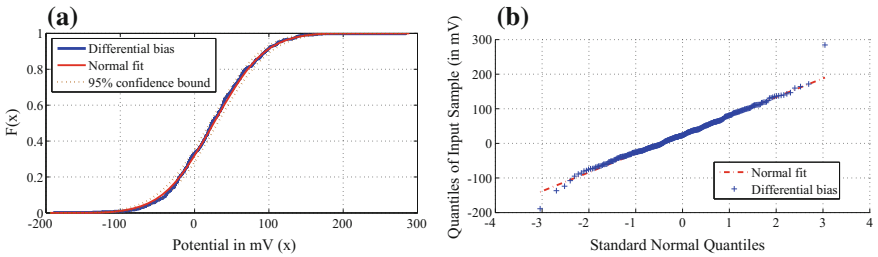


Fig. 6. **a** Cumulative distribution function (cdf) and **b** quantile–quantile (QQ) plots of differential bias

other. The QQ plot also shows that till the second quantile, the data almost follows the normal probability fit line. Therefore, it can be said that the parameter μ_{diff} is a valid statistical parameter, which varies within 25.76 ± 109.16 mV for 95% of the generally healthy population considered in this study. Hence, it can be said that in a healthy person, the dynamically changing LH and RH biopotentials will not be identical. This is typically characterized by a nonzero differential bias.

4 Conclusions

An endosomatic EDA-based technique has been adopted to acquire biopotential signals from both hands of 23 human subjects. Total 426 sets of data from each hand, which have been acquired over a period of 7 months, have been analyzed to determine their similarity and dissimilarity.

It has been observed that the pairs of signals exhibit all four characteristic patterns possible, namely converging, diverging, crossover, and parallel. Of these, the converging pattern is observed to be most prevalent, followed closely by the parallel and then the diverging patterns, while crossover patterns occur rarely. However, since individual signals do not have large deviations, these patterns can be ascribed to differences in the bias of the underlying signals.

Using various statistical tools, it has been established that these raw biopotentials have unimodal distributions that are left skewed and more peaked than an almost normal distribution. Thereafter, it has been established that neither the signals nor their mean values, which can be considered to be representative of the signals, are distributed normally.

A new parameter, termed as differential bias (μ_{diff}), has been proposed in this work to characterize the nominal gap between the pairs of signals in terms of the difference between their mean values. It has been found that μ_{diff} follows a normal distribution with 97.5% confidence bound. Chi-square test performed on this data confirms the same and the corresponding QQ plot shows linearity for the ± 2 quantile.

In view of these findings, it can be said that these LH and RH biopotentials are related to each other in terms of the differential bias, which is typically

nonzero for a healthy person. However, it remains to be investigated how different infirmities cause a change in this parameter. Various filtering techniques might be used and/or developed for this purpose. Since the electrodynamics of the human system is a complex phenomenon, hence, more such parameters can possibly be determined in order to characterize the system dynamics further. Investigations are being done in order to explore both these aspects.

5 Declaration

An informed consent has been signed by the participants. The study is totally noninvasive, passive, and optically isolated. Therefore, there is no known risk associated with the research. Confidentiality of subjects has been made as per protocol.

Acknowledgements. The first author is very thankful to CSIR for funding this research and to Jadavpur University for providing the research facilities.

References

1. Behnke RS (2012) Kinetic anatomy with web resource. Human Kinetics
2. Tomkinson GR, Olds T (2000) Physiological correlates of bilateral symmetry in humans, vol. 21, no. 08, Int J Sports Med, Georg Thieme Verlag, Stuttgart, New York, pp 545–550
3. Trivers R, Fink B, Russell M, McCarty K, James B, Palestis BG (2014) Lower body symmetry and running performance in elite jamaican track and field athletes, vol. 9, no. 11. Public Library of Science, p e113106
4. Serdyuk V (2013) Scoliosis and Spinal Pain Syndrome: New understanding of their origin and ways of successful treatment, 1st ed., vol 12. Byword Books
5. van Putten M.J, Peters JM, Mulder SM, de Haas JA, Bruijninckx CM, Tavy D (2004) A brain symmetry index (BSI) for online eeg monitoring in carotid endarterectomy, vol 115, no 5. Clin neurophysiol, Elsevier, pp 1189–1194
6. Yan M et al. (2011) A bilateral brain symmetry index for analysis of eeg signal in stroke patients. In: 4th International Conference on Biomedical Engineering and Informatics (BMEI), Shanghai, Oct 2011, pp 8–11
7. Cromwell L, Weibell FJ, Pfeiffer EA (1980) Biomedical instrumentation and measurements. Prentice Hall
8. Boucsein W (2012) Electrodermal activity. Springer Science & Business Media
9. RISH Multi 18S datasheet. http://www.rishabh.co.in/index.php/products/product/118
10. RISH Multi SI232 datasheet. http://www.rishabh.co.in/index.php/products/product/121
11. Webster JG, Eren H (2014) Measurement, instrumentation, and sensors handbook: spatial, mechanical, thermal, and radiation measurement. CRC press, pp 2045
12. Lipman EA (2007) Electrical safety information. http://web.physics.ucsb.edu/phys13CH/electrical_safety.pdf

Real-Time Bottle Detection Using Histogram of Oriented Gradients

Mahesh Jangid(✉), Sumit Srivastava, and Vivek Kumar Verma

SCIT, Manipal University Jaipur, Jaipur, India
mahesh_seelak@yahoo.co.in

1 Introduction

The object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class in digital images and videos. The involvement of the computer vision is rapidly increased around us and also used for many purposes like pedestrian detection [3–5], vehicle detection [6], traffic signal recognition [7], fire detection [8], etc. The computer vision technology [9] is also being used with the robots to sense the environment and perform tasks accordingly. We are working on to develop a robot to serve a water bottle at the desk of person in the office to minimize the human affords that is basically object detection problem [10]. The bottle is to be detected and classified from a video feed from a nonstationary camera mounted on the top of the robot.

Owing to its age, this problem has a lot of literature published. Two of the most important approaches include using HAAR wavelet descriptors [11, 12] as input parameters or using part-based method containing detectors for various objects. An extension of this algorithm can be seen as the skeleton modeling done by Kinect using RGBD images. The approach we propose to implement is much simpler than the abovementioned methods and is proven to provide significantly higher performance in the real world.

The basic idea behind this approach is capturing the object appearance and shape by characterizing it using local intensity gradients and edge directions. The image is densely divided into small special regions called cells (Fig. 1).

For each cell, a 1-D histogram of gradient directions/edge directions is computed and later all cell data is combined to give a complete HOG descriptor of the window. The variety of colors and illumination in the surrounding make normalization inevitable. We further describe the normalization technique as a part of our approach later in the report. In their work, we make our own dataset for training purpose, which has a sufficiently large negative set by sampling out patches from bottle-free images. Figure 2 shows the direction information in each cell and next picture shows the feature descriptor information per block.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_13

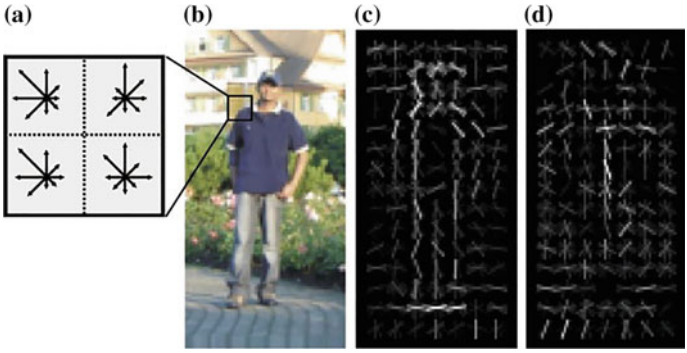


Fig. 1. Histogram of oriented gradients

2 Problem Statement

We aim to implement an object detection system for detecting and marking one or more bottles in a scene. This project is to serve the purpose of bottle detection and classification in a video feed captured by a robot in the office to serve the needs of person. This will help in automating the servant work and reduce human involvement as well as dependency.

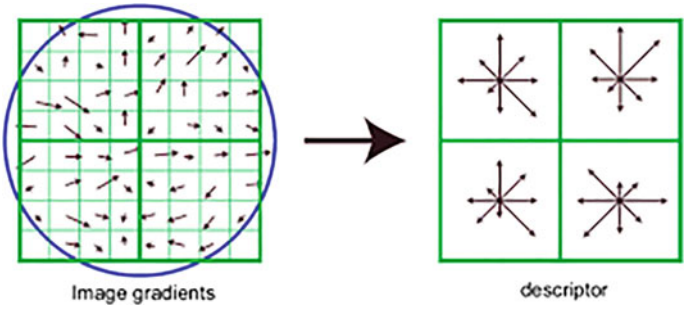


Fig. 2. Image gradients and spatial orientation binning

3 System Overview

3.1 Database

There is no standard dataset for bottles. The dataset has been prepared by capturing the images in the office. The positive and negative images (samples) were captured at the same time, which includes 500 positive and 500 negative pictures. Entire pictures have been normalized in 64 × 128 dimensions.

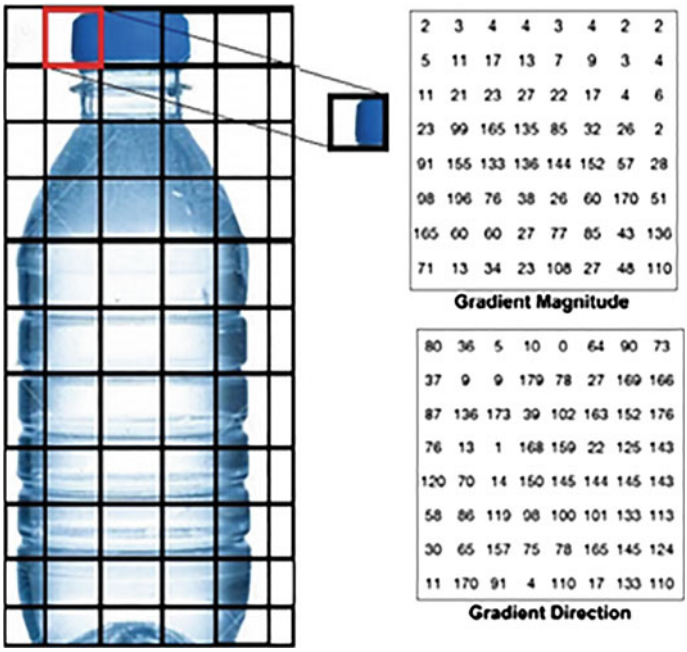


Fig. 3. Calculating the gradient direction and magnitude of the image and storing them in a vector

3.2 Preprocessing

We preprocessed the images in the grayscale space as the paper by Dalal and Triggs [3] gives no distinct advantage of using the RGB or LAB color spaces. Apart from that, we apply gamma normalization to improve the intensity of the image. This has been done as images clicked from camera devices that have low illumination.

3.3 HOG Feature Extraction

The following sections describe the HOG feature extraction procedure from scratch as implementation given by Dalal and Triggs [3]:

A. Gradient Computation
To compute the gradient of the image, we simply apply the point discrete derivative mask in both horizontal and vertical directions. This method requires filtering the intensity data of the image with the kernels [−1 01] and its transpose in both horizontal and vertical directions, respectively.

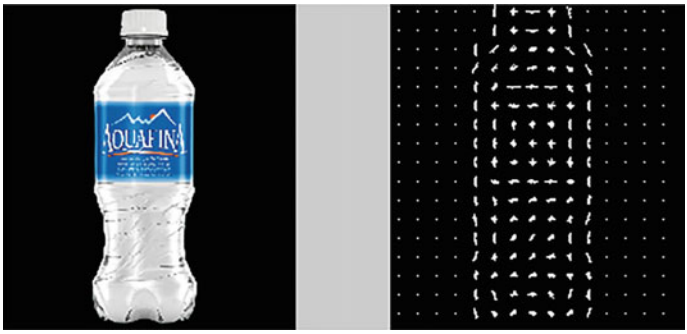


Fig. 4. HOG visualization of a bottle

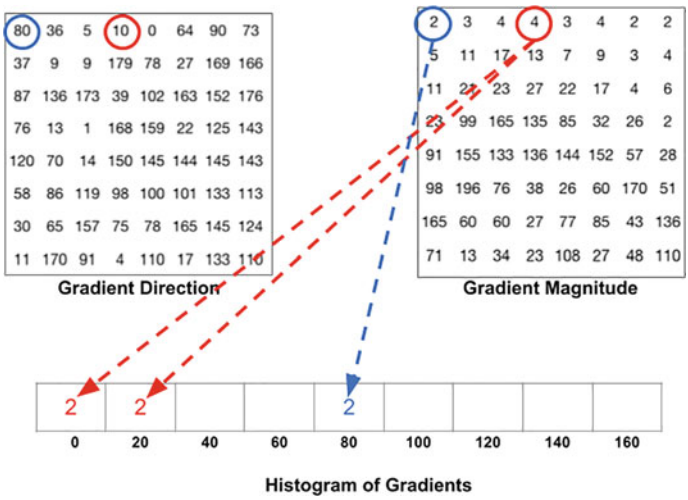


Fig. 5. Histograms calculation of each block

B. Orientation Binning

A cell histogram is created by weighted quantization of the orientation of each pixel of the cell into predefined orientation-based bins. The cells are usually square in shape (for convenience we will stick with rectangular) but they can be rectangular or circular. The weighting of the orientations can be either using the gradient magnitude itself. To calculate gradient magnitude and direction, we use

$$g = \sqrt{g_x^2 + g_y^2}$$

$$\theta = \arctan \frac{g_y}{g_x},$$

where g_x and g_y are the Sobel directional vectors. The gradient magnitude and direction are calculated by the given formulas and then stored in different vectors. Figure 3 shows the magnitude and direction information (Fig. 4).

C. Block Division and Normalization

The cells must be grouped together in order to factor in the changes in illumination and contrast. The complete HOG descriptor is then the vector of the components of the normalized cell histograms from all of the block regions as shown in Fig. 5. These blocks are normalized by four prominent methods: L1 norm, L1 norm square root, L2 norm, and L2 norm followed by clipping (L2 Hys). We experimented and choose the one that works best. Figure 4 shows the HOG visualization of a bottle.

The following formula is used for normalization of the blocks of the image:

$$f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}},$$

where "v" be the non-normalized vector containing all histograms in a given block, is its *k*-norm for k = 1, 2, and e is a small constant.

D. Calculation of HOG Feature Vector

The final step collects the HOG descriptors from all blocks of a dense overlapping grid of blocks covering the detection window into a combined feature vector for use in the window classifier. We calculate the final feature vector for the entire image patch, and the 36×1 vectors are concatenated into one giant vector. There are 7 horizontals and 15 vertical positions of the 16×16 blocks making a total of $7 \times 15 = 105$ positions. Each 16×16 block is represented by a 36×1 vector. So when we concatenate them all into one giant vector, we obtain a $36 \times 105 = 3780$ dimensional vector. Training the classifier: Based on the literature survey done, we chose a linear kernel SVM for the classification purpose. SVM is one of the best classifiers used for the computer vision area.

E. Feature Dimension Reduction

HOG method produced a high-dimensional feature vector which needs more memory and computational power. So we reduced the feature using principal component analysis (PCA) which has been widely used for the feature reduction. PCA helps us to reduce features 3.7–1 K.

3.4 Sliding Window

To detect the bottle in a given image, we applied a sliding window approach to predict the presence of a bottle in a window which kept sliding over the complete image as shown in Fig. 6. The window has been shifted with a step length equal to the length and width of the block size, respectively, in the vertical and horizontal directions. This process is computationally heavy, given that the gradient magnitudes and orientations for each patch needs to be computed for each window during sliding. To speed up the process, we divide the whole image into the blocks of the given block size, and compute gradient histograms over them before applying the sliding window. Thus, after computing these histograms beforehand, now while applying sliding window,

we just need to consider the subset of blocks which belong to the window and concatenate their histograms to get our feature vector (Fig. 7).

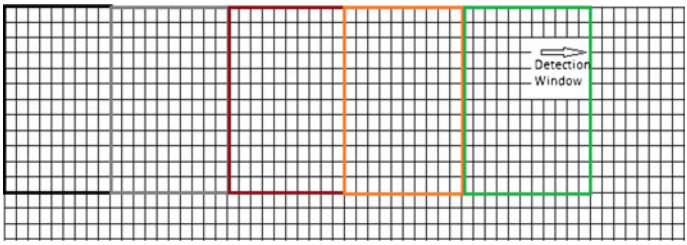


Fig. 6. Sliding window over the image

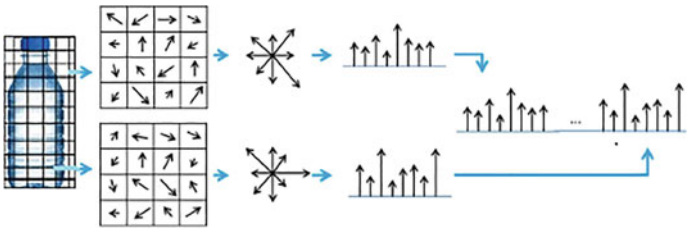


Fig. 7. HOG descriptor representation

4 Experimental Results

Our target is to detect bottles from an image taken from a camera. 100 experimental images were taken for the testing of the system. The images were processed on Intel(R) i5 processor at a clock speed of 1.60 GHz. During the calculation of HOG features, we tried several variations in its parameters.

First, we varied the number of bins used in the HOG descriptor. We used the values 5, 7, and 9 for the number of bins. With the bin number 5, the 180° gradient orientation range was divided into five segments of around 36° each. Out of the 100 test images, 85 successfully detected the bottle. The 7-bin system divided the 180° gradient orientation range into equal segments of around 25° each. This had a very good detection rate at 93 out of 100 bottles. The 9-bin system was tested. It divided the range into 9 segments of 20° each and detected 91 out of 100 bottles as shown in Table 1.

Table 1. Variation of accuracy with the variation of the bins

Bins	Block size	Cell Size	Overlap (%)	Accuracy (%)
5	2 × 2	8 × 8	50	85
7	2 × 2	8 × 8	50	**93**
9	2 × 2	8 × 8	50	91

Next, we changed the block overlap to 0%, which meant the blocks did not have any cells in common. The observation of this variation was very poor as many objects lay partly in multiple blocks, and they did not get detected. As shown in Table 2, the detection rate was 75 out of 100, whereas with 50% the detection rate was 93 out of 100. Lastly, we varied the cell size. We changed the cell sizes as 2×2, 8×8, 16×16, and 32×32 and calculated the HOG feature with 50% overlap. The results are shown in Table 3. The HOG value of each block is accumulated into a single value; here, HOG value of each block was too muddled to detect the bottles accurately. The 8×8 cell size provided a much better detection rate as compared to other cell size.

Table 2. Variation of accuracy with the variation of block overlap

Bins	Block Size	Cell Size	Overlap (%)	Accuracy (%)
7	2×2	8×8	50	**93**
7	2×2	8×8	0	75

Table 3. Variation of accuracy with the variation of the cell size

Bins	Block size	Cell size	Overlap (%)	Accuracy (%)
7	2×2	4×4	50	88
7	2×2	8×8	50	**93**
7	2×2	16×16	50	90
7	2×2	32×32	50	77

5 Conclusions

The computer vision is rapidly involving every sector owing to the security and the vision power as human being. This paper primarily focused on the water bottle detection to reduce the human involvement and dependency on him. We used the HOG features for this purpose and got the satisfactory results. We also found that the HOG feature works well with the cell overlap and performed badly without it. We considered the water bottles of different colors, sizes, and shapes that why the accuracy is around 93. Our future work will be in the same direction to improve the detection rate and also toward the experiment with other object detection approaches to make an independent hardware for the robot.

References

1. Mohan A, Papageorgiou C, Poggio T (2001) Example-based object detection in images by components. PAMI
2. Lowe DG (2004) Distinctive image features from scale-invariant key points. IJCV 60(2):91–110

3. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE computer society conference on computer vision and pattern recognition, 2005 (CVPR 2005), vol 1. IEEE, pp 886–893
4. Ye Q, Jiao J, Zhang B (2010) Fast pedestrian detection with multi-scale orientation features and two-stage classifiers. In: Proceedings of the IEEE international conference on image processing
5. Suard F, Rakotomamonjy A, Bensrhair A, Broggi A (2006) Pedestrian detection using infrared images and histograms of oriented gradients. IEEE
6. Gavrila DM, Philomin V (1999) Real-time object detection for smart vehicles. In: Conference on computer vision and pattern recognition (CVPR)
7. Kassani PH, Teoh ABJ (2017) A new sparse model for traffic sign classification using soft histogram of oriented gradients. Appl Soft Comput 52:231–246
8. Chen T-H, Wu P-H, Chiou Y-C (2004) An early fire-detection method based on image processing. In: 2004 international conference on image processing, 2004 (ICIP'04), vol 3. IEEE, pp 1707–1710
9. Ren X, Ramanan D (2013) Histograms of sparse codes for object detection. In: Proceedings of the IEEE international conference on computer vision and pattern recognition
10. Papageorgiou C, Poggio T (2000) A trainable system for object detection. IJCV 38(1):15–33
11. Amit Y (2002) 2D object detection and recognition: models, algorithms and networks. MIT Press, Cambridge, MA
12. Viola P, Jones MJ (2004) Robust real time face detection. Int J Comput Vis

Global Scenario of Solar Photovoltaic (SPV) Materials

Sandeep Gupta(✉) and Abhishek Sharma

Department of Electrical Engineering, JECRC University, Jaipur 303905, Rajasthan, India
{jecsandeep,abheeshek96}@gmail.com

1 Introduction

Since the last decade, the world is developing at a pace like never before. An important factor behind this evolution is the recent advancements in the field of the energy generation technology. With the growing energy demands, the need for technology which must be economically, environmentally, and socially compatible has also increased. The energy-generating technologies can be classified as conventional and nonconventional [1]. The conventional energy sources have been in use since a long time. These include petroleum, coal, wood, etc. These types of sources have proved to be very beneficial for the human race development but still they have many disadvantages. Most of them cause environmental pollution and also are costly as the energy is needed to be transmitted over long distances after conversion into electricity. On the other hand, there are nonconventional energy sources like wind, solar, and thermal energies. These are inexhaustible and environment friendly but still need to be developed more to be conventionally used. Solar energy as a nonconventional energy source is a very good option for not only bulk electricity production but also for the off-grid purposes. They are also helpful in avoiding the long-distance transmission costs [2, 3].

There are many emerging technologies which have been discussed in this paper. Classification in the PV technology is explained in Sect. 2. Different conventionally used material technologies are described with different structures in Sect. 3. In Sect. 4, modern different emerging PV technologies are clearly explained. Finally, Sect. 5 concludes this paper.

2 Classification

The solar cell technology can be characterized into three eras. Figure 1 shows the different types of the PV technologies based on different types of the materials used. First, generation cells were based on silicon wafers. Silicon is the second most abundant element present on earth and its nontoxic nature makes it suitable for the widespread use in the PV industry [4].

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_14

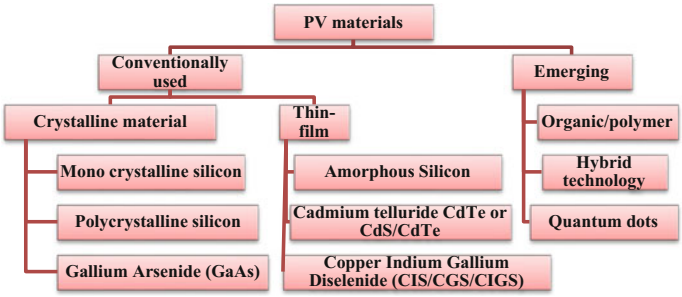


Fig. 1. PV technologies classification

However, due to its high cost an alternative path was required, so second-generation thin-film modules were invented. These modules reduced the material used and hence reduced the cost, but the efficiency was also less as compared to crystalline Si. The third-generation polymer technology is scotch as well as lightweight. It is helpful in meeting concerns regarding the environmental problems. But even this technology has lower efficiency as compared to Si-based ones. Hybrid technology involves the combination of both the crystalline and thin-film modules [5, 6].

3 Conventionally Used Material Technologies

3.1 Crystalline Material

This technology is considered to be the first generation of photovoltaic technologies. Modules are made by combining different silicon cells or GaAs cells. The crystalline silicon-based cells are as yet driving the PV market. The conversion efficiency of single-crystal silicon cell has hit the mark of 26.3% at STC (Standard test conditions, i.e., 25 ºC and 1000 W/m^2 sunlight intensity) [7].

3.1.1 Monocrystalline Silicon

It is the most prevalent material in PV modules. It utilizes a p–n junction for its course of action. The front of the cell is encrusted with a blanket of micrometer-sized pyramid structures. Solar cells based on this technology have immensely phosphorous-doped n+ section stacked over p-type boron-doped substrate to engender a p–n junction. Immensely doped p+ field (BSF) sections are framed on the back facade of the silicon substrate; its aim is to curtail the recombination of minority carriers. Its structure is shown in Fig. 2. The cells based on this technology are typically 5 inches squares [8, 9].

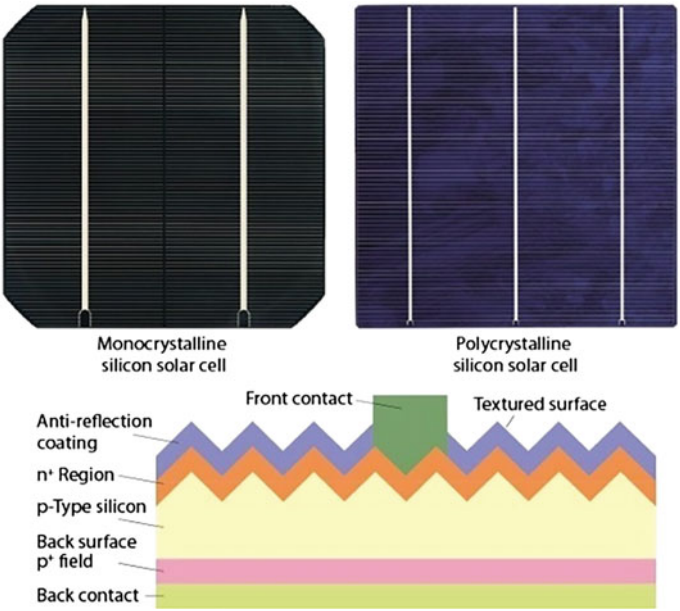


Fig. 2. Structure of crystalline solar cells [8]

3.1.2 Polycrystalline Silicon

This technology was introduced to reduce the production cost of silicon ingots. The wafers of these cells are formed by pouring the molten silicon in a cubical mold. The liquid silicon is then allowed to cool down and solidify. The solidified block is sliced to form perfectly square cells [10]. Polycrystalline silicon has high packing factor. Nowadays, new sorts of back-contact polycrystalline cells are created by different organizations. Among these, metal wrap through (MWT) cells and emitter wrap through (EWT) cells are mostly used for PV cells [10, 11]. The efficiency of the poly-c-Si-based modules by Trina solar corporation at STC is found to be 19.9% [12, 13].

3.1.3 Gallium Arsenide (GaAs)

GaAs cells are having high energy conversion efficiency as compared to mono-c-Si and poly-c-Si cells. But due to high cost, it is not commercially used. It has high-temperature coefficient and hence is suitable for use in the space applications and concentrated PV modules. It possesses lighter weight as compared to c-Si [14]. GaAs can be further alloyed with phosphorous (P), indium (I), aluminum (Al), or antimony (Sb) to improve the efficiency. The efficiency of alloying increases due to the formation of multi-junction structure [15–17].

3.2 Thin-Film Material

This technology is considered to be the second generation of photovoltaic technologies. Thin-film technology extensively reduces amount of semiconductor material used and hence reduces the production costs. But due to high radiation capture losses, its efficiency is lower than c-Si cells [18, 19]. Gallium arsenide (GaAs), CdS, and titanium dioxide (TiO_2) are the materials that are most commonly used [20, 21].

3.2.1 Amorphous Silicon

This material is having about 40 times superior absorptive rate of light as comparison to mono-c-Si. Due to high efficiency, it is most commonly used material in the thin-film cells. As shown in Fig. 3, a-Si cells due to high band gap of 1.7 eV absorb very broad range of the light spectrum [22].

The Tel solar corporation a-Si solar cell at STC is found to have 12.3% efficient. But when exposed to sunlight, its efficiency decreases by about 30 to 40%. This reduction is caused due to Staebler–Wronski effect (SWE) which can be minimized by thermal annealing at or above 150 ºC [22–24].

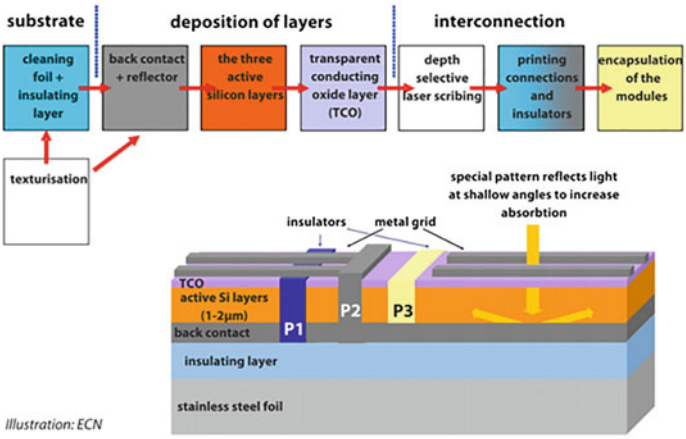


Fig. 3. Structure of a-Si solar cell

3.2.2 Cadmium telluride (CdTe or CdS/CdTe)

Photovoltaic solar cells based on CdTe contribute to the major part (about 5.1%) of commercial thin-film module production worldwide. CdTe-based solar cells are the second most normal PV innovation on the planet. The United States is the leading manufacturer of CdTe PV. In addition to high efficiency, these cells can be quickly manufactured and also costs low [25]. Typical CdTe thin-film deposition techniques are shown in Fig. 4.

3.2.3 Copper Indium Gallium Diselenide (CIS/CGS/CIGS)

It is fabricated by storing a thin layer of copper, indium, gallium, and selenide on plastic or glass backing. CIGS is a strong arrangement of copper indium selenide (CIS) and copper gallium selenide (CGS), having chemical composition as $CuIn_xGa_{(1-x)}Se_2$ [26]. As discussed earlier, the more suitable the band gap, the more is the range of the wavelength to be absorbed from the solar radiation. CIGS-based solar panels are the most elevated performing thin-film solar panels till date. These cells contain less of the toxic material cadmium as compared to CdTe cells [27].

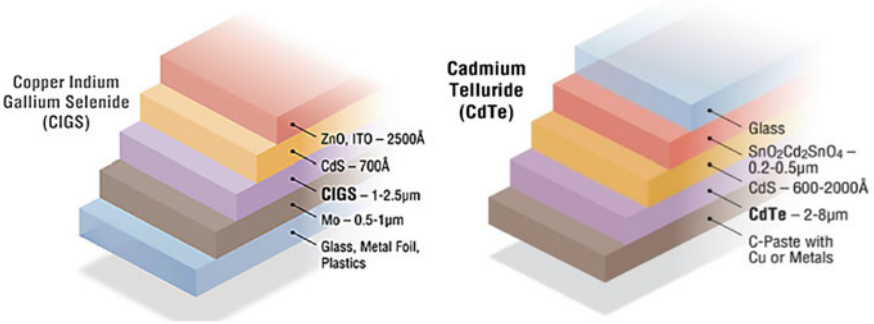


Fig. 4. CIGS and CdTe solar cell diagrams [45]

Photo-degeneration takes place in CIGS modules when subjected to sunlight just like CdTe modules. Additional barrier coating is required to mitigate this problem [28]. CdS layer protects the CIGS layer from further processes. Further, the ZnO layer as the transparent front contact is deposited by radio frequency sputtering or atomic layer deposition (ALD) [29]. The heterojunction is formed between ZnO and CIGS layer [30]. In June 2016, the Centre for Solar Energy and Hydrogen Research Baden-Württemberg (ZSW) held world record for highest efficient CIGS laboratory-sized (0.5 square centimeter) solar cell with 22.6% efficiency [31].

4 Emerging PV Technologies

4.1 Organic/Polymer Material

This underdeveloped technology is the part of the third generation of photovoltaic technologies. The organic polymers have high light absorption coefficient, and hence more light can be absorbed using less material [32]. The demerits of this technology include very low conversion efficiency and less stability due to photochemical degradation [32, 33]. Based on the junction types, these cells can be divided into three basic categories: (1) Single layer [34–36], (2) Discrete heterojunction [35–37], and (3) Bulk heterojunction (BHJ) [38].

4.2 Hybrid Technology

A hybrid solar cell can be a combination of organic and inorganic materials. Due to combination of high charge carrier mobility of the inorganic material and high light absorption capability of organic materials, this technology got much attention in the recent year. Following are some common hybrid technologies:

- Perovskite-based cells: This solar cell uses a perovskite-structured compound as active layer for light absorption. The problem with this technology is low sunlight stability of the perovskite material [39]. Shin et al. [40] using methylammonium lead iodide ($MAPbI_3$) perovskite as active layer and lanthanum (La)-doped $BaSnO_3$ (LBSO) as photoelectrode materials achieved photo-conversion efficiency of 22.1% and high photostability of about 1000 h.
- Multi-junction solar cells: These cells contain three or more p–n junctions using materials having different band gaps. The use of different semiconductors with different band gaps increases the range of the light spectrum absorbed by the cell [41]. Frank et al. [42] developed a 44.7% efficient four-junction GaInP/GaAs// GaInAsP/GaInAs tandem cell.

4.3 Quantum Dots

A quantum dot is a semiconductor crystal having size in nanometers. The band gap of these dots can be changed by changing their size. The change of the band gap changes the range of the solar spectrum radiation absorbed by the material. Hence, it is an attractive technology to be used in multilayer PV cells [43]. These cells are easy to synthesize and less costly. The highest conversion efficiency shown by quantum dots based solar cells till date is 11.3% only [44].

5 Conclusion

Various state-of-the-art solar photovoltaic materials have been discussed in this paper. The first half of the paper is mainly focused on the structure, efficiencies, and manufacturing processes of the conventionally used solar cells. Si-based solar cells still rule the PV industry. There are many emerging technologies which have been discussed in the later sections of the paper. These emerging technologies may prove to be competing with the conventionally used technologies in the near future. But currently the challenges for these emerging technologies are increasing conversion efficiencies and stability under direct solar radiation exposure. The race to develop highly efficient solar cells with low manufacturing cost is never ending. Therefore, further improvements are expected in the near future in the PV technologies.

References

1. Sen PK, Awtar K, Bohidar SK (2015) A review of major non-conventional energy sources. IJSTM 4(01):20–25
2. Tazvinga H, Thopil M, Numbi PB, Adefarati T (2017) Distributed renewable energy technologies. Handbook of distributed generation. Springer International Publishing, pp 3–67
3. Gupta S, Singh R (2011) Investigation of steady state performance of static synchronous compensator on transmission line. ELEKTRIKA J 13(1):42–46
4. An X et al (2016) Empirical and Quokka simulated evidence for enhanced VOC due to limited junction area for high efficiency silicon solar cells. In: 2016 IEEE 43rd photovoltaic specialists conference (PVSC)
5. De Azevedo Dias CL, Branco DAC et al (2017) Performance estimation of photovoltaic technologies in Brazil. Renew Energy 114:367–375
6. Goetzberger A, Knobloch J, Voss B (1998) Crystalline silicon solar cells. Wiley
7. Green MA et al (2015) Solar cell efficiency tables (Version 45). Prog Photovolt Res Appl 23 (1):1–9
8. Green MA, Emery K (1993) Solar cell efficiency tables. Prog Photovolt Res Appl 1(1):25–29
9. Saga T (2010) Advances in crystalline silicon solar cell technology for industrial mass production. NPG Asia Mater 2(3):96–102
10. Chu TL, Singh KN (1976) Polycrystalline silicon solar cells on metallurgical silicon substrates. Solid State Electron 19(10):837–838
11. Van Kerschaver E, Beaucarne G (2006) Back-contact solar cells: a review. Prog Photovolt Res Appl 14(2):107–123
12. Fabre E, Baudet Y (1978) Polycrystalline silicon solar cells. In: Photovoltaic solar energy conference, pp 178–186
13. Pandey AK et al (2017) Solar photovoltaics (PV): a sustainable solution to solve energy crisis. Green technologies and environmental sustainability. Springer International Publishing, pp 157–178
14. Knechtli RC, Loo RY, Kamath GS (1984) High-efficiency GaAs solar cells. IEEE Trans Electron Devices 31(5):577–588
15. Khanna V et al (2016) Statistical analysis and engineering fit models for two-diode model parameters of large area silicon solar cells. Sol Energy 136:401–411
16. Sivananthan S, Carmody M, Bower RW, Mallick S, Garland J (2016) Tunnel homojunctions in group IV/group II-VI multijunction solar cells. U.S. Patent 9,455,364, 27 Sept 2016
17. Kurtz SR et al (2001) InGaAsN/GaAs heterojunction for multi-junction solar cells. U.S. Patent No. 6,252,287, 26 June 2001
18. Dezfooli AS et al (2017) Solar pavement: a new emerging technology. Sol Energy 149:272–284
19. Chopra KL, Das SR (1983) Why thin film solar cells? In: Thin film solar cells. Springer, US, pp 1–18
20. Coutts Timothy J et al (2003) Critical issues in the design of polycrystalline, thin-film tandem solar cells. Prog Photovolt Res Appl 11(6):359–375
21. Aberle AG (2009) Thin-film solar cells. Thin Solid Films 517(17):4706–4710
22. Galloni R (1996) Amorphous silicon solar cells. Renew Energy 8(1):400–404
23. Kołodziej A (2004) Staebler-Wronski effect in amorphous silicon and its alloys. Opto-Electr Rev 12(1):21–32

24. Watahiki T et al (2016) Analysis of short circuit current loss in rear emitter crystalline Si solar cell. J Appl Phys 119–129
25. Wu X (2004) High-efficiency polycrystalline CdTe thin-film solar cells. Sol Energy 77 (6):803–814
26. Wieting RD et al (2011) Single junction CIGS/CIS solar module. U.S. Patent Application No. 13/086,135
27. Pollock GA, Mitchell KW, Ermer JH (1990) Thin film solar cell and method of making. U.S. Patent No. 4,915,745, 10 Apr 1990
28. Gwak J et al (2016) Method of fabricating copper indium gallium selenide (CIGS) thin film for solar cell using simplified co-vacuum evaporation and copper indium gallium selenide (CIGS) thin film for solar cell fabricated by the same. U.S. Patent No. 9,472,708, 18 Oct 2016
29. Lee SW et al (2014) Improved Cu2O-based solar cells using atomic layer deposition to control the Cu oxidation state at the p-n junction. Adv Energy Mater 4(11)
30. Metin B, Nayak D, Pinarbasi M (2011) Cigs based thin film solar cells having shared bypass diodes. U.S. Patent Application 13/163,485, 17 June 2011
31. Osborne M (2016) ZSW achieves world record CIGS lab cell efficiency of 22.6%. 2016-06-15. http://www.pv-teeh.org/news/zsw-achieves-world-record-cigs-lab-eell-efficiency-of-22.6
32. Duan C et al (2015) Wide-bandgap Benzodithiophene-Benzothiadiazole copolymers for highly efficient multijunction polymer solar cells. Adv Mater (Wiley Online Library) 27 (30):4461–4468
33. Gupta S, Tripathi RK (2015) Transient stability assessment of two-area power system with LQR based CSC-STATCOM. Automatika 56(1):21–32
34. Günes S, Neugebauer H, Sariciftci NS (2007) Conjugated polymer-based organic solar cells. Chem Rev 107(4):1324–1338
35. Chen J-D et al (2015) Single-junction polymer solar cells exceeding 10% power conversion efficiency. Adv Mater 27(6):1035–1041
36. Peng J et al (2017) Interface passivation using ultrathin polymer-fullerene films for high-efficiency perovskite solar cells with negligible hysteresis. Energy Environ Sci
37. Bagher AM (2014) Introduction to organic solar cells. Sustain Energy 2(3):85–90
38. Gupta S, Sharma AK (2010) STATCOM-Its control algorthim. I-Manager's J Electr Eng 3 (4):41–48
39. Burschka J et al (2013) Sequential deposition as a route to high-performance perovskite-sensitized solar cells. Nature 499(7458):316–319
40. Shin SS et al (2017) Colloidally prepared La-doped BaSnO3 electrodes for efficient, photostable perovskite solar cells. Science 356(6334):167–171
41. Dimroth F (2017) III–V solar cells–materials, multi-junction cells–cell design. In: Photovoltaic solar energy: from fundamentals to applications (Book Chapter), pp 373–382
42. Dai P et al (2017) Electron irradiation study of room-temperature wafer-bonded four-junction solar cell grown by MBE. Sol Energy Mater Sol Cells 171:118–122
43. Kamat PV (2013) Quantum dot solar cells. The next big thing in photovoltaics. J Phys Chem Lett 4(6):908–918
44. Gupta S, Tripathi RK (2014) Improved performance of LQR controller in CSC based STATCOM using genetic optimization. In: 6th IEEE power india international conference (PIICON), pp 1–6, Dec 2014
45. Abou-Ras D, Kirchartz T, Rau U (eds) (2011) Advanced characterization techniques for thin film solar cells. Wiley-VCH, Weinheim, Germany

A 50 MHz–4 GHz Low-Noise Amplifier for Wideband Applications

Vimal Kant Pandey(✉), Dhananjai K Verma, Sandeep Sharma, and Sonika Singh

Department of EECE, DIT University, Dehradun, India
{vimalpandey94, tek.learn, gsonika}@gmail.com,
djaiverma07@hotmail.com

1 Introduction

Wideband technology has attracted many researchers because a wide range of future and modern communication system has been proposed that operates over a bandwidth of several gigahertz such as cognitive radio, software defined radio, etc. Wideband systems are capable of transmitting data over a wide spectrum of frequency bands with very low power and high data rates up to gigabits per second. The broadband behavior of wideband transceivers is mainly determined by the low-noise amplifiers. Low-noise amplifier has significant impact on the overall performance of the receivers as they are the first block of the receiver. Therefore, to achieve good performance parameters like ultrawide bandwidth with reasonable noise figure (NF) and good input (close to 50 Ω) and output impedance matching across the bandwidth of interest, it poses a more stringent requirement on the LNA.

Trade-off between voltage headroom, noise figure, bandwidth, and linearity has been done for optimal low-noise amplifier design. Several topologies for wideband low-noise amplifiers exit in literature such as common gate amplifiers, negative feedback amplifiers, distributed amplifiers [1–7], and inductive source degeneration amplifiers. The topology used for a particular application begins with requirements of input matching. Most of the existing wideband LNAs are operated at high frequencies only. The proposed design operates at both high frequency and low frequency so it can be used for wideband applications and also for cognitive radio that operates in the frequency range of 50 MHz–10 GHz.

In this paper, a low-noise amplifier is designed based on cascode structure. The paper is organized in the following manner: In Sect. 2, wideband input matching techniques are described. Section 3 provides concept of proposed circuit. Section 4 discusses the simulation results. In the end, Sect. 5 presents the conclusion.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_15

2 LNA Topologies

There are several topologies to design a wideband LNA and the choice of topology begins with the input matching requirements, and thus the first task of LNA design is to create 50 Ω resistive input impedance. The noise figure of LNA (about 2–3 dB of overall noise figure of a typical receiver) directly adds to the receiver, and thus the LNA should be designed in a way that it adds minimum noise in the input signal path. The input matching topologies used for LNAs have several forms such as inductively degenerated common source (CS) stage, gain stage with resistive feedback, common gate (CG) stage, and combination of CS and CG stages [8, 9]. Figures 1 and 2 shown below are inductively degenerated common source stage and common gate stage, respectively. The CS LNA has superior noise performance because inductive degeneration is ideally noiseless and the RF input signal is pre-amplified by the input matching series resonant network, whereas the CG LNA uses a parallel resonant network to match the input impedance at resonance ($\frac{1}{g_m}$) to 50 Ω [9]. Also, CG LNA provides a wideband input match that is less sensitive to the input parasitic capacitances.

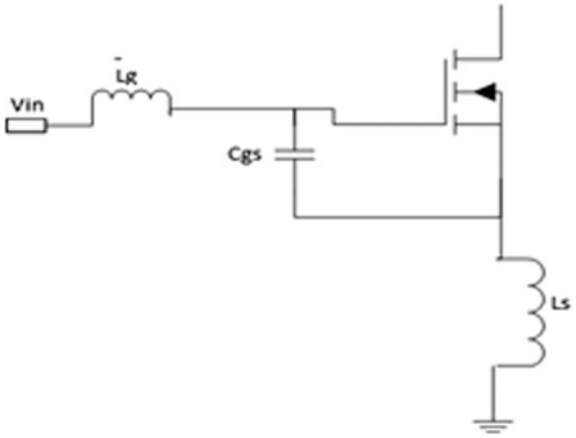


Fig. 1. Inductively degenerated common source

The inductive source degeneration topology provides perfect matching without adding noise and without imposing any restriction on the transconductance [10]. To obtain the perfect matching, two inductors L_s and L_g are used as shown in Fig. 1. The circuit proposed in the paper is designed using inductive degeneration. Figure 3 shows the small signal equivalent circuit of the inductively degenerated CS stage.

Since for a low-noise amplifier the input impedance is an important parameter and is set to be 50 Ω, we will find the expression for input impedance. The input impedance of the circuit shown in Fig. 3 is capacitive and is given by

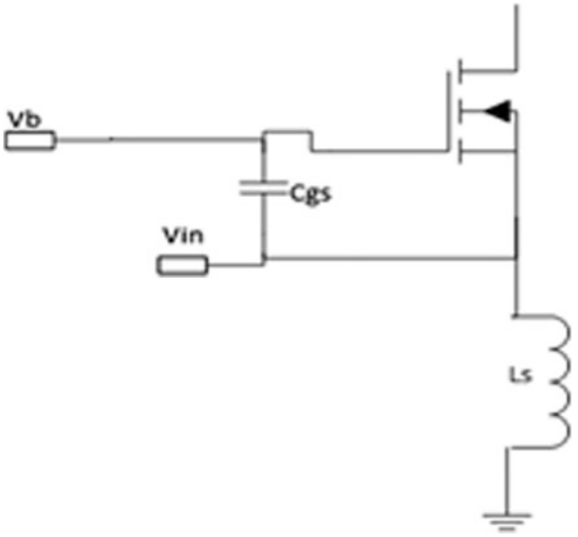


Fig. 2. Common gate stage

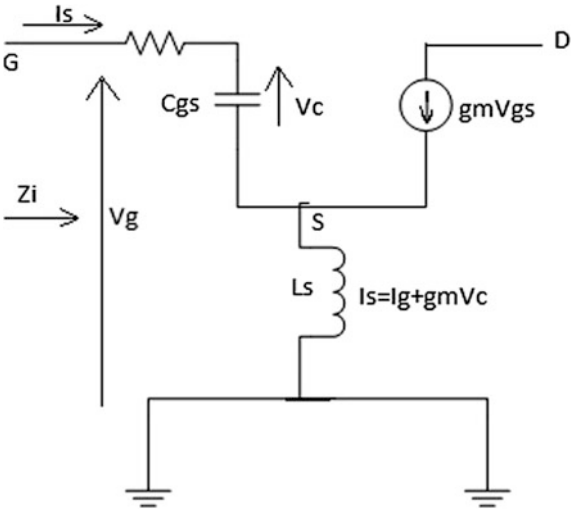


Fig. 3. Small signal equivalent circuit of inductively degenerated CS stage

$$Z_i = \frac{V_g}{I_g} = \frac{\left(I_g R_g + V_c + j\omega I_s L_s\right)}{I_g}, \tag{1}$$

where

$$V_c = \frac{I_g}{sC_{gs}}, I_s = I_g + g_m V_c \tag{2}$$

Using (2) in (1), we get

$$Z_i = \frac{I_g R_g + \frac{I_g}{sC_{gs}} + s\left(I_g + g_m \frac{I_g}{sC_{gs}}\right) L_s}{I_g} \tag{3}$$

Putting s = jω and on solving (3) we get

$$Z_i = R_g + \frac{L_s g_m}{C_{gs}} + j\left(\omega L_s - \frac{1}{\omega C_{gs}}\right) \tag{4}$$

For MOSFET, value of R_g is chosen as zero. Hence, the input impedance has only two components. For a low-noise amplifier, the input impedance should be resistive; hence, another inductance L_g is added to the gate terminal so that the capacitive effect is canceled out at resonant frequency, i.e.,

$$j\left(\omega L_s - \frac{1}{\omega C_{gs}}\right) = 0 \tag{5}$$

3 Proposed Circuit

The proposed circuit using cascode inductive source degeneration topology is shown in Fig. 4.

In this, M1 and L_s form a CS inductive degeneration amplifier which is the input stage of cascode circuit driven by the input signal source V_{in}. This input stage drives the M2 transistor act as common gate amplifier with output signal V_{out}. M2 reduces the Miller feedback capacitance C_{gd} of M1 by providing low input impedance to M2, thereby reducing voltage gain of M1. Thus, by introducing M2, isolation and low-frequency gain improve. However, the gain, linearity, and noise figure at high frequency degrade due to the parasitic capacitances of M2. To overcome these degradations, inductor L_D is used in between M1 and M2. L_D partially cancels the parasitic capacitance at the source of cascode transistor M2. The noise contributed by LD remains less than that of M1 even at relatively high frequencies. M3 and L2 are used to bias the transistor M1. R_D is the load impedance of M2.

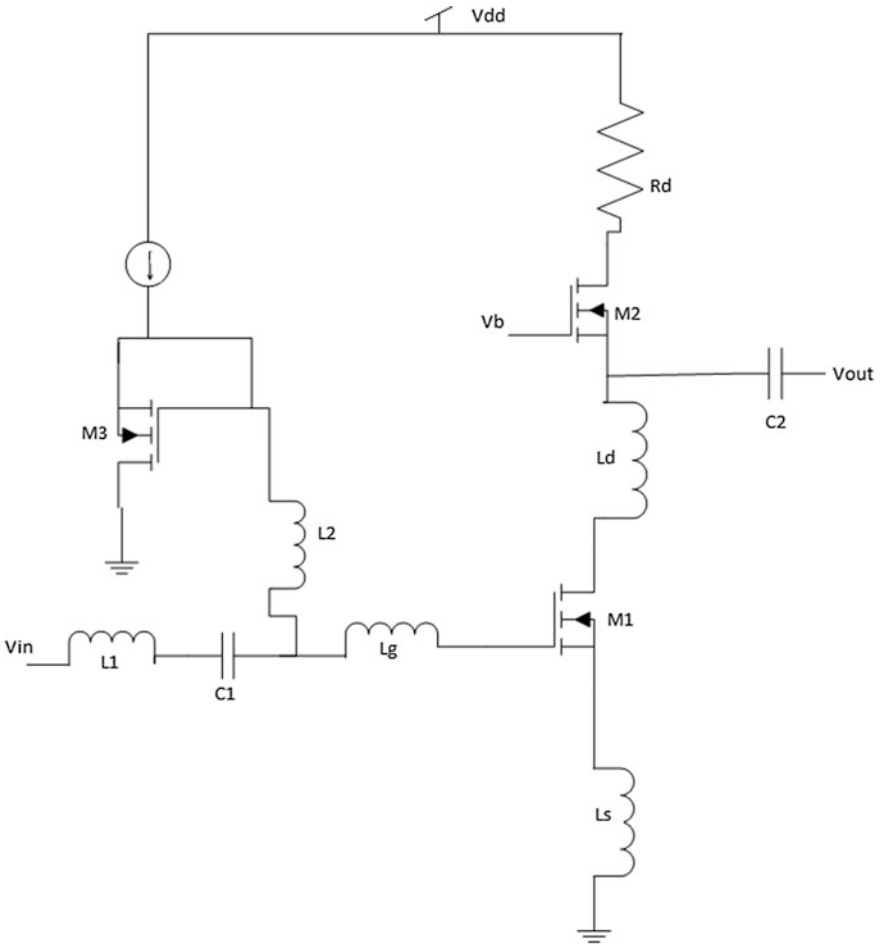


Fig. 4. Proposed LNA

4 Simulation Results

The proposed circuit is designed using 180 nm CMOS technology and simulations were carried out using EldoRF tool. The proposed circuit consumes 10.2 mW power from 1.8 V supply.

Figure 5a shows that the input return loss S11 is less than −10 dB over all the bandwidth. Figure 5b shows the gain versus the power curve where power ranges from −20 to 10 dB. The noise figure of the circuit is shown in Fig. 5c which lies between 2.45 and 2.62 dB. The noise figure of the proposed circuit is close to the desired noise figure of 2.5 dB. Figure 5d illustrates the third-order intercept point which is less than −14 dBm. The performance of the proposed LNA and comparison with the existing circuits are summarized in Table 1.

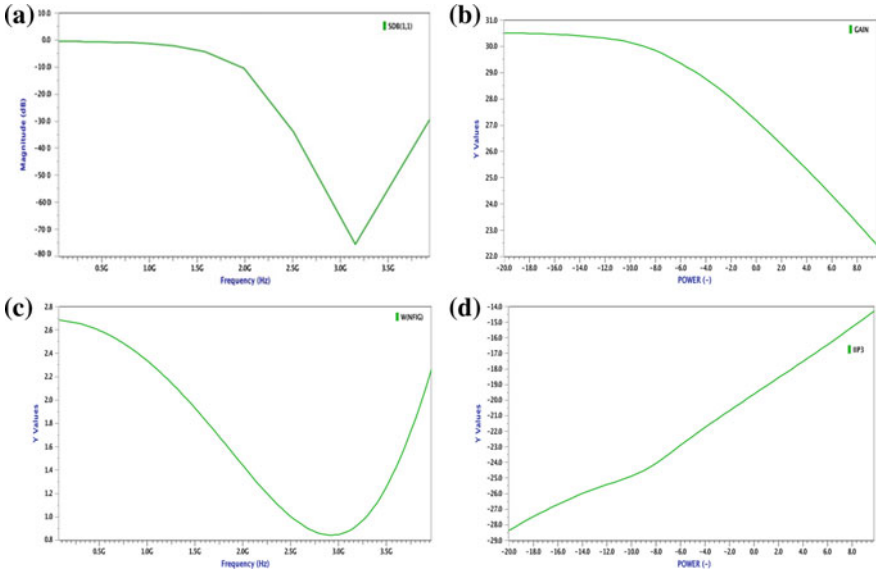


Fig. 5. **a** Input return loss (S11). **b** Gain versus power. **c** Noise figure. **d** IIP3 versus power

Table 1. Performance summary and comparison with different wideband LNAs

References	Frequency (GHz)	Gain (dB)	NF (dB)	IIP3 (dBm)	Power consumption (mW)	Process technology (nm)	Supply (V)
Hidayov et al.[1]	0.7–2.7	17	2.5	1.5	13.5	180	1.8
Guo et al. [2]	0.1–2	17.5	2.9–3.5	10.6–14.3	9.7	180	2.2
Liao et al. [3]	0.5–3.5	17–22	2.68	–	32.8	180	1.8
Ma et al. [4]	0.1–1.45	16.9	2.5	5.5	9.3	180	1.8
Ito et al. [5]	0.8–1.8	13.4	2.7	−7	6.5	180	1.8
Wang et al. [6]	0.47–3	13.6	2.5	−3.5	27	180	1.8
This Work	0.05–4	22.2	2.64	−14.2	10.02	180	1.8

5 Conclusions

Table 1 above helps to draw a conclusion that the present work offers better gain (22.2 dB) as compared to existing works, and also the power consumption (10.02 mW) and noise figure (2.64 dB) are competitive to the existing works. Further, these improvements occur for a wider frequency band (0.05–4 GHz).

Acknowledgements. This work is supported by the project entitled "Study and Design of Low Power Circuits for Cognitive Radio" under the Faculty Research Scheme, DIT University, Dehradun, India (Ref. No.: DITU/R&D/2015/ECE). The author would like to thank Prof. K. K. Raina, Vice-Chancellor, DIT University for encouragement and support during the present research work. The author would also like to thank the anonymous reviewers for their valuable comments and suggestions.

References

1. Hidayov OA, Nam NH, Yoon G, Han SK, Lee SG (2013) 0.7–2.7 GHz wideband CMOS low-noise amplifier for LTE application. Electron Lett 49(23):1433–1435
2. Guo B, Chen J, Li L, Jin H, Yang G (2017) A wideband noise-canceling CMOS LNA with enhanced linearity by using complementary nMOS and pMOS configurations. IEEE J Solid State Circuits. IEEE
3. Liao W-R, Yang J-R (2016) A 0.5–3.5 GHz wideband CMOS LNA for LTE application. IEEE
4. Ma L, Wang Z-G, Xu J, Amin NM (2016) A high linearity wideband common-gate LNA with differential active inductor. IEEE
5. Ito T, Okada K, Matsuzawa A (2011) A 0.8–1.8 GHz wideband low noise amplifier with capacitive feedback. In: Proceedings of the Asia-Pacific microwave conference
6. Wang S-F, Hwang Y-S, Yan S-C, Chen J-J (2011) A new CMOS wideband low noise amplifier with gain control. Integr VLSI J 44:136–143
7. Moezzi M, Bakhtiar MS (2012) Wideband LNA using active inductor with multiple feed-forward noise reduction paths. IEEE Trans Microw Theory Tech 60(4)
8. Razavi B (2002) Design of analog CMOS integrated circuits. Tata McGraw-Hill. ISBN: 978-0-07-052903-9
9. Razavi B (2012) RF microelectronics. Prentice Hall. ISBN: 978-0-13-713473-1
10. Li RC-H (2009) RF circuit design. Wiley. ISBN: 978-0-470-16758-8

DEM Reconstruction for Mizoram Area Using CARTOSAT Data and Preserving it by Cloud Computing Method

Srayashi Mallick[1(✉)] and Aishwarya Ramachandran[2]

[1] ICAR-National Research Centre for Orchids, Pakyong, India
srayashimallick@gmail.com
[2] Nissan Renault Financial Services India Pvt Ltd, Chennai, India
aishwarya141219 89@gmail.com

1 Introduction

Digital elevation models (DEMs) are considered as one of the most important data used for geospatial analysis, as the detailed information we are getting from it. Thus, elevation modeling has become important for geospatial analysis. The large-scale mapping system is the result of development of remote sensing technology.

Cloud computing provides a solution to create, store, retrieve, and analyze data. Implementation of cloud computing is a transformative paradigm that enables scalable, convenient, and on-demand access to a shared pool of configurable computing and network resources for efficiently delivering applications and services over the Internet. Cloud computing involves provisioning of computing, networking, and storage resources on demand. Geographic information system has huge data and same is increasing day-to-day. CARTOSAT-I and II are the first two satellites of ISRO designed to provide high-resolution stereoimagery of earth's surface for the purpose of cartographic applications including generation and updation of topographic maps. CARTOSAT-I and II provide stereoscopic data with spatial resolution of 2.5 m and 0.8 m, respectively, and are uniquely designed with a high radiometric resolution of 10 bits and spectral bandwidth of 500–850 nm. In local language, the word Mizoram means "The land of highlanders". Basically, Mizo hill dominates the majority of the topography. The steep hills are separated by rivers flowing on either direction of north and south, creating gorges between the hills. Hence, high spatial resolution data is to enhance the capability of identification of best possible detail of features on image, improve accuracies and cartographic quality of mapping, produce accurate digital elevation models of the terrain, and extract relevant terrain parameters from DEM.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_16

2 Objective of the Study

- The objective of the study is to reconstruct DEM using stereogrammetry and radargrammetry techniques on stereopairs of CARTOSAT 1 and RADARSAT 1 data products and subsequently the accuracy assessment using the available reference SRTM DEM.
- The objective is to attain the above result and to store huge data storage, processed data, and spatial data analysis, and is to develop the new application by implementing the Cloud Computing technologies.

3 Terrain Characteristics of the Mizoram

Mizoram is situated between 21° 58′ and 24° 29′ north latitude and 92° 29′ to 93° 22′ east longitude. It shares its geographical borders with Assam, Manipur, and Tripura extended over, respectively. The present investigation falls in the north Mizoram which is located nearly 30–35 km away from Aiwal, the capital of Mizoram (Fig. 1). The corresponding toposheets are 84A/9, 84A/13 and part of 84A/10, 84A/14, 83D/12, 83D/16 [1]. The geographical coordinates of the test site are as follows.

92° 27’ 40.52’’ E 24° 03’ 28.01’’N, 92° 59’ 42.27’’ E 24° 03’ 28.01’’N

92° 27’ 40.52’’ E 23° 38’ 52.45’’N.92° 59’ 42.27’’ E 23° 38’ 52.45’’N

Fig. 1. Study area, North Mizoram

4 Methodology

The methodology adopted in the present study can be categorized into two parts:

- Reconstruction of DEM and
- Accuracy assessment.

4.1 DEM Generation Using OPTICAL Data

Using PCI Geometica software, the automatic DEM extraction module allows you to create digital elevation models (DEMs) from stereoimages (CARTOSAT Data). Processes involved are as follows:

- **Data input**: A project file has been created with specified math model (Rational Function) and projection parameters (UTM, 43S D000). B and A and B and F along with its RPC (Rational Polynomial Coefficients) files are used for registration process.
- **GCP/TP collection**: Tie points were located carefully to compute transformation in order to improve the fit between two images, i.e., B and A and B and F. In the present investigation, 79 GCP/tie point and its RMS error are collected.
- **Epipolar pair generation**: Epipolar image reprojects the left and right images to a common orientation and the common features of the images to the same axis. B and F were selected as left image and B and A as right image.
- **Automatic DEM extraction**: Once the epipolar pair was constructed, then DEM has been extracted specifying the input parameters such as minimum and maximum elevation value of the region, failure and background value with the desired pixel spacing of the DEM. However, in this analysis, 100 and −150 are considered as the default failure value and background value, respectively. As the spatial resolution of the raw data is 2.5 m, hence, the spatial resolution of the reconstructed geo-coded DEM of 20 m can be obtained.

4.2 DEM Generation Using Field GCP

DGPS data are used to rectify the flight path/geometric error for scene no 5 data products. In order to achieve online positioning with high accuracies, the GPS is used in differential mode or differential GPS (DGPS). Since the field area is thick vegetation and undulation, the transmitting signals and line of sight are the major problems in real time, so data has been taken in static mode. Base station was established at Assam rifle camp, Seling [2]. This is nearly 30 km s from Aizwal city. Info/details of base station are

Job: 031106,
az Point ID,
ref: 1 antenna height 1100, and
GDOP: 2.3.

DGPS survey was carried out from Tuiaini suspension bridge to Darlawn. Total 34 GCP were collected which are homogeneous and well identifiable on the ground and records its location using the GPS receiver. These locations were plotted over an image, and DEM has been reconstructed as mentioned earlier.

4.3 Mosaic

Six scenes of CARTOSAT DEMs have been mosaic Using Ortho Engine of PCI Geomatica. Mosaic file had been defined and then by adding adjacent products.

4.4 DEM Generation Using RADARSAT 1 Data

The reconstruction of stereo-DEM has been carried out using the stereopairs of RADARSAT images with the help of Radar Mapping Suite of ERDAS imaging software for the overlapping regions of the two images: correlation image and constructed DEM.

4.5 Data Fusion or Merging

Scene no 3 (path/row no 0611-0286) has a huge cloud cover area; DEM has not come properly, and hence RADARSAT data has been used to mask the cloud cover area to get better DEM. Data merging or fusion procedure is as follows:

Cloud covered area has been masked using bitmap layer in which bitmap layer acts as a graphical layer having the same pixel and line dimensions as the image data. The masked region is assigned by value 1. Three new image channels, CARTOSAT (Bitmap), RADARSAT (translate data), and merge of both the data (EASI), were created, respectively. Translate utility is used to translate RADARSAT (.img) supported file format to CARTOSAT (.img) file from a geo-gateway format. While translating file, geo-referenced source and destination files are defined and then the layer information between the two files is shared.

The **Engineering Analysis and Scientific Interface** (**EASI**) is a full-featured interpreted programming language. EASI supports control structures such as for, while, and if statements, the definition of variables and structures, the definition of functions taking any number of arguments and providing a return value, and a powerful set of built-in intrinsic functions [3]. For present data fusion, small program has been used:

```
if %1 = 1 then
%3 = %2
end if
if %1 > 1 then
%3 = %1
end if
```

Here, %1 represents CARTOSAT product, %2 represents RADARSAT product, and %3 represents fused product.

5 Architecture of the New Application

To deploy this, application will not take more time and the application can be used for a more number of users in a short. SaaS model due to its varied advantage is quite attractive to enterprises. SaaS cloud is the latest model in deployment of software, where a GIS application is hosted as a service provided to customers across the Internet [4]. It is defined as a software application delivery model, where the same is deployed and hosted on a cloud platform for its customers to operate the application [5]. This innovative technology gives a full access to GIS tools such as creation of data, analysis of data, editing, and visualizing the data. When we compare the conventional GIS software with cloud, application has some unique features, i.e., this application hosted as web application and not being installed on-premise as it does not require the deployment of a large infrastructure at the client's location [6]. For the integration of SaaS, web services have become a widely adopted interface. SaaS applications represent their functions through web services, so that they can be easily accessed by on-premise applications and the same can be accessed from outside [7] (Figs. 2 and 3).

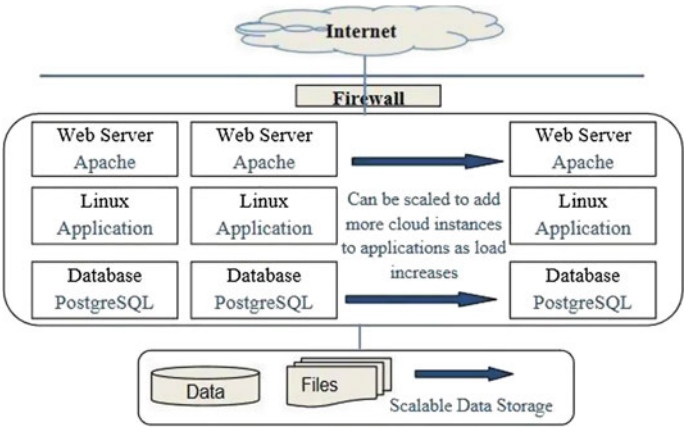


Fig. 2. Shows multiuser SaaS model

6 Flow Diagram for the Application

See Fig. 3.

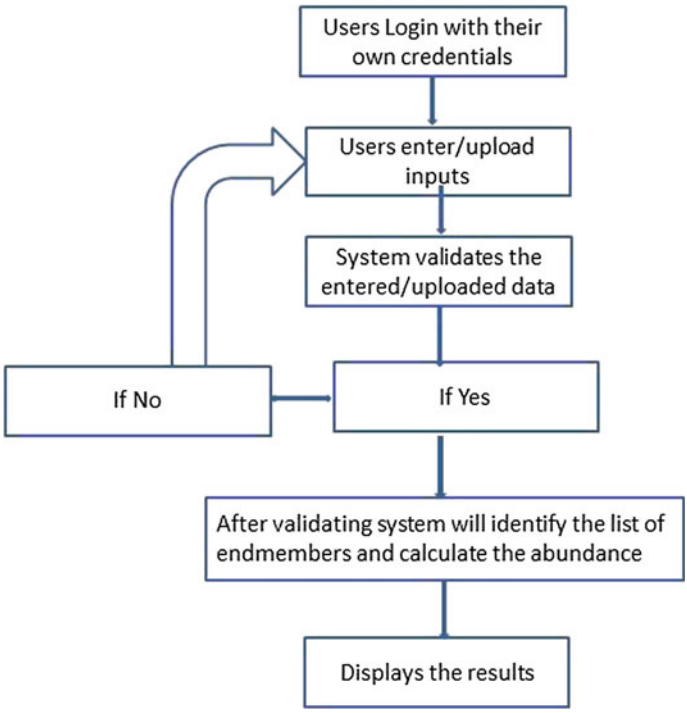


Fig. 3. Flow diagram of application

7 Result and Discussion

In order to check the accuracy of the reconstructed DEMs with respect to the reference SRTM DEMs, an accuracy assessment scheme is developed in-house. The validation of constructed high-resolution CARTOSAT DEM and SAR DEM was done with respect to **SRTM DEM** as "**Reference DEMs**" [7], using ERDAS Imagine s/w. The pixel spacing of the reference DEMs was resampled from 90 to 20 m to match with the spatial resolution of the reconstructed DEMs. Subsequently, the difference images were generated for common area, which are the product of pixel-to-pixel change detection of reference image from the corresponding subset of the target image under evaluation. When the height error is plotted against frequency of observation as histogram, it is observed that the histogram peak is shifted by few meters. This is because of datum shift as no external GCP was used in the initial images for registration. Hence, the plot

was shifted in X direction to account for highest number of pixels at 0 height error. Moreover, a correction factor was given to the reconstructed image based on the regression analysis results to scale the image with respect to "Reference DEM" and histogram was plotted based on the new difference image data. Subsequently, the error estimation was carried out on the difference images with respect to relief, slope, and slope-aspect categories. In below figure, the pixel deviation shown in red color has huge cloud patches. After replacing cloud patches with SAR data, the deviation has been improved which is represented in blue color (Fig. 4).

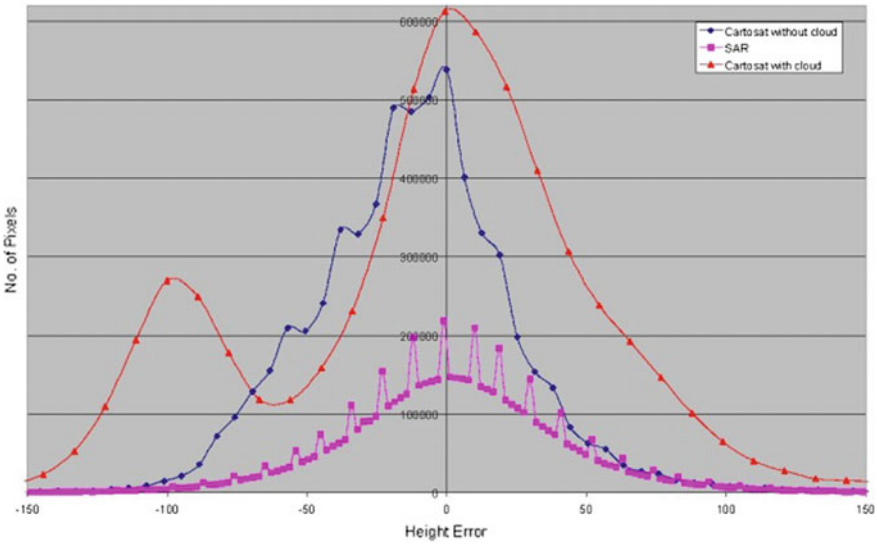


Fig. 4. Error histogram of CARTOSAT and SAR stereo-DEM

8 Conclusions

The extracted elevation values are compared with a derived topography of DEM available with DTRL and SRTM DEM [8]. While comparing all the possible combinations of the stereo-DEMs and CARTOSAT DEM, it is observed that the elevation results are better with less error in comparison with reference DEM using stereopairs of CARTOSAT data and the error standard deviation is consistently lesser for all types of terrain and slope conditions such as relative relief, slope, and slope aspect. The analysis of accuracy assessment revealed that the quality of DEM marginally degrades with increase in height of the terrain (SD 22.18 m, 31.13 m, and 33.41 m, respectively, for low, moderate, and high relief terrain). Comparison of DEM with reference to slope categories indicated that higher the slope, the poorer the results (SD 5.50, 19.93, and 19.66 m for Gentle, Moderate, and Steep Slope) although it demonstrated better results than absolute height ranges. The orientation of slope also plays a role in generation of

DEM [9]. The DEM of the region facing NW-N and W-NW has the most accurate pixels with least deviation (SD 9.59 and 11.52, respectively) from the reference height values, whereas S-SW and NE-E oriented slopes could generate least accurate DEM and their mean standard deviation in the difference image are 31.59 and 27.09, respectively. Better DEM can be achieved using field GCP, when compared to manually GCP/tie points collected. Comparative study of DEM accuracy shows more accurate that can be achieved using field GCP with respect to elevation and slope. Different data products can be used to remove bad data using data fusion/ merging techniques. Fusion of RADARSAT DEM with CARTOSAT DEM gives better result.

Acknowledgements. We are thankful to Dr. M. Bhutiyani, Director, DTRL DRDO, Dr. Sujata Dash Sc 'F' DTRL, DRDO, and other scientists DTRL, DRDO for providing us the data and guidance during our training period at their esteem organization.

References

1 www.gis.unbc.ca
2 Lee JS et al Polarimetric radar imaging: from basics to application, pp 5–7
3 Benz UC Multi resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS J Photogrammetry Remotee Sens
4 Farina A Introduction to radar signal & data processing: the opportunity, NATO, 063 bis
5 European Space Agency, ASAR user guide, Glossary
6 Burkholder RJ et al Comparision of monostatic and biastatic radar images
7 www.whereincity.com
8 www.north-east-india.com
9 Ballester JD et al (2010) Applying the Freeman Durden decomposition concept to polarimetric SAR interferometry. (IEEE) 8

Statistical Viability Analysis and Optimization Through Gate Sizing

K. Sreenath(✉) and S. R. Ramesh

Department of ECE, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Amrita University, Coimbatore, India
sreenathkl99@gmail.com, sr_ramesh@cb.amrita.edu

1 Introduction

The ultimate goal in the design of an integrated circuit is meeting some timing constraints. Through identification of critical paths, this goal can be achieved. The presence of false paths makes it very difficult to find true critical paths. If these false paths are the longest path in a circuit timing analysis tools estimate the maximum path delay based on these false paths. Hence, the resulting analysis reports would be too optimistic. This causes degradation in system performance. Several algorithms exist for the detection of false paths [1–3]. In order to detect false paths, each method has a particular sensitization condition. Static sensitization and dynamic sensitization are the main sensitization methods.

The two properties that timing analysis algorithm should satisfy are correctness and robustness. The property that a path sensitization method does not depreciate the longest path delay is correctness, which means true path should not be treated as false path. Robustness is defined such that once a sensitization condition detects a true critical path X in a slower die, there should not be any other path Y in any other faster die which is having delay more than the detected true critical path delay. Since timing properties vary from die to die, there can be some long paths in smaller dies which are true paths. Normal corner-based analysis fails to detect these paths.

Viability analysis [4] method which satisfies both correctness and robustness properties is one of the most appropriate false path detection methods so far. Since the analysis uses only the worst-case timing values, some of the essential timing information is missed while considering process variations. As the process parameter deviation increases, the circuit delay variation also increases. This gives pessimistic timing values using traditional static timing analysis (STA). This can be overcome using a statistical distribution of gate values while considering process variations. Statistical static timing analysis [5–8] (SSTA) is one of the fields where researches are going on from the last few years. Several algorithms exist to extract false path statistically [9]. However, in order to extract false paths under delay variations [10], only few works are present.

A method for false path detection and timing analysis using SSTA has been proposed by the authors of [11]. This consists of two phases. In the first phase, it removes all false path problems using the conventional false path filtering methods in [4]. In the second phase, it performs SSTA. In the first phase, it uses only worst-case values for

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_17

false path filtering. Hence, the results are still pessimistic. The viability analysis method [4] is enhanced to use under delay variation which can handle statistical gate delay distribution. It is used in this work. This method is able to identify true paths and false paths. Also, it gives the probability that a particular path being true or false under delay variation. The major application of this viability analysis technique is statistical timing analysis by considering the false paths under delay variations.

For further optimization of circuit delay, gate sizing is used. Based on the criticality of a particular gate to be in longest path, a set of gates is selected for sizing. In the existing works of gate sizing, the concept of false paths is not considered. Sizing without considering false path does not meet the actual yield expected. In this work, true critical paths are identified through viability analysis and gates for sizing are selected from these true critical paths.

2 False Path Detection

2.1 False Path

False paths are defined as paths through which signals cannot propagate through it. By identifying such paths and performing the timing analysis of a design after false path filtering gives the set of timing constraints for a circuit to be met, as shown in Fig. 1, in the path $\{a, v, w, x, z\}$, in order to propagate signal v to w, $b = 0$, to propagate w to x, $c = 0$, and to propagate x to output z, $y = 1$. But since $b = 1$ and $c = 1$, the output of y will be always 0. Therefore, the path under consideration $\{a, v, w, x, z\}$ is not excited. This path is an example of false path.

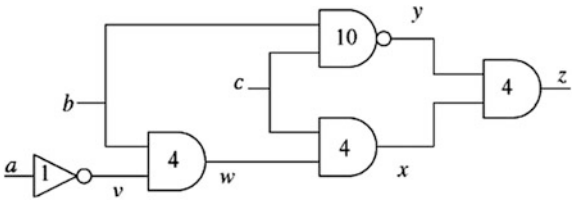


Fig. 1. Example for circuit containing false path [2]

2.2 Static Sensitization

Static sensitization is carried out by considering that all signals are stable. The signals from the input should propagate to output without getting masked by side inputs. For a path $P = \{f_0, \ldots, f_m\}$ to be a true path, when the value in f_i changes, then the value of f_{i+1} should also change. This condition can be expressed as a Boolean expression: $\frac{\delta f_{i+1}}{\delta f_i} = f_{i+1_{f_i}} \wedge f_{i+1_{\bar{f}_i}}$. All nodes in a path should satisfy this condition. Static sensitization may violate the property correctness by underestimating the critical time path delay. The method explained above is static sensitization.

2.3 Dynamic Sensitization

While static sensitization checks the condition $\frac{\delta f_{i+1}}{\delta f_i}$ statically (i.e., t = ∞), dynamic sensitization checks this condition at the actual time in which signal flows through the circuit. Consider Fig. 1, in which the values written inside each gate are its delay. Initially, *b* and *c* are 0 and at time *t* = 0, it changes to 1. Then, the value of *y* = 1 from *t* = −∞ to 10 and *y* = 0 after *t* = 10. Also, when the input "*a"* changes, it takes 9 (1+4 +4) unit time to reach *x*. Since *y* = 1 till *t* = 10, *x* could be propagated to z from *t* = 9 to *t* = 10 (i.e., *z* changes from *t* = 13 to *t* = 14). Therefore, the path {*a, v, w, x, z*} is true path under this method.

From this, it is clear that static sensitization violates correctness property. Dynamic sensitization follows correctness but it may violate robustness. Consider the circuit in Fig. 2.

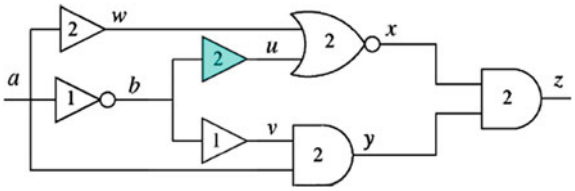


Fig. 2. Circuit to show robustness failure [2]

When "*a*" changes from 0 to 1 at *t* = 0, we have *u* = 1 from *t* = −∞ to 3, and *x* = 0. In the time period *t* = 2 to ∞, *w* = 1, again *x* = 0. As a result *x* = 0 all the time and thus *z* = 0 all the time. Now, when "*a*" changes from 1 to 0, *y* is 0 from *t* = 0, as a result *z* = 0 always. Hence, as per dynamic sensitization, there are no paths that can be sensitized in this circuit.

While considering delay variation, there can be faster die in the wafer where the circuit is fabricated. Assume the buffer shaded in Fig. 2 is having delay of 0 units. When a changes from 0 to 1 at *t* = 0, *u* = 0 from *t* = 1. Also, *w* = 0 from *t* = −∞ to 2. Therefore, *x* = 1 in the period *t* = 3 to *t* = 4. Also, during *t* = 2 to *t* = 4, *y* = 1. Thus, the output *z* = 1 during *t* = 5 to *t* = 6. It can be concluded that there can be slower paths in a circuit containing faster chips. Hence, dynamic sensitization fails to follow robustness property.

2.4 Viability Analysis

Viability analysis is the most accurate method for path sensitization which satisfies both correctness and robustness. In Fig. 2, the path {*a, v, w, x, z*} is not using static sensitization. However, this path can be sensitized through dynamic sensitization. The basic idea of viability analysis is as follows; if the side input having delay greater than path under consideration is presented to a particular gate, this input might sensitize $f_0, \ldots, f_m$ that gate which is driving the signal. The input which can propagate through the circuit is obtained by finding a set of path through viability analysis, which contains the paths through which the input vector $\forall$ can propagate. Such paths are viable paths.

The conditions for a path P = {f_0, ..., f_m} to be viable when an input vector $\forall$ is applied are such that for every f_i and for every $g! = f_{i-1}$ it is either of the following [4]:

1. g is set to its sensitizing value by $\forall$ or
2. g terminates a viable path under $\forall$ of delay length at least the delay length of partial path {f_0, ..., f_{i-1}}.

The viability function which is in the form of logic equations [4] uses dynamic method, which improves the accuracy of viable path identification. This function is modified so as to get the viable paths under delay variation. The notations for AND and OR operations in viability functions [4] are replaced with statistical min and max operations, respectively. While considering delay variation, the equations for viability are as follows.

$$\varphi_p = min_{i=0}^{m} \varphi_p^{f_i}, \tag{1}$$

where

$$\varphi_p^{f_i} = \max_{U \subseteq S(f_i, P)} \left(S_U \frac{\delta f_i}{\delta f_{i-1}} \cdot \min_{g \in U} \varphi^{g, \tau_i - 1}\right) \tag{2}$$

And $\varphi^{g,t}$ is

$$\varphi^{g,t} = \max_{Q \in P_g} (\min(\varphi_Q, \mathrm{d(Q)} - \mathrm{t})) \tag{3}$$

P_g represents those paths ends in g, while $d(Q)$ and t are timing distributions, such that $d(Q)$ is the delay of the path till Q. It can be defined that under delay variation, viability of path is the probability of φ_p to be greater than zero. Viability of a path depends on the input vector applied. Thus, viability of a particular path under all possible inputs in ($\forall_0$, …, $\forall_n$) can be expressed as

$$\text{Prob } [(max_{i=0}^{n} \varphi_p(\forall_i) > 0] \tag{4}$$

Algorithm for fast viability analysis [2]

```
Function Evaluate (Circuit Ckt, Input Vector Y)
    1: for every gate g ∈ Ckt do
    2: if visit[g] = false then
    3: Depth First Search (g,Y);
Function Depth First Search (Gate g, Input Vector Y)
    4: for every gate i ∈ input[g] do
    5: if i≠ a primary input and visit[i] = false then
    6: Depth First Search (i);
    7: for every gate i ∈ input[g] do
    8: for every path P ∈ Pi do
    9: update φ_P^g (Y)
    10: visit[g] ← true;
```

3 Gate Sizing

Gate sizing is a promising method used for timing optimization from many years ago. In order to meet the slack, gates in critical paths are upsized so that its delay reduces. Through this method, required timing constraints could be met. The major work in gate sizing is the selection of gates for sizing. Previous works have taken statistical timing approach for gate selection [11, 12]. However, concept of false paths is not taken care. This results in pessimistic timing values. Viability analysis gives the probability which can be directly related to the criticality of paths. Those gates in the long paths with high viability are the best candidates for upsizing. Gate sizing algorithm is as follows [11]:

```
1: Set all gates ←minimum size.
2: G = selected gates for sizing
3: for g ∈ G
4: Set, S_best = 0, g_best = 0
5: for each cell sizes g
7: Calculate reduction in delay Δdelay
8: Calculate increment in cost (area and power) Δcost
9: S = Δdelay/ Δcost
10: if S_cur>S_best then
11: S_best= S ,g_best = g
13: end if
14: end for
15: Increase size of best gate.
```

4 Results

The viability algorithm is implemented in C++, and the program is run on a Linux machine with Intel Core i5 processor (3 GHZ) and 8 GB RAM. This viability technique is experimented on ISCAS'85 benchmark circuits. Cell characterization has been done using Monte Carlo simulation. Statistical min and max operations are used for evaluating viability in Algorithm 1. Table 1 shows the timing distribution after SSTA and the number of critical paths. Critical paths are those paths having $\mu+3\sigma$ delay more than the $\mu-3\sigma$ delay of longest path. The results obtained through viability analysis are compared with SSTA after false path filtering using static sensitization. Table 2 shows these comparison results. It is clear that the results obtained in static sensitization violate the correctness property. Thus, it shows less number of true critical paths compared to viability analysis.

Gates for sizing are selected after doing viability analysis. The candidate gates which give best slack variation with less power and area variation are selected for sizing using Algorithm 2. Table 3 shows the improvement in delay after sizing. Reduction of delay in the range of 5–13% is obtained with a slight increase of power 0.4–1.3%, and an area overhead of 0.4–1.7%.

Table 1. SSTA results

Circuit	Total paths	μ (ns)	3σ (ns)	Critical paths
C432	83926	102.274	0.145077	1944
C880	8642	303.103	7.28919	84
C1908	729057	97.8307	2.98984	23296
C2670	679960	67.307	297982	9504
C3540	28676671	296.86	5.94956	2156
C5315	1341305	139.843	4.381595	98304
C7552	726494	136.419	2.390519	6144

Table 2. Comparison of viability with static sensitization

Circuit	Static sensitization			Viability analysis		
	True critical paths	μ (ns)	3σ (ns)	True critical paths	μ (ns)	3σ (ns)
C432	1462	94.02	1.6784	1506	94.02	1.6784
C880	71	267.56	2.3256	73	270.23	3.6545
C1908	19689	79.89	6.0564	20654	79.89	5.0564
C2670	7902	59.98	2.1045	8232	62.05	2.6548
C3540	1645	257.23	4.3656	1720	269.36	6.0125
C5315	67532	128.23	4.6547	70102	130.88	4.0023
C7552	4954	124.66	2.3256	5203	131.66	2.6545

Table 3. Gate sizing results

Circuit	Delay reduction (%)	Power increase (%)	Area overhead (%)
C432	13.6	0.89	1.71
C880	13.05	1.29	1.38
C1908	6.79	0.75	0.59
C2670	10.99	0.65	0.63
C3540	12.19	0.79	0.80
C5315	5.14	0.41	0.39
C7552	10.7	0.34	0.37

Gate sizing without false paths under consideration reduces the actual yield obtained. Without false path analysis when the gates in the longest paths are upsized, the result obtained in the actual true longest path is lesser than the expected result.

5 Conclusion

The timing analysis of VLSI circuits under delay variation is done using viability analysis. This method with SSTA framework is the best method for timing analysis by considering false paths. Analysis is carried out on ISCAS benchmark circuits, and results are compared with the values obtained by false path filtering using static sensitization. From the results it is clear that, while static sensitization underestimates true paths, viability analysis is the best method which satisfies both correctness and robustness. Using the criticality results obtained through viability, gate sizing is done. Through gate sizing up to 13%, delay reduction is obtained with very small power (1.3%) and area overhead (1.71%). In future, this viability analysis could be used for false path elimination in design, and thus area and power could be optimized. Also, similar to gate sizing, this analysis can be used for multiple threshold voltage assignment.

References

1. Brand D, Iyengar VS (1988) Timing analysis using functional analysis. IEEE Trans Comput 37(10):1309–1314
2. Jung J, Kim T (2013) Statistical viability analysis for detecting false paths under delay variation. IEEE Trans Comput Aided Des Integr Circuits Syst 32(1):111–123
3. Du DHC, Yen SHC, Ghanta S (1989) On the general false path problem in timing analysis. In: Proceedings of the IEEE design automation conference, pp 555–560
4. McGeer PC, Brayton RK (1989) Efficient algorithms for computing the longest viable path in a combinational network. In: Proceedings of the IEEE design automation conference, pp 561–567
5. Chang H, Sapatnekar SS (2003) Statistical timing analysis considering spatial correlations using a single pert-like traversal. In: Proceedings of the IEEE international conference on computer-aided design, pp 621–625
6. Bhaghath PJ, Ramesh SR (2014) A comparison on timing analysis using probabilistic approaches. In: Proceedings of the international conference on communication and signal processing, pp 419–423
7. Ramesh SR, Jayaparvathy R (2016) Improved statistical static timing analysis using refactored timing graphs. J Comput Theor Nanosci 13(11):8879–8884
8. Agarwal A, Blaauw D, Zolotov V (2003) Statistical timing analysis for intra-die process variations with spatial correlations. In: Proceedings of the IEEE international conference on computer-aided design, pp 900–907
9. Visweswariah C, Ravindran K, Kalafala K, Walker SG, Narayan S (2004) First-order incremental block-based statistical timing analysis. In: Proceedings of the IEEE design automation conference, pp 331–336
10. Liou JJ, Krstic A, Wang LC, Cheng KT (2002) False-path-aware statistical timing analysis and efficient path selection for delay testing and timing validation. In: Proceedings of the IEEE design automation conference, pp 566–569
11. Guthaus MR, Venkateswaran N, Visweswariah C, Zolotov V (2005) Gate sizing using incremental parameterized statistical timing analysis. In: Proceedings of the IEEE international conference on computer-aided design, pp 1029–1036
12. Jacobs ETAF, Berkelaar MRCM (2000) Gate sizing using a statistical delay model. In: Proceedings of the conference on design automation test in Europe, pp 283–291

A Decision Support System in Healthcare Prediction

Ashish Kumar Mishra, Pankaj Kumar Keserwani(✉), Shefalika Ghosh Samaddar, Harka B. Lamichaney, and Abhishek Kumar Mishra(✉)

Computer Science and Engineering Department, National Institute of Technology Sikkim, Ravangla, South Sikkim, India
pankaj.keserwani@gmail.com, am67633@gmail.com

1 Introduction

In the age of big data analysis, the analytical techniques of data get changed. Data mining and web mining are the related computing processes of data treatment in large data sets involving the methods capable of applicability over large data sets. A number of techniques are available which are amalgamation of artificial intelligence, machine learning, statistics and database systems [1]. The techniques developed being an interdisciplinary product are applicable to computer science as well. In computer science, such techniques are collected within the larger set of data mining and web mining methodologies. Data mining is used to extract information from a large data set (may not be big data) and derivation of newly revealed facts with using comprehensive structure. Data mining process involves a raw analysis step prior to taking care of database and data management requirement, data preparation and data preprocessing, corresponding model and interface generation, metric of interest complexity (polynomial or exponential both) consideration and post-processing for suitable inference according to the applicability in the discipline. Visualization of the inferences drawn and further assumption of the newly generated knowledge into history database, etc. are further derivation. Data mining particularly relates the analytical procedure which includes knowledge discovery in database (KDD) [2] process, or KDD (Fig. 1). The data mining methods are typical analytical treatment for generation of new knowledge from previously understood patterns after identification. At times, the processes may not be fully automatic but practitioners and scientist have to satisfy themselves with semi-automatic analysis of data giving interesting patterns of utility. The patterns can be a group of data having similar features as obtained in cluster analysis or some data that do not fit into any cluster or association between various data records as obtained in case of association rule mining, sequential pattern mining, etc. The techniques of spatial indices are used in data mining just like the database technique of spatial indices.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_18

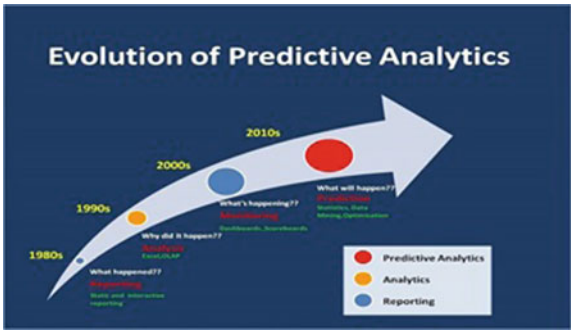


Fig. 1. Evolution of data mining [3]

In large data set, patterns are generally unidentified, resulting in the irrelevant output even when input data sets are correct. The data patterns are known for their repetitiveness, conditional occurrences and divergence of inference of interest. The patterns in a way are summary of input data for a particular preposition. The preposition proved using data mining techniques can further be applied in data mining and web mining. Such kind of pattern analysis may take machine learning and predictive analytical treatment of data. Identification of various clusters of data can go through the process for further refinement resulting in a prediction that may be utilized by a decision support system (DSS). The total process involves the stages such as data collection, data preparation, inference drawn from results and report generation. However, these stages are not the parts of the core data mining process though, and they contribute to the process of knowledge discovery of databases (KDD).

The term data mining was used and discovered in 1990s. The methods of data mining evolved through a number of processes resulting in pattern generation [4] of a field with a long history. Generation of patterns can be obtained by applying various techniques such as Bayes theorem (1700s), regression analysis where data follows a continuity [5]. The ever-increasing computing power contributed a lot towards optimized data collection, storage and processing. Data sets get enlarged with addition of newer dimension and become complex having multiple dimensions, analysis of data sets by hands-on direct processing. The automatic and semi-automatic data analyses become a part of data processing. Slowly analytical attempts were made through new data mining analytical techniques including the techniques of neural network, cluster analysis, genetic algorithm which evolved in 1950s, decision tree that involved in 1960s and support vector machine evolved later [6].

The generation and capturing of data for such experiment requires special mention. Amarjeet Singh Cheema et al. in the design of Blood Bank Management System discussed one of the challenges of data capturing in a uniform manner [7,8]. Creation of health records has been taken care for a cloud-based generic medical equipment interface [9]. After the introduction of the field of problem, Sect. 2 divulges different parameters of decision support system (DSS) in health care and their prediction capability. Section 3 proposes the actual DSS based

on various techniques already available; rather an intelligent assembly of the techniques has been proposed after running several experimentation to obtain the maximum probability of true predictive results.

The proposed system also contains the system and software used and total experimental setup. Section 4 is used for concluding remarks and future direction of work.

2 Background Study of Decision Support System (DSS) in Healthcare Prediction

Data mining extensively used in different kinds of predictive analysis, health care being one of the present sources of information gathering from 1990s. The techniques applied are support vector machine, neural network, regression analysis, naive Bayesian classifier, etc. and the patient databases are collected from all over the world [10]. The patient databases are aggregated as a matter of preparation and formats are processed in order to be consumed by all such techniques. In short, the databases are aggregated in nature and different authors identified different parameters from database for training and testing for supporting their propositions. The error function is also generated by various methods and compared. Xing et al.'s [11] survey of 1000 patients provided a pertinent result. They used SVM and obtained an accuracy of 92.1%. The same database while undergoing through ANN and decision trees produce accuracy of 91% and 89.6%, respectively. The parameters used are TNF, IL6, IL8, HICRP, MPOI, TNI2, sex, age, smoke, hypertension, diabetes and survival rate. A similar comparison of techniques has been conducted by Chen et al. [12]. They considered SVM, ANN, Bayesian classification, decision tree and logistic regression over a data set of 102 cases. The accuracy rates are 90.5%, 88.9%, 82.2%, 77.9% and 73.9%, respectively. Such comparative analysis is not very uncommon for the purpose of selection of best method to be applied to get best support for decision-making. Thus the level of automation of decision-making remains confined to decision support rather than automated decision-making [13]. Data mining is also known by the process of knowledge discovery that comes from analysing large data sets having different dimensions and in other one way summary information that can also be used for proposition testing such predictive analytic capabilities. Data mining is desired predictive component in all the aspects related to human life. The hidden patterns in such aggregated data sets are identified. Supervised and unsupervised learning techniques in classification classify items into predefined class label. The total process allows to build classification models from an aggregated data set revealing future data trends.

The present paper considers different data mining applications and classification algorithms for treatment analysis in health care for finding the efficiency of algorithm and improves upon the performance by applying data preprocessing and minimization of error function. Minimization of error function is inherent in all the methods of data mining and has been considered earlier. The prediction of new class level helps in the healthcare decision support system for further investment of efforts and resources by the authorities.

Following concepts are useful for understanding the algorithms and the working of knowledge generation from the designed DSS of healthcare prediction.

2.1 Information Gain

A large data set of patient record is partitioned into finite number of classes C_1, C_2, C_3, ..., C_i. These classes are the application of some categorical attributes for identification. The information needed to identify the class of an element of a partitioned set S is given by the formula:

$$I(S) = -(P \downarrow (1log \downarrow 2[P1+])P \downarrow (2log \downarrow 2P2) + P \downarrow (3log \downarrow 2P3).....P \downarrow (ilog \downarrow 2Pi))$$

P_i denotes the probability distribution of the partitions C_i. From such information of the entropy can be calculated as

$$\sum_{i=1}^{n} P_i = \frac{|C_i|}{|S|} C_i$$

This is accurate information gain in performance while branching with attribute A. A can be calculated with equation:

$$Gain(A) = I(S) - E(A)$$

where S is the complete set and A is that distributional category.

The data preprocessing techniques applied in the healthcare database is not novel, and it comes from computing the information gain for contribution decision tree, i.e. the part of proposed healthcare decision support system [14]. The data obtained after running the algorithms and various types of gain factors for sepal length and width are 0.44 and 0.0 and for petal 1 and 2, respectively [14].

Depending on the type of output generated, there are two types of algorithms in classification problem in data mining:

1. **Class output**: SVM and KNN typically create class outputs. A binary classification generates output either 0 or 1. The class output can be converted to probability output though its acceptability is questionable.
2. **Probability output column**: Generation of probability output column is an outcome of logistic regression, random forest, gradient boosting and Adaboost technique. Probability output can be easily converted into output by detecting a suitable value of threshold probability and providing discrete value for class output.

2.2 Area Under the ROC Curve (AUC ROC)

Calculation of the area under ROC curve is one of the interfacing metrics used in various decision support systems. The proposed paper also makes use of metrics for the training and testing of healthcare data set. The ROC curve is always independent of the change in proportional of responders. As long as the set of responders grows, proportionality of the decision support system proposed will generate a suitable predictive analysis.

Receiver operating characteristic (ROC) curve can be understood with the help of a probabilistic model having confusion matrix. In case of confusion matrix applicable in case of a probabilistic model, different values for each metric can be obtained from the same large data set. The occurrence of such confusion matrix may be because of the equal number of false positive and false negative which are difficult to determine using the designed error functions. Therefore, such sensitivity and different specificity are required by building designing error function. The sensitivity and specificity vary as follows (Figs. 2 and 3).

Confusion Matrix		Target			
		Positive	Negative		
Model	Positive	a	b	Positive Predictive Value	a/(a+b)
	Negative	c	d	Negative Predictive Value	d/(c+d)
		Sensitivity	Specificity	Accuracy = (a+d)/(a+b+c+d)	
		a/(a+c)	d/(b+d)		

Fig. 2. Confusion matrix

Hence, for each sensitivity, a different specificity is obtained. The sensitivity and specificity vary as presented by the following curve:

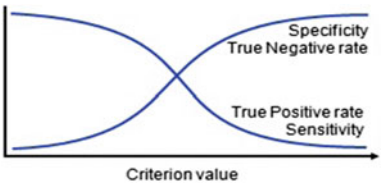


Fig. 3. ROC curve

The ROC curve is the typical plot between sensitivity and (1-specificity). 1-specificity is also known as a false positive rate, and sensitivity is also known as false negative rate. Knowing of such false positives and true positives is not difficult in case of training data set. The same can be translated into false negative and true negative in case of training data set. The data set that is organized is capable of giving the results in similar manner with the testing data set. The error function determined in a training data set will continue to give results in case of the testing the data. Therefore, the minimization of error function is not only a necessity but also efficiency of DSS depend on optimal value of error function.

3 Proposed Healthcare Decision Support System

The predictive analysis can be applied in healthcare system especially in case of prediction and monitoring. The data regulated and generated are again a feedback to the system producing required decision support in realizing health condition for working professionals. A medical camp is considered to take care of the stressful work life and other negligence to health issues by constant monitoring of health-related data and providing healthcare guidance on the basis of the decision support made available from the DSS installed at the medical camp that runs as a not-for-profit organization.

The basic structure for data modelling and extracting inferences can be understood as below (Fig. 4).

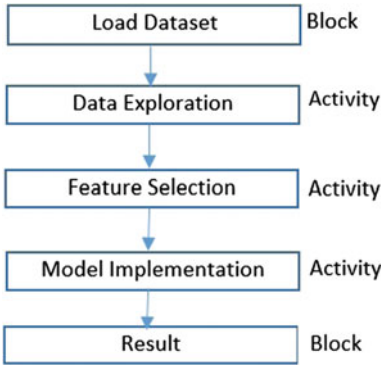


Fig. 4. Activity structure

Different types of machine learning algorithms for healthcare prediction are surveyed for their applicability and evaluated the individual performance by comparison. A comparative analysis of the performance has been obtained. The various algorithms that are used are presented below:

1. **GLM (generalized linear model)**: GLM, as the name suggests, is a generation of linear regression [15]. Linear regression considers response variables after data preprocessing. Therefore, this stage can be considered very much under data preparation. These response variables are having error distribution models. The error distribution models may vary, but never becomes a normal distribution. The GLM considers the linear model to be related to the response variable. The establishment of relation uses a magnitude link function. Each measurement is allowed to have magnitude of variance as a function of its predicted value. The authors of the paper [15] proposed an iterative re-weighted least square method for maximum likelihood estimation of the model parameters. Many statistical computing packages use maximum likelihood estimated method such as GLM.

2. **GBM (gradient boosting model)**: Gradient boosting [16] is a machine learning method. GBM has been widely used in regression and classification problems. The prediction model resulted from GBM is an ensemble of a number of weak prediction models. These weak prediction models are obtained by the method of decision trees. From the earlier comparative study, it was shown that decision tree produces the lowest level of accuracy. Thus, the weak prediction from these models can be understood. GBM contrue the model for usage in data mining in stage-wise manner and further generalizes it by using the optimisation gradient boost of an arbitrary differentiable loss function. It typically prepares the data set for minimal collection of false positive and false negative. Leo Breiman [17] observed the gradient boosting and he interpreted that boosting can be considered as an optimization algorithm with respect to a suitable cost function. The functional gradient view of boosting has been utilized in many boosting algorithms used in machine learning and statistics. The scope of use is not limited to regression and classification. GBM happens to be less accurate.
3. **XGBoost**: XGBoost [18] is an open-source library based on the above concept that provides gradient boosting framework for popular high-level languages and platforms. It supports distributed processing frameworks such as Apache Hadoop, Apache Spark and Apache Flink. Traditionally, it is slower than GBM.
4. **Deep learning**: The mechanism of deep learning consists of a number of machine learning algorithms known by its class. Together, they are known as deep structured learning. The depth of deep learning comes from the uses of a number of layers cascading one after another of non-linear processing units that are typically used for feature extraction. They are also used in case of required transformation [19]. Each layer uses the output of an earlier layer as an input successively. The technique is applicable in both supervised and unsupervised learning. The unsupervised learning gives interesting results in case of such pattern analysis. Even in case of supervised classification, deep learning is able to produce one of the higher rank predictions, and high-level features can be extracted from the low-level feature forming a hierarchical representation in case of deep learning.
5. **Ensemble model**: Ensemble model is not a single model but an assembly of models to generate an optimal model for a specific database. However, the assembling of models in one way is not an ultimate for all designed DSS; rather, it leaves a wide scope of assembling in a variety of ways adhering to the specificity of DSS requirements. Ensemble modelling can improve the performance of any model if a number of models taken together are paired and ordered in an optimized fashion. The built up of ensemble model is depends on the function of aggregation using the model in order to find empirically the well-ranked assembly of models. Ensemble models are used in competitions like Kaggle and there was obvious benefit from its usage. Ensemble modelling is the art of collecting an assembling diverse set of learners (individual model) together so that there can be stable model capable of generating improvement with specificity and sensitivity of an application. Such DSS supported by

ensemble modelling has been used in the proposed methodology of the paper. The derivation of an interesting assembly of learning models in the proposed DSS is one of the prominent outcomes of the paper.

3.1 Experimental Setup of Proposed DSS

1. **Software used for the development of DSS**: R Studio is the platform for obtaining experimental results and implementation. R Studio is a free and open source integrated development environment. For R, a programming language for statistical computing and graphics is available. R Studio is written in C++ language and uses the Q_t framework for its graphical users interface.
2. **R Studio setup**: The setup was installed in windows platform along with Java runtime environment and Java Development Kit (JDK). The different types of packages were installed from tools in R Studio for different algorithms such as Random Forest, deep learning, h2o, etc.
3. **Data collection**: All the real data have been collected from analytics Vidya Knockobter-2016 [20]. It was training data set and test data set. The link of the website from which database is collected: https://datahack.analyticsvidhya.com/contest/knocktober-2016/ [20].

3.2 Implementation of Decision Support System

Following are the depiction of the running of the experiment and the results obtained. The parameters used in the DSS are start data difference, category, next data difference, count patient, end data difference, donation flag, previous data difference, age, count patient data, education score, city type and income that is already being shown by a screenshot (Fig. 5).

	Feature	Gain	Cover	Frequency
2	Start_Date_Diff	0.190376507	0.202525785	0.17317612
3	Category2	0.164409213	0.063574689	0.04495210
4	Next_Date_Diff	0.077420878	0.127393566	0.13559322
5	Count_Patient	0.050844061	0.090040681	0.09395726
6	End_Date_Diff	0.039000877	0.101574297	0.14406780
7	Donation_Flag	0.038965203	0.033214802	0.02173913
8	Prev_Date_Diff	0.025103120	0.110155753	0.13154016
9	Age	0.018337879	0.051534341	0.07258659
10	Count_Patient_Date	0.012156870	0.046224263	0.03426676
11	Education_Score	0.011789702	0.041909051	0.04863670
12	City_Type	0.007531363	0.025454396	0.04495210
13	Income	0.005699562	0.007409346	0.02284451

Fig. 5. The parameters used in the DSS

Several sets of experiments are run, which are not presented here. The algorithms such as deep learning and GLM have not been utilized for the purpose of experimentation. However, an ensemble model can also be created using these models. The present paper presents the success scenario with the help of performance analysis using various techniques described earlier such as XGB, GBM and ensemble models of XGB and GBM.

In DSS with the help of XGBoost Algorithm, the above result shows that in Y-axis (i.e. Health campID) in between 6550 and 6590, a number of patients have visited for the maximum number of times. It also shows the probability to visit in respective camps is higher while patients with health camp ID except the range are not having more interest or willingness to attend the health camps as indicated by the probability of the learner (weak models) (Figs. 6, 7 and 8).

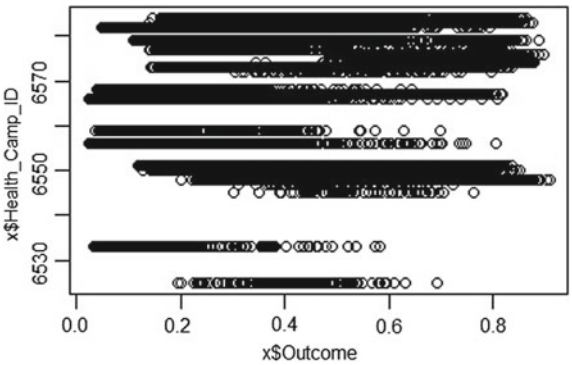


Fig. 6. Healthcare prediction performance using XGBoost

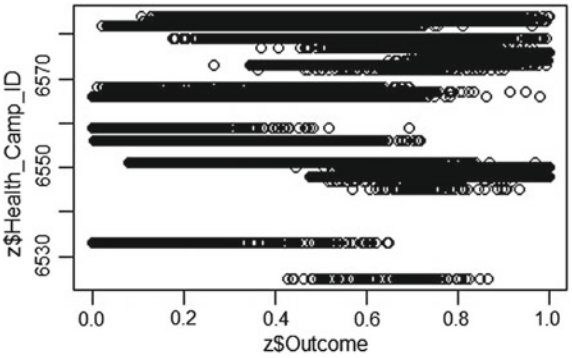


Fig. 7. Healthcare prediction performance using XGB + GBM

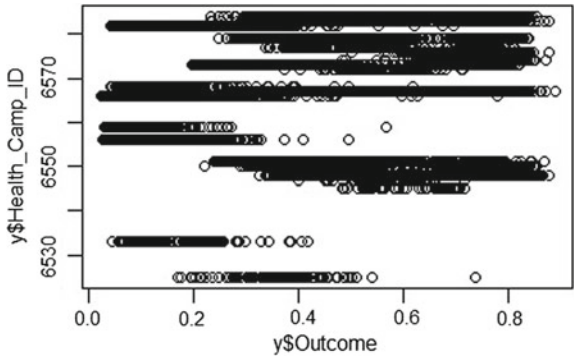


Fig. 8. Healthcare prediction performance using GBM

In DSS having the capacity to run different algorithms with the help of ensemble model of XGB+GBM Algorithm, the result shows that in Y-axis (i.e. Health campID) in between 6560 and 6580, a number of patients have visited for the maximum number of times.

In the designed DSS with the help of GBM Algorithm, the experimentation run, result shows that in Y-axis (i.e. Health campID) in between 6540 and 6580, number of patients have visited for the maximum number of times.

The different machine learning algorithms are chosen. The selection of the algorithm is considered on the basis of a number of rudimentary experimentation on combination. Only the results successfully obtained are reported with the success combination.

3.3 Comparative Analysis

The screenshots and results for comparative analysis are displayed using graph technique (Figs. 9 and 10).

S.I	Algorithm	Accuracy (%)	Ratio
1	XGB + XGB	78.16	0.48 + 0.52
2	GBM	77.15	1
3	XGB + GBM	76.85	0.5 + 0.5

Fig. 9. Comparative analysis of healthcare prediction

The first predictive model and two XGBoost models are used where the models differ on the set of parameters. The weighted average of the results obtained is equal to 1 (0.48 + 0.52) which is maximum probability output. The second predictive model uses GBM algorithm and is not ensembled in any other

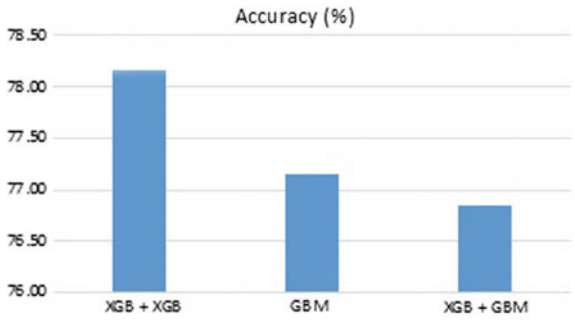


Fig. 10. Comparative analysis of healthcare prediction in bar chart

algorithm. The third model is a pure ensemble model of XGB and GBM, and the weighted average is 1 (0.50 + 0.50) which is again maximum but percentage accuracy is lower than the earlier ensemble model of probability output 1 (first model).

The comparative analysis of the healthcare prediction shown is able to achieve a result that is all-time high when compared to the study of the algorithms such as GBM, XGB and their ensemble models, and thus proved to be useful for healthcare prediction.

4 Conclusion and Future Direction of Work

Decision support system for healthcare prediction [21] using various data algorithms has gone through tremendous advances from the last few years but is yet to be practically implemented in an efficient manner. Hence, the individual and ensemble model performances of decision support system for healthcare prediction using various data algorithms such as GBM (gradient boosting model) and XGBoost are calculated, displayed using available visual techniques for easy understanding and compared. A comparative analysis of decision support system for healthcare prediction using different algorithms is able to show that ensemble modelling of XGB and XGB (well known for preventing overfitting of predictions) algorithm gives the best result as its time complexity is very less as well as the accuracy is comparatively higher in comparison to other algorithms and can be used as a premier technique in healthcare prediction as a decision support.

References

1. Tan P-N, et al (2006) Introduction to data mining. Pearson Education India
2. Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) The kdd process for extracting useful knowledge from volumes of data. Commun ACM 39(11):27–34
3. Smyth P, Pregibon D, Faloutsos C (2002) Data-driven evolution of data mining algorithms. Commun ACM 45(8):33–37

4. Teller A, Veloso M (1995) Program evolution for data mining. Int J Expert Syst Res Appl 8(3):213–236
5. Berry MJ, Linoff G (1997) Data mining techniques: for marketing, sales, and customer support. Wiley
6. Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Philip SY et al (2008) Top 10 algorithms in data mining. Knowl Inf Syst 14(1):1–37
7. Cheema, AS, Srivastava S, Srivastava P, Murthy B (2015) A standard compliant blood bank management system with enforcing mechanism. In: 2015 International Conference on Computing, Communication and Security (ICCCS), IEEE, pp 1–7
8. Ekanayaka E, Wimaladharma C (2015) Blood bank management system. Technical session-computer science and technology & industrial information technology, vol 7
9. Srivastava S, Gupta R, Rai A, Cheema A (2014) Electronic health records and cloud based generic medical equipment interface. arXiv:1411.1387
10. Shazmeen, SF, Baig MMA, Pawar MR (2013) Performance evaluation of different data mining classification algorithm and predictive analysis. IOSR J Comput Eng (IOSR-JCE) e-ISSN 2278
11. Xing Y, Wang J, Zhao Z et al (2007) Combination data mining methods with new medical data to predicting outcome of coronary heart disease. In: 2007 international conference on convergence information technology. IEEE, pp 868–872
12. Chen J, Xi G, Xing Y, Chen J, Wang J (2007) Predicting syndrome by nei specifications: a comparison of five data mining algorithms in coronary heart disease. In: Life system modeling and simulation, pp 129–135
13. Mythili T, Mukherji D, Padalia N, Naidu A (2013) A heart disease prediction model using svm-decision trees-logistic regression (sdl). Int J Comput Appl 68(16)
14. Tomar D, Agarwal S (2013) A survey on data mining approaches for healthcare. Int J Bio-Sci Bio-Technol 5(5):241–266
15. Elder J (2009) Handbook of statistical analysis and data mining applications. Academic Press
16. Seni G, Elder JF (2010) Ensemble methods in data mining: improving accuracy through combining predictions. Synth Lect Data Min Knowl Discov 2(1):1–126
17. Breiman L, Friedman J, Stone CJ, Olshen RA (1984) Classification and regression trees. CRC press
18. Chen T, Guestrin C (2016) Xgboost: a scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 785–794
19. Witten IH, Frank E, Hall MA, Pal CJ (2016) Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann
20. Vidhya V, Analytics. https://www.analyticsvidhya.com/
21. Sales AE, Liu C-F, Sloan KL, Malkin J, Fishman PA, Rosen AK, Loveland S, Nichol WP, Suzuki NT, Perrin E et al (2003) Predicting costs of care using a pharmacy-based measure risk adjustment in a veteran population. Med Care 41(6):753–760

Random Forests in the Classification of Diabetic Retinopathy Retinal Images

Amrita Roychowdhury(✉) and Sreeparna Banerjee

Maulana Abul Kalam Azad University of Technology, BF-142, Sector 1, Salt Lake City, Kolkata 700064, West Bengal, India
amrita.me.cse@gmail.com, sreeparnab@hotmail.com

1 Introduction

Rapid advances in health sciences have resulted in the spooling of a huge amount of data, clinical information, and generation of electronic health records [1]. Machine learning and data mining methods are being used to perform intelligent transformation of data into useful knowledge [1]. Application of machine learning in medical imaging has become one of the most interesting challenges of the researchers. In the medical domain, description of a set of diseases in terms of features is supplied to the machine learning classifiers as knowledge base. Depending on the supplied training set, the classifier has to identify the disease for a test set. Performance measures specify the correctness of the classifier in identification of disease from an unknown situation. In this research work, signs of the diseases at different stages of Diabetic Retinopathy (DR) are considered as classes. The first stage of DR is silent in nature as no such clear symptoms are noticeable. The starting of DR includes deformation of retinal capillary resulting in very small spots known as microaneurysms (MA). The next stage contains hard exudates (HE) which are lipid formations from the fragile blood vessels. As DR advances, cotton wool spots (CWS) are formed which are micro-infarcts caused by obstructed blood vessel. Hemorrhages can also occur with further progression of the disease when blood vessels leak blood into retina causing hemorrhages (HAM). Due to poor oxygen supply, new vessels are formed in this state, which challenges the patients eyesight. These stages of DR which comprise the different classes of a machine learning classifier are referred to as Non-proliferative Diabetic Retinopathy leading to Proliferative Diabetic Retinopathy. This is a supervised learning problem owing to the fact that the number of objects to be classified are finite and predetermined [2]. The research aims to find the significance and acceptability of Random Forest classifier to the semiautomated detection of different cases of Diabetic Retinopathy. The performance of the classifier is calculated in terms of number of correct and incorrect classification, and the average accuracy is 99.275%.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_19

2 Related Work

Machine learning can be classified as supervised and unsupervised. In supervised machine learning, every input field is associated with its corresponding target value. On the contrary, unsupervised machine learning deals with only input fields. We use Random Forest, a supervised machine learning technique, to classify the different signs (i.e., lesions) of DR. Random Forest (RF) also known as random decision forests are an ensemble learning method used for classification. This ensemble classifier is a combination of tree predictors. Each tree depends on the values of a random vector which is independently sampled with the same distribution for all the trees of the forest [3]. A number of such decision tree classifiers are applied on subsamples (with same size of original input sample size) of the dataset and averaged in order to improve the predictive accuracy as well as controlling overfitting. A scheme of fusing ill-focused images using Random Forest classifier is proposed in the research work of Kausar et al. [4]. The well-organized images are useful for image enhancement and segmentation. This work aims to generate all-in-focus images by Random Forest classifier. Visibility, spatial features, edge, and discrete wavelet transform are considered as features for machine learning. This scheme outperforms previous approaches like principal component analysis and wavelet transform. Saiprasad et al. [5] used Random Forest classifier to classify pixel-wise an image of abdomen and pelvis into three classes: right adrenal, left adrenal, and background. For this purpose, a training set is formed with a dataset of adrenalin gland images. Manual examination and labeling of a radiologist are used as ground truth. The classification phase is combined with the histogram analysis phase for more accurate result. A Random Forest-based approach for classifying lymphatic disease diagnosis is developed in the report of Almayyan [6]. Segmentation and classification of very high-resolution Remote Sensing data with Random Forests are mentioned in the work of Csillik [7]. Random Forest has been used successfully in the classification of Diabetic Retinopathy Diseases in one of our earlier papers [8], which concludes Random Forest as better than Naive Bayes classifier and support vector machine.

3 Proposed Methodology

A database of 69 images containing different stages of DR is considered as input dataset. The dataset is formed collecting the images from DIARETDB0 [9] and DIARETDB1 [10]. These images contain bright lesions as hard exudates (HE) and cotton wool spots (CWS) as well as dark lesions like hemorrhages (HAM) and microaneurysms (MA). The retina image database contains 25 images of HE, 15 images of CWS, 15 images of HAM, and 14 MA images. In Fig. 1, the pictorial representation of the system is given.

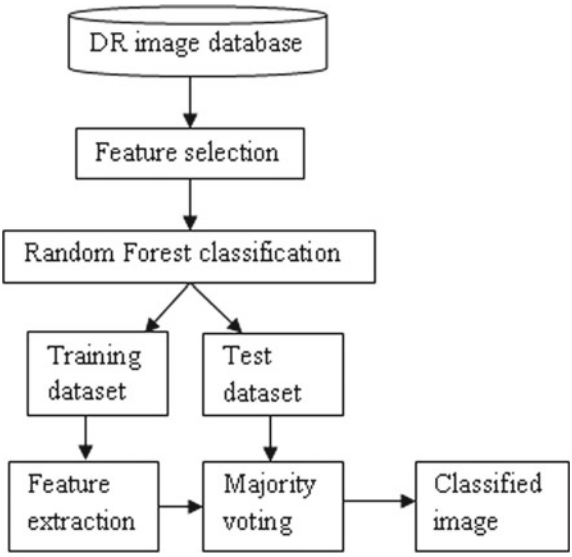


Fig. 1. Graphical representation of the proposed system

3.1 Feature Selection

Several consultations with retina specialists were carried out to identify the process of distinguishing a variety of DR. According to those suggestions, nine features among which six texture-based features and the remaining statistical measure-based features are considered for constructing the training set for Random Forest classifier. The texture-based features are as follows: color, frequency, shape, size, edge, and visibility. Color feature is useful for differentiating dark lesions from bright lesions. Among bright lesions also, for HE, it is generally bright yellow and for CWS, it is generally whitish yellow. In dark lesion cases, MA is normally light red in contrast to the dark red color of HAM. The frequency of occurrence is also important as HE comes in large numbers compared to CWS. Shape and size of MA vary from HAM in a significant manner. The sharpness of edge is also significant as edge of CWS is blurred whereas the edge of HE is sharp. Visibility in terms of human eye is considered as another feature. HE and HAM have a high visibility compared to CWS and MA. Figure 2 shows some variety of DR images. The texture-based features are manually extracted from the original input image. No preprocessing is done on the original input image before feature extraction as the texture-based features are directly accessible from the input image without losing their originality. Table 1 describes the features along with their values for different diseases. Three statistical measure-based features are as follows: mean, standard deviation, and entropy. Mean is the average brightness of an image. Standard deviation describes the distance between ith pixel from mean. Another statistical measure is entropy which is the randomness of an image to realize the texture of the image. These features are defined in equations [1–3].

$$Mean(\mu) = \frac{1}{N}\sum_{i=0}^{N-1} x_i \tag{1}$$

$$Standarddeviation(\sigma) = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}(x_i - \mu)^2} \tag{2}$$

$$Entropy = -\sum(p. * log_2 p) \tag{3}$$

In Eqs. (1) and (2), N is the total number of pixels in the input image, and x_i is the gray value of ith pixel. In Eq. (3), p contains the histogram counts of the image and operator.* indicates element-by-element multiplication. In Eqs. (1) and (2), N is the total number of pixels in the input image, and x_i is the gray value of ith pixel. In Eq. (3), p contains the histogram counts of the image and operator.* indicates element-by-element multiplication.

The feature values of each lesion of an input image are evaluated along with the disease type and this contributes a row for the training set. This training set is supplied to the RF classifier. RF classifier is not used for selection of features. Feature selection is totally based on the working principle of retina specialists and literature survey.

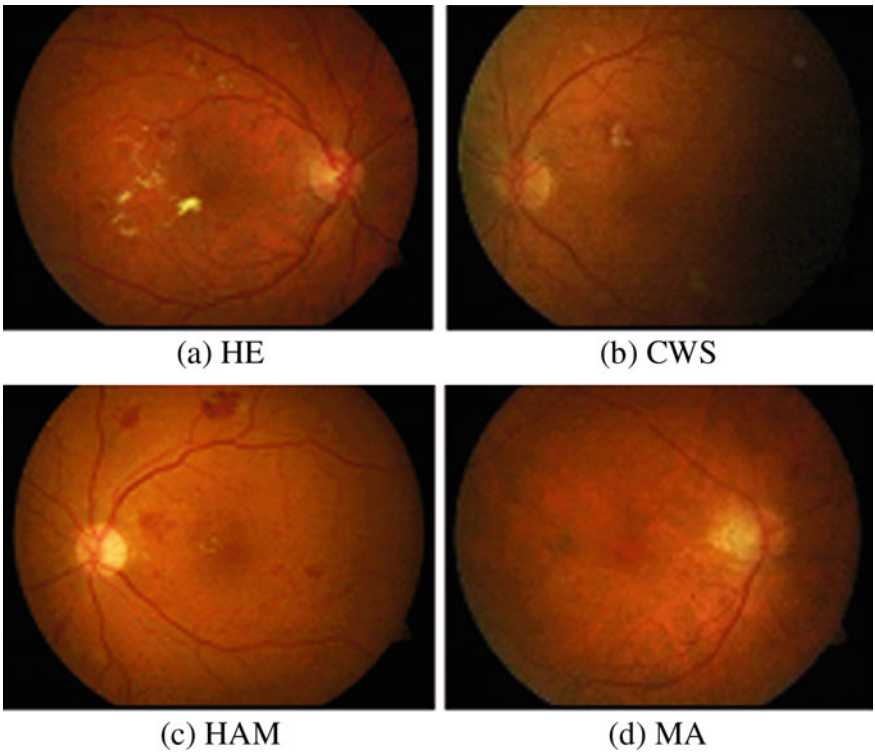


Fig. 2. Various types of DR images

Table 1. General feature values of different DR images

Disease	Color	Frequency	Shape	Size	Edge	Visibility
HE	Yellow/bright yellow	Many	Circular/not defined	Small/ medium	Sharp	High
CWS	Whitish yellow/white	Few	Not defined	Medium	Blunt	Medium/ high
HAM	Dark red	Less	Not defined	Large/ medium	Sharp	High
MA	Light red	Many	Circular	Small	Blunt	Low

3.2 Random Forest Classifier

In Random Forest classification technique, several trees are grown together. For classifying a test data, the new data is added down each of the trees. Each tree generates a class for the test data which is called voting for the class. The most voted class is selected by the classifier as final class of the test data. Random Forest classification is the most popular ensemble model classifier as it runs on large database in a time-saving approach. This model is not suitable to the cases dealing with completely new data.

Random Forest classifier with percentage split (66%) is used for machine learning. Two-third data of the total dataset is selected as training dataset which helps the growing of tree. Cases are selected at random with replacement, that is, the case considered for a tree can be reassigned to another tree. Generally, the square root of total number of feature variables is selected at random from all the prescribed feature values. This value remains fixed during the growth of the forest. The best split on these selected feature variables is used to split a node. The remaining 1/3 data is considered as test dataset. This dataset is called out-of-bag (OOB) data. For each test data, each tree generates a class that is counted as the vote for that class of the test data. The class with maximum votes is assigned to the test data. The term Random is associated in two ways with this classifier: random selection of sample data and random selection of feature variables. It is the characteristic of Random Forest classifier that it needs no separate test dataset. The OOB data is used for calculating the error internally at the time of construction. When each tree of the forest is grown, then OOB cases are put down the tree and the number of votes for the correct class is calculated.

3.3 Algorithm of Random Forest Classifier

Step 1: Initialize total number of classes to L and total number of feature variables to M.
Step 2: Let m be the number of selected feature variable at a node (generally $m = \sqrt{M}$).
Step 3: For each decision tree, randomly select with replacement a subset of

dataset containing L different classes.
Step 4: For each node of a decision tree, randomly select m feature variables to calculate the best split and decision at this node.

4 Results and Analysis

The performance analysis of Random Forest classifier is carried out with the basis of confusion matrix. In Weka 3.7 [11] machine learning classifier, Random Forest model is selected for analyzing the input dataset. The database contains 69 images. As Random Forest classifier automatically selects 2/3 data as training set and the rest as test set, 46 data are treated as training dataset, and 23 data are considered as test set. After analyzing the dataset, a confusion matrix is generated by the classifier itself.

TP_i (True Positive): number of members of class X_i correctly classified as class X_i,

FN_i (False Negative): number of members of class X_i incorrectly classified as not in class X_i,

FP_i (False Positive): number of members which are not of class X_i but incorrectly classified as class X_i,

TN_i (True Negative): number of members which are not of class X_i and correctly classified as not of class X_i,

Accuracy (Acc): the ability of the classifier to correctly classify a member

$$Acc_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \times 100 \tag{4}$$

Sensitivity (Sn): the ability of the classifier to detect the positive class

$$Sn_i = \frac{TP_i}{TP_i + FN_i} \times 100 \tag{5}$$

Specificity (Sp): the ability of the classifier to detect the negative class

$$Sp_i = \frac{TN_i}{TN_i + FP_i} \times 100 \tag{6}$$

F-measure: statistical test of prediction accuracy of the classifier in terms of Precision and Recall

$$F - measure_i = \frac{(2 \times Precision_i \times Recall_i)}{(Precision_i + Recall_i)} \tag{7}$$

where $Precision_i = \frac{TP_i}{TP_i + FP_i}$ and $Recall_i = \frac{TP_i}{TP_i + FN_i}$.

Mathew Correlation Coefficients (MCC): estimate of over prediction and under prediction

$$MCC_i = \frac{(TP_i \times TN_i) - (FP_i \times FN_i)}{\sqrt{(TP_i + FP_i) \times (TP_i + FN_i) \times (TN_i + FP_i) \times (TN_i + FN_i)}} \tag{8}$$

MCC generates three different types of values: −1, 0, and 1. The value −1 means the classifier's prediction is incorrect, 0 means random prediction, and 1 means fully correct prediction. In Table 2, TP, TN, FP, and FN are calculated using the values of confusion matrix. Table 3 represents different performance measures for each class. For each class, receiver operating characteristic (ROC) is plotted. An ROC curve is plotted with false positive rate along X-axis against true positive rate along Y-axis. Figure 3 shows ROC for class CWS.

Table 2. Calculation of TP, TN, FP, and FN

Class	Total	TP	FP	TN	FN
CWS	15	14	0	54	1
HE	25	25	1	43	0
HAM	15	15	0	54	0
MA	14	14	0	55	0

Table 3. Performance measures

Class	Sensitivity (%)	Specificity (%)	Accuracy (%)	F-measure	MCC
CWS	93.33	100	98.55	0.966	0.957
HE	100	97.73	98.55	0.98	0.969
HAM	100	100	100	1	1
MA	100	100	100	1	1
Average	98.33	99.43	99.275	0.986	0.98

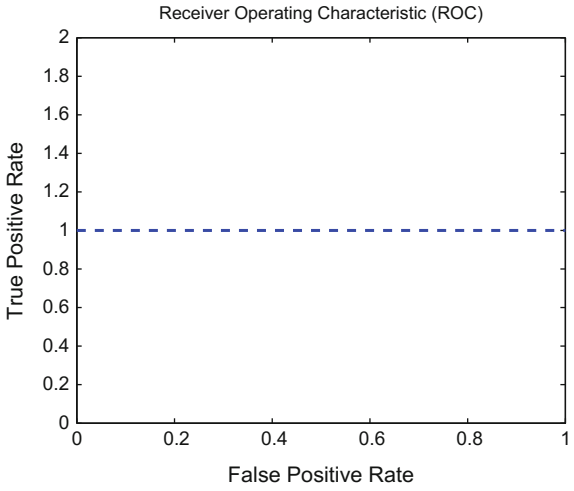


Fig. 3. ROC for class CWS

5 Conclusion

In this research work, Random Forest classifier is used for determining different stages of retinal abnormalities due to DR using machine learning techniques. Being an ensemble classifier, Random Forest constructs several decision trees at training time and generates the classification for each tree. In this research work, a dataset containing several retinal images having abnormalities is formed. The images are collected from various sources like DIARETDB0 [9] and DIARETDB1 [10]. A set of nine features including three statistical features and six texture-based features are selected for the machine learning. For each input image, the feature values are calculated, and thus, a dataset is formed for 69 images. In Weka 3.7 [11], the dataset is supplied as input to the Random Forest classifier. Performance measures like accuracy, sensitivity, and specificity are calculated depending on the classification result. The accuracy of HAM and MA classes is 100% each as the feature size distinctly separates these two classes. The size of HAM is considered to be medium to large while MA is very small. In future, a collection of large database with more features can be added to the feature set, and the number of images can be increased to get more complex training set for the classifier. The average accuracy is 99.275% which is promising.

References

1. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I (2017) Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J 15:104–116

2. Jingar R, Verma M (2012) Semantic segmentation of an image using random forest and single histogram class model. Technical report, Indian Institute of Technology Kanpur
3. https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf
4. Kausar N, Majid A (2016) Random forest based scheme using feature and decision levels information for multi-focus image fusion. Pattern Anal Appl 19:221–236
5. Saiprasad G, Chang C, Safdar N, Saenz N, Seigel E (2013) Adrenal gland abnormality detection using random forest classification. J Digit Imaging 26:891–897
6. Almayyan W (2016) Lymph disease prediction using random forest and particle swarm optimization. J Intell Learn Syst and Appl 8:51–62
7. Csillik O (2017) Fast segmentation and classification of very high resolution remote sensing data using SLIC superpixels. Remote Sens 9(3):243
8. Roychowdhury A, Banerjee S, Machine learning in the classification of diabetic retinopathy lesions from abnormal retina images (submitted)
9. Kauppi T, Kalesnykiene V, Kamarainen J-K, Lensu L, Sorri I, Uusitalo H, Klviinen H, Pietil J, DIARETDB0: evaluation database and methodology for diabetic retinopathy algorithms, Technical report
10. Kauppi T, Kalesnykiene V, Kamarainen J-K, Lensu L, Sorri I, Raninen A, Voutilainen R, Uusitalo H, Klviinen H, Pietil J, DIARETDB1: diabetic retinopathy database and evaluation protocol, Technical report
11. Frank E, Hall MA, Witten IH (2016) The Weka Workbench, Online Appendix for Data Mining : Practical Machine Learning Tools and Techniques, 4th edn. Morgan Kaufman

Performance Evaluation of M2M and H2H Communication Coexistence in Shared LTE: A DLMS/COSEM-Based AMI Network Scenario

Rajesh Chintha[1(✉)] and V. Kumar Chinnaiyan[2]

[1] Central Power Research Institute, Bangalore, India
rajeshchintha@gmail.com
[2] KPR Institute of Engineering and Technology, Coimbatore, India

1 Introduction

A typical architecture of an AMI network is shown in Fig. 1, which is composed of three main blocks: (1) the local data collection units, (2) the communications network, and (3) the management and control center, the energy company. According to Fig. 1, communication networks constitute the "backbone" of an AMI architecture in which AMI is the core device and acts as a gateway between Home Area Network (HAN) and Neighborhood Area Network (NAN).

With the rapid deployment of smart meters into the AMI networks; M2M in LTE, relying on cellular networks imposes overload issues, the network congestion, increases delays, packet loss, and even causes service interruptions [2].

In this paper, we investigate on the performance of LTE communications for smart meter traffic, in the presence of LTE background traffic such as VoIP, HTTP browsing, FTP, video streaming, and gaming. We assume a public, shared LTE network for data transfer; therefore, the amount of background traffic has a big impact on the latency of the smart metering traffic and background Internet traffic. A series of experiments were conducted with the increase in number smart meters and LTE background traffic applications to verify the quality requirements as specified in [3]. The performance metrics for LTE such as throughput, packet loss ratio, and point-to-point (P2P) delay are investigated using network simulator version 3.

The remainder of the paper is organized as follows: In Sect. 2, we describe the related work in the literature followed by AMI component models and LTE Internet traffic models implementations in NS-3. In Sect. 4, the results obtained from these simulation experiments are presented, followed by implications of the findings and future work.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_20

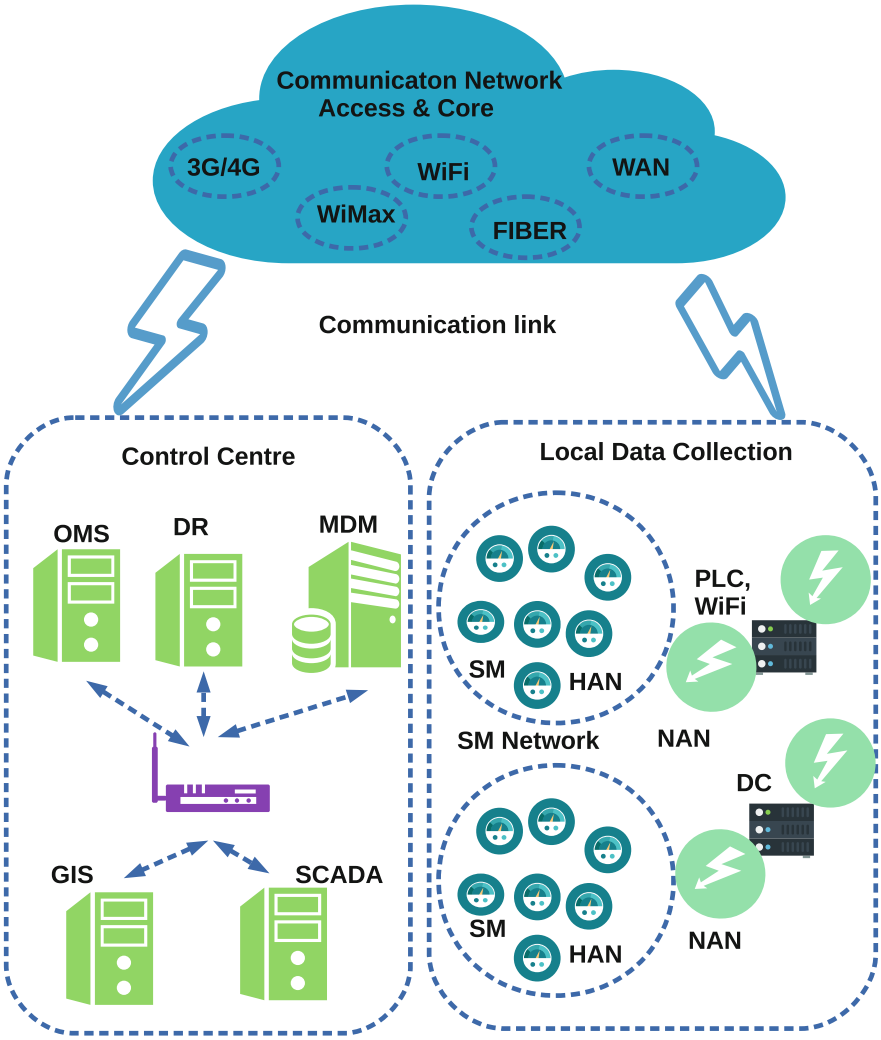


Fig. 1. Architecture of an AMI network [1]

2 Related Work

Reviewing the literature on the subject, we found similar works focusing on the evaluation of the performance of communication schemes and protocols for measurement reading networks or AMR (automated meter reading)/AMI implemented on the communication technologies through PLC (power line communication) [4], cellular [5], and the DMLS/COSEM protocol. In [6] the performance of ZigBee, LTE, and WiFi under different sizes, node layouts, and different transmission sequences is investigated to compare the maximum distance for

transmitting data between DCU and AMI devices under log-distance channel propagation of three wireless technologies. In machine-to-machine (M2M) networks, a scheme is proposed in [7] which dynamically partitions and allocates the random access channel (RACH) resource to overcome the overload problem caused by random access requests from massive M2M devices defined by the Third Generation Partnership Project (3GPP).

To improve the accuracy of routine that simulates the LTE random access channel (RACH) was proposed in [8], which was subsequently used in our work to assure the QOS requirements are satisfied for the video services. So authors have realized that so far, there was no real attempt has been made to study the impact of background traffic on the performance of smart meter functions and vice versa, especially in public, shared LTE networks. Therefore, present work is an attempt to evaluate the mutual impacts between smart meter traffic and background Internet traffic, and is aimed at network/mobile operators who want to know the impact of AMI traffic through cellular communications infrastructure especially in the presence of existing background Internet traffic.

3 AMI Component Models in NS-3

The basic COSEM model proposed by Juan M. Aranda [9] for an intelligent electric meter with total active energy (kWh) Extended register was used. All smart meters (Fig. 2) work behind the DCs, so we have created WiFi links with the help of WifiHelper and lower layers are configured with the help of YansWifiPhyHelper and NqosWifiMacHelper. As shown in Fig. 2, the LTE stack is also installed on the DC that allows it to connect to the MDM to relay (route) the traffic of the smart meters from the WiFi link. The MDM host is implemented as a node connecting to the P-GW of LTE network through a point-to-point link

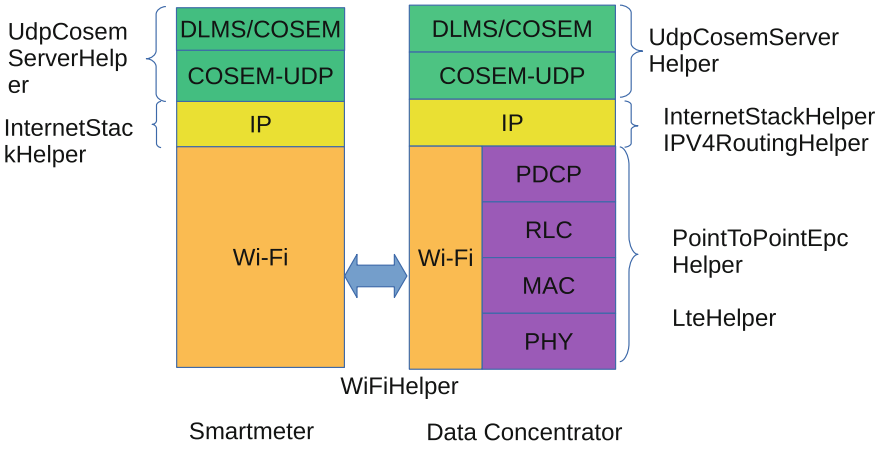


Fig. 2. Implementation of the WiFi smart meter and WiFi/LTE DC in NS-3

(using PointToPointHelper). At the application level, meter data management (MDM) system has the functionality intended for the management of meter reading data. Similar to the SM and MDM node implementations, the background UEs have LTE network interface, IP stack, and LTE stack installed. After that, different applications mentioned in Table 1 are installed on the nodes with the help of GeneralUdpServerHelper [10]. A CSMA network is implemented with MDM host and remote hosts of different traffic types using C++ classes CsmaHelper and GeneralUdpClientHelper.

Table 1. Background traffic mix

Application	Traffic	% users
VOIP	Real-time	30
FTP	Best effort	10
HTTP	Interactive	20
VIDEO	Streaming	20
GAMING	Interactive real-time	20

4 Experiments and Evaluation

4.1 The Simulation Scenario in NS-3

The simulation scenario as shown in Fig. 3 was proposed to evaluate and analyze the performance of communications for an AMI network.

The simulation experiments run on NS-3 version 3.26 and Ubuntu 16.04 LTS. Simulation experiment assumes point-to-point connection between S-GW/P-GW and remote hosts (including MDM). Simulation parameters defined by [11]

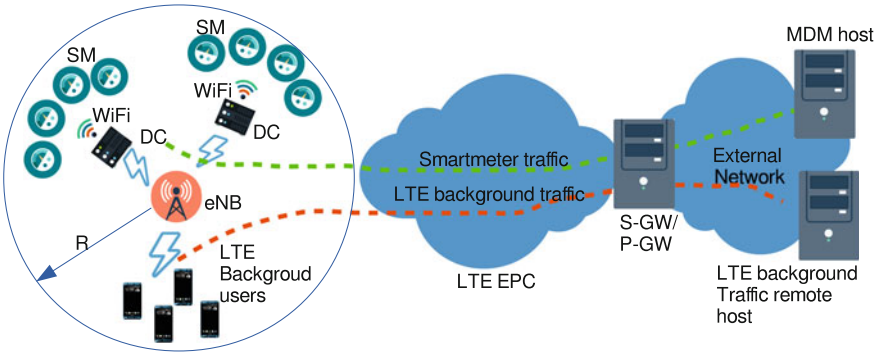


Fig. 3. Single cell simulation topology

Table 2. Simulation parameters

Parameters	Values
Uplink bandwidth	5 MHz
Downlink bandwidth	5 MHz
Transmission mode	MIMO 2 × 2
UE transmission power	26 dBm
eNB transmission power	49 dBm
Cell radius(R)	800 m
Simulation time	2000 s
No of eNB	1
No of DC	5
DC polling time	1 min
MDM request to DC	3 min

were used in the experiments and are shown in Table 2. The proposed scenario consists of the following:

- Five local networks of smart meters (SMs), belonging to a low voltage distribution network, managed by data concentrators (DCs).
- Five DCs with a capacity of up to 1,000 SMs distributed over an urban area (mostly residential and commercial) and with the ability to execute polling mechanisms to request the measurements from SMs every 60 s.
- A single cell of radius R is configured with the models and parameters of Table 2. It has the ability to simultaneously service the DCs and several fixed users that demand Internet applications.
- A metropolitan network implemented over optical fiber, capable of transporting the traffic generated by SMs and remote hosts.
- Five remote hosts (VOIP, HTTP, FTP, video streaming, and gaming) and the smart grid control center (MDM) are connected to a router. The control center (MDM) runs the RequestingAPP application, which is responsible for requesting and receiving DC data every 180 s.

4.2 Simulation Results and Analysis

This section presents the simulation results and analysis for the conducted experiments. All plots were drawn using GNU Octave, version 4.0.0. Three different cases for simulation were considered.

Fixed Background Traffic and Variable AMI Traffic Clearly, it can be seen in Fig. 4 that both the average throughput and the average point-to-point

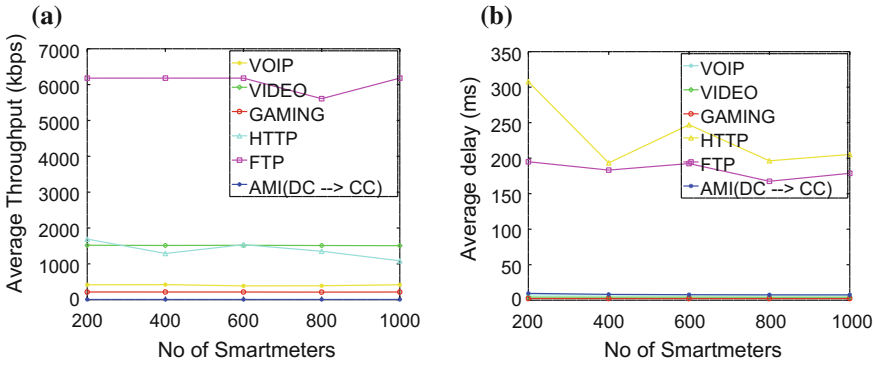


Fig. 4. Performance metrics per application with fixed (100 users) background nodes. **a** Throughput and **b** point-to-point delay

delay per application are kept practically constant in the face of the increase in the number of SMs per network.

With respect to the AMI traffic through the LTE network, it is observed that the average throughput in the DC and control center (DC -> CC) direction increases as the number of SMs increases.

On the other hand, the total average AMI delay (Fig. 5a), i.e., the time spent from the moment the request is issued by the control center until later receives the data transferred by the DC, is 10.79 ms, during which time an average of 7.75 Kb was transferred for each request made by the CC (for a total of 1,000 SMs per DC) (Fig. 5b). This delay is below the limit established by the QoS (Table 3). The packet loss ratio (PLR)for different applications including AMI (Fig. 5c) is constant with increase in number of smart meters and offers a high reliability for all applications. Therefore, under this scheme, a high percentage of coverage is achieved without overloading the network and guarantees the quality of the services offered. This fact may motivate telephone operators to offer their telecommunications infrastructure services to AMI network operations without much infrastructural changes.

Fixed AMI Traffic and Variable Background Traffic Figure 6 shows the simulation results obtained for different traffic loads of mobile Internet applications. Clearly, performance metrics for video applications were deteriorated as traffic increases within the cell, and a point was reached where it is no longer possible to guarantee the quality requirements for the video service. With 150 users within the cell, an average point-to-point delay of more than 250 ms was observed, i.e., the QoS requirement was not satisfied (Table 3) and a percentage of received packets of 98%, i.e., reliability of packet transfer was reduced by 5.88% compared to the confidence obtained in the previous scenario (Fig. 5c). Additionally, the total average AMI delay was increased to 10.85 ms (5% increase). Therefore, by not guaranteeing the quality of service (QoS) requirements of the

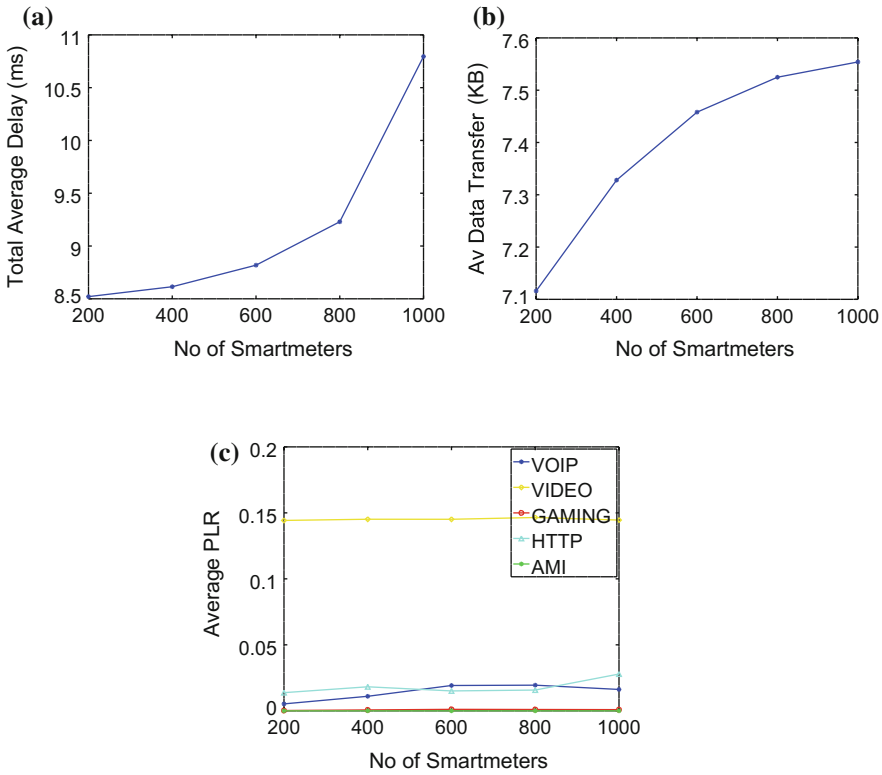


Fig. 5. Average total delay, AMI data transferred to the CC, and packet loss ratio with fixed (100 users) background nodes. **a** Average total delay, **b** Av total AMI data transferred, and **c** packet loss ratio (PLR)

video application, it is no longer possible to continue servicing more than 150 background users and 5000 SMs (1000 SMs per DC).

Table 3. Quality requirements for the applications offered [3]

Application/Parameter	VOIP	VIDEO	HTTP	FTP	AMI
End-to-end delay	<150 ms	<250 ms	<400 ms	No limit	<15 s
PLR	<0.01	<0.1	<0.001	<0.001	<0.001
Jitter	<50 ms	<2 s	N.A	No limit	N.A

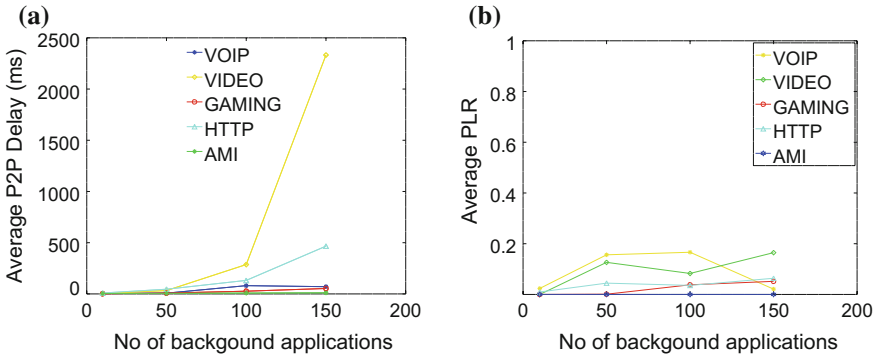


Fig. 6. Average total delay and PLR for fixed (150 users) background nodes. **a** Average total delay and **b** packet loss ratio

RACH for Fixed AMI Traffic and Variable Background Traffic An NS-3 patch implemented by Michele Polese [8,12] was used to demonstrate the LTE RACH for evaluating the background traffic with fixed AMI traffic.

Clearly, the average P2P delay of 120.0741 ms (Fig. 7a) for video application was reduced by large amount and is well within the QOS requirements (Table 3). From Fig. 7b, AMI total delay (11.2514 ms) is also within QOS requirements.

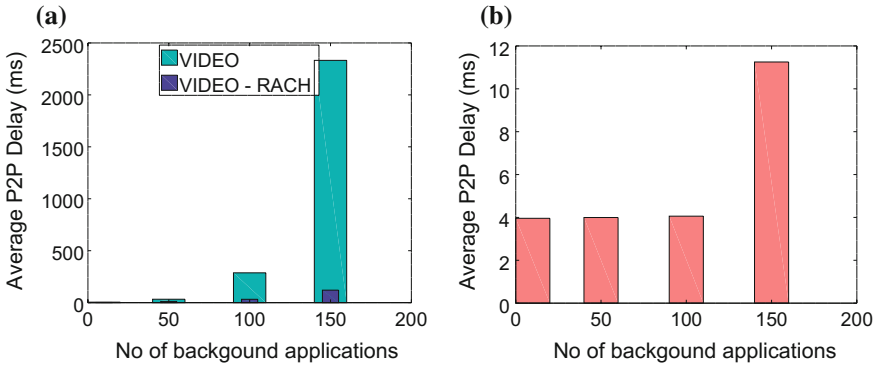


Fig. 7. Average delays for video and AMI applications using LTE RACH. **a** Average P2P delay comparison for video applications and **b** average total delay for AMI

5 Conclusions

This paper presents the results of performance evaluation of a communication scheme for an AMI network as well as H2H applications when they coexist inside

a public, shared LTE network using the NS-3 network simulator. The scenarios with large numbers of simultaneous Internet users (along with 1000 SMs per DC) within the cell (up to 150), it was observed that the quality requirements of the video services (streaming) were affected. An LTE RACH procedure has been successfully demonstrated to restore the same. The authors have uploaded the entire NS-3 source code of the simulation experiments including Octave scripts for plotting to Github, a global supporting community for developers.[1]

5.1 Future Work

As a future work, it is proposed to carry out the simulation experiments using 6LowPAN and LR-WPAN in NS-3 for AMI networks.

References

1. Sui H, Wang H, Lu M, Lee W (2009) An AMI system for the deregulated electricity markets. IEEE Trans Ind Appl 45(6):2104–2108
2. Chen Y-S, Chiang WL (2015) A spiderweb-based massive access management protocol for M2M wireless networks. IEEE Sens J 15:5765–5776
3. US Department of Energy (2010) Report: Communications Requirements for Smart Grid Technology
4. Zaballos A (2009) Survey and performance comparison of AM Rover PLC standards. IEEE Trans Power Deliv 24(2):604–613
5. Ruz AM, Narvez HG (2011) Performance evaluation of a network of smart meters, implemented on HSDPA cellular technology. Master's thesis, Department of Electrical and Electronics Engineering, Universidad de los Andes, Bogot, pp 1–154
6. Kheaksong A, Prayote A, Lee W (2016) Performance evaluation of smart grid communications via network simulation version 3. In: 13th international conference on electrical engineering/electronics, computer, telecommunications and information technology (ECTI-CON), Chiang Mai, 2016, pp 1–5
7. Du Q, Li W, Liu L, Ren P, Wang Y, Sun L (2016) Dynamic RACH partition for massive access of differentiated M2M services. Sensors 16(4):455.1455.19
8. Polese M, Centenaro M, Zanella A, Zorzi M (2016) M2M massive access in LTE: RACH performance evaluation in a smart city scenario. In: IEEE international conference on communications (ICC), Kuala Lumpur, 2016, pp 1–6
9. Aranda JM (2012) Evaluation of the performance of an AMI network implemented on the PLC and HSDPA technologies. Bogota: Master thesis. Universidad de los Andes, 2012, pp 1–148
10. Khan F (2009) LTE for 4G mobile broadband: air interface technologies and performance. Cambridge University Press, Cambridge (Print)
11. https://github.com/tgpham/gen-udp/tree/master/model
12. https://github.com/signetlabdei/lena-plus

[1] https://github.com/rajeshchintha.

Smart Device for Ensuring Women Safety Using Android App

V. Mareeswari(✉) and Sunita S. Patil

ACS College of Engineering, Bangalore, India
{mareesh.prasanna, sunita.chalageri}@gmail.com

1 Introduction

We come across many issues regarding the safety and security in day-to-day life. "Arakshaka dal", what we call as police force which is only meant for the protection, security, and safety of citizens. If the unsafe situation like this type is known to police like force, then only they can give protection to needy women. Unfortunately, nowadays the scenario is quite different; after the incident happens, such forces come to know. As such, cases are increasing day by day; what a woman can do in such kind of situation? Making phone calls to her well-wishers for help. Constraint is distance of her well-wisher from her place. If the place is unknown, the identity of place is also the constraint. Moreover, in trouble (unsafe situation), it is difficult to call anybody for help. Even a woman may be unaware of the place where she is and cannot give place identity.

In recent days, the incidents like physical harassments are more common. As one of the solutions to this problem, this proposed system is being developed using the VR (voice recognition) technology. Voice recognition could more generally be called as speech recognition or sound recognition, which is used to identify words and phrases in spoken language. At the same time, it converts them to a machine-readable format. This technology was first used in 1952 to convert spoken words into machine-readable format digit by digit using the ASR (automatic speech recognition) algorithm. The performance of this technology is measured by accuracy and speed. It is easy to handle and comfortable.

This paper is organized as follows: The second section briefs the related work, i.e., problem statement, motivation, and objectives. The third section presents the materials and methods. The fourth section describes the results and discussion. Section 5 concludes this paper and highlights some further work.

2 Literature Review

Koss et al. [1] studied the change in the memory pattern of physically abused women. A study was conducted on women working in medical centers and universities, etc. Results proved that after the harassment, the victim memory is affected: "clarity" and "alertness". Authors portrayed that post-harassment, physical symptoms and mental

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_21

pressure are high. Mezey [2] published a paper on treating mechanism for harassment victim. Psychological post-harassment effect includes depression, anxiety, and sexual dysfunction. Hence, the aim of the author is to recover the health of the victim to the normal condition. The authors conclude saying that mental health professional can be aware of the history of past physical abuse and chattering. Sudderth's [3] study is based on qualitative analysis of interview with 30 rape survivors, who had been harassed by men. The core of this survey is that the victim people are more feeling negative while discussing their bitter moments that they faced in harassment. They feel hesitant to share their experience with others. Lots of surveys say that a victim person, especially younger victims, does not go and complaint in police station. Even they do not feel good when shares their experience till considerable time. The ultimate purpose of this research is "recovery process". Victimizing survivors avoid exposing inducing emotion connected with harassment. Victim people hesitate to share their experience because of stigma attached to being a harassment victim.

The objective of the research by Shipherd and Gayle Beck [4] is to overcome post-traumatic stress disorder because of harassment. The survey report says that the people become more anxious, depressed, and distressed throughout. They are not able to get rid of harassment related thoughts. This study by Brad Ford (2000) says that people who had experienced child abuse are more depressed alcohol abuse and dating stress. In the survey by Campbell et al. [5], 102 harassment survivors were interviewed about social reaction and family and friends support post-harassment. The support from society and friends are negligible. Instead of giving care and help regarding physically and mentally, they (society and friends) lead them into misery by giving negative complaints and blaming. The survey concludes that the support from society as well as the friend and family did not adequate. The main purpose of Schewe (2002) study is developing cocurricular from Middleware School to high school. Wasco [6] highlighted the limitation of trauma response and applications of post-traumatic stress in people who are harassed. The study by Sochting [7] enforced to provide harassment prevention programs for girls. Most of the North American Universities give priority to change women attitude, by educating them about harassment prevention techniques.

2.1 Problems in Existing System

The drawbacks of all these applications are able to intimate their beloved once but are that loved once are able to protect them from that harassment? Probably, they need to ask help from some protecting force like police, etc.; in addition, their functioning is restricted to a particular city or town, so our application is the only solution for this drawback.

2.2 Motivation

A woman is a symbol of love, purity, knowledge, sacrifice, etc. peace and prosperity lie in the society where a woman is happy and honored. Remembering the goddesses you bow down to in the same house where you raise your hand at the actual goddess of

your home. The universal truth is that in a house, all the deities are pleased, where women are honored and no meritorious deed will not yield any result where they are abused. Crime against women is a crime against humanity. One-third of the women may suffer from abuse and violence in her lifetime. Such incidents are more common in recent days, which motivated us to this proposed system. This proposed system will be helpful to all the women's around the globe which will not require any extra efforts for carrying the device.

2.3 Objectives

This paper summarizes current safety device available for women's self-protection.

- By implementing and using our proposed system, not only safety of women but also valuable things will be just a click away at cheaper rate.
- The device will be activated through the voice commands.
- The panic switch can be used as an alternative to the voice commands.
- Buzzer will be used to alert the surrounding people in emergency situations.
- The details of the victim will be stored in the device, which will also be mailed along with the location and captured images.

3 Materials and Methods

Personal safety has become an important issue for everyone, but especially for women. The current women safety device in the market has to push the button to send a message. It is difficult to press buttons in critical situation when the keypad is locked. This paper deals with recognizing voice of victim and converting it into text to send email. The email consists of the following details:

a. stored details of the victim such as image of the victim, parent/guardian's contact no, and address of the victim,
b. surrounding captured images, and
c. GPS location information about the victim along with longitude and latitude.

The device gets activated once the voice kit is enabled through voice commands. Email can be sent even if the keypad is locked. Many embedded systems have substantially different designs according to their functions and utilities. In this proposed system, structured, modular design concept is adopted, and the system is mainly composed of a single microcontroller (RL78), LCD, Bluetooth, panic switch, voice kit, light source and buzzer, and Android smart device.

The microcontroller placed at the center of the unit forms the control unit for the entire device. Embedded within the microcontroller is a program that helps the microcontroller to take action based on the inputs provided to it.

The attack by the victim is indicated through the voice commands, when the voice commands provide through voice kit the unit will get activated, as an instant protection light source is used to give electric shocks to the person who is harassing her and

buzzer is activated to alert surrounding people. A command is sent to Android device through Bluetooth and as it receives the command, the app is activated and captures the victims face and surrounding image, GPS location, and stored details which are fetched from the SQLite database and are sent as email to police and sms to parent/guardians. If the system fails to recognize the voice command, then an alternative way to activate the system by pressing the panic switch through which the police and their guardians are altered. The system architecture is shown in the Fig. 1.

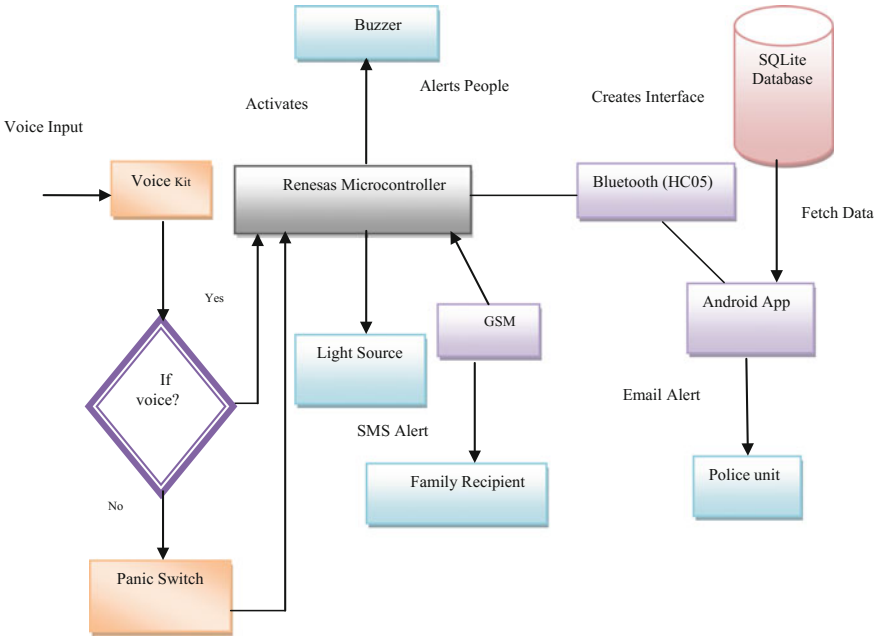


Fig. 1. System architecture

4 Results and Discussion

Table 1 shows the relationship between input and corresponding output parameters, which deal with the proposed system. The input is either voice signal or panic switch signal. The outputs are buzzer, light source, sms alert to parent, and email alert to nearby police station.

Table 1. Input and output parameter

S.no	Input	Output
1	Voice signal	Buzzer, light source, sms, and email alert
2	Panic switch	Buzzer, light source, sms, and email alert

Table 2 shows the efficiency of the proposed system, which deals with different test cases.

Table 2. Efficiency of different test cases

S.no	Input	No. of cases	Output occurrence	Accuracy (%)
1	Voice	10	Buzzer—yes	95
			Light source—yes	
			SMS—delay	
			Email—delayed due to N/W	
2	Panic switch	10	Buzzer—yes	95
			Light source—yes	
			SMS—delay	
			Email—delayed due to N/W	

In general, results are the outcome of the paper after the complete work is being carried out. These results can be shown individually from the start of the work to the end how the paper works. Here, the results are shown in the form of snapshots and explained in brief.

A login is the screen which takes the details of the username and the password to log in. After creating the interface between the controller and the application, the application being developed is opened and the login screen appears. To log in to the application, the necessary username and the password are to be entered and if the entered input is valid, a pop-up box comes with the message login success. Once the login is successful, the application moves to the registration. In this screen, the victim provides the details like email id of any one person for whom the mail containing the message help, victim's details like her name, phone number, parent phone number, and the current fetched location address. As the name indicates, this is the screen where all the important activities of this paper will be carried out. After registering, the user can enter into this screen which is used to create interfaces between the hardware tool and the Android application. Before connecting to the Bluetooth component, this screen shows the message connect. After enabling the Android Bluetooth device, the screen displays the available devices to get connected with. As we have incorporated the HC-05 hardware, Bluetooth device within the microcontroller the screen displays the device name to connect. After the pairing is done, the message HELP, which is to be sent to the person, is being sent and received will be displayed on the screen.

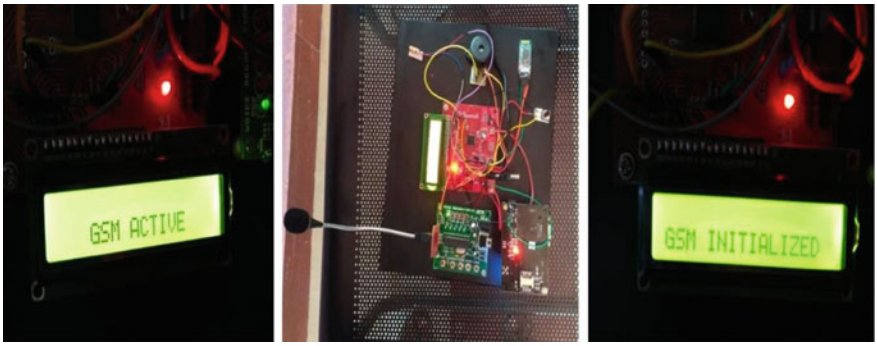


Fig. 2. GSM activation and initialization

When the voice kit and the Android application get activated, the GPS of the Android mobile should be enabled in order to fetch the location of the victim's when she is in some unsafe situation. This screen shows the message as fetching location once the device starts to fetch the location which is to be sent to the person who approaches toward her to paper. The above screen displays the latitude and longitude once the GPS of the mobile is enabled, which will be sent as the messages to the person who can approach her and help her. After the latitude and longitude of the current location are fetched through the GPS module, the current location address of the victim is fetched and displayed.

When the voice is recognized by the voice module, the voice kit gets activated and instantly buzzer and laser get activated and display the message. Instantly, GSM gets activated as shown in Fig. 2. The SIM is correctly inserted into the appropriate port of the GSM hardware. Later, when the interface is being created between the hardware and the application, GSM gets initialized and helps in sending the normal message containing the information of latitude and longitude along with the help message.

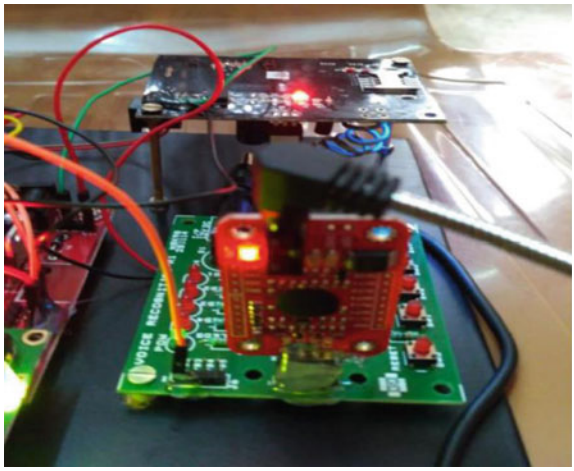


Fig. 3. Training of the voice kit

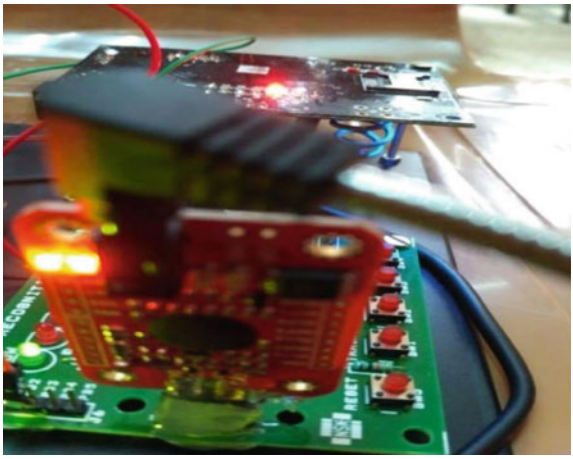


Fig. 4. Successfully trained

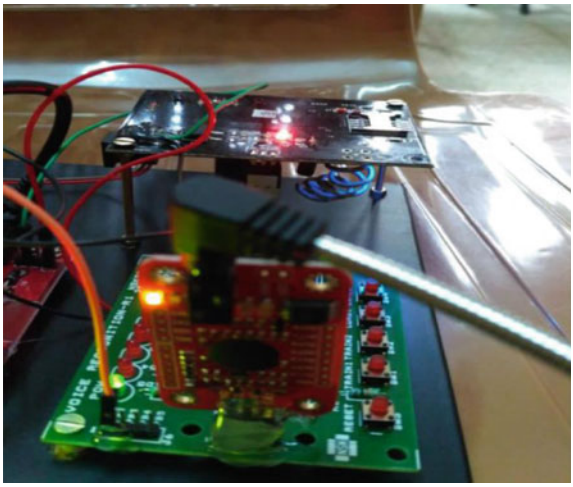


Fig. 5. Providing the voice input

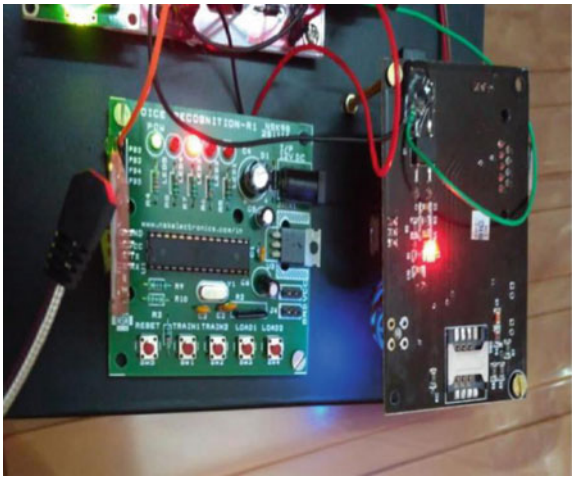


Fig. 6. LED glows when the voice matches

Initially, the voice recognition module must be provided with the power supply to train the voice which may be any of the following: characters, words, or numbers. Once it is trained, it must be loaded to the voice kit as shown in Fig. 3. It remains in the kit till the module is reset. Whenever a woman is in danger, she spells the trained voice, when the yellow light is blinking as shown in the Fig. 5. It will be matched with the loaded voice as shown in Fig. 4. If the match occurs, then the LED in the kit glows indicating that the match is found as shown in Fig. 6.

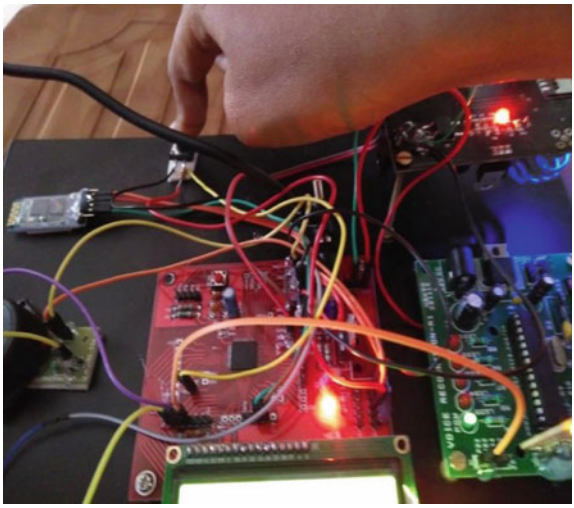


Fig. 7. Pressing the panic switch

Fig. 8. LCD displaying help message

Fig. 9. LCD displaying data sent message

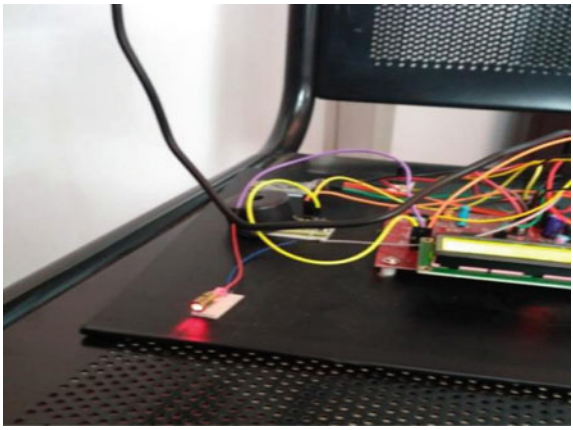


Fig. 10. Laser emitting light

When the voice command given does not match with the trained voice, as an alternative to the voice kit, panic switch is pressed to give as input to the microcontroller which similarly acts as the voice module (Fig. 7). The last step in designing is generating the output of the system either through the voice kit or the panic switch following output actions occurs simultaneously, buzzer buzzes a sound and the laser emits the light. The request for help will be sent to the Android application through the created Bluetooth interface and that is displayed on the LCD as shown in Fig. 8.

Once the request is received, the application fetches the current location of the victim through the enabled GPS of the mobile and also dynamically captures the surrounding location images. And then, the stored image of the victim along with the current location, dynamically captured image, and the data that is stored in the database

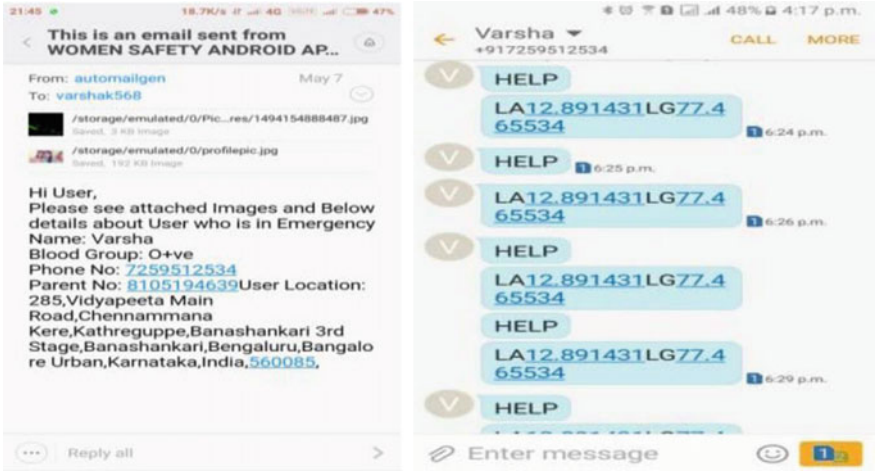


Fig. 11. Snapshots of the email and SMS

will be sent to the control room through the previously saved email id and phone number to make them aware that a woman is in danger and to help her faster. When the data is sent, the notification of the message is displayed on the LCD as shown in Fig. 9.

The last step in designing the output of the system either through the voice kit or the panic switches, simultaneously the following two, i.e., the buzzer and the laser get activated and the laser starts to emit the light as shown in Fig. 10. Using the developed application in smartphone, email is sent to the police unit and then via GSM alert, the message is sent any other recipient as shown in Fig. 11.

5 Conclusion

As one of the solutions to this sexual harassment, this proposed system is being developed using the VR (voice recognition) technology. Voice recognition could more generally be called as speech recognition or sound recognition, which is used to identify words and phrases in the spoken language, and then it converts them to a machine-readable format.

The proposed design is used to solve the critical situation with technologically sound equipment's and ideas by the women. This system can give the self-confidence to the women to face the incident which is against her safety and security. This paper attempts to deal with a community concern that has been destroying the lives of uncountable individuals and their families. A device like this improves the level of safety of women and girls. As technology is advancing, there are better means of getting on hand information about unsafe women. The unit testing and integration that were performed clearly indicate that the system provides security to all the women across the globe without any major complications. Since this device is smaller in size, all the women can carry it without much effort. The main purpose of the proposed system is to save the women before any intolerable incidental occurs, and this is achieved through this work.

Acknowledgements. The authors thank the management and principal of ACS College of Engineering, Mysore Road, Bangalore for permitting and supporting us to carry out the research work.

References

1. Koss MP, Figuerdo AJ, Bell I, Tharan M, Tromp S (1996) Traumatic memory characteristics: a cross validate mediational model of response to rape among employed women. J Abnormal Psychol 105(3):421–432
2. Mezey GC (1997) Treatment of rape victims. Adv Psychiatr Treat 3:197–203
3. Sudderth LK (1998) It ll come right back at me. Violence Against Women 4(5) October 1998
4. Shipherd JC, Beck JG (1999) The effects of suppressing trauma-related thoughts on women with rape related posttraumatic stress disorder. Behav Res Ther 37(2):99–112
5. Campbell R, Ahrens CE, Sefi T, Wasco SM, Barner HE (2001) Social reactions to rape victims: healing and hurtfull effects on psychological and physical health outcomes. Violence Vict 16(3)

6. Wasco SM (2003) Conceptualizing the harm done by rape applications of trauma theory to experiences of sexual assault. Trauma, Violence, & Abuse October 2003
7. Sochting I, Fairbrother N, Cock WJ (2004) Sexual assault of women (Prevention efforts and risk factors). 10(1):73–93. Doi:10.1177/107780120325568
8. Descamps MJ, Rothblum E, Bradford J, Ryan C (2000) Mental health impact of child sexual abuse, rape, intimate partner violence, and hate crimes in the national lesbian health care survey. J Gay Lesbian Soc Serv June 2000
9. Moser C, Mcilwaine C (2006) Latin American urban violence as a development concern: towards a framework for violence reduction. World Dev 34(1):89–112
10. Hill R, Temin J, Pacholek L (2007) Building security where there is no security. J Peace Build Dev 3(2):3851
11. Muggah R, Krause K (2009) Closing the gap between peace operations and post-conflict insecurity: towards a violence reduction agenda. Int Peacekeep 16(1):136–150
12. Rathmell A (2009) Security and justice development—what next? J Secur Sect Manage 7:2
13. Jatti A, Kannan M, Alisha RM, Vijayalakshmi P, Sinha S (2016) Design and development of an IOT based wearable device for the safety and security of women and girl children. In: IEEE International conference on recent trends in electronics information communication technology, May 20–21
14. Supriya HS, Patil SR, Pyati S, Annigeri V (2014) Pathipa Dharshinyaha Paalana. Karnataka State Council for Science and Technolgy, 2013
15. Hussain SM, Nizamuddin SA, Asunico R, Ramaiah C, Vikram Singh A (2016) Prototype of an intelligent system based on RFID and GPS technologies for women safety. In: IEEE 5th international conference on reliability, infocom technologies and optimization (ICRITO)
16. Miriyala GP (2016) Smart intelligent security system for women. Int J Electron Commun Eng Technol (IJEET) 7(2):41–41
17. Bhardwaj N, Aggarwal N (2014) Design and development of "Suraksha". A women safety device. Int J Inf Comput Technol 4(8):787–792 (2014). ISSN 0974-2239
18. Shipherd JC, Gayle Beck JG (2003) A preliminary examination of treatment for posttraumatic stress disorder in Chronic pain patients: a case study. J Traumatic Stress 16(5): 451–457 November 2003
19. Campbell et al. (2001) Healing or hurtful: sexual assault survivors' interpretations of social Reactions from support providers

Characteristics Analysis of Si0.5Ge0.5 Doping-Less PNPN TFET

Sudakar Singh Chauhan[1(✉)], Gaurav Verma[1], and Vinod Naik[2]

[1] Department of ECE, National Institute of Technology, Kurukshetra, Haryana, India
sudakar@nitkkr.ac.in, gauravnitkl3@gmail.com
[2] Department of School of VLSI Design & Embedded Systems, National Institute of Technology, Kurukshetra, Haryana, India
naikvinod094ll@gmail.com

1 Introduction

The day by day reduction in size of MOSFETs has made them prone to several short channel effects [1, 2] such as drain induced barrier lowering (DIBL), surface scattering, velocity saturation, impact ionization, and hot electron effect. Introduction of tunnel field effect transistor (TFET) has made it possible to overcome these effects and achieved low OFF-state currents and low subthreshold slope (SS). In spite of all these merits, TFETs has its own side effects like low ON-state current due to poor band-to-band tunneling (BTBT) efficiency at source–channel junction. Several remedies have been proposed to overcome this deficiency of TFETs like low band gap materials, use of high-k dielectrics, strained silicon, nanowires, use of dual material gate work functions, formation of N^+ source pockets [3, 4], etc. However, the use of high-k dielectric increases the effect of ambipolar current. One important short come of TFET is its variation in performance due to random dopant fluctuations (RDFs) [5]. In an attempt to overcome these effects of TFET, a novel device using charge plasma technique has been introduced [6]. Since the obviation of dopants to form source and drain regions made it a doping-less concept, namely, doping-less TFET (DL-TFET). The use of different work function metal electrodes ensures the formation of electron and hole concentrations using charge plasma technique. Thus, obviating the need of dopants not only avoids the effect of RDFs but also made it an easy way to realize low thermal budget TFETs. However, the DL-TFETs shown low ON-state current similar to that of doped TFETs, this is due to the nature of lateral electric field which is related to BTBT rate (low) exponentially. To improve BTBT rate, a TFET with N^+ source pocket is formed which modulates the energy band profiles and increases ON-state current by reducing tunneling width and is known as PNPN TFET.

Similar technique of formation of N^+ source pocket to enhance the ON-state current is applied in DL-TFET by using dual material gate concept [7]. This again avoids the use of ion implantation by ensuring the unusability of dopant atoms in DL-TFETs. In this paper, we present a low bandgap Si0.5Ge0.5 DL-PNPN TFET which is designed

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_22

by using two gate materials with different work functions to form PNPN regions. As an addition, to enhance the performance of the device hetero-gate dielectric is also used. As compared to Si DL-PNPN TFETs, Si0.5Ge0.5 DL-PNPN TFET provides enhanced ON-state current and reduced SS.

2 Device Structure and Simulation Parameters

The schematic shown in Fig. 1 depicts a hetero-gate dielectric DL-PNPN TFET. The parameters used in the simulation are shown in Table 1.

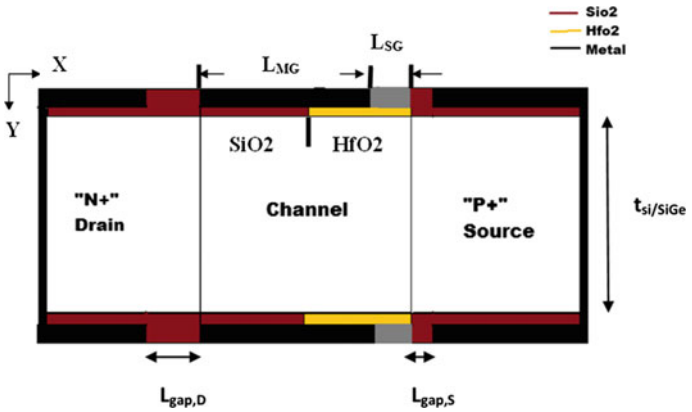


Fig. 1. Hetero-gate dielectric DL-PNPN TFET

Table 1. Simulation parameters

Parameters	DL-PNPN TFET
Film thickness (Si, Si0.5Ge0.5) (nm)	10
Film length (nm)	250
Dielectric thickness (nm)	3
Length of HfO_2 (nm)	25
Length of SiO_2 (nm)	25
Main gate length (L_{MG}) (nm)	46
Short gate length (L_{SG}) (nm)	4
Main gate work function (eV)	4.4
Short gate work function (eV)	3.9
Channel doping (N_A)	1×10^{-15} cm^3

The film thickness is chosen to be 10 nm to avoid effects of thin film. The charge plasma technique is used to design a DL-TFET by using electrode work functions of 3.9 eV (Hafnium), 5.96 eV (Platinum) and 4.4 eV (Metal) to form drain, source, and gate, respectively. To improve the performance of DL-TFET more tunneling width is required. To increase the tunneling width, an N^+ source pocket needs to be formed to have an abrupt junction which enhances the performance of TFETs. In case of DL-TFETs, it may be attained by using a different gate material with a work function similar to that of drain electrode (3.9 eV). Thus, two gates, namely, main gate (LMG) and short gate (LSG) form a DL-PNPN TFET. Two gate dielectrics SiO_2 and HfO2 are used. The high-K dielectric (HfO2) is used on source side of gate to reduce ambipolar effects. The spacers of SiO_2 are inserted between gate-source and gate-drain interfaces. The length of spacer L_(gap,S) between gate and source electrode is 3 nm. The length of spacer between gate and drain electrode is 15 nm. To avoid silicide formations at source/drain electrodes, an oxide thickness of 0.5 nm SiO_2 or 3 nm HfO2 is used. The DL-PNPN TFET is developed using the SDE EDITOR tool of Synopsys Sentaurus TCAD Version J2014.09.

3 Results and Discussion

Using charge plasma technique, the carrier concentrations are induced in source and drain regions. These induced carrier concentrations are same as that of conventional TFET doped carrier concentrations. Figure 2 shows the thermal equilibrium electron carrier concentration of Si and Si0.5Ge0.5 DL-PNPN TFETs at a gate and drain

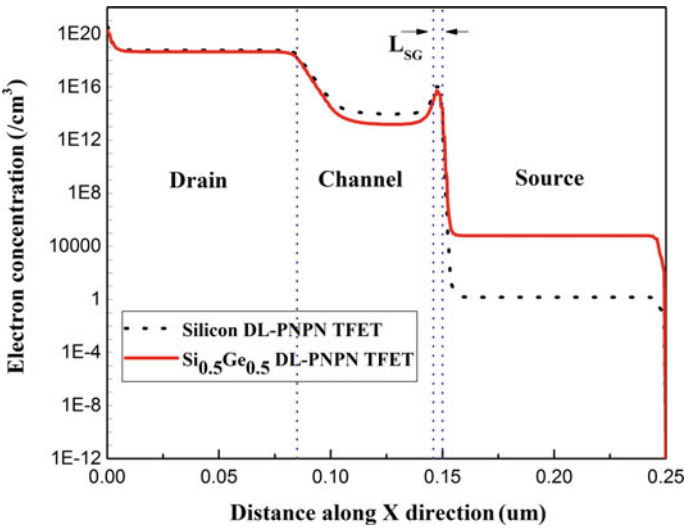


Fig. 2. Thermal equilibrium electron concentration in the Si DL-PNPN TFET and a Si0.5Ge0.5. DL-PNPN TFET at 1 nm below the Si–SiO2 interface

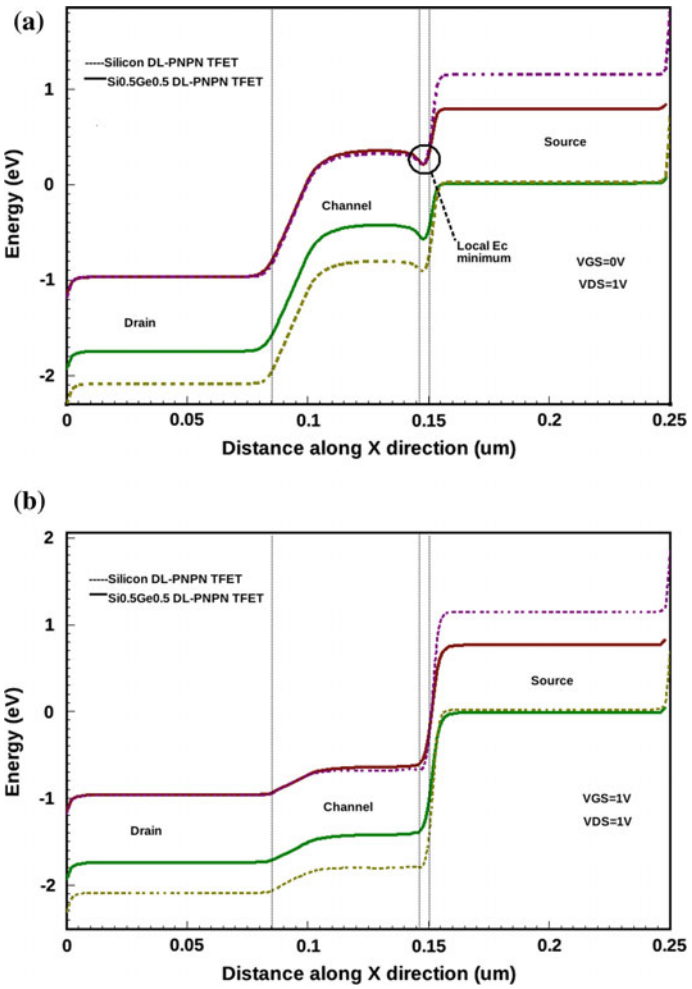


Fig. 3. Energy band structure of Si and Si0.5Ge0.5 DL-PNPN TFET in **a** OFF state ($V_{GS} = 0$ and $V_{DS} = 1$ V) and **b** ON state ($V_{GS} = V_{DS} = 1$ V)

voltages of zero volts. The plot is taken along Si-SiO_2 interface cross section of 1 nm. It can be seen from Fig. 2 that there is a short peak between source–channel interfaces. This peak is due to the use of short gate (LSG) with different metal work function than main gate (LMG), forming a small N-type region between source–channel interface. Thus, forming a DL-PNPN TFET. The energy band diagrams of Si and Si0.5Ge0.5 DL-PNPN TFETs are shown in Fig. 3 under bias conditions of (a) OFF state, where gate to source voltage is at zero volts (V_{GS} = 0 V) and drain to source voltages is at one volt (V_{DS} = 1 V) and (b) ON-state, where (V_{GS} = V_{DS} = 1 V). From Fig. 3a, it can be seen that a notch is present at source–channel interface which is key to modulate the energy band profile of DL-PNPN TFETs. We can also observe that the notch in conduction band of Si0.5Ge0.5 DL-PNPN TFET is closer to the valence band than the

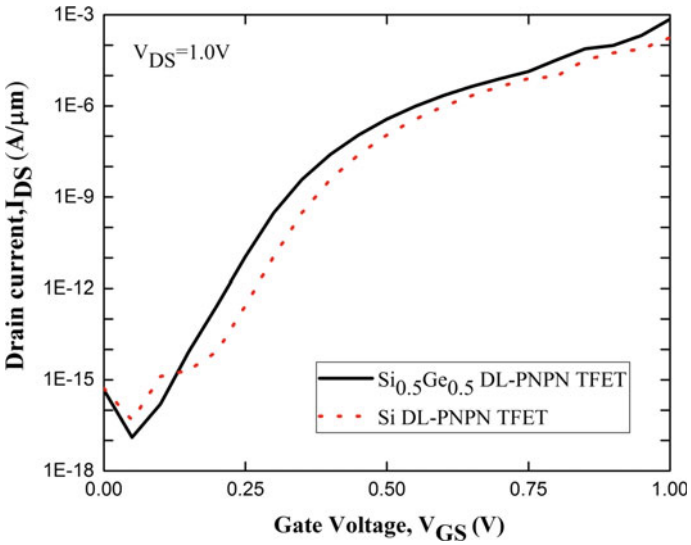


Fig. 4. Transfer characteristics of Si0.5Ge0.5 DL-PNPN TFET compared with Si DL-PNPN TFET

Table 2. Comparison of proposed device with Si DL-PNPN TFET

	$Si_{0.5}Ge_{0.5}$ DL-PNPN TFET	Si DL-PNPN TFET
I_{OFF}	4.66×10^{-16}	3.19×10^{-16}
I_{ON}	7.2×10^{-4}	1.78×10^{-4}
I_{ON}/I_{OFF}	1.54×10^{12}	0.55×10^{12}
Subthreshold swing	20 mV/*decade*	25 mV/*decade*

Si DL-PNPN TFET. Thus, achieving reduced tunneling width than Si DL-PNPN TFET and leading toward the enhancement of ON-state current of Si0.5Ge0.5 DL-TFET. The transfer characteristics of Si and Si0.5Ge0.5 DL-PNPN TFETs are shown in Fig. 4. The results of Si DL-PNPN TFET are replicated for comparison of Si0.5Ge0.5 DL-PNPN TFET in terms of Ion, Substhreshold Swing and Ion/Ioff ratio Table 2. The Si0.5Ge0.5 DL-PNPN TFET due to its reduced tunneling width shows 74.3% enhancement in ON-state current and increment of 64.2% in Ion/Ioff ratio than Si DL-PNPN TFET.

4 Conclusion

In this paper, we presented hetero-gate dielectric Si0.5Ge0.5 DL-PNPN TFET. All the simulations were carried out using Synopsys Sentaurus TCAD. The proposed device shows enhanced performance than Si DL-PNPNP TFET in terms of I_{ON} and I_{ON}/I_{OFF}

ratio. Nevertheless, a DL-TFET is free from doping problems which makes it a highly suitable candidate for future low-cost low-power applications.

References

1. Seabaugh AC, Zhang Q (2010) Low-voltage tunnel transistors for beyond CMOS logic. Proc IEEE 98(12):2095–2110
2. Sunny A, Chauhan SS (2016) An enhancement mde GaN MOSFET with AlGaN/GaN hetero structure. In: IEEE international conference on microelectronics, computing and communications (MicroCom2016), Jan 23–25
3. Krishnamohan T, Kim D, Raghunathan S, Saraswat K (2008) Double gate strained Ge heterostructure tunneling FET (TFET) with record high drive currents and < 60 mV/dec subthreshold slope. IEDM Tech Dig, 947–949
4. Gopi C, Chauhan SS (2016) Double-gate Ge, InAs-based tunnel FETs with enhanced ON-current. In: 5th IEEE international conference on communications and signal processing (ICCSP'16), Apr 6–8
5. Leung G, Chui CO (2013) Interactions between line edge roughness and random dopant fluctuation in nonplanar field-effect transistor variability. IEEE Trans Electron Devices 60 (10):3277–3284
6. Juyal R, Chauhan SS (2016) TCAD simulation of Ge source dopingless Tunnel FET. In: IEEE international conference on advances in computing and communications (ICACCE'16), Apr 8–9
7. Cui N, Liang R, Xua J (2011) Heteromaterial gate tunnel field effect transistor with lateral energy band profile modulation. Appl Phys Lett 98, 14, 142105

Logical Implication to Reduce Run Time Memory Requirement and Searches During LZW Decompression

Subrata Roy(✉)

Ghani Khan Choudhury Institute of Engineering and Technology, Narayanpur, Malda, West Bengal, India
Subrata@gkciet.ac.in

1 Introduction

LZW data compression algorithm maintains in dictionary code for every distinct pattern; this results in unnecessary searches for decoding a code during decompression. Many authors have proposed improvements to LZW algorithm in literature for software and hardware implementation. Authors in [1] have presented a parallel LZW decompression algorithm and implemented it with the help of NVDIA's CUDA enabled GPU. This algorithm achieves speedup factor up to 69.4. The authors in [2] have presented an implementation of LZW compression with the help of a CUDA enabled GPU. Two approaches has been taken in this work: (i) an image is compressed using GPU but written in storage through the CPU and (ii) An image is transferred to CPU, compressed by the CPU and written in storage by the CPU. The authors have proved that the first approach is three times faster than the second approach. The author in [3] has presented two mechanisms to improve LZC variation of LZW; this results in better compression for files of small size. Test suit size and download time of test vectors for System-on-chip design can be reduced using LZW algorithm [4]. The authors in [5] have presented a joint LZW and LDBP data compression algorithm for a congested network. This algorithm can save at least 30% of network bandwidth when text data transmitted. The authors in [6] have improved LZW algorithm for use with Global Nevigational Satellite System (GNSS) data. Three approaches has been considered to improve LZW compression: (i) increased dictionary size, (ii) efficient storage of codeword and (iii) using hash function to find and match strings. The OLZW algorithm [7] assigns optimal variable-length codes to patterns. In this technique, the LZW adds a symbol to the dictionary only when it is required. This algorithm shows higher performance for small size files as compared to other variations of LZW known as LZW12 and LZW15V. The authors in [8] have presented a modified OLZW algorithm and its two variants that achieves better compression ratio for large files also. The performance of LZW compression can be increased if it is applied on block sorted data [9]. The decompressor output must be reverse block sorted to get actual output. The author in [10] describes the working principle of LZW, its implementation issues and some of its variants. The authors in [11] have compressed ROI medical images using LZW, LZ78 and LZ77 algorithm. Result shows that LZW compress such images more

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_23

efficiently than the other two algorithms and also LZW acieves higher compression ratio. BD-LZW algorithm [12] is a parallel LZW algorithm. This algorithm decomposes a source file into data blocks. Each block is assigned a pair of threads. One thread is responsible for compressing the data block from one side, the other thread from the other side. Threads create their own dictionary. As a result, two threads compress the same data block at the same time. The authors in [13] have improved efficiency of LZW algorithm by efficient use of dictionary and bi-mode encoder. LZ77 algorithm is a dictionary based algorithm [14]. It uses the previously encountered strings as dictionary. A sliding window consists of search buffer and look-ahead buffer is maintained during compression. Search for a matching pattern is performed using a length-distance pair in the sliding window. LZ78 algorithm uses dynamic dictionary [15]. Initially, the dictionary is empty. The encoder outputs a pointer to the dictionary and the code of symbol. LZW is a LZ78 variant. The authors in [16] have presented an efficient technique for implementation of LZW decompression. This technique uses a pointer table and a code table. Efficient indexing and searching in compressed files can be done using LZW algorithm [17]. LZW and its existing improvements in literature maintain one code for each pattern in its dictionary and hence in memory.

However, hardly there have been attempt to reduce the total number of searches while decoding a code during decompression and run time memory requirement of LZW algorithm by changing the standard LZW pattern encoding scheme (the way codes are assigned to patterns).

The MSED technique uses logical implication to construct LZW dictionary. In MSED technique, dictionary organization of LZW decompressor is different from that of compressor. However, LZW compressor and decompressor agree on the way they form and encode patterns. We have proved in this work that when we use logical implication to encode patterns, we need to store in dictionary (hence in memory) effectively one code for two consecutive patterns. Thus, only 50% of the actual number of codes is maintained in its dictionary by LZW with MSED technique as compared to standard LZW algorithm. As a result, MSED technique reduces both required number of searches to decode a code during decompression and memory requirement of LZW algorithm.

Organization of the rest of the paper is as below. We define the necessary notation and propositions needed for MSED technique in Sect. 2. We present the MSED technique with an example in Sect. 3. We prove improvements in standard LZW algorithm due to MSED technique in Sect. 4 and conclusion is given in Sect. 5.

2 Notation and Propositions

Following notation and propositions form the basis of MSED technique.

2.1 Complement Operation

The complement operation denoted by $\sim$ (tilde) inverts every bit in a binary number from 0 to 1 and 1 to 0.

2.2 Logical Implication for Encoding and Decoding Patterns

Logical implication denoted by $X \rightarrow Y$ says that if X is true then Y is also true. For instance, consider two sets X and Y, and $y \in Y$, then, $\{y \in Y \wedge y \in (X \cap Y)\} \rightarrow y \in X$.

Proposition 1 *For a set C of N-bit (N > 0) binary codes when we perform $C \times C$, where $\times$ is the Cartesian product operation, then there exists a relation R_{code} on $C \times C$ such that $\forall$ (x, y) $\in R_{code}$, $\{(x = \sim y) \rightarrow (y = \sim x)\}$ holds.*

Proof. We know that with *N*-bits we have a set of 2^N binary codes. Let us call this set *S*. Out of these 2^N codes, $\frac{2^N}{2}$ codes have most significant bit (MSB) value 0. Let us call this set S_0 ($S_0 \subset S$). Rest of the $\frac{2^N}{2}$ codes have MSB value 1. Let us call this set S_1 ($S_1 \subset S$). We can build a set *C* ($C \subseteq S$) consisting of *n* number of codes ($2 \leq n \leq 2^N$ and *n* is even) with the elements of S_0 and S_1 such that there exists a relation R_{code} as $R_{code} = \{(x, y) \mid x \in S_0 \wedge y \in S_1, \wedge\, y = \sim x \wedge x = \sim y\}$. For example, consider we want to construct set *C* with six 4-bit codes. With 4-bits, we have set S_0 = {0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111} and set S_1 = {1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111}. One possible selection of set *C* from the elements of S_0 and S_1 is *C* = {0000, 0001, 0010 1101, 1110, 1111}. This gives us R_{code} = {(0000, 1111), (0001, 1110), (0010, 1101)}. Relation R_{code} can be constructed for any even value of *n* over a set of *N*-bit codes. Since $\forall$ (x, y) $\in R_{code}$, $\{(x = \sim y) \rightarrow (y = \sim x)\}$ holds, this proves Proposition 1. Therefore, we conclude that $\forall$ (x, y) $\in R_{code}$, given any one between x and y, the other can be determined unambiguously.

Let us now extend this concept to encode and decode patterns. Consider we have a non-empty set M of *n* (consider *n* is even) number of distinct patterns $\{P_1, P_2, \ldots, P_n\}$ to encode. Also consider, for every ordered pair of adjacent patterns $(P_{k,}\, P_{k+1}) \mid k = 1$ to *n*, P_k is the previous pattern of P_{k+1} and a pattern appears only in one ordered pair. Let us logically decompose set M into two subsets M_{odd} containing odd-numbered patterns as $M_{odd} = \{P_1, P_3, \ldots, P_k \mid k = 1, 3, \ldots, n-1 \wedge |k| = \frac{n}{2}\}$ and M_{even} containing even-numbered patterns as $M_{even} = \{P_2, P_4, \ldots, P_m \mid m = 2, 4, \ldots, n \wedge |m| = \frac{n}{2}\}$, where $|k|$ and $|m|$ are number of elements in M_{odd} and M_{even}, respectively. Since *n* is even, therefore $|k| = |m|$ and $|k| + |m| = n$. Now, we define a relation $R_{pattern}$ on $M_{odd} \times M_{even}$ as below.

$R_{pattern} = \{(P_k, P_m) \mid m-k = 1 \wedge$ any P_k or P_m appears exactly once in $R_{pattern}\}$

An ordered pair $(P_k, P_m) \in R_{pattern}$ can be encoded with another ordered pair of codes $(x, y) \in R_{code}$ as below:

$$\text{Code}[P_k] = x$$
$$\text{Code}[P_m] = y$$

This implies that an ordered pair of patterns $(P_k, P_m) \in R_{pattern}$, when encoded as above, then, Code[P_m] can be uniquely determined from Code[P_k] with the help of Proposition 1 and the vice versa. We define Proposition 2 as below.

Proposition 2 *An ordered pair of patterns* $(P_k, P_m) \in R_{pattern}$ *encoded with an ordered pair of binary codes* $(x, y) \in R_{code}$ *as* $Code[P_k] = x \wedge Code[P_m] = y$, *results in* $\{(Code[P_k] = \sim Code[P_m]) \rightarrow (Code[P_m] = \sim Code[P_k]), \forall Code[P_k], Code[P_m]) \in R_{code} \wedge (P_k, P_m) \in R_{pattern}\}$.

Hence, maintaining only one between Code[P_k] and Code[P_m] in both LZW compressor's and decompressor's dictionary (hence in memory) is sufficient.

3 The MSED Technique

In this section, before we discuss the MSED technique, we present modifications in dictionary organization of LZW compressor and decompressor. An example of the MSED technique is given later in this section.

3.1 Modified Dictionary of LZW Compressor

Considering N-bit fixed length codes, the compressor generates a dictionary that can hold 2^N entries. For simplicity, let us consider $N = 2$. With 2-bits, we have set S = {00, 01, 10, 11}. Proposition 1 gives us {(00, 11), (01, 10)} $\in R_{code}$. The dictionary can hold four entries. Consider we want to store patterns: 'Sh', 'hi', 'iv' and 'va' in dictionary. We have sets M_{odd} = {Sh, iv} and M_{even} = {hi, va} and the two ordered pairs are {(Sh, hi), (iv, va)} $\in R_{pattern}$. Proposition 1 and Proposition 2 allows us to maintain only one code for one ordered pair of patterns in dictionary. Consider we maintain in dictionary Code[P_k] for every ordered pair (P_k, P_m). We call Code[P_k] an *rpattern_code* and Code[P_m], a *cpattern_code* (since Code[P_m] = $\sim$Code[P_k]). We call P_k a regular pattern or *rpattern* in abbreviated and P_m a complementary pattern or *cpattern* in abbreviated. Thus, an ordered pair of patterns in dictionary is <*rpattern*, *cpattern*>. Proposition 1 and Proposition 2 gives us the dictionary as in Table 1 below. Binary values of codes are given inside parentheses.

Table 1. Compressor's dictionary

rpattern_code	rpattern	cpattern
0 (00)	Sh	hi
1 (01)	iv	va

3.2 Modified Dictionary of LZW Decompressor

Dictionary of the decompressor is similar to that of compressor except it has one extra 2-bit field called *recent*. The *recent* field has four possible values: 00, 01, 10 and 11. Each dictionary entry is a quadruple of the form <*rpattern_code*, *recent*, *rpattern*, *cpattern*>. The *recent* value 00 indicates absence of meaningful *rpattern* and *cpttern* in a quadruple. The *recent* value is 01 when a quadruple has meaningful *rpattern* but meaningful *cpattern* is missing; it is 10 when meaningful *rpattern* and

cpattern is present. The *recent* value 11 represents that a quadruple has meaningful *cpattern* but meaningful *rpattern* is absent.

A meaningful pattern is a pattern discovered by the compressor and hence also discovered by the decompressor. When n is odd ($n \geq 1$) there will be exactly one *rpattern* which will not form any ordered pair. We call such a pattern an isolated pattern. An isolated pattern can be grouped with a dummy pattern. A dummy pattern may be any pattern that indicates the absence of a meaningful pattern. The absence of a meaningful pattern will not affect the correctness of the algorithm. During compression, the code of a dummy pattern will never be output by the compressor. During decompression, the *recent* information will tell us whether a meaningful *rpattern* or *cpattern* exists in a quadruple or not.

3.3 Example of MSED Technique

We shall consider the same example string as in [18]. Let us consider an alphabet A = {a, b, c} and the string to be compressed be ababcbababaaaaaaa. Let us build the LZW compressor's dictionary considering 4-bit codes as per MSED encoding technique as above.

With 4-bits, we have R_{code} = {(0000, 1111), (0001, 1110), (0010, 1101), (0011, 1100), (0100, 1011), (0101, 1010), (0110, 1001), (0111, 1000)}. The first ordered pair (a, b) of alphabet characters is encoded with (0000, 1111) as per Propositions 1 and 2. The final alphabet character 'c' is encoded with the first code 0001 of code pair (0001, 1110). The first pattern 'ab' of the input string is encoded with 1110 of the code pair (0001, 1110) and so on.

After processing the entire input string, the compressor will generate the dictionary shown in Table 2 below.

Table 2. Dictionary for compression of string: ababcbababaaaaaaa

rpattern_codae	rpattern	cpattern
0 (0000)	a	b
1 (0001)	c	ab
2 (0010)	ba	abc
3 (0011)	cb	bab
4 (0100)	baba	aa
5 (0101)	aaa	aaaa

We have the compressed string: 0 15 14 1 2 12 0 11 5 0.

Let us now discuss decompression. Table 3 below shows codes assigned to the alphabet characters.

Table 3. Initial dictionary for decoding of 0 15 14 1 2 12 0 11 5 0

rpattern_code	recent	rpattern	cpattern
0 (0000)	10	a	b
1 (0001)	01	c	dummy

The decompressor reads the first code 0 (0000) from compressed string and finds the MSB to be 0. Thus, code 0 is an *rpattern_code*. Next, decompressor finds code 0 in the quadruple <0, 10, a, b> with *recent* value 10. Thus, meaningful *rpattern* and *cpattern* are present in <0, 10, a, b>. Decompressor outputs the *rpattern* entry 'a'. The next code to be decoded is 15 (1111). Decompressor determines 15 as *cpattern_code*. Hence, decompressor performs ~15 to get *rpattern_code* 0. Decompressor finds code 0 in <0, 10, a, b> and *recent* value is 10. Thus, decompressor outputs the *cpattern* entry b. The new pattern 'ab' is inserted in <1, 01, c, dummy> as *cpattern_code* and *recent* value is updated. This is shown in Table 4 below.

Table 4. Dictionary after insertion of pattern 'ab'

rpattern_code	recent	rpattern	cpattern
0 (0000)	10	a	b
1 (0001)	10	c	ab

The codes 14, 1 and 2 are decoded in similar fashion. Table 5 below shows the dictionary up to this time.

Table 5. Dictionary after decoding of 0 15 14 1 and 2

rpattern_code	recent	rpattern	cpattern
0 (0000)	10	a	b
1 (0001)	10	c	ab
2 (0010)	10	ba	abc
3 (0011)	01	cb	dummy

The next code to be decoded is 12 (1100), a *cpattern_code*. Thus, decompressor performs ~12 to get 3 (0011). Decompressor finds 3 in <3, 01, cb, dummy> but the recent value is 01. Thus, the *cpattern* entry in <3, 01, cb, dummy> is not a meaningful pattern. This special case of LZW occurs whenever a string of the form KwKwK appears as input during compression. Therefore, the decompressor will generate the appropriate *cpattern* 'bab', output it and will insert it into the dictionary in <3, 10, cb, bab>. We may encounter this exception for *rpattern* also. However, the absence of only a meaningful *rpattern* in a quadruple for any *rpattern_code* will be indicated by

recent value 11. This way, the entire compressed string is decoded. The final dictionary is given in Table 6.

Comparing Table 6 (decompressor dictionary) with Table 2 (compressor dictionary), we see that all patterns and their corresponding codes are identical.

Table 6. Final dictionary for decoding: 0 15 14 1 2 12 0 11 5 0

rpattern_code	recent	rpattern	cpattern
0 (0000)	10	a	b
1 (0001)	10	c	ab
2 (0010)	10	ba	abc
3 (0011)	10	cb	bab
4 (0100)	10	baba	aa
5 (0101)	10	aaa	aaaa

4 Improvements in Standard LZW Algorithm

In this section, we prove two possible improvements of standard LZW algorithm due to MSED technique: (i) reduced memory requirement of LZW and (ii) reduced total number of searches in order to decode any code during decompression. Let us prove (i) first. Number of bits required for *rpattern* and *cpattern* will be same for both standard LZW and LZW with MSED. Therefore, we compare the two approaches based on *rpattern_code*, *cpattern_code* and *recent* information. For N-bit fixed length code, we have 2^N entries in dictionary. Total number of code bits stored by LZW with MSED technique during compression is $\left(\frac{2^N}{2} \times N\right)$. During decompression, $\left(\frac{2^N}{2} \times N\right)$ *rpattern_code* bits and $\left(\frac{2^N}{2} \times N\right)$ *recent* bits are stored. Thus, total $\left(\left(\frac{2^N}{2} \times N\right) + 2^N\right)$ bits required for *rpattern_code* and *recent* information. As against this, standard LZW will store $(2^N \times N)$-bits for codes during compression and decompression. For instance, consider $N = 12$. LZW with MSED technique will store 24,576-bits for codes during compression and 24,576 + 4096 = 28,672-bits during decompression. As against this, standard LZW will store $(2^{12} \times 12) = 49152$-bits for codes both during compression and decompression. Therefore, LZW with MSED technique saves 49,152–24,576 = 24,576-bits during compression and 49,152 − 28,672 = 20,480-bits during decompression. For practical values of N used in LZW, $(2^N \times N)$ is greater than $\left(\frac{2^N}{2} \times N\right)$ and $\left(\left(\frac{2^N}{2} \times N\right) + 2^N\right)$. Therefore, it is proved that LZW with MSED technique uses memory more efficiently as compared to standard LZW algorithm.

Let us now prove improvement (ii). Let us consider that by time t during decompression, L out of n $(0 \leq L < n)$ distinct patterns has been discovered by the LZW decompressor with MSED technique and the next code to be decoded be CODE_i. To decode CODE_i, the decompressor will first find the type of CODE_i (*rpattern_code* or *cpattern_code*). If CODE_i is a *cpattern_code*, it will be complemented and searched. If CODE_i is an *rpattern_code*, it will be searched only. In both cases, the decompressor

has to search over a set of $\frac{L}{2}$ codes. As against this, standard LZW decompressor will search over a set of L codes in order to decode CODE_i. For any positive value of L, $\frac{L}{2}$ is less than L. This proves that the total number of searches required by LZW decompressor with MSED technique in order to decode any CODE_i is 50% less than the total number of searches required by standard LZW decompressor.

5 Conclusion

We have presented a new pattern encoding and decoding technique called MSED. We have proved that Logical Implication reduces the overall run time memory requirement of LZW algorithm and it also reduces the total number of searches during decompression by 50%. The MSED technique is also applicable for variable-length LZW.

Acknowledgements. The author is thankful to Dr. Nilkanta Barman, Associate Professor, GKCIET, India for his valuable guidance on paper writing. The author is thankful to Mr. Gopal Bandyopadhyay, System Manager, GKCIET, India for his help in collecting study materials for this research work. The author is also thankful to Mr. Debanjan Konar, Assistant Professor, SMIT, India for his help in understanding the author's guidelines.

References

1. Funusaka S, Nakano K, Ito Y (2015) A parallel algorithm for LZW decompression with GPU implementation. In: International conference on parallel processing and applied mathematics, pp 228–237
2. Funusaka S, Nakano K, Ito Y (2015) Fast LZW Compression Using a GPU. In: Third international symposium on computing and networking, pp 303–308
3. Horspool RN (1991) Improving LZW. In: Proceedings of data compression conference (DCC'91). pp 332–341
4. Keneiser MJ, Wolff FG, Papachriston CA, Weyer DJ, McIntyre DR (2003) A technique for high ratio LZW compression. In: Proceedings of the conference on design, automation and test in Europe and exhibition, vol 1, pp 116–121
5. Kho LC, Tan Y, Lim Y (2015) Joint LZW and lossless dictionary-based bit-packing compression techniques for congested network. In: International conference on computer, communications, and control technology, pp 196–200
6. Ma X, Xu C, Zheng P, Hu C (2012) The application of improved LZW algorithm in the data processing of GNSS simulation. In: Fourth international conference on computational and information sciences, pp 160–163
7. Nandi U, Mandal JK (2012) A compression technique based on optimality of LZW code. In: Third international conference on computer and communication technology, pp 166–170
8. Nandi U, Mandal JK (2013) Modified compression technique based on optimality of LZW code. In: International conference on computational intelligence: modelling techniques and applications, vol 10, pp 949–956
9. Saha SK, Rahaman M (2007) Boosting the performance of LZW compression through block sorting for universal lossless data compression. In: 10th international conference on computer and information technology, pp 1–5

10. Saloman D Data compression: the complete reference. http://apachetechnology.in/ati/www/KC/dw/Saloman%20%20Data_Compression_Complete_Reference.pdf
11. Singh S, Pandey P (2016) Enhanced LZW technique for medical image compression. In: 3rd international conference on computing for sustainable global development, pp 1080–1084
12. Zhan J, Zhou Q, Bai S, Cuihua L, Hu B, Li L (2008) BD-LZW picture compression algorithm for WSN system. In: Third international conference on pervasive computing and applications, vol 1, pp 146–150
13. Singh P, Duhan M, Priyanka (2006) Enhancing LZW algorithm to increase overall performance. In: Annual IEEE India conference, pp 1–4
14. Ziv J, Lampel A (1977) A universal algorithm for sequential data compression. In: IEEE transactions on information theory IT-23, vol 3, pp 337–343
15. Ziv J, Lampel A (1978) Compression on individual sequences via variable-rate coding. IEEE transactions on information theory IT-24, vol 5, pp 520–540
16. Zhou X, Ito Y, Nakano K (2016) An efficient implementation of LZW decompression in the FPGA. In: IEEE IPDPSW, pp 599–607
17. Zhang N, Tao T, Satya RV, Mukherjee A (2005) A flexible compressed text retrieval system using a modified LZW algorithm. In: Proceedings of the data compression conference IEEE computer society
18. Welch TA (1984) A technique for high performance data compression. In: IEEE computer 17, vol 6, pp 8–19

Detection of Schizophrenia Disorder from Ventricle Region in MR Brain Images via Hu Moment Invariants Using Random Forest

M. Latha(✉), M. Muthulakshmi, and G. Kavitha

Department of Electronics Engineering, MIT Campus, Anna University, Chennai, India
{lathakaran, lakshmingm.2, kavithag_mit}@gmail.com

1 Introduction

Schizophrenia (SZ) is a chronic and persistent brain disorder with onset in late adolescence. It is characterized by disintegration in perception of reality and cognitive domains, which leads to loss of intellectual functions and impairs quality of life in humans. Positive symptoms are hallucinations, delusions disorganized speech, and disturbed behavior. Negative symptoms are associated with disruptions to normal emotions and behaviors. Cognitive symptoms include problems related to attention, working memory [1]. Currently, the diagnoses of psychiatric disorders such as SZ are made on the basis of clinical manifestations and related psychosocial disturbances [2]. However, still, there is instability in diagnose of psychotic patients at an early stage of illness.

Magnetic resonance (MR) imaging is used to examine the anatomical change in the brain and examine the underlying neuropathology of SZ. Recently, machine learning techniques are used to classify SZ based on the structure of various regions in MR brain image. Structural findings show reduced cerebral volumes and enlarged ventricular volume in schizophrenic subjects [3–5]. The identification of imaging biomarkers enables early diagnosis, treatment, and track of illness progression in such disorders.

Diagnosis of schizophrenia is challenging one as its heterogeneous pattern that is analogous to other psychosis. Ventricle enlargement is a sensitive biomarker of neuropathological change associated with SZ [6, 7]. The ventricle is located in the center of the brain. Hence, any change in the surrounding structures affects its shape. Hence, segmentation of ventricle region is needed to identify the disease condition in SZ, as it is one of the biomarkers.

The segmentation of ventricle from brain MR images is a tough task. This is because of the existence of noise, the irregular intensity between tissues and boundaries. Edge-based methods suffer from blur and uncertain boundaries [8]. Recent literature shows that different methods such as multiphase level-set model and multiplicative intrinsic component optimization are used for tissue segmentation [9, 10]. The bias correction and segmentation of ventricles are carried out using variational level-set method and diffusion level set [11, 12].

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_24

Recently, feature-based analysis is utilized in the study of complicated disorders associated to brain. Hu moment invariants are used for pathological brain detection, to recognize the tooth in dental radiograph and iris recognition [13–15]. The invariance property of Hu moments describes the spread of the pixel intensities along the image. In this work, identification of pattern change in ventricle region in schizophrenic subjects is performed using Hu moment invariant feature. Recently, various machine learning techniques are applied for classification. Among them, random forest is widely used in EEG classification, dementia, and Alzheimer [16, 17]. Here, random forest classifier is adopted for classification of normal and SZ.

In this proposed work, a study is made to interpret the anatomical and pattern changes in ventricle using region-based energy minimization method and Hu moment invariants. First, images are subjected to segmentation using this method for detection of ventricles. The validation with the ground truth is done with the help of similarity measures. The texture pattern change in ventricle region is studied and analyzed with Hu moment invariants. Finally, the random forest classifier is used to classify the SZ subjects and normal. These are compared with state-of-the-art methods such as the support vector machine (SVM) with linear and RBF kernel, decision tree, logistic regression, and linear discriminant analysis (LDA) algorithm.

2 Database

2.1 Image Database

The brain MR images used in this work are collected from the collaborative Informatics and Neuroimaging Suite Data Exchange tool and funded by a Center of Biomedical Research Excellence (COBRE). Each subject has the demographic description that includes age, gender, educational qualification, and handedness [18].

3 Segmentation

The segmentation of different tissues from MR brain images is complex, due to inhomogeneity in the tissue texture [9]. Bias field blurs the images and thus decreases high-frequency information such as edges and contours. Hence, multiplicative intrinsic component optimization (MICO) method is used for simultaneous bias correction and segmentation of ventricle region in this work. Initially, the original raw images (normal and SZ) are subjected to skull stripping to delineate the non-brain tissues using non-parametric region-based active contour method [19].

3.1 Multiplicative Intrinsic Component Optimization (MICO)

The bias correction and segmentation of ventricle is adopted simultaneously with energy minimization framework in this work [10]. Energy minimization approach is

used to optimize the components bias field (B) and true image (J). The energy function is represented as

$$F_q(u,a,w) = \int_\Omega \sum_{i=1}^{N} \left| I(x) - w^T G(x) a_i \right|^2 u_i^q dx \tag{1}$$

where I is an image, $J(x) = \sum_{i=1}^{N} a_i u_i$ with "a" as a constant and u_i is the fuzzy membership function. G(x) and w is basis function vector and optimal vector, respectively.

The optimization of bias field and true image could be done by minimizing the energy F with reference to u, a, and w. The optimization of a, w, and u is carried out as

$$a_i = \frac{\int_\Omega I(x) B(x) u_i^q(x) dx}{\int_\Omega B^2(x) u_i^q(x) dx}, \; i = 1, \ldots N \tag{2}$$

$$w = \left(\int_\Omega G(x) G^T(x) \left(\sum_{i=1}^{N} a_i^2 u_i^q(x) \right) dx \right)^{-1} \int_\Omega G(x) I(x) \left(\sum_{i=1}^{N} a_i u_i^q(x) \right) dx \tag{3}$$

$$u_i = \frac{(\zeta_i(x))^{\frac{1}{1-q}}}{\sum_{j=1}^{N} (\zeta_i(x))^{\frac{1}{1-q}}}, \; i = 1, \ldots N \tag{4}$$

where $\zeta_i(x) = |I(x) - w^T G(x) a_i|^2$.

In this work, the number of iterations considered for the contours to converge at ventricle boundaries is 20. The ventricle is extracted from the other tissue structures using connected component analysis.

4 Validation Measures

The segmented ventricle area is correlated with the ground truth. The degree of likeness between the ground truth and the segmented ventricle regions are quantitatively validated with the similarity measures. The dice similarity (DS), Tanimoto coefficient (TC), Jaccard coefficient (JC), and volume similarity (VS) are used to validate the degree of correspondence between ground truth and segmented ventricle images [20]. All these measures range from low (0) to high (1).

5 Hu Moment Invariants

Moments and the related invariants are investigated to distinguish the structure in images. Hu developed six absolute and one skew orthogonal invariants. The zero-, first-, and second-order moments reveals the area, center of mass, and axis length and its

orientation angle, respectively. The higher order describes the image fine details [13]. Hu moments describe the pixel intensities and their locations. Hu moments are represented as a statistical expectation of variable [15]. These rotation invariant moments extract the characteristic attribute of the image.

6 Random Forest Classifier

A random forest (RF) classifier is an ensemble classifier that consists of many decision trees. It is a mixture of tree predictors, and the tree relay on random vectors that are individually sampled [21]. The simplification error depends on the solidity of the distinct trees and their correlations in the forest. The ensemble of trees is constructed by two levels. In the first level, based on bootstrapped form of the training samples each tree is grown-up. In the second level, the best split is chosen by tree predictors in a random manner at each node. If the forest is constructed, the allocation of a new instance to a class is done by linking the trees, with a majority voting. The low correlation of trees is guaranteed by bagging and selected random variable [16]. The algorithm is as follows:

T: Training set consists of (X1, Y1), (X2, Y2), …(Xn, Yn). Here X and Y are the input features and the labels.

N: number of trees.

M: number of variable for separating every node.

1 For i = 1:N // Training Task
2 The bootstrap sample (X) is drawn from the training set data T
3 RF tree is grown by repeating the procedure for every terminal node, until the least node is attained.
   a. Pick M variables randomly and determine the best split.
   b. Split the node into two daughter nodes
4 Classify by majority of voting between the N tress by averaging the predictions

Here, M is considered as the square root of the number of input variables. The number of trees is selected as 200.

7 Results and Discussion

The MR images (normal and SZ) in axial view are considered for this study. Here, 110 (50 normal and 60 abnormal) images were considered from COBRE database. The segmentation of ventricle region is implemented using MICO algorithm. The results based on MICO segmentation method for ventricle segmentation from normal and SZ

subjects are shown in Fig. 1. The skull stripped normal and SZ subjects are represented in Fig. 1a, b. The segmented and bias-corrected image is shown in Fig. 1c, d, e, f, respectively. Then, the ventricle is extracted from tissue structures using connected components analysis and is shown in Fig. 1g, h, and the ground truth is shown in Fig. 1i, j. It is observed that there is an enlargement of ventricle region in SZ subjects compared to the normal image. The segmented ventricle region is validated with the ground truth.

Figure 2a, b depicts the correlation plot of the normal and SZ images with the ground truth. The result implies that the segmented ventricle region highly correlates with the ground truth for both the normal and abnormal images. This shows that the pixels extracted from ventricle region by the MICO method are comparable with the ground truth. The segmentation algorithm is validated using similarity measures.

Figure 3 shows the similarity measures for the segmented region. The Dice coefficient, Tanimoto coefficient, Jaccard coefficient, and volume similarity are 0.967, 0.915, 0.936, and 0.983, respectively. This shows the efficacy of the algorithm in segmentation of ventricle region.

The Hu moments were extracted from the segmented ventricle region. These features are analyzed with translation along the vertical direction with an increment of 2 in the range of (−8, 8), scaling in the range (0.8, 1.20) over an increase of 0.05, and angle of rotation (−120, 120) with the increase of 30. The validation of Hu moment invariance with translation, scaling and rotation is shown in Fig. 4. The errors are reduced with no transform by keeping translation and rotation as 0 and scaling as 1. The Hu moment invariants such as H3, H5, and H7 show lesser error for no transformations.

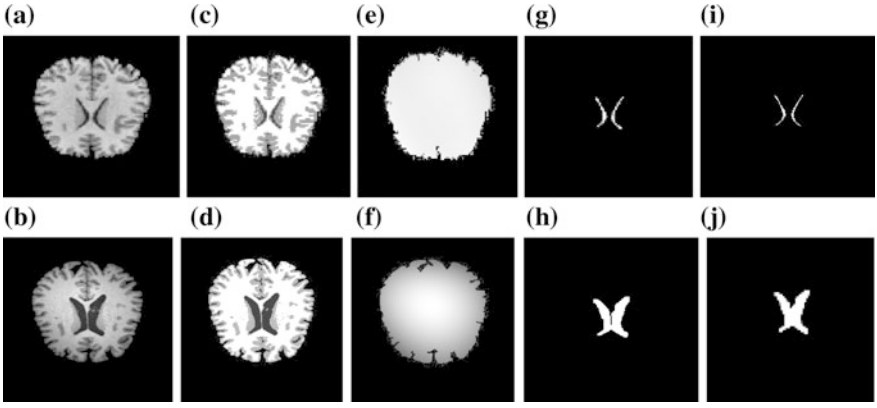


Fig. 1. Segmentation using MICO method (first and second rows are normal and SZ): **a**, **b** normal and SZ image; **c**, **d** segmented regions; **e, f** bias field estimation; **g, h** segmented ventricle; and **i, j** ground truth

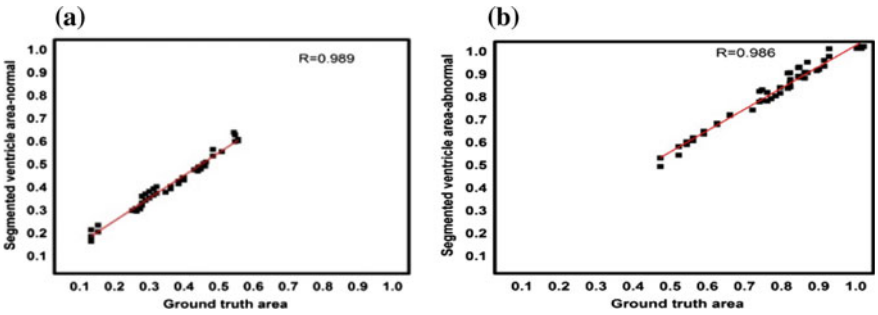


Fig. 2. Correlation plot for segmented normal ventricle image: **a** normal and **b** abnormal image

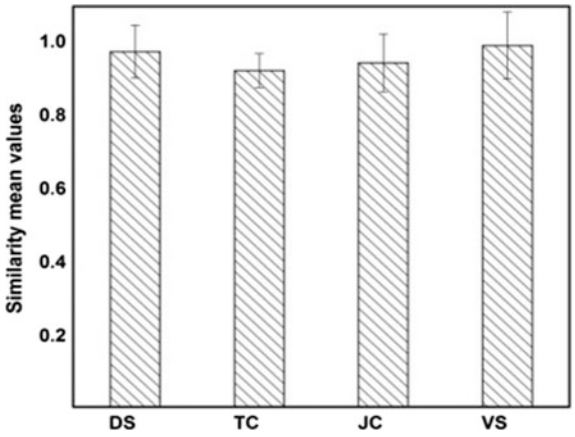


Fig. 3. Similarity measures for segmented ventricle region

This could be due to the capturing properties with axis and the angle of alignment along the principal axis. In addition, the higher moments outline the fine information of an image. The translation error lies in the range of 34–40% for all the Hu invariants. In H3, the maximum error for scale and rotation is 32.73% and 33.64%, respectively. In H5 and H7 the error for scaling is 36.36 and 30%. The rotation error for the same is less than 30%. In total, the Hu moment invariants of the image that is transformed are closely comparable to the original image.

Then, the classification performance measure such as accuracy, sensitivity and specificity are measured for the individual feature and entire feature set. The classification results using Hu moments with random forest classifier is shown in Table 1. The feature such as H3, H5, and H7 shows better accuracy. The classification accuracy (82.73%) is improved considerably for the entire feature set compared to the individual feature set. The sensitivity and specificity are 82% and 83.33%, respectively.

Figure 5 shows the ROC of Random forest classifier with Hu moment invariants. The area under the ROC is 0.827. Figure 6 demonstrates the variation in error with growing forest size. It is observed that the error is 17.27% with 200 trees and remains constant beyond 300. Hence in this analysis, the number of trees (N) considered is 200 with M as 3.

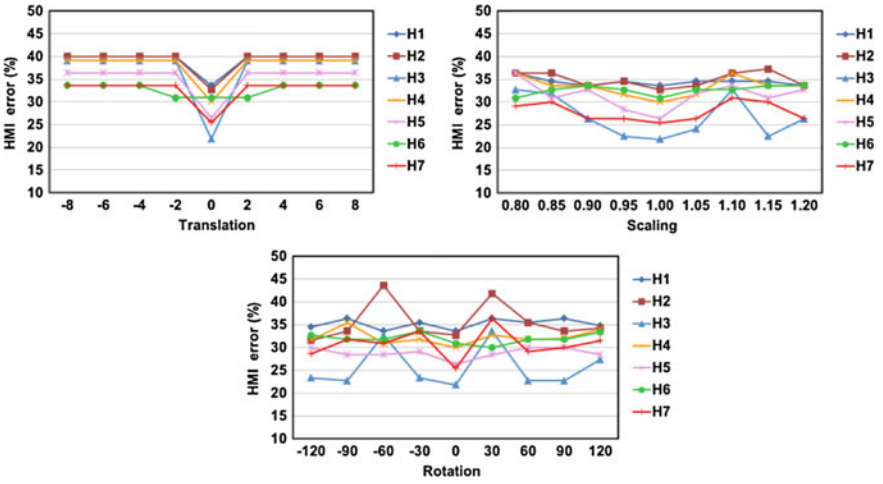


Fig. 4. Validation of Hu moment invariance (HMI) with translation, scaling, and rotation

Table 1. Performance measures of Hu moment invariants (HMI) using random forest classifier

Hu moment invariants	Accuracy (%)	Sensitivity (%)	Specificity (%)
H1	66.36	62	70
H2	67.27	72	63.33
H3	78.18	72	83.33
H4	70	70	70
H5	73.64	70	76.67
H6	69.09	66	71.67
H7	74.55	68	80
All features	82.73	82	83.33

The classification performance of the random forest is compared with commonly used classifiers, such as SVM (used as linear and RBF), decision tree, logistic regression and linear discriminant analysis (LDA). Here, the prediction with RF classifier is evaluated using tenfold cross-validation. The comparison results of classifiers are given in Table 2. The random forest gives better results compared to other classifiers.

Discussions

Segmentation of a region from MR images becomes difficult due to overlapping intensities present in the region. It is also identified from the results that MICO

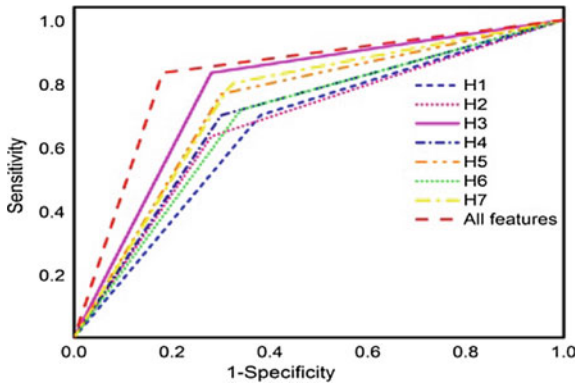


Fig. 5. ROC analysis of Hu moment invariants (HMI) using random forest classifier

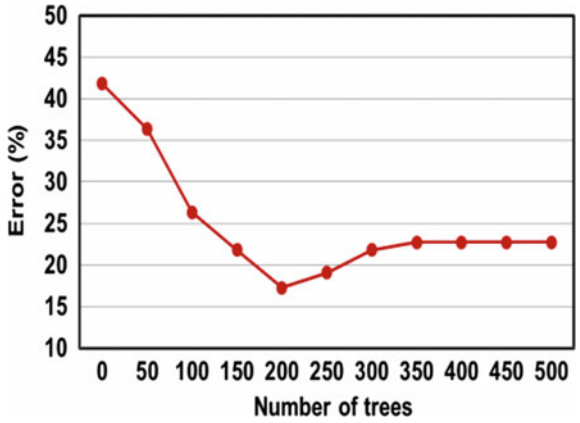


Fig. 6. Average error for different forest sizes with M

Table 2. Performance of the random forest classifier with state-of-the-art classifiers

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)
Random forest	82.73	82	83.33
SVM-linear	70	70	70
SVM-RBF	77.27	60	91.67
Decision tree	80	70	88.33
Logistic regression	76.36	68	83.33
LDA	67.27	72	63.33

simultaneously performs bias correction and segmentation of regions. After 20 iterations the energy decreases and converges. The high correlation of 0.989 and 0.986 illustrates that the proposed method is able to extract ventricle region more precisely. The high similarity measures such as DS, TC, JC, and VS corresponds to the closeness factor of the delineated ventricle results with ground truth.

It is noticed that without transformation the error is low for HMI features using random forest classifier. In addition, Hu moment invariants such as H3, H5, and H7 show a lesser error for the same. HMI with random forest method correctly predicts the instances with an accuracy of 82.73% compared to other classifiers. The AUC was 0.827 This shows that Hu moments that are extracted from ventricle region seems to be an important discriminating feature between SZ and normal subjects. This framework could be used to study the neuropsychiatric disorder such as schizophrenia.

In future, this work could be extended by considering other regions such as caudate, putamen, and thalamus from different views with more number of samples.

8 Conclusions

In this work, the classification of SZ from normal is identified using random forest classifier with Hu moments derived from ventricle region. At the beginning, the ventricle segmentation from MR brain images was implemented using multiplicative intrinsic component optimization method. The analysis shows that a multiplicative intrinsic component optimization method is capable to improve the bias correction and achieve better segmentation of ventricle from the skull stripped images. The calculated area from segmented ventricle region with ground truth shows high correlation (>0.98) for both normal and abnormal images. The high similarity measure shows the ability of the proposed method in segmenting the ventricle region. The Hu moment invariant along with random forest classifier gives better results compared to conventional methods. The classification accuracy is 82.73% for the entire feature set. The sensitivity and specificity are 82% and 83.33%, respectively. The AUC is 0.827. Hu moment invariant features extracted from segmented ventricle gives significant variation between the normal and SZ subjects. Hence, the Hu moment invariants extracted from ventricle along with random forest classifier could aid the physician in better diagnosis of schizophrenia subjects.

Acknowledgements. Data was downloaded from the Collaborative Informatics and Neuroimaging Suite Data Exchange tool (COINS; http://coins.mrn.org/dx) and data collection was performed at the Mind Research Network, and funded by a Center of Biomedical Research Excellence (COBRE) grant 5P20RR021938/P20GM103472 from the NIH to Dr. Vince Calhoun.

Informed consent. This study includes the images from the publicly available database, and the database has been cited and acknowledged.

References

1. Pawan KS, Ram Sarkar A (2015) Simple and effective expert system for schizophrenia detection. Int J Intell Syst Technol Appl 14(1):27–49
2. Nuechterlein KH, Dawson ME (1984) A heuristic vulnerability stress model of schizophrenic episodes. Schizophr Bull 10:300–312
3. Eleni Z, Thomas WJM, Stephen ML (2013) Towards the identification of imaging biomarkers in schizophrenia, using multivariate pattern classification at a single-subject level. NeuroImage Clin l(3):279–289
4. Haijma SV, Haren NV, Cédric WCP, Koolschijn MP, Hulshoff Pol1 HE, Kahn RS (2013) Brain volumes in schizophrenia: a meta-analysis in over 18,000 subjects. Schizophr Bull 39(5):1129–1138
5. Wright IC, Rabe-Hesketh S, Woodruff PW, David AS, Murray RM, Bullmore ET (2000) Meta-analysis of regional brain volumes in schizophrenia. Am J Psychiatry 157(1):16–25
6. Del Re EC, Konishi J, Bouix S, Blokland GAM, Mesholam-Gately RI, Gold-stein J, Kubicki M, Wojcik J, Pasternak O, Seidman LJ, Petryshen T, Hirayasu Y, Niznikiewicz M, Shenton ME, McCarley RW (2015) Enlarged lateral ventricles inversely correlate with reduced corpus callosum central volume in first episode Schizophrenia: association with functional measures. Brain Imaging Behav 10(4):1264–1273
7. Liu J, Huang S, Nowinski WL (2009) Automatic segmentation of the human brain ventricles from MR images by knowledge-based region growing and trimming. Neuroinform 7: 131–146
8. Chan TF, Vese LA (2001) Active contour without edges. IEEE Trans Image Process 10:266–277
9. Li C, Huang R, Ding Z, Gatenby C, Metaxas DN, Gore JC (2011) A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI. IEEE Trans Image Process 20(7):2007–2016
10. Li C, John CG, Christos D (2014) Multiplicative intrinsic component optimization (MICO) for MRI bias field estimation and tissue segmentation. Magn Reson Imaging 32(7):913–923
11. Kayalvizhi M, Kavitha G, Sujatha CM, Ramakrishnan S (2015) Analysis of anatomical regions in Alzheimer's brain MR images using level sets and Minkowski functional. J Mech Med Biol 15(2):1540024(1–7)
12. Anandh KR, Sujatha CM, Ramakrishnan S (2016) A method to differentiate mild cognitive impairment and Alzheimer in MR images using eigen value descriptors. J Med Syst 40(25): 1–8
13. Zhang Y, Jianfei Y, Shuihua W, Zhengchao D, Preetha P (2017) Pathological brain detection in MRI scanning via Hu moment invariants and machine learning. J Exp Theor Artif Intell 29(2):299–312
14. Zhang HF, Zhang X (2011) Shape recognition using a moment algorithm. In: International conference on multimedia technology, pp 3226–3229
15. Yaser DK, Sher AK, Farooq A, Saeed I (2014) Iris recognition using image moments and k-means algorithm. Sci World J 2014(723595):1–9
16. Chen W, Wang Y, Cao G, Chen G, Qiufang G (2014) A random forest model based classification scheme for neonatal amplitude-integrated EEG. BioMed Eng OnLine 13(Suppl 2): S4
17. Meenakshi D, van der Zandec JJ, van Dellena E, Sommer IEC, Scheltensc P, Lemstrac AW, Stama JC (2016) Random forest to differentiate dementia with Lewy bodies from Alzheimer's disease. Alzheimer's Dement Diagn Assess Dis Monit 4:99–106

18. Çetin M, Christensen F, Abbott C, Stephen J, Mayer A, Cañive J, Bustillo J, Pearlson G, Calhoun VD (2014) Thalamus and posterior temporal lobe show greater inter-network connectivity at rest and across sensory paradigms in schizophrenia. NeuroImage 97:117–126
19. Ni K, Bresson X, Chan T, Esedoglu S (2009) Local histogram based segmentation using the Wasserstein distance. Int J Comput Vision 84(1):97–111
20. Abdel AT, Allan H (2015) Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging 15(29):1–28
21. Breiman L (2001) Random forests. Mach Learn 45:5–32

Non-invasive Anaemia Detection by Analysis of Conjunctival Pallor

Medha Sharma(✉) and Bindu Garg

Department of Computer Science and Engineering, Bharati Vidyapeeth's College of Engineering, New Delhi, India
{medhall04, bindugarg80}@gmail.com

1 Introduction

The word anaemia owes its origins to the Greek word 'ἀναιμία *anaimia*', which means 'lack of blood'. It is a condition that is characterised by a reduction in the number of Red Blood Cells (RBCs) in the body. As per the statistics of the World Health Organisation (WHO), 1.62 billion people are affected by anaemia globally. This corresponds to approximately 24.8% of the total population. Southeast Asia has the maximum percentage of anaemic individuals [1].

Depending on the concentration of haemoglobin, anaemia is of three main types: mild, moderate and severe. The repercussions of anaemia tend to worsen with dwindling of haemoglobin concentration. While mild anaemia may result in weakness and lethargy, severe anaemia could lead to death, if no precautions are taken. What makes anaemia tough to diagnose is the fact that in the initial stages, it usually goes unnoticed. It is only when the concentration of haemoglobin undergoes a substantial reduction that people begin to notice it.

The diagnosis of anaemia is usually carried out via a blood test. This technique is invasive in nature since it necessitates injecting surgical instruments into the patient's body. While this may find application in most parts of the world, there are certain regions where this technique isn't of much use. One such example is the remote areas of our nation, which face acute shortage of healthcare facilities. Conducting blood tests isn't quite feasible in such a context. In such situations, doctors tend to rely on non-invasive techniques. This encompasses clinical evaluation of the patient, wherein the patient is tested for pallor in some specific regions like nailbed, tongue, palm, conjunctiva, etc.

2 Related Work

Diagnosing anaemia by means of non-invasive techniques isn't a recently discovered technique. It has been used by doctors as a preliminary assessment before the actual result is determined by means of a blood test. Several researches have been carried out to determine which region yields the best results, and also to ascertain the accuracy of this technique. After performing a detailed analysis of different researches, it was found

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_25

that a common consensus, regarding the best site to check for pallor, hadn't been arrived upon. Also, most of these studies [2–8] were based upon children inhabiting the developing nations, where anaemia is highly prevalent. These studies thus offer a rather narrow view of the problem, as they focus solely on children and ignore the other vulnerable age groups such as teenage girls and pregnant ladies. A compilation of the findings of a few such investigations is mentioned herewith.

Butt et al. [9] conducted a study that aimed at determining the accuracy of pallor for detecting mild and severe anaemia. The sites selected for the study were palm, nailbed and conjunctiva. The results of the study indicate that for detecting mild anaemia (Hb < 11.5 g/dl for females, Hb < 13 for males), pallor isn't a reliable sign. But in the case of severe anaemia (Hb < 7), pallor at all the three regions is accurate to a fairly decent level. Also, the absence of pallor at the palm and the conjunctiva is sufficient to rule out severe anaemia. The limitation of this study was the fact that anaemia was narrowed down to just two categories (mild and severe), instead of accommodating diversity by including three to four variants (severe, moderate, borderline and accurate). While most of the studies were performed by selecting several sites to determine the most accurate region, a few studies were based on the selection of one particular site (e.g. conjunctiva) and determining the accuracy of the selected site in the diagnostic procedure [10–12].

Chowdhury et al. [13] used pallor, in conjunction with medical history and a colour scale, to improve the accuracy of the diagnostic procedure. For this purpose, a new colour scale, consisting of 13 hues of red (from whitish to deep pink), was developed. The diagnosis was carried out in three steps. To begin with, the patient was examined for pallor in nine sites. This was followed by incorporation of patient's medical history. The last step was to assign a colour tint that was the closest to the shade of the palpebral conjunctiva. This technique classified patients into one of the four categories, viz. severe, moderate, mild and non-anaemic. The findings indicate that the incorporation of medical history improved the sensitivity and specificity to 78% and 85%, respectively. The drawback of this approach is the fact that paramedics might require training in order to use the colour scale effectively. The colour of the scale may fade over time, so due care must be kept in this regard. WHO had released a haemoglobin colour scale which was employed in a few other studies [14, 15]. The outcomes of such researches clearly indicate that usage of the colour scale enhances the accuracy.

Owing to the advances in technological domains, digital imaging-based techniques are also being proposed [16, 17]. Such studies use digital photography, followed by the analysis of various parameters of the captured picture, to estimate the level of haemoglobin. The major drawback associated with these techniques is to take into account various lighting conditions, as well as the properties of the camera being used for photography.

3 Objective

Diagnosis of anaemia, along with its therapeutic monitoring is generally carried out via blood test. This requires the presence of skilled staff and adequate resources. Moreover, if this process is to be repeated several times, it could be resource intensive, both in terms of monetary cost and the amount of blood used in the diagnosis. Furthermore, in resource-limited settings, proper healthcare facilities are scarce. This acts as a hindrance to the diagnostic procedure.

The objective of this study is to devise a non-invasive way of diagnosing anaemia in which merely a picture of the conjunctival region could be scanned for pallor and the result obtained post the image processing would be enough to diagnose whether a patient is anaemic. The limitations associated with the previous studies were taken into account while deciding the research methodology. The conjunctiva was chosen over the other sites since it gives a more accurate account of the disease, irrespective of skin pigmentations or any other factors. ImageJ tool has been used for image processing.

4 Proposed Solution

It's a well-known fact that haemoglobin absorbs lights of smaller wavelength like green and blue, while it reflects lights of higher wavelength, such as red. Thus, the concentration of haemoglobin in the body does leave an impact on the colour of the tissues located in the conjunctival region. This seemingly trivial piece of information served as the backbone of the proposed solution. The foremost step is to analyse the intensities of red, green and blue colours in the palpebral conjunctival region. The palpebral conjunctiva was chosen over the forniceal conjunctiva since the former is more erythematous as compared to the latter.

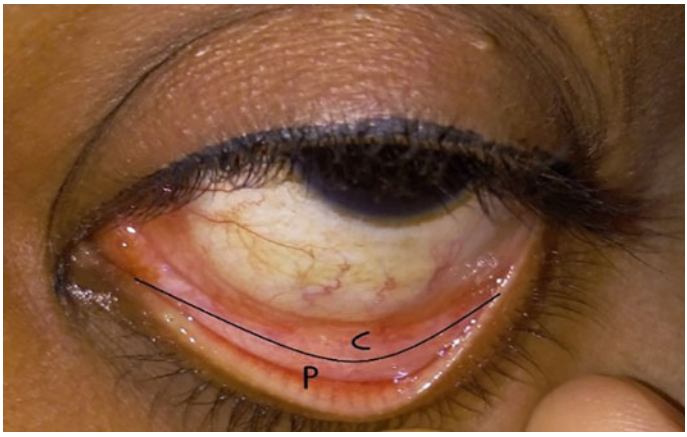


Fig. 1. Diagram depicting palpebral and forniceal conjunctiva (P: palpebral conjunctiva; C: forniceal conjunctiva)

This is followed by the calculation of 'Erythema Index' (EI), based on the analysis and calculations performed on the RGB components of the image. This index acts as a measure of the degree of anaemia (Fig. 1).

5 Research Methodology

5.1 Research Design

A cross-sectional study was conducted in the outpatient department of a few private hospitals in south and south-west Delhi, between August and September 2016. The inclusion criteria were as follows: age between 15 and 65 years and haemoglobin measured within 48 h of being photographed via a clinical laboratory test. The exclusion criteria were: use of eye drops while the study was being carried out, participant affected by any ocular disease, use of kajal, eyeliner or any other cosmetic that acted as a hindrance towards the study, being subjected to blood transfusion after being photographed but before the haemoglobin measurement was conducted, and unwillingness to provide consent to be a part of the study.

5.2 Conjunctival Sample Collection

A smartphone camera was taken to collect the conjunctival samples. The rear camera of Samsung Galaxy Grand Quattro (5 MP, 2592 × 1944 pixels) was used for this purpose, by means of the default camera application. The pictures were clicked in natural illumination or in ambient lighting, without using any flash. To ensure that the photograph obtained was of decent quality, 3–4 pictures were taken from each participant.

5.3 Image Processing

The Fiji (Fiji Is Just ImageJ) version of ImageJ was used for image processing. As mentioned on the official site of ImageJ (www.imagej.net/ImageJ), 'it is a distribution of ImageJ with many plugins useful for scientific image analysis in fields such as life sciences'.

Prior to being processed, each image was visually analysed so as to ascertain that sufficient portion of the palpebral conjunctival region is captured in the picture. The equation proposed by Yamamoto et al. [18] was modified to include the contribution of the blue colour channel as well. This led to the formation of Eq. 1. Here, red, green and blue denote the conjunctival brightness in the corresponding colour channels.

$$\mathrm{EI} = \log(\mathrm{Red}) - \log(\mathrm{Average}(\mathrm{Green},\ \mathrm{Blue})) \tag{1}$$

To select the palpebral conjunctival region, the Trainable Weka Segmentation tool was used. Two classes were created, the first corresponding to the region of interest, i.e. the palpebral conjunctival region while the second corresponded to the other extraneous region. Traces were added to both the classes and the classifier was trained by clicking on the 'Train Classifier' button. Upon the completion of the training process,

the segmented image overlayed with the corresponding class colours is displayed. To enhance the degree of accuracy of this process, the process can be repeated several times, by adding more traces as per requirements, till the desired level of accuracy is reached (Fig. 2).

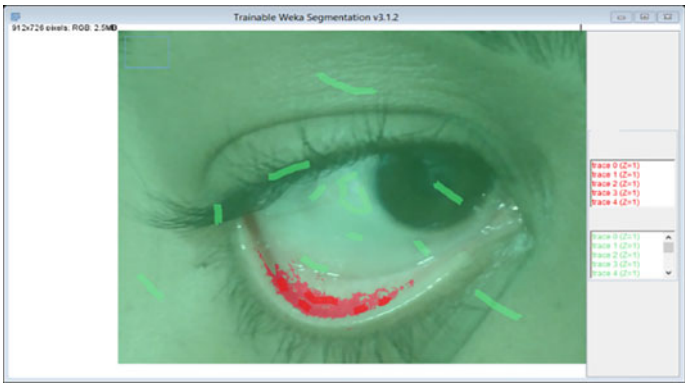


Fig. 2. Training Weka Segmentation output

Once the desired region is segmented, the image split into RGB channels. The average of the green and blue channels is calculated thereafter, using the imageCalculator tool, and the resulting channel is named as average. Then, the log function is used on the red channel and the average channel. The imageCalculator is again used to subtract the log average channel from the log red channel. The EI is given by the magnitude of the pixel brightness values, in the resulting image. Though the entire process is carried out manually, one could write a macro for the same that would ease things considerably and automate the process to some extent.

5.4 Laboratory Haemoglobin Measurement

Haemoglobin test conducted in the laboratory is used as gold standard. BeneSphera 3-part haematology analyser was employed to analyse the blood samples. The sample tube is placed underneath the sampling needle and the aspiration key is pressed. A WBC histogram would be generated after the sample has been tested. There's an option to adjust the thresholds in case of any anomaly in the threshold. Upon generation of the result, the report could be created via the print command.

6 Results

For analysing the credibility of the proposed method, 24 participants were tested over a period of 6 weeks. The mean age of the participants was 35.625 years. Females comprised 67% of the participants. The mean haemoglobin was 10.633 g/dL. As per

the equation mentioned in the previous section, the EI values for different individuals were calculated. On the basis of those computed EI values, the EI cut-off threshold for different haemoglobin concentration ranges was formulated. Table 1 illustrates the same.

Table 1. EI cut-off values for different Hb values

Haemoglobin value (g/dL)	EI cut-off values
Hb $\leq$ 7	0.15
7 < Hb $\leq$ 9	0.23
9 < Hb $\leq$ 11	0.30
11 < Hb $\leq$ 15	0.46
15 < Hb < 18	0.60

A graph depicting the relationship between the palpebral conjunctival EI and the measured haemoglobin was also drawn. Linear regression was employed to find the best fit, which was denoted by the equation: $y = 0.0383x - 0.127$ (where y represents the EI, and x represents the haemoglobin value). The R^2 value obtained was 0.9899. The sensitivity was 87.5% while the specificity was 88.9% (Fig. 3).

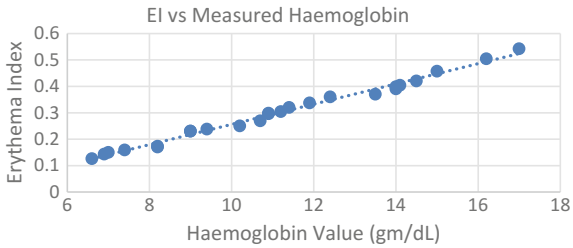


Fig. 3. A graph depicting the relationship between the Erythema Index and the measured haemoglobin

7 Limitations

This study has a few limitations. The foremost limitation would be that it's a semi-automatic process. This necessitates human intervention for its functioning. While the macro does automate things to some extent, it doesn't eliminate the need for human input. If the input given to the software isn't correct, it may lead to anomalous result generation. Thus, a qualified and knowledgeable individual is needed to carry out this task. The quality of the camera being used would, in turn, affect the image quality. So a minimum of 5 MP camera is necessary to obtain decent results. Furthermore, the illumination while the picture is being clicked does play an important role in the process. The best results are obtained in natural lighting, devoid of any artificial source

of light such as flash. The research was tested on a limited dataset, so it might be possible that certain modifications to the algorithm need to be made while applying this to some other dataset.

8 Conclusion and Future Scope

A reliable non-invasive way to diagnose anaemia can be made possible by augmenting the currently followed procedure with technologies like classification and image processing. Such modern technologies increase the reliability and accuracy of the existing methodologies. The limitations associated with the existing non-invasive techniques have also been alleviated using this technique.

The study could find applications in several areas. The foremost application would be in those regions where a major section of the population is anaemic and where there are no proper healthcare facilities that enable its timely detection. It could also be of great use for individuals who need to undergo periodic blood tests. Through this technique, there won't be any need to take blood samples each time the haemoglobin concentration is to be determined, thereby saving some amount of blood. Another important issue worth noticing is that blood test generally requires skilled professionals and sterilised equipment. The absence of either of these could cause harm to the individual who opts for the blood test. However, in case of non-invasive anaemia detection, there is no such constraint. So individuals who are extremely particular about the proficiency of the doctors or about the sterility of the surgical equipment being used can surely rely on this method. In the future, a dedicated smartphone app catering to this process could be implemented. Since smartphones have become a common commodity, it would ensure that a large number of people could be benefitted.

Note: Prior permission was taken from the CMS of the hospital. All the patients were duly informed about the study and only those individuals who agreed to be a part of it were photographed.

References

1. World Health Organisation (2008) Worldwide prevalence of anaemia report 1993–2005. World Health Organisation, Geneva
2. Kalter HD, Burnham G, Kolstad PR, Hossain M, Schillinger JA, Khan NZ, Saha S, De Wit V, Kenya-Mugisha N, Schwartz B, Black RE (1997) Evaluation of clinical signs to diagnose anaemia in Uganda and Bangladesh, in areas with and without malaria. Bull World Health Organ 75(Suppl 1):103
3. Zucker JR, Perkins BA, Jafari H, Otieno J, Obonyo C, Campbell CC (1997) Clinical signs for the recognition of children with moderate or severe anaemia in western Kenya. Bull World Health Organ 75(Suppl 1):97
4. Thaver IH, Baig L (1994) Anaemia in children (part I): can simple observations by primary care providers help in diagnosis? J Pakistan Med Assoc 44:282

5. Luby SP, Kazembe PN, Redd SC, Ziba C, Nwanyanwu OC, Hightower AW, Franco C, Chitsulo L, Wirima JJ, Olivar MA (1995) Using clinical signs to diagnose anaemia in African children. Bull World Health Organ 73(4):477
6. Weber MW, Kellingray SD, Palmer A, Jaffar S, Mulholland EK, Greenwood BM (1997) Pallor as a clinical sign of severe anaemia in children: an investigation in the Gambia. Bull World Health Organ 75(Suppl 1):113
7. Stoltzfus RJ, Edward-Raj A, Dreyfuss ML, Albonico M, Montresor A, Thapa MD, West KP, Chwaya HM, Savioli L, Tielsch J (1999) Clinical pallor is useful to detect severe anemia in populations where anemia is prevalent and severe. J Nutr 129(9):1675–1681
8. Chalco JP, Huicho L, Alamo C, Carreazo NY, Bada CA (2005) Accuracy of clinical pallor in the diagnosis of anaemia in children: a meta-analysis. BMC Pediatr 5(1):46
9. Butt Z, Ashfaq U, Sherazi SF, Jan NU, Shahbaz U (2010) Diagnostic accuracy of pallor for detecting mild and severe anaemia in hospitalized patients. J Pakistan Assoc 60(9):762
10. Sheth TN, Choudhry NK, Bowes M, Detsky AS (1997) The relation of conjunctival pallor to the presence of anemia. J Gen Intern Med 12(2):102–106
11. Wurapa FK, Bulsara MK, Boatin BA (1986) Evaluation of conjunctival pallor in the diagnosis of anaemia. Journal Tropic Med Hyg 89(1):33–36
12. Ekunwe EO (1996) Predictive value of conjunctival pallor in the diagnosis of anaemia. West Afr J Med 16(4):246–250
13. Chowdhury ME, Chongsuvivatwong V, Geater AF, Akhter HH, Winn T (2002) Taking a medical history and using a colour scale during clinical examination of pallor improves detection of anaemia. Tropical Med Int Health 7(2):133–139
14. Van den Broek NR, Ntonya C, Mhango E, White SA (1999) Diagnosing anaemia in pregnancy in rural clinics: assessing the potential of the Haemoglobin Colour Scale. Bulletin World Health Organization 1(77):15–21
15. Montresor A, Albonico M, Khalfan N, Stoltzfus RJ, Tielsch JM, Chwaya HM, Savioli L (2000) Field trial of a haemoglobin colour scale: an effective tool to detect anaemia in preschool children. Tropical Med Int Health 5(2):129–133
16. Suner S, Crawford G, McMurdy J, Jay G (2007) Non-invasive determination of hemoglobin by digital photography of palpebral conjunctiva. J Emergen Med 33(2):105–111
17. Collings S, Thompson O, Hirst E, Goossens L, George A, Weinkove R (2016) Non-invasive detection of anaemia using digital photographs of the conjunctiva. PLoS One 11(4): e0153286
18. Yamamoto T, Takiwaki H, Arase S, Ohshima H (2008) Derivation and clinical application of special imaging by means of digital cameras and ImageJ freeware for quantification of erythema and pigmentation. Skin Res Technol 14(1):26–34

Design and Implementation of Portable and Compact Human Heartbeat Rate Monitoring System

Kishor Kumar Das(✉), Ram Kishore Roy, Hidam Kumarjit Singh, and Tulshi Bezboruah

Department of Electronics and Communication Technology, Gauhati University, Guwahati, India
{kkdas0ll2, r.kore5lguece}@gmail.com, kumarjit_hidam@yahoo.com, zbt_gu@yahoo.co.in

1 Introduction

In our body, heart is responsible for providing blood circulation continuously. A cardiac cycle consists of one contraction (systole) and one relaxation (diastole) state of heart. During systole state, the heart collects blood and at the end of diastole state, the purified blood pumps out to the blood vessels. Contraction of atria starts the cardiac cycle and the relaxation of ventricles completes the cardiac cycle of heart. One heartbeat is one complete cardiac cycle. Heartbeat rate (HBR) can be defined as the cardiac cycle per minute or beats per minute (bpm). Adult person can have average HBR of 72 bpm. Normal HBR for resting adult person is in the range of 60–100 bpm. If the HBR falls below 60 bpm, it is called Bradycardia, and if it is faster than normal range, i.e., 100 bpm, then it is called Tachycardia. The HBR changes during sleep, stress, anxiety, illness, physical activities, and cardiovascular disease. There are two techniques for HBR measurement, namely: (i) electrocardiography (ECG) and (ii) photoplethysomography (PPG) [1]. During each heartbeat, some electrical changes occur in our body which is detectable by placing suitable sensors over some of our body part. ECG is used to detect this changes that uses electrodes on the skin over our chest. This method requires direct electrical connection with our body through the electrodes. It is a commonly used method for medical purpose and requires medical experts for operating the device. The PPG works on the optical properties. It is non-invasive method which does not require direct electrical connection. It is a technique which is based on the light reflection and transmission properties. It uses light source and light detector. The PPG can measure blood volume changes in an organ, e.g., fingertip. In a cardiac cycle, the heart contracts and expands, which causes a change in blood volume in the finger. Blood volume inside the fingertip increases during the diastole state and decreases during the systole state of a cardiac cycle. Blood volume synchronously changes with the heartbeat. The pulse of blood volume inside the fingertip is directly proportional to the HBR. We can determine the HBR in bpm by counting the number of pulses per minute. The PPG signals can be obtained in two modes, namely: (a) light reflection mode (LRM) and (b) light transmission mode

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_26

(LTM). In LRM, the medium is placed over the sensor system or vice versa. A photodetector (PD) detects the reflected light from the medium. The medium may be fingertip, forehead, etc. In LTM, the medium is placed between the light source and the PD. A PD detects the transmitted light in the medium. The medium can be fingertip, earlobe, etc. [1]. The PPG signal can be obtained by using simple electronic circuit. Though PPG has many advantages, it has some limitations too. They are as follows: (i) it has large DC component compared to the actual PPG signal, (ii) the PPG signal is undetectable for reduced blood level, (iii) other sources of light affects the PD, and (iv) the reflected light received by the PD changes due to motion.

2 Related Works

In the year 1991, Y. Iyriboz et al. had studied the noninvasive measurement of heart rate accurately at rest and during exercise. But during heavy exercise, this method shows some error about 9% [2].

In the year 1992, Joseph M. Schmitt et al. had proposed a noninvasive method that can measure hemoglobin concentration and oxygen saturation continuously [3].

In the year 2008, Fezari et al. had developed a real-time heart rate monitoring system by using microcontroller (μC). Doctor can access patient's pulse rate file sent through email every 24 h. Athletic persons can use this system which can read, write, and analyze the pulse rate [4].

In the year 2009, Jubadi et al. had proposed a heart rate monitoring system which can send the resultant HBR as SMS to a family member or to doctor. This is based on the PPG technique which is explained above. The sensor output was processed by using PIC16F87 μC to count the HBR per minute. An alert was given to medical experts or family members via SMS. With the help of this system, doctors could monitor and diagnose patient's condition continuously and could suggest them precautions if any [5].

In the year 2012, Verma et al. had developed a system by using which patients are able to measure their own heart rate and body temperature. Health professionals can examine the status of the patient through messages sent by GSM module connected to the system along with time, date, name of the patient, location of the patient, etc. [6].

In the year 2012, D.J.R. Kiran Kumar et al. had developed a health monitoring system and necessary data acquisition system to study remotely the parameters of health of the patient like heart rate, ECG, body temperature, blood oxygen saturation level, and blood pH level. Using this system, a doctor can monitor a patient on PC using Zigbee wireless module [7].

In the year 2012, Chi Kin Lao et al. had proposed a portable pulse rate detector system. In this system, the heart rate was calculated from the PPG signal and transmitted to computer or smartphone by using radio frequency (RF) transmitter section. To view the HBR on smartphone, they have also developed an Android-based application [8].

In the year 2013, Deshmukh had developed a real-time patient monitoring system which helps medical professionals to see status of patient's health through website.

This system has temperature sensor and pulse rate sensor along with a Wi-Fi module [9].

In the year 2013, N. Singh et al. had designed sensor nodes for measuring heart rate and temperature, which have remote monitoring capability along with wireless sensor module. This system can transmit the sensor data wirelessly to the remote monitoring station and to the controller by using RF transmitter and receiver module [10].

In the year 2013, Venugopal et al. had developed a centralized heart rate monitoring system. Here, the sensor data is collected from various patients and transmitted over a Wireless Local Area Network (WLAN). This system can be used in medical purposes [11].

In the year 2014, S. Saravanan had proposed a system that can monitor heartbeat of a patient and capable to send the data by using computer communication networks, which support Wi-Fi, Bluetooth, etc. This system can monitor ECG, pulse, blood pressure, arterial oxygen saturation, blood glucose concentration, etc. They constructed a peer-to-peer messaging system which can send to the doctor where the monitoring section receives the data via Wi-Fi, Bluetooth, and Internet [12].

In the year 2014, Purnima et al. had proposed a GSM and Zigbee technology-based health monitoring systems which can monitor and transmit ECG, temperature, and heartbeat signals continuously. Every patient monitoring system was connected to a Zigbee node. Doctors can receive the signals through computer or mobile via Zigbee as well as GSM technology [13].

In the year 2014, Chandana et al. had proposed a system by using which one can stay connected with doctor. In emergency, one can take immediate action by using this system. This system can measure blood pressure, drug level, and HBR. This data can be sent via GSM module to the doctor's mobile [14].

In the year 2014, M.V.N.R. Pavan Kumar et al. had developed a GPS-based system for health monitoring and tracking, which focuses on tracking a soldier's location, so that control stations can guide them in emergency. Using this system, central base station can get information about the body temperature, heart rate, and the location of a soldier [15].

In the year 2014, S. Bae et al. had proposed a health management system based on self-organizing software platform that uses smart devices. One can add new devices and services in this plug-in system without modification. They proposed a smart watch that can measure blood pressure of the person who is wearing the watch. The blood pressure monitor identifies the smart watch by transmitting a low-frequency signal. In order to record the blood pressure, the data collected by the blood pressure monitor is transmitted to the watch. Then, the watch sends this data to the router and to the main server [16].

3 Proposed Methodology

The objective of the proposed work is to design, fabricate, and implement an HBR monitoring system. Here, we have used one infrared (IR) light-emitting diode (LED) (wavelength: 930 nm) as light source and one IR detector in light reflection

mode. While we place our index finger over the IR sensor, the IR LED emits light which can pass through our finger. A portion of light is reflected back. The intensity of reflected light depends on the blood volume inside the fingertip. The intensity of reflected IR wave slightly alters with each heartbeat in a harmony. The IR detector detects this change of reflected IR signal. The pulsating reflected signal detected by the IR detector is converted to current or voltage signal by using the sensor circuit. The sensor output is visualized in digital storage oscilloscope (DSO: Model: Tektronix, TD20014C) and is processed by using μC, and the recorded data is then transferred to PC for storage. The proposed block diagram of the system is shown in Fig. 1.

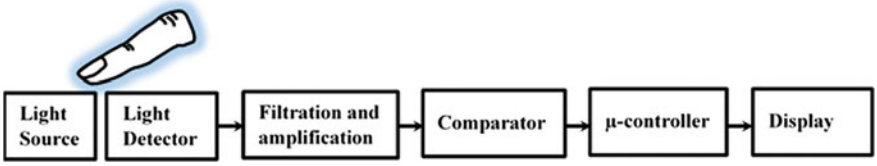


Fig. 1. Block diagram of the system

4 Circuit Implementation

4.1 The Sensor Circuit

An optical sensor TCRT5000 (from Vishay Semiconductors) is used to obtain the PPG signal in reflection mode. It consists of an IR emitter and a phototransistor. It can also block visible light. The sensor system is connected to +5 V power source as shown in Fig. 2. The signal detected by the sensor is fed to the amplifier circuit by using a coupling capacitor.

4.2 The Filter and Amplifier Circuit

The heartbeat signal frequencies lie between 0.5 and 4 Hz. We have used a low pass filter to bypass the high-frequency components above the required frequency level. The sensor output is fed to the non-inverting terminal of the first amplifier through a coupling capacitor. The capacitor C_1 and the resistor R_3 form a low-pass filter. The cutoff frequency of the filter is calculated as

$$F_s = \frac{1}{2\pi R_3 C_1} \tag{1}$$

Here, the cutoff frequency is almost 2.4 Hz. An amplifier is used to amplify the amplitude of the PPG signal. We have used IC 741 operational amplifier (OPAMP) based amplifier in non-inverting mode. The gain for the first stage of the amplifier is fixed at 101. We used a two-stage amplifier circuit for the system, and the overall gain of amplifier is $\sim 10^4$. The amplified output from the first stage is fed to the second stage of the non-inverting amplifier as shown in Fig. 2. The cutoff frequency and the

gain are same for the second stage filter and amplifier as in the first stage. The final amplified output is fed to a comparator.

4.3 The Comparator Circuit

The Comparator gives a binary output signal indicating which is larger by comparing two voltages or currents. It has two analog input terminals V_{IN} and V_{REF}. The output of a comparator can be formulated as

$$V_O = \begin{cases} 1, & V_{IN} > V_{REF} \\ 0, & V_{IN} \leq V_{REF} \end{cases} \tag{2}$$

Here, we have used OPAMP to design a comparator circuit as shown in Fig. 2. One adjustable reference voltage is applied to the non-inverting terminal. We have used one Trimpot (Bourns 3296) to adjust the reference voltage. The amplified sensor signal is applied to the inverting terminal. The resistance, R_9, and capacitor, C_5, are used to remove unwanted noise components from the input signal, which also increases the stability of the signal. The comparator gives logic level output by comparing the input signal with the applied reference voltage. The comparator output is fed to μC 8051 for further processing and display.

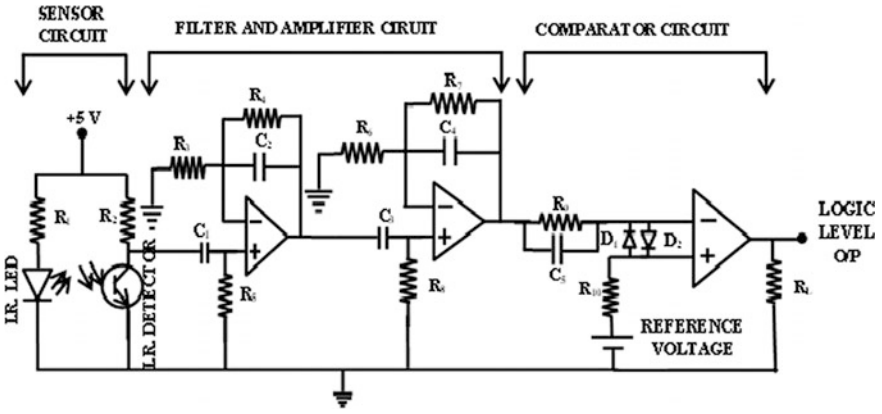


Fig. 2. Signal conditioning circuit diagram

4.4 Interfacing and Processing

The complete setup of the system has been shown in the Fig. 3. The comparator output is directly interfaced to Timer 1 (pin 15) of 8051 μC. We have used timer 0 for 1 min delay generation and Timer 1 for counting the pulse rate from the PPG signal, since the HBR is measured as the number of heartbeats in 1 min. We have created one window of 1 min duration to calculate HBR that occurs during this interval. The timer and counter start from 0 and count the pulses and stop after 1 min duration. The data are

stored in memory location. The count value gives the HBR per minute. To display the result in LCD, the resultant data which are in Hexadecimal code are converted to Decimal and then the Decimal to ASCII code. Then, the data are sent to the LCD display JHD162A through Port 1 (pin 1–in 8) of 8051 μC continuously. The DSO is interfaced to the sensor circuit for capturing the output data. The sample output PPG signal captured by DSO is shown in Fig. 5a. Also, an algorithm has been developed for monitoring the HBR at the interval of 5 s.

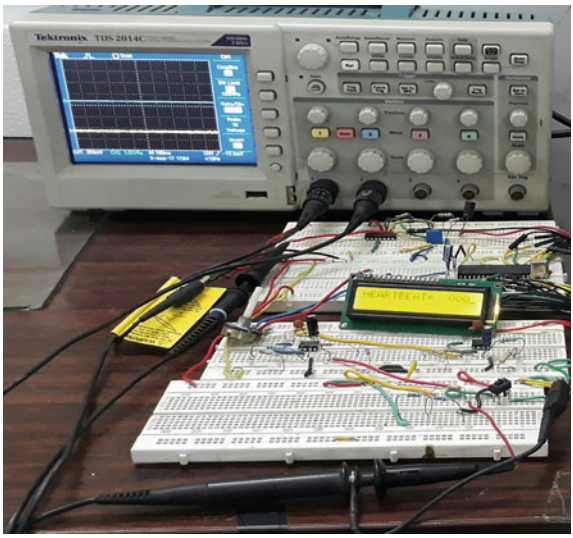


Fig. 3. Snapshot of the experimental setup

The theoretical HBR is calculated by using the period from peak to peak of the PPG signal. Reverse of the time period gives the frequency *(f)*. Then, the HBR can be calculated as

$$\text{HBR} = 60 \times f \text{ bpm} \tag{3}$$

5 System Calibration

The system has been calibrated and tested by using a test signal from function generator. Signals with different frequencies have been applied to the signal conditioner circuit, and the data are displayed in LCD after processing through the DAQ system. The plot of standard and measured HBR at different frequencies of the input signal is shown in Fig. 4a. The plot of % error between standard and measured HBR versus frequency of the input signal is shown in Fig. 4b. Now, we have applied the PPG signal to the input of the system which can process the PPG signal and give HBR per minute.

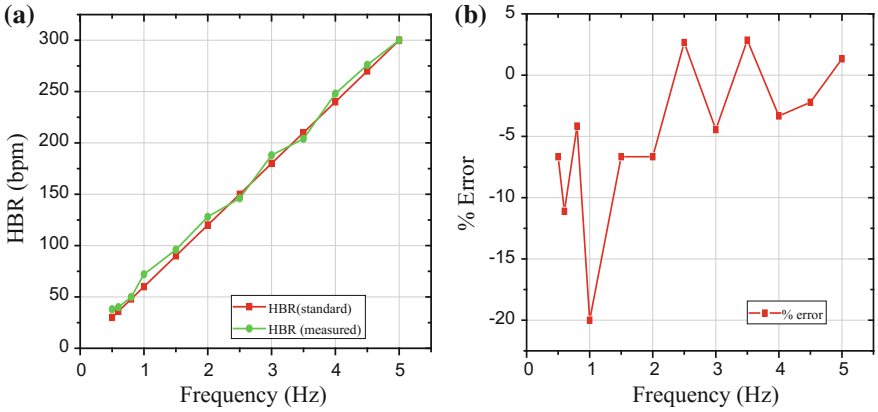


Fig. 4. **a** HBR standard and measured versus input frequency curve, and **b** % error versus frequency curve

6 Result and Discussion

The PPG signal output and the comparator output are shown in Fig. 5a. Theoretically, the HBR is calculated by using Eq. (3). In our experiment, the time period for the PPG signal is 800 ms, and the frequency is 1.25 Hz. Therefore,

$$\begin{aligned} \text{HBR} &= 60 \times f \text{ bpm} \\ &= 60 \times 1.25 \text{ bpm} \\ &= 75 \text{ bpm} \end{aligned}$$

which is equal to the value calculated by using the algorithm, and the HBR is displayed in the LCD screen as shown in Fig. 5b.

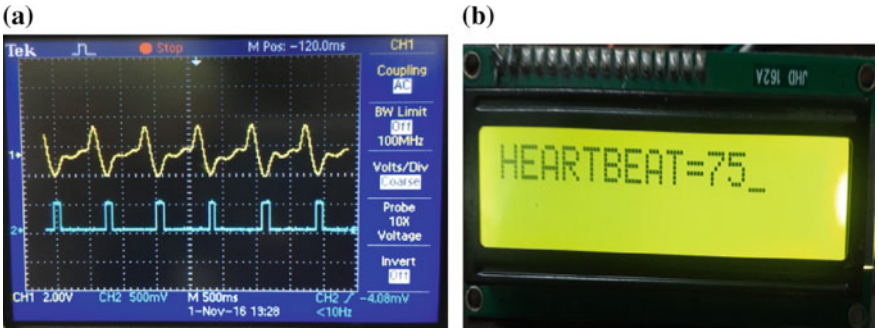


Fig. 5. **a** Snapshot of sensor output versus comparator, and **b** snapshot of HBR displayed in

The recorded HBR for different persons is given in Table 1.

Table 1. Experimental and theoretical values of HBR

Sl. No.	Person	Theoretical HBR (bpm)	Practical HBR (bpm)	(%) Error
1	A	76	75	1.31
2	B	62	65	4.84
3	C	62	64	3.23
4	D	70	68	2.86

7 Conclusion

In the proposed work, we have developed an HBR monitoring system. The system is tested and found that it is capable of monitoring the HBR of patient and other persons at home or during indoor activities. The system is noninvasive, because it is desired to be used during outdoor exercise or other physical activities. The system is desired to be run by the body power generated using piezoelectric sensor. Increasing movement of person leads to decrement of the accuracy of HBR measurement. A proper filter should be incorporated into the system to remove the unwanted signal components, which affect the PPG signal during motion.

Acknowledgements. The authors are thankful to Assam Science Technology and Environment Council (ASTEC), Govt. of Assam, India for providing financial support towards the work (Grant no. ASTEC/S&T/192(152)/14-15/1037 dated 19.05.2015).

References

1. Das KK, Roy RK, Singh HK, Bezboruah T (2016) An embedded system for monitoring pulse rate during indoor exercise. Advanc Res Electric Electron Eng 3(5):354–357
2. Iyriboz Y, Powers S, Morrow J, Ayers D, Landry G (1991) Accuracy of pulse oximeters in estimating heart rate at rest and during exercise. Br J Sports Med 25:162–164
3. Schmitt JM, Xiong ZG, Miller J (1992) Measurement of blood hematocrit by dual-wavelength near-IR photoplethysomography. SPIE Proc 1641
4. Fezari MD, Salah MB, Bedda M (2008) Microcontroller based heart rate monitor. Int Arab J Informat Technol 5(4)
5. Mat JW, Sahak S, Faridatul A (2009) Heartbeat monitoring alert via SMS. In: IEEE symposium on industrial electronics and applications
6. Verma S, Gupta N (2012) Microcontroller based wireless heart rate telemonitor for home care. IOSR J Eng (IOSRJEN) 2(7):25–31
7. Kumar DJRK, Kotnana N (2012) Design and Implementation of Portable health monitoring system using PSOC mixed signal Array chip. Int J Recent Technol Eng (IJRTE) 1(3):59–65
8. Lao CK, Che UK, Chen W, Pun SH, Mak PU, Wan F, Vai MI (2012) Portable heart rate detector based on photoplethysmography with android programmable devices for ubiquitous health monitoring system

9. Deshmukh RS (2013) Wi-Fi based vital signs monitoring and tracking system for medical parameters. Int J Eng Trends Technol (IJETT) 4(5):1935–1938
10. Singh N, Mishra R (2013) Microcontroller based wireless temperature and heart beat read-out. IOSR J Eng (IOSRJEN), 01–06
11. Venugopal K, Kumar A (2013) Centralized heart rate monitoring and automated message alert system using WBAN. Int J Sci Res Publicat 3(9)
12. Saravanan S (2014) Remote patient monitoring in telemedicine using computer communication network through bluetooth, Wi-Fi, internet android mobile. Int J Advanc Res Comput Commun Eng 3(7):7590–7596
13. Purnima P, Singh P (2014) Zigbee and GSM based patient health monitoring system. Int Confer Electron Commun Syst (IECS)
14. Chandana D, Hema Latha B (2014) A tele-medicine system for measuring heart rate, blood pressure, and drug level detection. IJEDR 2(1):23–29
15. Pavan Kumar MVNR, Vijay GR, Adhikrao PV, Vijaykumar BS (2104) Health monitoring and tracking of soldier using GPS. Int J Res Adv Technol 2(4)
16. Bae S, Kim K, Lee DY (2014) A scalable health management system based on a self-organizing software platform. IEEE, 70–80

Implementation of a Secure and Efficient Routing Algorithm for Vehicular Ad Hoc Networks

M. N. Prashanth(✉) and Annapurna P. Patil

Ramaiah Institute of Technology, Bangalore 560054, Karnataka, India
mnprashu@gmail.com, annapurnap2@msrit.edu

1 Introduction

1.1 Vehicular Ad Hoc Networks

Vehicular ad hoc network is a form of ad hoc network wherein intelligent vehicles take part in communication. Ad hoc network does not rely on infrastructure for communication with this principle; VANETs have additional constraint, highly dynamic environment.

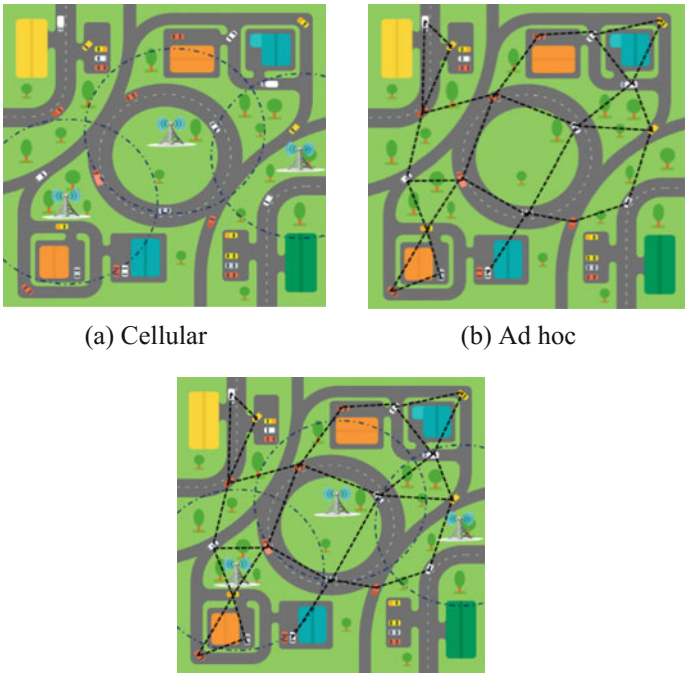
The architecture of VANETs falls under three categories cellular, ad hoc, and hybrid. As shown in the Fig. 1a, the access points are used for communication among the vehicles. These access points are installed at traffic intersections. Access points have their own advantages and limitations, and they have a fixed range for communication; moreover, the fixed infrastructure incurs more cost and also the maintenance cost. Best way to overcome is to have ad hoc network as shown in Fig. 1b; the vehicles form a network and communicate among themselves. Finally, the third kind is combination of both cellular and ad hoc shown in Fig. 1c; this incorporates the features of both cellular and ad hoc architecture. Namboodiri et al. [1] proposed such a hybrid architecture, which uses some vehicles with both WLAN and cellular.

1.2 Basic Characteristics

Vehicular ad hoc networks comprise vehicles which act as mobile nodes and also routers for other vehicles. VANETs have unique characteristics in addition to ad hoc networks characteristics such as dedicated short-range communication, self-organizing, and configuring, which distinguish it from other ad hoc network. The basic characteristics are as follows [2]:

1. Highly dynamic environment: The topology of VANETs changes frequently due to high-speed movement between vehicles. For example, the link among vehicles exists if they fall under the transmission range. In case of vehicle moving in opposite direction with high speed, the link lasts only few seconds; this makes the topology to break.
2. Frequent link failure: Due to high dynamic topology, the connection among the vehicles get disconnected frequently. Especially in the low density, the probability of link failure is higher.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_27



(a) Cellular (b) Ad hoc

(c) Hybrid

Fig. 1. Vehicular ad hoc network architecture

3. Communication environment: Typically, VANETs' communication falls under two scenarios: highway traffic and street in a city, where the communication is complex and no direct communication as it comprises more obstacles (buildings, trees). In the latter case, it is simple and straightforward as it has one-dimensional movement.
4. Sufficient energy and processing: VANETs have sufficient processing power and energy compared to the small mobile handheld devices.
5. Geographical type of communication: To forward packets in geographical area, VANETs have a new type of communication unlike other networks where multicast and unicast are used.
6. Delay constraints: For safety applications, the messages should be communicated within certain time. In this case, the network need not to have high throughput but the constraint is on delay. Network should have minimum delay.

2 Related Work

In [3], author considers the vehicles which are few hops away in the network for the routing. Considering the small-scale network with few hops, the traditional algorithm AODV is applied. However, the routes determined by the algorithm prone to frequent

breaks due to dynamic mobility. The author proposed prediction-based algorithms which are PRAODV and PRAODVM; both the algorithms use acceleration and position of the vehicles to predict the path lifetime. The simulation results of the proposed work showed considerable improvement with respect to packet delivery ratio (PDR). The algorithm performance highly depends on the outcome of prediction accuracy.

In [4] AODV protocols has been altered to forward the request within the zone, i.e., zone of relevance (ZOR). ZOR has a fixed range for the communication. The range is determined by the nature of the application [5]. The idea is similar to that of location aided routing [6].

In [7], the author proposed a location-based algorithm called LORA_CBF, which uses flooding concept. In this, each node can be a gateway, cluster head, or member. Similar to the greedy algorithm, the packets are forward. If the destination location is not known, the source node initiates the location request (LREQ) packet. This is similar to that of AODV route discovery phase, but in this case, the dissemination of the packet, i.e., LREQ and LREP, is done by the cluster heads and gateway only. The simulation results showed that performance of AODV and DSR was significantly lower than LORA_CBF due to the constraint scalability and mobility.

3 Proposed Algorithm

In VANETs, because of highly dynamic topology, discovering and maintaining route is a challenging task. Many routing protocols were proposed, which are classified under the following category ad hoc, position based, broadcast, geocast, and cluster-based routing. As stated earlier, the VANETs incorporate the features of ad hoc network. The routing protocols like AODV [8] and DSR [9] are still applicable for VANETs, but they are designed for general ad hoc network; the route maintenance is not done unless it is required. However, in case of VANETs, due to dynamic topology, it is unable to find the route quickly and also, it is difficult to update and maintain the route information. This degrades the performance as packets are lost due to frequent route failure. Therefore, the routing protocol must possess the following features: (1) better route convergence time and (2) rate of link failure should be minimal.

The bio-inspired routing protocol AntHocNet [10] uses ant agents to discover and maintain the route. The algorithm has proven its performance with respect to route convergence time and rate of link failure for mobile ad hoc networks (MANETs). But for VANETs, the direction and speed of moving vehicle need to be considered. In this paper, we describe how the AntHocNet algorithm is modified to improve the performance considering the direction of the moving vehicle.

In the proposed algorithm, two cases are considered during routing phase: (1) destination vehicles moving in the same direction as of source and (2) destination vehicle moving in opposite direction. Figure 2a, b shows the proposed algorithm for destination vehicle moving in the same and opposite direction to that of source. In the proposed algorithm, the forward ants are generated at the intermediate nodes only if the criterion is satisfied. In case of destination moving in the same direction as of source, the axis of moving nodes is compared with that of source; if it matches, then that node

will take part in the routing as an intermediate node. The process continues until the destination is reached. In case of destination vehicle moving opposite to source, nodes moving in the direction of destination are considered to take part in the routing process as an intermediate node.

```
Algorithm: Nodes moving in same direction
Begin:
        Generate the ants to find the route between the pair of nodes.
        If node is destination
        Begin:
                Generate the backward ants
                Halt
        End
        Else if it is an intermediate node
        Begin:
                If the node is moving in the same lane
                Begin:
                    Generate the forward ants
                End
                Else
                Begin:
                    Discard
                End
        End
End
```

```
Algorithm: Nodes moving in opposite direction
Begin:
        Generate the ants to find the route between the pair of nodes.
        If node is destination
        Begin:
                Generate the backward ants
                Halt
        End
        Else if it is an intermediate node
        Begin:
                If the node is moving in the opposite lane
                Begin:
                    Generate the forward ants
                End
                Else
                Begin:
                    Discard
                End
        End
End
```

Fig. 2. **a** Destination node moving in the same direction as of source, and **b** destination node moving in opposite direction to that of source

The networks are prone to attack, which often leads to poor performance. There are different types of attacks; few of them are attacks on identification and authentication, attacks on privacy, attacks on availability, network denial of service (DoS), and attacks

on data trust. In this paper, the attack on availability of the resource is described. In the hybrid architecture of VANETs, the roadside units (RSUs) act as a communication end point. The vehicles use the RSUs to transfer messages to other vehicles. In a safety application and smart transport system, RSUs act as a focal point wherein the crucial and traffic-related information are disseminated from RSUs. Therefore, RSUs act as a crucial resource which needs to be secured from attack. As stated earlier, the attack on availability is addressed. RSUs have the capability to process the requests that are sent from different vehicles and meantime, it also performs its own computation.

In VANETs, vehicles have sufficient energy and processing capability. But RSUs are devices that are installed at the road junction points to disseminate the messages during the critical situation; therefore, we assume that it has minimal energy and processing capacity. RSUs may get detached from the network due to low energy during the DOS attack, as it consumes more energy to process junk requests that are generated from the malicious nodes, which impact on the availability of RSUs.

In the proposed algorithm (Fig. 3), the minimum and maximum threshold values for energy are defined. The algorithm periodically monitors the network; if the nodes energy fall below the minimum threshold value, then that node is treated as victim of DOS attack. The algorithm identifies the malicious nodes by considering the maximum threshold values of energy and deactivates those nodes from the network. The pseudocode of the algorithm is shown in Fig. 3.

```
Algorithm: Identifying the malicious node
Begin:
        Periodically monitor the network nodes
        Check for nodes energy level
        If energy of the node is below the min threshold value
        Begin:
                Identify the nodes that has energy level more than the max threshold
                value
                  If found
                  Begin:
                          Deactivate the node
                  End
        End
End
```

Fig. 3. Identifying the malicious node

4 Simulation Environment

Table 1 summarizes the QualNet configuration setup under which the proposed algorithm is simulated and evaluated its performance with respect to the existing protocols.

The simulation is executed for 660 s, during which vehicles were configured to move in a fixed mobility path using fixed waypoint mobility model. The density of the vehicles is increased in the interval of 20 nodes (20, 40…). The granularity of the movement has been varied from the range 0.1–1.0 m. The simulation has been set up for the 2500 × 2500 urban terrain; totally, 12 RSUs were set up for the better

Table 1. Simulation configuration setup

Simulation tool	QualNet 7.4
Terrain (dimension)	2500 × 2500
Number of nodes (vehicles)	20, 40, 60, 80
Number of RSUs	12
Granularity (in meters)	0.1, 0.5, 1.0
MAC protocol	MAC 802.11p
Routing protocol	AntHocNet, Proposed_AntHocNet, AODV
Simulation time	660 s

communication at the road junction points. All the vehicles and the RSUs were configured to use MAC 802.11p protocol.

4.1 Simulation Results

The proposed algorithm has been compared with the existing algorithm with respect to varies performance metric, viz., throughput, average end-to-end delay, average jitter, network lifetime, average energy consumption, and number of link failures, after running extensive test cases. The primary focus of the proposed algorithm is to reduce the frequent link failure and to defend the network during the attack on availability. The performance of the proposed algorithm with respect to link failure and the energy consumption with attack and without attack are described in this section.

Frequent Link Failure

VANETs are highly dynamic in nature; the topology of the network changes frequently due to its high mobility; because of this, the communication link breaks frequently which degrades the performance of the network.

Figure 4 shows the number of link failures under the environment of 20 nodes with the varying granularity. It is observed that with low granularity the proposed and existing AntHocNet algorithm has no change in the link breaks as the granularity has been increased the proposed algorithm has better performance compared to existing algorithm. Both the existing and proposed algorithms have better performance compared to traditional AODV algorithm.

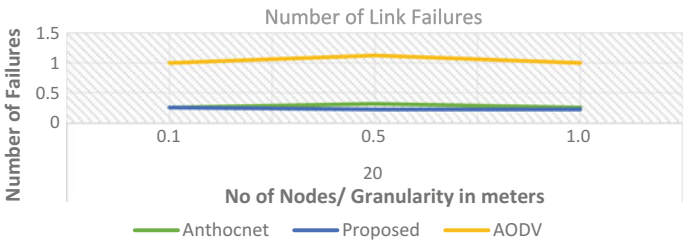


Fig. 4. Number of link failures under 20 nodes and varying granularity

Figure 5 shows the overall performance of the network with respect to number of link failures under the sparse, moderate and dense environment. From the figure, it is inferred that, when the granularity and also the number of nodes participating in the network are low, the number of link failures is less. As the granularity increases, the number of link failures also increased in the proposed algorithm but considerably less compared to the existing AntHocNet. In case of high granularity and moderate environment, the proposed algorithm has less number of link failures compared to the existing algorithm.

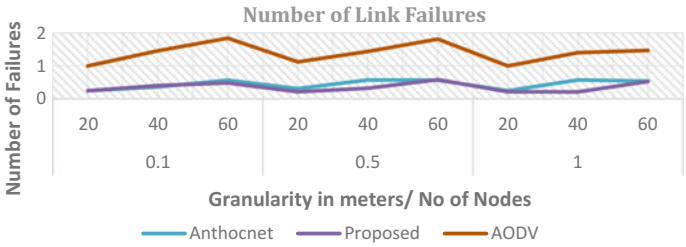


Fig. 5. Number of link failures under 20, 40, 60 nodes and varying granularity

From Fig. 6, it is observed that the enhanced algorithm has less delay compared to the traditional AntHocNet. It is also observed that there is no significant change in delay for both the enhanced and traditional AntHocNet over the network consisting of 40 nodes.

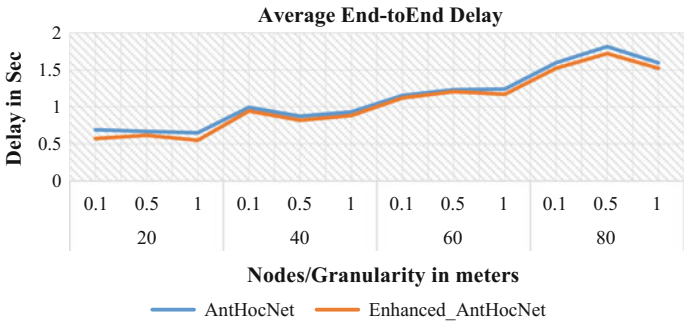


Fig. 6. Average end-to-end delay

Figure 7 shows that the proposed algorithm has shown the significant difference over the traditional AntHocNet with respect to average jitter. In the dense environment (consisting of 80 nodes), the proposed algorithm as less jitter compared to the existing AntHocNet.

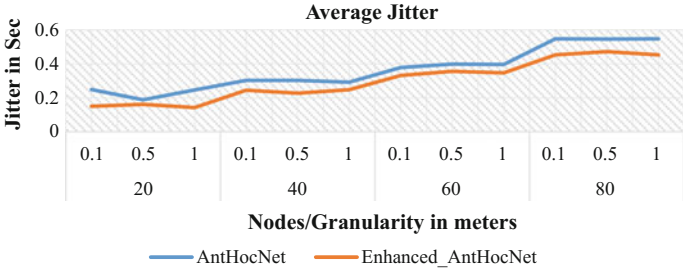


Fig. 7. Average Jitter

5 Conclusion

Vehicular ad hoc networks are particular form of wireless networks formed by vehicles. VANETs comprise vehicles which act as mobile nodes and also routers for other vehicles. Vehicular communication contributes to safer and more efficient roads by providing timely information to drivers and also to make travel more convenient. In this paper, the routing and security aspects of VANETs are addressed. Based on the direction, the paper focused on two cases: (1) destination vehicle moving in same direction as of source and (2) destination vehicle moving in opposite direction to that of source. The performance of the proposed algorithm has been evaluated with existing algorithms; the proposed algorithm has outperformed the existing algorithm with respect to metric number of link failures, end-to-end delay, average jitter, energy consumption, and network lifetime. The observation shows that in the proposed algorithm, there is 18% reduction in link failures compared to the traditional AntHocNet algorithm. With respect to jitter, compared to the existing algorithm, the proposed algorithm has 18.5% reduction and it has less jitter compared to the traditional AntHocNet. There is 6% reduction in the end-to-end delay in the enhanced algorithm. For future enhancement, the velocity of the vehicle can be considered to improvise the performance of the algorithm.

References

1. Namboodiri V, Agarwal M, Gao L, A study on the feasibility of mobile gateways for vehicular ad-hoc networks. In: Proceedings of the first international workshop on vehicular ad hoc networks
2. Wang Y, Fan L, Vehicular ad hoc networks book chapter guide to wireless ad-hoc networks
3. Namboodiri V, Agarwal M, Gao L (2004) A study on the feasibility of mobile gateways for vehicular ad-hoc networks. In: Proceedings of the first international workshop on vehicular ad hoc networks, pp 66–75
4. Ooi C-C, Fisal N (2004) Implementation of Geocast-enhanced AODV-bis routing protocol in MANET. In: Proceedings of the IEEE Region 10 Conference, vol 2, pp 660–663
5. Briesemeister L, Schäfers L, Hommel G (2000) Disseminating messages among highly mobile hosts based on inter-vehicle communication. In: Proceedings of the IEEE intelligent vehicles symposium, pp 522–527

6. Ko Y-B, Vaidya NH (2000) Location-aided routing (LAR) in mobile ad hoc networks. Wireless Netw 6(4):307–321
7. Aquino SR, Arthur E, Robert E, Luke S (2005) Performance evaluation of routing protocols in vehicular ad hoc networks. Int J Ad hoc Ubiquitous Comput 1(1/2) 80–91
8. Perkins CE, Royer EM Ad-hoc on demand distance vector routing. In: Proceedings of the 2nd IEEE workshop on mobile computing systems and applications, pp 90–100
9. Johnson DB, Maltz DA Dynamic source routing in ad hoc wireless networks. In: Imielinski T, Korth HF (eds) Mobile computing. Springer
10. Caro D, Gianni, Frederick D, Luca MG (2005) AntHocNet: an adaptive nature-inspired algorithm for routing in mobile ad hoc networks. European Trans Telecommun 16 (5):443–455

Design and Analysis of Optimized Hybrid Active Power Filter for Electric Arc Furnace Load

Rajkumar Jhapte[1(✉)], R. N. Patel[1], and D. D. Neema[2]

[1] Department of EEE, Shri Shankaracharya Technical Campus, SSGI, Bhilai, Chhattisgarh, India
{jhapte02, ramnpatel}@gmail.com
[2] Department of EEE, Yugantar Institute of Technology and Management, Rajnandgaon, Chhattisgarh, India
neemadd@gmail.com

1 Introduction

The power quality aims to develop and maintain the standard voltage and standard current with proper frequency for the normal operation of the customer equipment. In last few decades, usage of computers, adjustable speed drives, and process control has increased exponentially which led to an increase in power quality problems. This research paper states a confirmation for the current and voltage harmonic minimization induced by nonlinear loads particularly for Electric Arc Furnace (EAF) load. A random variation in nonlinear loads, generally large and continuous, refers mainly to electric metal-melting arc furnaces connected directly to the transmission network. It consists of the carbon electrodes in contact with iron having dissimilar impedances between the positive and negative flows of current. An electric arc furnace produces a quantity of interfering emissions, such as flicker voltage and stress wear caused by arc phenomena. Voltage cutting complicates the construction of an accurate filter system in the 0–200 kHz frequency range [1].

Here, an analysis of current and voltage harmonic issues arising from the operation of furnace load and the applicability of the proposed systems to mitigate those problems are being carried out.

The objective of this paper is that the Hybrid Active Power Filter (HAPF) is used to compensate harmonic content in voltages and currents, besides Selective Reactive Power Compensation (SVC). The proposed HAPF has the objective of up gradation of different types of APF compensation performance [2], reduction of cost and complexity of compensation systems, etc. From previous investigations, it seems that the electric arc furnace issues are compensated with the help of SVC and series inductance [3–5]. In response to these hybrid/composite filter factors, one of the best options is to be considered by reducing cost, simple design, high reliability, and better control compared to other options for power quality refinement [5–7]. The impact of harmonics and voltage flicker can occur in the electrical power distribution system in the industry and in another power supply unit of the substation [8]. Minimization of the impact of EAFs

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_28

can improve electric efficiency and reduce power fluctuations in the system. The importance of mitigation devices, such as passive filters, active line filter, and hybrid filtering, has been proposed [9]. It has been observed that Genetic Algorithm (GA) tuned controllers give high Total Harmonic Distortion (THD) value and also, Particle Swarm Optimization (PSO) tuned controllers take more time for specific iterations [10]; to overcome these problems, the proposed algorithm is implemented. This paper presents the performance optimization of the conventional PI controller and optimized PI controller for three phase HAPF with Harmony Search Algorithm (HSA). Harmony search method is inspired by harmony improvisation; it has two distinct operators: one is Harmony Memory Considering Rate (HMCR) and another one is Pitch Adjusting Rate (PAR) [11].

2 System Modeling

2.1 Modeling of Electric Arc Furnace

An EAF not only uses power strongly but also uses reactive one. The typical process in the furnace is irregular in nature, in which one or more electrodes with the electric arc are present between the furnace and scrap. In order to know the proper characteristics of an arc furnace, it is needed to design an appropriate three-phase electric arc furnace system for voltage and current harmonic analysis. The voltage drop due to reactive power current acts by reacting to spurious flickering and into the blinking light of fluorescent lamps supplied from the contaminated feeder [12–14]. The proposed arc furnace model is designed in a random time-varying signal in the form of band-limited white noise. It is used to obtain the arc furnace voltage fluctuations in terms of voltage harmonics [9].

2.2 Vigorous Nature of Electric Arcs

The unrestrained, multivalued characteristics of an arc are determined by using an adopted dynamic furnace model in the form of the equations derived in [13]. The arc parameters on different phases are found in [13, 14]. For an arc, power balance equation is as shown in (1).

$$p_1 + p_2 = p_3 \tag{1}$$

where p_1 depicts the power transferred to the external environment in the form of heat; p_2 shows the power, which enhances the inherent energy of the electric arc, and which in turn affects the radius; and p_3 is the gross power produced in the arc and transformed into heat. An assumption is considered that the cooling effect is a function of the electric arc radius (r). Thus, we get (2).

$$p_1 = k_1 r^n \tag{2}$$

where $n = 1$ and k_1 represents an arbitrary constant.

The dependence of arc on temperature may be ignored for low voltage circuits. Thus, the arc radius r only describes the mathematical state of arc.

For $n = 2$, p_2 is analogous to the differentiation of energy in the arc, given by (3). Hence,

$$p_2 = k_2 r \frac{dr}{dt} \tag{3}$$

Again, from (1), we get the sum as (4). Finally,

$$p_3 = vi = \frac{k_3}{r^{m+2}} \tag{4}$$

where k_2 and k_3 are also arbitrary constants.

In (4), the resistance of the arc is considered to be varying inversely with r^m, where $m = 0 \ldots 2$. Now, the arc differential equation is given by (5).

$$k_1 r^n = k_2 r \frac{dr}{dt} = \frac{k_3}{r^{m+2}} \tag{5}$$

Here "r" the arc radius is taken as a variable function instead of arc conductance or resistance. Hence, the voltage of arc is as shown in (6).

$$v = \frac{i}{g} \tag{6}$$

where i is the arc current and g is conductance of arc described by (7).

$$g = \frac{r^{m+2}}{k_3} \tag{7}$$

2.3 Design of Hybrid Filter

The performance of the hybrid active filter, designed by HSA, turns out to be excellent at eliminating the line harmonics as compared with Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). Figure 1 displays the simplified configuration of the hybrid active power filter. The proposed technique is based on the theory of introduction of harmonic current into the alternating current system of the same dimensions of load current harmonics to meet the IEEE 519-1992 limits [15]. The system configuration of hybrid filter comprises a shunt active filter with a shunt passive filter as shown in Fig. 2.

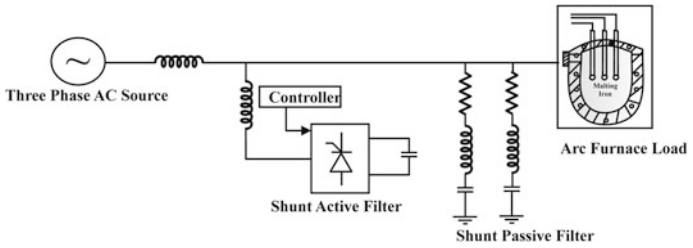


Fig. 1. Test diagram of HAPF with furnace load

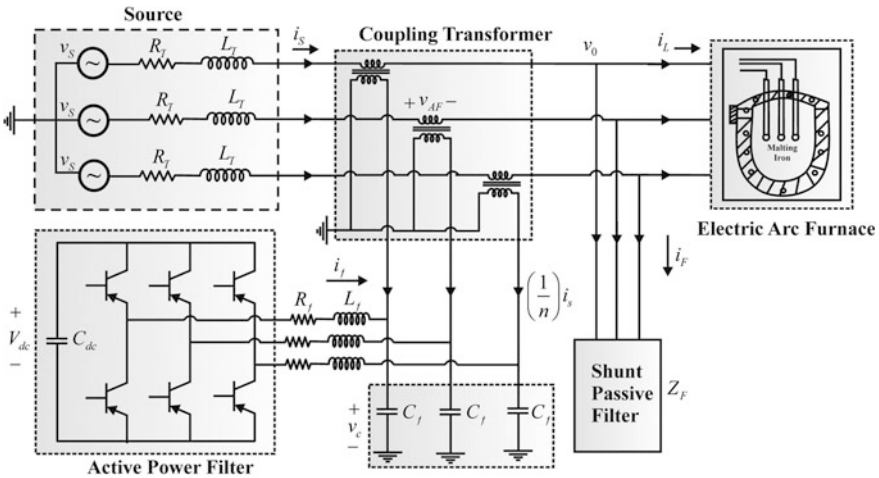


Fig. 2. Configuration of HAPF

2.4 Controller Design

The control input receives an error wave with respect to reference voltage and the RMS magnitude of measured supply voltage. The PI controller ensures that the incorrect signal, generated from the desired angle to zero, i.e., the error signal which is the difference between load RMS voltage and reference voltage, is minimum [16]. The values of K_p and K_i are selected iteratively as discussed in the literature [17, 18] using any of the methods available for tuning it. Here, K_p and K_i are selected arbitrarily and iterated intuitively. The THD value minimization is considered as the objective, and the issue becomes nonlinear due to the variation of THD with respect to various K_p and K_i values.

2.5 Objective Function

The mean value of THD data is achieved from time-domain simulation, and the values are passed to the harmony search solver as objective function. The K_p and K_i values are

randomly populated, and the minimization of objective function is made. The PI controller equation at DC link voltage control part in hybrid active power filter is defined as shown by (8).

$$I_{af} = K_p \times (V_{dcref} - V_{dcmesured}) + K_p \times \int_0^t (V_{dcref} - V_{dcmesured})dt \tag{8}$$

where

I_{af} = compensation current, $V_{dcmesured}$ = DC measured voltage at DC link, V_{dcref} = DC reference voltage, K_p = proportional constant, and K_i = integral constant. The objective function is to minimize the sum of mean values of THD as given by (9).

$$\text{Objective Function} = \text{Minimized} \sum_{i=0}^{n} \text{Mean(THD)} \tag{9}$$

where i represents the number of data samples for mean values of THD.

Subjected to

$$P_1 \leq K_p \leq P_2 \tag{10}$$

$$I_1 \leq K_i \leq \mathrm{I}_2 \tag{11}$$

Here,

P_1= lower limit of K_p, P_2= upper limit of K_p, I_1= lower limit of K_i, and I_1= upper limit of K_i.

3 Optimization Technique

3.1 Genetic Algorithm

Genetic algorithm was first introduced by John Holland. This is the optimization technique which uses the natural selection procedure for calculating its parameters [10].

3.2 Particle Swarm Optimization

Particle Swarm Optimization (PSO) is an optimization technique designed by Dr. Eberhart and Dr. Kennedy in [10]. It is an optimization technique based on the social behavior of fish schooling and bird flocking.

3.3 Harmony Search Algorithm

The HSA is in reality inspired by the functioning principles of the harmony characterization. Harmony search is a music-based metaheuristic optimization algorithm [11].

4 Results and Discussion

The implementation of the proposed system is analyzed for its application in the electric arc furnace load for shunt hybrid active filters. In the current study, the fifth matched passive filter branches are taken into consideration.

4.1 Simulation Result of Shunt Hybrid Active Filter

The harmonic content of source voltage and current using the FFT (Fast Fourier Transform) waveform analysis without any filter at the Point of Common Coupling (PCC) and then with tuned hybrid active power filters with GA, PSO, and HSA optimization techniques has been analyzed.

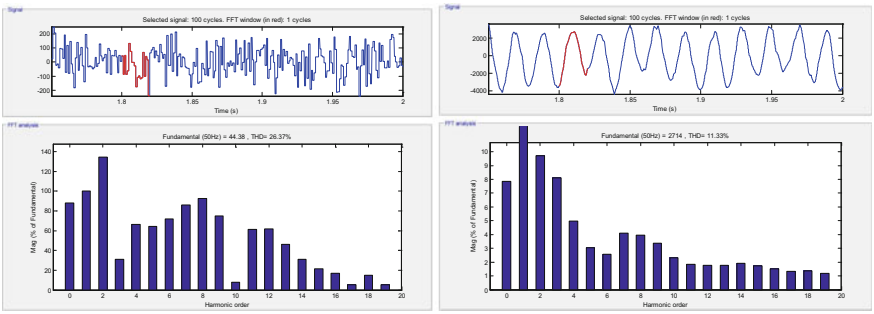


Fig. 3. Voltage and current FFT waveforms at PCC with arc furnace load without any filter

Figure 3 shows the simulation results of system without shunt hybrid active power filter. Here, the THD in source voltage and current is 26.37% and 11.33%, respectively.

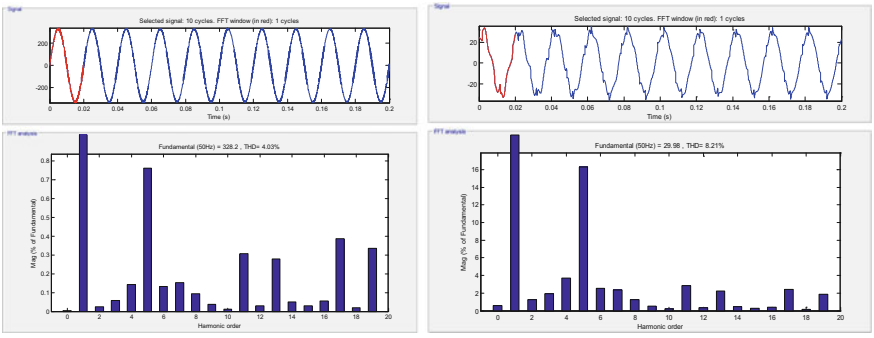


Fig. 4. Voltage and current FFT waveforms at PCC for HAPF with optimized PI using GA

Figure 4 shows the simulation result of system with shunt hybrid active power filter optimized by genetic algorithm. Here, the THD in source voltage and current is 4.03% and 8.21%, respectively.

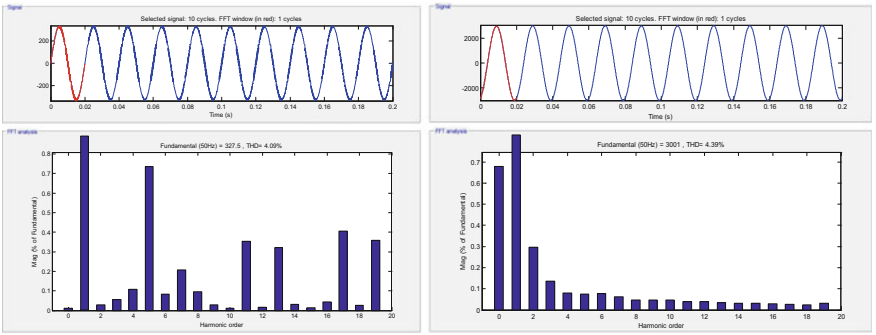


Fig. 5. Voltage and current FFT waveforms at PCC for HAPF with optimized PI using PSO

Figure 5 shows the simulation result of system with shunt hybrid active power filter optimized by particle swarm optimization. Here, the THD in source voltage and current is 4.09% and 4.39%, respectively.

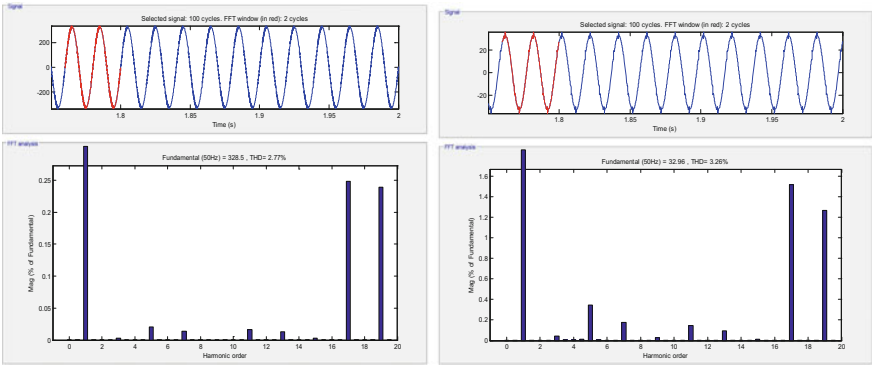


Fig. 6. Voltage and current FFT waveforms at PCC for HAPF with optimized PI using HSA

Finally, Fig. 6 shows the simulation result of system with shunt hybrid active power filter optimized by harmony search algorithm. The THD in source voltage and current is 2.77% and 3.26%, respectively.

5 Conclusion

In this paper, the issue of harmonic reduction is comparatively performed and analyzed. PI tuning using GA, PSO, and HSA optimization method is carried out and compared with the THD values for a shunt hybrid active power filter. It has been observed that THD in the magnitudes of current and voltage has been reduced significantly by using the HSA optimization as compared to GA and PSO optimized PI controllers as shown in Table 1. Hence, the proposed analysis verifies the applicability of the optimized hybrid power filtering method for electric arc furnace load variations.

Table 1. Comparison of THD with different conditions

Type of controller	K_p	K_i	Comparison analysis of THD	
			Voltage THD (%)	Current THD (%)
Nonlinear load connected without filter	–	–	26.37	11.33
HAPF with optimized PI using GA	6.854	3.1771	4.03	8.21
HAPF with optimized PI using PSO	4.231	1.695	4.09	4.39
HAPF with optimized PI using HSA	7.262	2.6833	2.77	3.26

References

1. Esfahani MT, Vahidi B (2012) A new stochastic model of electric arc furnace based on hidden Markov model: a study of its effects on the power system. IEEE Trans Power Deliv 27(4):1893–1901
2. Akagi H (2005) Active harmonic filters. Proc IEEE 93(12):2128–2141
3. Wu C, Liu YW (2014) Reactive power compensation for unbalanced fluctuating loads by using two-dimensional space vector and a static var compensator. Appl Mech Mater 533:397–400 (2014)
4. Surapong C, Yu CY, Thukaram D, Nipon T, Damrong K (2000) Minimization of the effects of harmonics and voltage dip caused by electric arc furnace. In: IEEE power engineering society winter meeting, vol 4, pp 2568–2576
5. Montanari GC, Loggini M, Pitti L, Tironi E, Zaninelli D (1993) The effects of series inductors for eduction in electric power systems supplying arc furnaces. In: IEEE industry applications society annual meeting, vol 2, pp 1496–1503
6. Kolagar AD, Shoulaie A (2011) Reduction of undesired harmonic components in a steel industrial plant with DC electric arc furnaces. In: Proceeding of 2011 2nd power electronics, drive system technologies conference, Tehran, Iran, pp 522–528
7. Peng F (1990) A new approach to harmonic compensation in power systems—the combined system of shunt passive and series active filters. PhD thesis, Nagaoka University of Technology, Nagaoka, Japan, Dec 1990
8. Acha E, Semlyen A, Rajakovic N (1990) A harmonic domain computational package for nonlinear problems and its application to electric arcs. IEEE Trans events Relat Power Qual Power Deliv 5:1390–1397
9. Ozgun O, Abur A (2002) Flicker study using a novel arc furnace model. IEEE Trans PWRD 17(4):1158–1163

10. Gowtham N, Shankar S (2016) PI tuning of shunt active filter using GA and PSO algorithm. In: 2nd international conference on advances in electrical, electronics, information, communication and bio-informatics (AEEICB). IEEE
11. Geem ZW, Kim JH, Loganathan GV (2001) A new heuristic optimization algorithm: harmony search. Simulation 76(2):60–68
12. Beites LF, Mayordomo JG, Hernandez A, Asensi R (2001) Harmonics, interharmonics and unbalances of arc furnaces: a new frequency domain approach. IEEE Trans Power Deliv 16(4)
13. Hamadi A, Rahmani S, Al-Haddad K (2010) A hybrid passive filter configuration for VAR control and harmonic compensation. IEEE Trans Ind Electron 57(7):2419–2434
14. Collantes Bellido R, Gomez T (1997) Identification and modeling of a three phase arc furnace for voltage disturbance systems. IEEE Trans PWRD 12:1812–1817
15. IEEE Std 519-1992 (1993) IEEE recommended practices and requirements for harmonic control in electrical power systems. IEEE Inc
16. Verma V, Shanker S, Singh B (2009) Design and implementation of current controlled parallel hybrid power filter. IEEE Trans Ind Appl 45(5):1910–1917
17. Dixon JW, Venegas G, Luis AM (1997) A series active power filter based on a sinusoidal current-controlled voltage-source inverter. IEEE Trans Ind Electron 44(5):612–620
18. Singh M, Jhapte R, Patel RN (2016) Performance comparison of optimized controller tuning techniques for voltage stability. In: IEEE first international conference on control, measurement and instrumentation (CMI), Kolkata

To Detect the Influencers in a Dynamic Co-authorship Network Using Heat-Diffusion Model

Ritam Sarkar[1], Debaditya Barman[2], and Nirmalya Chowdhury[1(✉)]

[1] Department of Computer Science and Engineering, Jadavpur University, Kolkata, India
nirmalya_chowdhury@yahoo.com
[2] Department of Computer and System Sciences, Visva-Bharati, Santiniketan, India

1 Introduction

The field of Social Network Analysis (SNA) emerged from social science, network analysis, and graph theory. With the help of SNA, we can analyze various complex social structures and the relations or interactions among individuals or groups using the concepts of network and graph theory in order to uncover the explanations for various social behaviors. SNA has been extensively used in diversified areas like social influence analysis in social network, social network advertisement [1], potential friend and communities in social networks [2], improvement of communication flow in business organization [3], information diffusion model of social networks [4], identification of criminal or terrorist groups [5] etc. Growing popularity of various social networking sites has provided the opportunity to study how a person or a community can influence other person or community and examine the results. A large number of recent works have focused on analyzing different research aspects of influence like influencer detection [6], influence diffusion [7] in social network. Sara and Mckeown [8] have detected influencers by analyzing the mode of interaction in various online discussion forums. Trusov et al. [9] have determined a user as influencer when their activity has some significant effect on other user's activity on internet social network. Fang et al. [10] have developed a topic-sensitive influencer detection model where several sets of influencers have been detected based on some topics. Social networks are very dynamic in nature, since their topologies evolve in accordance with time. Most of the aforesaid research works are not suitable for a dynamic network. Recently, researchers have proposed various methods [11, 12] to find influencers in a dynamic network. In this work, in order to address the problem of detecting influencers in a dynamic social network, we have proposed a method which is based on heat-diffusion model. Since a heat-diffusion model is time dependent, our model can adapt itself to the evolving nature of the social network. We have applied our proposed method on co-authorship network which has gained attention from scholars because of certain factors like its ability to represent the evolution of research communities with time, and it provides the opportunity to study the role of social interactions in knowledge generation [13, 14].

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_29

We have organized rest of the paper as follows. Section 2 presents statement of the problem. Description of the proposed method is presented in Sect. 3. Experimental result can be found in Sect. 4. Finally, concluding remarks and scope for the future work are given in Sect. 5.

2 The Problem Statement

Consider a dynamic graph $G = \{G_i | 1 \leq i \leq p\}$ where G_i is the graph at timestamp i which ranges from 1 to p, where p indicates the last timestamp. Moreover, $G_i = \{V_i, E_i\}$ where V_i is the set of nodes and E_i is the set of undirected edges at ith timestamp. Each node represents an author. The co-authorship relation between two authors has been represented by assigning an undirected edge between the corresponding nodes. Here graph G_1 is the initial topology and it contains the node set V_1, which includes the initial collection of authors and co-authors. Now at each successive timestamps, new nodes and edges will be created if some new authors collaborate with the existing authors at V_1. Note that the connectivity of the nodes would change from one timestamp to the next based on the current collaborations between the authors and co-authors. For this reason, we have termed G as an evolving dynamic graph and its formal definition has been stated below.

Definition (Evolved Dynamic Graph)

An evolved dynamic graph $G = \{G_i | 1 \leq i \leq p\}$ where G_i is the graph G at timestamp i. $G_i = \{V_i, E_i\}$ where $V_i = \{ V_{i-1} \cup V_i' | 1 < i \leq p\}$ which implies at timestamp 1, the node set is V_1 and for i_{th} timestamp, $(1 < i \leq p)$, V_i is the union of node set at timestamp $i - 1$ and V_i', i.e., the newly introduced node set at timestamp i according to the co-authorship relations with the node set V_1 and E_i is the edge set based on co-authorship relations among the node set V_i at timestamp i.

It seems that, there is a correlation between an author's extent of influence and his/her number of collaborators or co-authors, since a collaborator always wants to collaborate with an author possessing a high level of expertise on a specific domain. Thus, we can safely assume that a higher number of collaborations of an author implies a higher level of influence. Given an evolving dynamic co-authorship graph $G = \{G_i | 1 \leq i \leq p\}$, our objective is to identify influential authors from V_1 by utilizing the dynamics of the authors of V_1 in the graph G at consecutive timestamps.

3 The Proposed Method

Heat diffusion is the process of propagation of heat from an object having high temperature to an object having low temperature. Heat diffusion has been applied on various research fields. Belkin et al. [15] have used heat kernel for nonlinear dimensionality reduction. Sun et al. [16] have proposed Heat Kernel Signature (HKS) in order to detect the repeated structure in a single and a set of shapes. Ma et al. [17]

have applied heat diffusion for selecting top k candidates for marketing in a static graph. Here in this work, we have applied the heat-diffusion model based on [17] in order to address the problem of finding influential authors in the aforesaid evolving dynamic co-authorship network by considering the heat distribution of the authors in each consecutive timestamps.

In order to define heat diffusion in a dynamic evolving graph, we have first explained heat diffusion in initial static graph, i.e., G_1.

3.1 Heat Diffusion in Initial Static Graph

At any instance of time t, the heat consumed by a node v_{j1} ($v_{j1} \in V_1$) is denoted as $f_{j1}(t)$ and the node v_{j1} receives heat from another node $v_{k1}(v_{k1} \in V_1)$, if there exists an edge between v_{j1} and v_{k1}. Therefore, the total heat consumed by v_{j1} between the time interval (t) and $(t+\Delta t)$ from all other nodes in the network can be expressed as

$$\frac{f_{j1}(t+\Delta t) - f_{j1}(t)}{\Delta t} = \alpha \sum_{k1:\left(v_{k1},v_{j1}\right)\in E_1} \left(f_{j1}(t) - f_{k1}(t)\right) \tag{1}$$

In Eq. 1, E_1 is the edge set at timestamp 1, and α is the heat-diffusion coefficient which indicates the rate at which the heat will diffuse.

Equation 1 can be rewritten to find the amount of heat consumed by node v_{j1} at time instant $(t+\Delta t)$ in matrix form as below:

$$\begin{aligned} &\frac{f(t+\Delta t)-f(t)}{\Delta t} = \alpha H f(t) \\ &or, f(t+\Delta t) = (f(t) + \alpha H \Delta t f(t)) \end{aligned} \tag{2}$$

In Eq. 2, H is the heat matrix which has been derived in the following manner:

$$h^{\wedge}_{j1k1} = \begin{cases} \frac{M_{j1k1}}{\sum_{l1:(v_{j1},v_{l1})\in E_1} M_{j1l1}}; & (v_{j1}, v_{k1}) \in E_1 \\ 0; & otherwise \end{cases} \tag{3}$$

In Eq. 3, M_{j1k1} is the number of manuscripts that the authors v_{k1} and v_{j1} have written together and E_1 is the set of edges at timestamp 1.

$$H_{j1k1} = \begin{cases} h^{\wedge}_{k1j1}; & k1 \neq j1 \\ -1; & otherwise \end{cases} \tag{4}$$

Finally, from Eq. 4, we can construct the heat matrix at timestamp 1. The intuition is to build the heat matrix in such a way that the amount of heat diffusion depends on the number of collaborations among authors. For instance, author v_{k1} collaborates with two authors (v_{l1}, v_{y1}) and has written three manuscripts with v_{l1} and five manuscripts with v_{y1} at timestamp 1. In this scenario, if v_{k1} diffuses h_0 amount of heat then v_{l1} will receive $(h_0 * \frac{3}{8})$ amount of heat and v_{y1} will receive $(h_0 * \frac{5}{8})$ amount of heat according to our Heat matrix in Eq. 4.

It may be noted that, in this work, all the authors have been considered as the heat source irrespective of the number of co-authors with whom they have worked.

3.2 Heat Diffusion in Dynamic Evolving Graph

At the end of timestamp 1, every node in the set V_1 has consumed some amount of heat say $h_{m1}(0 \leq h_{m1} \leq h_s)$, where h_s is the amount of heat which has been given to source node as input heat. At timestamp 2, the topology has been changed into G_2 from G_1 due to the introduction of new co-authors (i.e., $\acute{V}_2$) and the changes in node connectivity according to the changes in co-authorship relations at timestamp 2. Moreover at timestamp 2, the old nodes (i.e., the nodes belong to V_1) hold some amount of heat that they have consumed at previous timestamp, and the new nodes (i.e., the nodes belong to $\acute{V}_2$) have zero amount of heat. So at this timestamp, the old nodes act as the heat source and the amount of heat of each source node will be decayed according to Eq. 5 before applying them into G_2 as input heat. Now, we can apply Eq. 2 where heat matrix has also been updated according to the changes in topology at timestamp 2. This process will repeat till timestamp p when the graph evolved to G_p.

If an author v_{ki-1} ($v_{ki-1} \in V_{i-1}$) consumes h_{ki-1} amount of heat at timestamp $i-1$, then according to Eq. 5, $\beta(h_{ki-1})$ is the decayed heat and c_{ki-1} is the total number of co-authors of the author v_{ki-1} at timestamp $i-1$ and w is a constant greater than zero.

We have divided each timestamp into several equal time intervals Δt $(0 < \Delta t < 1)$. Therefore, we have repeated the process $\left\lceil \frac{1}{\Delta t} \right\rceil$ times at each timestamp, in order to reach the heat diffusion into a steady state. At the end, we have calculated influence score for each author based on the number of authors those have been influenced at each timestamp. Influence score has been described in the following definition:

$$\beta(h_{ki-1}) = \begin{cases} h_{ki-1} - \frac{h_{ki-1}}{c_{ki-1}} \; ; \; h_{ki-1} \geq 0 \; \textit{and the number of manuscripts of the author is less than average number of manuscripts among his / her co} - \textit{authors at time stamp } i-1 \\ h_{ki-1} - \frac{h_{ki-1}}{w} \; ; \; h_{ki-1} \geq 0 \; \textit{and the author has no co} - \textit{author at timestamp } i-1 \\ h_{ki-1} \quad ; \quad \textit{otherwise} \end{cases} \tag{5}$$

Definition (Influence Score)

Influence score of a node v_{k1} ($v_{k1} \in V_1$) is used to measure the influence ability of that node. So, the influence score of an author v_{k1}, i.e., $IS(v_{k1})$ has been calculated using the following equation:

$$IS(v_{k1}) = \sum_{i=1}^{p} \left(\frac{IN(v_{k1}(i))}{n_i} * log_{10}^{(10+i)} \right) \tag{6}$$

In Eq. 6, $IN(v_{k1}(i))$ is the number of authors which are influenced by author v_{k1} at timestamp i and n_i is the total number of authors at timestamp i.

Algorithm 1: *dynamic evolving graph Heat-Diffusion* $(\alpha, p, N, H, H_S, H_0, \Delta t)$

Input: Heat Co-efficient $\alpha = 1$, Number of time stamp p,
$N = \{n_1, n_2, n_3, \ldots, n_i, \ldots, n_p\}$; where n_i is the total number of nodes
At timestamp$i$ and $n_i = |V_i|$,
$H = \{h_1, h_2, h_3, \ldots, h_i, \ldots, h_p\}$; where h_i is the heat matrix at timestamp i,
Heat Source H_s,
Initial Heat H_0,
Δt is the time interval.
Output: Author influence score IS.

1. $IN = \{in_1, in_2, in_3, \ldots, in_i, \ldots, in_p\}$ where in_i is the total number of
Influenced nodes at timestamp i and initially it is empty.
2. $Th = \{th_1, th_2, th_3, \ldots, th_i, \ldots, th_p\}$ where th_i is the threshold
Of timestamp i and initially it is empty.
3. $IH = \{ih_1, ih_2, ih_3, \ldots, ih_i, \ldots, ih_{n1}\}$ where ih_i is the input heat of node i
And initially it is empty.
4. t_{temp} is a temporary time variable.
5. $I_{temp} = 0.0$ is a temporary variable for calculating influenced score IS.
6. for $i = 1$ to p
7. $\quad th_i = (H_0/n_i)$
8. for $i = 1$ to n_1
9. $\quad$ if $i == H_s$
10. $\quad\quad {}_ih_i = H_0$
11. $\quad$ else:
12. $\quad\quad {}_ih_i = 0$
13. I_1, I_2 are the temporary matrices for calculating the input heat.
14. for $i = 1$ to p
15. $\quad$ if $i == 1$
16. $\quad\quad I_1 = IH$
17. $\quad$ else:
18. $\quad\quad$ for $j = 1$ to n_i
19. $\quad\quad\quad$ if $j \leq n_{i-1}$
20. $\quad\quad\quad\quad I_2[j] = I_1[j]$
21. $\quad\quad\quad$ else:
22. $\quad\quad\quad\quad I_2[j] = 0$
23. $\quad\quad I_1 = \beta(I_2)$
24. $\quad t_{temp} = 0.0$
25. $\quad$ while $t_{temp} < 1.0$
26. $\quad\quad t_{temp} = t_{temp} + \Delta t$
27. $\quad\quad I_1 = (I_1 + \alpha * h_i * \Delta t * I_1)$
28. $\quad in_i =$ The total number of nodes having heat greater than th_i.
29. for $i = 1$ to p
30. $\quad I_{temp} = I_{temp} + (\frac{in_i}{n_i} * log_{10}^{(10+i)})$
31. $IS = I_{temp}$
32. Return IS.

Algorithm for heat diffusion in a dynamic evolving graph has been described in Algorithm 1. It generates the influence score for the author who has been taken as input heat source H_s. The following section narrates the experimental results.

4 The Experimental Results

In this section, we have demonstrated our experimental results and their evaluation to measure the effectiveness of our proposed method.

4.1 Dataset

We have applied our proposed algorithm on a co-authorship network called *DBLP-citation-network V6* which has been taken from www.arnetminer.org. This dataset contains 2,084,055 papers up to 29/09/2013. We have built three different dynamic co-authorship networks of varying size, based on the co-authorship relations from 2010 to 2012. We have assumed that if an author collaborates with another author (i.e., a co-authorship relation), then they have common interests in a research domain. This assumption holds transitive property up to a certain predefined depth d. Initially, V_1 contains the set of authors of a randomly selected paper from the year 2010. Then, we have added some new authors to V_1 based on the co-author relationship with previously selected authors. This process is continued up to a depth d for their paper published in 2010. V_2 consists of the authors from V_1 and some new authors collaborated with the authors of V_1 in the year 2011. V_3 has been constructed similarly for the year 2012 from V_1.

4.2 Experimental Evaluation

We have obtained the influence score of an author, considering the said author as the input heat source H_s to our proposed algorithm, one author at a time. The influence score depends on the number of co-authors who have been influenced by the source author at each consecutive timestamps. If h_0 is the amount of heat given to H_s, then H_s will influence a node v_{ki} in timestamp i if the amount of heat consumed by v_{ki} is greater than a threshold $th, th = \frac{h_0}{n_i}$, where n_i is the total number of nodes at timestamp i.

To measure the efficacy of our proposed method, we have divided the author's influence score into three classes (low influencer, mid influencer, and high influencer) as well as two classes (low influencer and high influencer). For this purpose, we have considered the entire range as $(\mu - 3 * \sigma)$ to $(\mu + 3 * \sigma)$ where μ is the mean, and σ is the standard deviation. We have taken the ranges for the three classes scenarios as: low influencer $((\mu - 3 * \sigma)$ to $(\mu + \sigma))$, mid influencer $((\mu + \sigma)$ to $(\mu + 1.8 * \sigma))$, and high influencer $((\mu + 1.8 * \sigma)$ to $(\mu + 3 * \sigma))$ and for the two classes scenario as: low influencer $((\mu - 3 * \sigma)$ to $(\mu + 1.5 * \sigma))$, and high influencer $((\mu + 1.5 * \sigma)$ to $(\mu + 3 * \sigma))$ for both author's influence score and author h-index score.

We have taken the h-index as the ground truth and this has been derived from the total citation count and the number of publications of each author. It's ranges have been

described previously. The h-index values are not normally distributed. For this reason, we have transformed the data by the following manner [18] as shown in Eq. 7.

In Eq. 7, $\gamma(h{-}index)$ is the transformed value of h-index and λ is a vector which defines the transformation.

$$\gamma(h{-}index) = \begin{cases} \frac{(h{-}index^{\lambda})}{\lambda}; & for\ \lambda > 0 \\ \log(h{-}index); & for\ \lambda = 0 \end{cases} \tag{7}$$

As shown in Table 1, for dataset 1, the precision value of "*low influencer*" class is greater than the other two classes which implies that more number of authors in "*low influencer*" class have been correctly classified, out of the total number of authors who have been classified as "*low influencer*" class based on author's influence score. Similarly, a higher recall value for "*low influencer*" class indicates that a higher number of authors have been correctly classified, out of total number of relevant authors assumed as "*low influencer*" based on h-index value. We have observed that the number of authors that belong to "*low influencer*" class is significantly higher than that of the other two classes. It seems that, if the number of classes is high, more data from the rest two classes (i.e., "*mid influencer*" and "*high influencer*") are required to obtain the desired level of accuracy. It may be the reason that our proposed algorithm has given an overall accuracy in the range of 0.67–0.69 (as shown in Table 1) for three class scenarios and that for two class scenarios, it has been raised in the range of 0.90–0.93 (as shown in Table 2) which is quite satisfactory. The confusion matrix for Dataset 1 has been shown in Fig. 1. Due to the limitation of space, we have not shown the confusion matrices for other datasets.

Table 1. Experimental results for three class scenarios

Dataset no.	Class name	Precision	Recall	Overall accuracy
Dataset 1	Low influencer	0.73	0.79	0.69
	Mid influencer	0.64	0.55	
	High influencer	0.55	0.67	
Dataset 2	Low influencer	0.62	0.76	0.67
	Mid influencer	0.75	0.60	
	High influencer	0.58	0.64	
Dataset 3	Low influencer	0.78	0.74	0.68
	Mid influencer	0.53	0.59	
	High influencer	0.56	0.56	

To evaluate the ranking capability our proposed algorithm, we have shown in Fig. 2 that the major portion *h*-index value passes through the Bollinger band of the influence score. In this Bollinger band, lower band is $(ma_n - k * \sigma_n)$ and the upper band is $(ma_n + k * \sigma_n)$, where ma_n is the n period moving average, and σ_n is the n period

Table 2. Experimental results for two class scenarios

Dataset no.	Class name	Precision	Recall	Overall accuracy
Dataset 1	Low influencer	0.96	0.96	0.93
	High influencer	0.58	0.55	
Dataset 2	Low influencer	0.96	0.96	0.93
	High influencer	0.54	0.57	
Dataset 3	Low influencer	0.95	0.94	0.90
	High influencer	0.55	0.58	

	h-index data				
		Low Influencer	Mid Influencer	High Influencer	Classification Overall	Precision
Influencer score results	Low Influencer	106	39	1	146	0.73
	Mid Influencer	27	51	2	80	0.64
	High Influencer	2	3	6	11	0.55
	Truth Overall	135	93	9		
	Recall	0.79	0.55	0.67		

Fig. 1. Confusion matrix for dataset 1

standard deviation of author's influence score which has been sorted according to the h-index value. The values of n and k are taken as 15 and 2, respectively. We have not shown the Bollinger band for other datasets due to the limitation of space. According to Fig. 2, almost all the parts of h-index value pass through the Bollinger index.

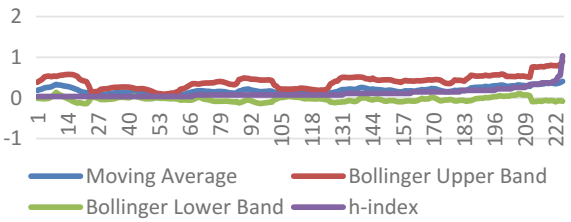


Fig. 2. Bollinger band for dataset 1

5 Conclusion and Scope for Further Work

Here in this work, we have developed a heat diffusion-based method to detect influential persons in social network. Our proposed method can handle a network which evolves dynamically. The topology of this network constantly changes with time by the introduction of new nodes and edges and the deletion of existing edges based on the current topology. Our proposed method has some distinctive features. First, the heat

matrix changes dynamically in accordance with the topology of the network at each consecutive timestamps. Second, the amount of heat consumed by each node in G_{i-1} has been considered as the heat source for the next timestamp (i.e., G_i) except the first timestamp where the input heat is being given to the heat source externally. Third, the method is not dependent on the amount of heat given to heat source initially, since the threshold value will get updated according to the amount of initial input heat and the threshold value also changes dynamically. Finally, if an author v_{ki} does not collaborate with any other author at timestamp i, then instead of wasting the heat, the author v_{ki} will just preserve his or her previously consumed heat and then it will get used in later timestamp when he or she will collaborate with someone. However, our proposed algorithm cannot deal with negative influence, since it is based on *heat-diffusion* model where the amount of heat can never be negative. In addition to that, we need to know the position of two nodes to measure the amount of heat diffuses from one node to another node. Since it is hard to derive exact position of two nodes in a social network, we have assumed that the graph is an underlying manifold of social network [17].

The experimental results show the novelty of our proposed approach. We can further extend our work by predicting the h-index value of an author. We can also try to build a model which can handle a dynamic heterogeneous network. Extracting the most influential set of authors in a co-authorship network can also be considered as a probable future work.

Acknowledgements. This paper is an outcome of the work carried out for the project titled "Development of some efficient techniques for applications in the field of Business Analytics and Business Intelligence" under "Mobile and Innovative Computing" under the UGC UPE Phase II scheme of Jadavpur University.

References

1. Moneypenny NF, Flinn SD. Influence-based social network advertising. U.S. Patent Application No. 12/172,236
2. Leskovec J, Singh A, Kleinberg J (2006) Patterns of influence in a recommendation network. In: Pacific-Asia conference on knowledge discovery and data mining. Springer, Berlin, Heidelberg
3. Carneiro A (2000) How does knowledge management influence innovation and competitiveness? J Knowl Manag 4(2):87–98
4. Kempe D, Kleinberg J, Tardos É (2005) Influential nodes in a diffusion model for social networks. In: International colloquium on automata, languages, and programming. Springer, Berlin, Heidelberg
5. Al-Zaidy R et al. Mining criminal networks from unstructured text documents. Digit Investig 8(3):147–160 (2012)
6. Puigbo J-Y et al (2014) Influencer detection approaches in social networks: a current state-of-the-art. CCIA
7. Li Y et al (2013) Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. In: Proceedings of the sixth ACM international conference on web search and data mining. ACM

8. Rosenthal S, Mckeown K (2017) Detecting influencers in multiple online genres. ACM Trans Internet Technol (TOIT) 17(2):12
9. Trusov M, Bodapati AV, Bucklin RE (2010) Determining influential users in internet social networks. J Mark Res 47(4):643–658
10. Fang Q et al (2004) Topic-sensitive influencer mining in interest-based social media networks via hypergraph learning. IEEE Trans Multimed 16(3):796–812
11. Subbian K, Aggarwal CC, Srivastava J (2016) Querying and tracking influencers in social streams. In: Proceedings of the ninth ACM international conference on web search and data mining. ACM
12. Gomez Rodriguez M, Leskovec J, Schölkopf B (2013) Structure and dynamics of information pathways in online media. In: Proceedings of the sixth ACM international conference on web search and data mining. ACM
13. Uddin S, Hossain L, Rasmussen K (2013) Network effects on scientific collaborations. PLoS ONE 8(2):e57546
14. Wallace ML, Larivière V, Gingras Y (2012) A small world of citations? The influence of collaboration networks on citation practices. PLoS ONE 7(3):e33339
15. Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput 15(6):1373–1396
16. Sun J, Ovsjanikov M, Guibas L (2009) A concise and provably informative multi-scale signature based on heat diffusion. In: Computer graphics forum, vol 28, no 5. Blackwell Publishing Ltd
17. Ma H et al (2008) Mining social networks using heat diffusion processes for marketing candidates selection. In: Proceedings of the 17th ACM conference on information and knowledge management. ACM
18. Box GEP, Cox DR (1964) An analysis of transformations. J R Stat Soc Ser B (Methodol) 211–252

Efficient Word2Vec Vectors for Sentiment Analysis to Improve Commercial Movie Success

Yash Parikh(✉), Abhinivesh Palusa, Shravankumar Kasthuri, Rupa Mehta, and Dipti Rana

Sardar Vallabhbhai National Institute of Technology, SVNIT, Surat 395007, India
{yashparikhnitsurat, pabhinivesh, kasthurishravankumar}@gmail.com, {rgm, dpr}@coed.svnit.ac.in

1 Introduction

Movie industry has turned huge today, with a lot of money put at stake by the producer. Along with this, marketing strategies can be planned and improved dynamically according to the sentiments of the users available through their reviews on prerelease events like trailer, music launch, and other promotions. Better marketing can guarantee the producer at least a good opening whatever the story might be. This is where sentiment analysis [1] becomes useful. Analyzing the sentiments of the reviews has been worked upon since long, and the algorithms for sentiment polarity classification used include tf-idf [2], word2vec [3], and doc2vec [4].

Doc2vec provides pretty high accuracy each time, considering the area of sentiment polarity classification. But there are two limitations to it, namely, high space complexity to store paragraph vectors and high running time. In this paper, both these limitations are overcome by using a modified approach built on top of word2vec algorithm, which improves the classification accuracy considerably as compared to word2vec and gives comparable and sometimes even better results than doc2vec [4].

In this paper, in the first phase, gross is predicted by taking the attributes that the producers have, just after finishing the shooting of the movie, which can help them to plan the marketing strategies initially. In the second phase, sentiment analysis of the prerelease reviews is done at regular intervals using our proposed modified approach on word2vec and is compared with other techniques, which will help producers to plan and change their marketing strategies later on and will guide the distributors as to which movies are worth investing, considering the interest of people.

The rest of the paper is organized as follows. In Sect. 2, previous and related works are discussed, which are employed in the model. In Sect. 3, the implementation of the gross prediction is discussed followed by the implementation of sentiment analysis on the user reviews of movies in Sect. 4. In the same section, details of our new approach to improve the classification accuracy of word2vec vectors are given. The conclusion and possible future work are mentioned in Sect. 5.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_30

2 Related Work

While a lot of work has been done in movie prediction in terms of hit or flop [4, 5], what actually matters to the producer is the profit that he earns on the movie. The proposed model in this paper is created using various machine learning techniques for classification. First, the data obtained from Internet Movie Database (IMDB) [4] are cleaned using techniques of binning, global value replacement and are combined into one file [6]. The techniques then used for classification include decision trees [6], random forest classifier [6], gradient boosting [6], Gaussian Naïve Bayes [7], logistic regression [8], and linear Support Vector Machine (SVM) [7].

The second phase focuses on sentiment analysis of movie reviews. Formerly, the major focus was on the analysis of the reviews after the release of the movie, whereas we focus a very different practical application that helps a producer as well as a distributor to increase their profits [5]. While sentiment analysis is done at word-level, phrase-level, and document-level, our focus is on document-level sentiment analysis as we want to delineate the sentiment of one entire movie review at once. For that, initially we model the words into vectors by utilizing various vectorization techniques like term frequency-inverse document frequency (tf-idf) [2], word2vec [3], and the latest doc2vec [4] utilizing paragraph vectors. In this paper, the expected rating obtained by Potts [9] is integrated on top of word2vec into the final vectors, which is utilized as input to various classifiers used in the first phase along with neural networks [10] and stochastic gradient descent [7]. As the expected rating obtained by Potts is signed, i.e., is negative for negative sentiment words and vice versa, this approach will be beneficial to sentiment polarity classification problems.

3 Proposed Model for Gross Prediction

This section focuses on the preparation of the model before the release of the movie.

3.1 Data Collection and Preprocessing

The data are collected from the open source repositories by imdb.com. All the files are available as X.list files, which are preprocessed and converted in the required form. The data use the following attributes for model preparation (Fig. 1):

- Actors' list
- Actors' popularity
- Director's popularity
- Genre
- Budget
- Release year
- Gross

Fig. 1. "Gross prediction" model workflow

Gross attribute is utilized as a training element for the model. All the above data are converted in the form of a.csv file. Next, preprocessing of the data aggregated is done. The preprocessing workflow is shown in Fig. 2.

In addition to the workflow shown in Fig. 2, genres attribute is also preprocessed by splitting and considering an optimized set of genres, which gives better classification results without increasing dimensionality much. Now, the term profit is subjective for different producers. But, one thing that the distributors as well as producers have their eye on is the ratio of Gross attribute to that of Budget as everyone wishes to recover what they spent and earn profit in its multiples.

Thus, using the above logic, a new attribute "Norm_Gross" is created that would act as our target attribute to be predicted, i.e., the ratio of "Gross" attribute to that of "Budget" attribute. This would as a result be an attribute with continuous distribution. But it can be modified to be distributed in different classes to estimate commercial success at the very beginning.

(1) Considering four classes

- Class 0: Range [0, 0.5)
- Class 1: Range [0.5, 1)
- Class 2: Range [1, 5)
- Class 3: Range [5, ∞)

(2) Considering three classes

- Class 0: Range [0, 1)
- Class 1: Range [1, 5)
- Class 2: Range [5, ∞)

Outliers tend to distort the model by overfitting. If the value of Norm_Gross is greater than 5, this value corresponds to an extraordinary or overwhelming success, which cannot be accounted for previously. Hence, these are outliers and are to be removed before applying modeling to the dataset. Thus, class 3 can be omitted from way number 1 of classifying and class 2 from way number 2 of classification. But just to ensure that the removal of outliers yields better results, the accuracy of a model is tested on it.

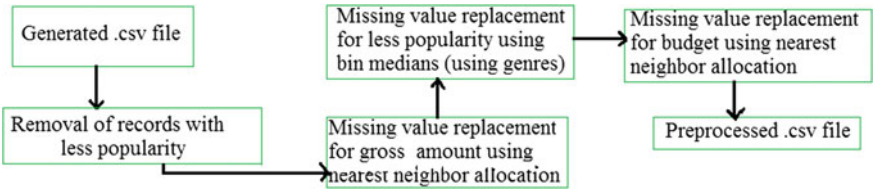


Fig. 2. Preprocessing workflow

3.2 Modeling

Here, we use a subset (around 5,000 instances) which is obtained from the big data split into various X.list files. The first algorithm that was implemented is logistic regression, followed by Gaussian Naïve Bayes' algorithm, decision trees, random forest classifier, gradient boosting, artificial neural networks, and support vector machine. The parameters of the algorithms are kept as the default as obtained from the scikit-learn library available in python. The accuracies obtained are shown in Table 1.

As evident from Table 1, random forest classifier gives the highest accuracy, both for three classes and two classes. Also, it is easy to implement random forest classifier in distributed systems environment which will be useful when data are very large to analyze. Thus, the model obtained by it is considered along with its accuracy.

Table 1 Gross prediction accuracies by various classifiers

Algorithm	Three classes [0–0.5, 0.5–1, 1 +] (%)	Two classes [0–1, 1 +] (%)
Decision tree	60.62	76.37
Random forest	**67.72**	**78.74**
Gradient boosting	66.14	77.95
Gaussian Naïve Bayes	50.39	62.99
Logistic regression	56.69	59.05
Support vector machine (kernel = "linear")	53.54	62.20

4 Proposed Review Model

Now, the sentiment analysis part of our technique is focused that will aid the producers to improve their marketing strategies catering to the user requirements. The workflow diagram is shown in Fig. 5 (Figs. 3, 4).

Fig. 3 Review model preparation

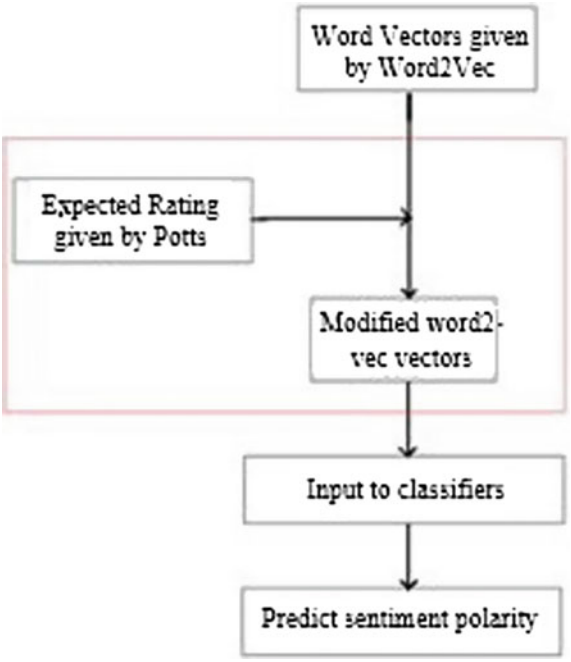


Fig. 4 Modified approach to word2vec

4.1 Data Collection and Preparation

The review corpus is taken from Large Movie Review Dataset, available from Stanford.edu [1]. It contains 25,000 labeled polar reviews for training and 25,000 for testing. Along with this, there is also unlabeled data available for use. This data is already preprocessed removing extra spaces and other delimiters. Emoticons are not considered in the model for now but they may be included in future work.

4.2 Vectorization

Here, the application of various techniques used for vectorization, including tf-idf, word2vec (both standard and modified), and doc2vec, is discussed.

For tf-idf, *25,000* rows are obtained in the co-occurrence matrix and the unique words in the training corpus will act as features. After removing most occurring and least occurring words (features) from the dataset, around *18,000* unique words are left, which will be the dimensionality for each row (document). Next, doc2vec is applied. The paragraph vectors are trained using 100 dimensions of the feature vectors. Different variants of doc2vec are used, i.e., the distributed bag-of-words (PV-DBOW), averaging of the word vectors and concatenation of the word vectors. The vectors are built using all training data, testing data, and unsupervised data as well. Now, these vectors are input to various classifiers, and the accuracies obtained are shown in Table 2. Followed by doc2vec, standard word2vec's two variants [3] are applied.

Proposed approach. Traditional model of tf and tf-idf does not capture the information about the context. Thus, they are in essence incapable of predicting any novice review even if it uses all the words from the trained model. This becomes even difficult when the review is written in a sarcastic tone. Also, word2vec algorithm can predict words pertaining to the particular topic. But for the sentiment classification, words in the same topic may be adhering to different emotions, i.e., both positive and negative. Also word2vec algorithm models vectors of words "good" and "great" nearby each other. But they have different intensities. This can be captured by incorporating the expected rating by Potts [9], in which negative words lie away from the positive ones as they have negative signed expected rating. Also, mod weights, i.e., expected ratings, are also taken into account.

This is implemented for both the variants of word2vec, i.e., continuous bag-of-words and skip-n-gram technique. The technique is delineated out in steps as follows:

- Obtain word vectors using word2vec and store them.
- Multiply each of the word vectors with the corresponding weights as obtained by Potts [9]. This multiplication is done in two ways:
  - First, just multiply the mod weights.
  - Second, multiply weights along with their polarity signs.
- Apply classification algorithms for sentiment classification using the modified vectors.

The accuracies obtained by using this technique shows great results with word2vec, giving accuracies comparable to doc2vec for Stanford dataset. Now, word2vec takes

less amount of space to store the vectors, and the time taken by classifiers is also considerably low. The running times are shown in Table 3 for reference. Also, the accuracies obtained by different classifiers are shown in Table 2, for both the variants of word2vec and even for tf-idf. Results obtained before and after the inclusion of expected rating are included for comparison. The accuracies have shown considerable improvement as shown.

In Table 3, it can be observed that training the vectors of words obtained using *tf* and *tf-idf* takes significantly large time to fit in the classifiers when compared to both *word2vec* and *doc2vec*. This is because of its dimensionality.

Also, doc2vec vector representations need paragraph vectors along with the word vectors whereas word2vec vectors are just word vectors. Thus, the space complexity is definitely reduced by using word2vec vectors in place of doc2vec vectors. Also, for predicting a new review's analysis, doc2vec needs to generate a paragraph vector unique to that review and then analyze its sentiment in contrast to word2vec vectors that just need vector representations of words, and analyzes the review by concatenating them. This even reduces the time required for sentiment analysis. Thus, word2vec is better than doc2vec in terms of space and time complexity. Our approach built on top of word2vec vectors gives two major advantages:

- It outperforms doc2vec in most cases and gives comparable accuracies in others.
- Running time is significantly low as compared to doc2vec, as we do not have to create a new paragraph vector to predict each of the new reviews, only word vectors are required for analysis, which is also depicted in Fig. 5.

For comparison, the same technique is used to classify reviews as positive and negative on Pang and Lee's dataset [11]. This dataset consists of 2000 reviews in all, with 1000 positive reviews and 1000 negative reviews. The ratings were determined from the star-ratings explicitly given by the users on IMDB while mentioning their reviews. For a 5-star rating system, reviews with 3.5 stars and above are considered positive reviews and 2 stars and below are considered negative reviews.

The accuracies obtained for some of the classifiers are shown in Table 4 for comparison. The execution time to specifically convert test documents to vectors is around 43.49 s for doc2vec for the initial phase and 1426.307 s for 10 iterations (performed so that the order of reviews does not affect the paragraph vectors) which is very high when compared to our proposed approach, which takes only 2.59 s for vectorization of Pang and Lee dataset.

Paragraph vectors in doc2vec technique are built on top of word vectors obtained by word2vec technique through incorporating context of the document. The paragraph vectors require a space complexity proportional to $O(n * |V|)$ where n is the number of documents (here review files) and $|V|$ is the size of the input vocabulary; whereas the space complexity for word vectors is proportional to $O(|V|)$. Our approach maintains the space complexity proportional to $O(|V|)$ as it is essentially a word vector representation. The time complexity of our modification applied on top of word2vec technique has time complexity linear to that of the size of the input vocabulary. And the time complexity of word2vec technique varies quadratically with the size of input vocabulary. Thus, the time complexity remains the same as word2vec, remaining less than doc2vec technique whose time complexity even depends on the number of the

Table 2 Accuracies for review analysis by various classifiers on different vectorization techniques (in %)

	tf	tf-idf	w2v (cbow)	w2v-mod (cbow)	w2v-pol (cbow)	w2v (sg)	w2v-mod (sg)	w2v-pol (sg)	d2v	d2v (bow)
RFC	78.05	77.62	74.18	83.08	84.98	78.05	84.75	86.33	73.34	71.81
DT	72.27	70.94	67.09	77.1	78.78	71.16	79.02	80.78	66.22	63.76
GBC	72.43	72.24	72.80	81.97	84.38	76.48	83.7	86.60	72.74	69.90
LR	85.39	88.09	85.29	88.09	88.55	86.43	83.53	87.48	86.24	88.62
SGD	83.73	87.8	80.18	87.98	88.58	85.94	86.91	87.34	79.44	86.46
LSVM	82.80	86.16	85.34	88.1	88.68	87.28	88.74	***89.01***	86.22	88.708

Table 3 Running times comparison of various vectorization techniques for sentiment polarity determination (in seconds)

	tf	tf-idf	W2v (cbow)	w2v- mod (cbow)	w2v-pol (cbow)	w2v(sg)	w2v-mod (sg)	w2v-pol(sg)	d2v	D2v (cbow)
Vectorization time	6.423	9.957	171.92	171.92 + 90	171.92 + 90	891.71	891.71 + 63	891.71 + 63	1837.6	5224.2
RFC	7.857	10.237	3.583	3.921	2.899	3.827	2.813	3.415	4.377	5.896
DT	33.72	49.02	4.780	5.784	6.324	5.724	5.421	6.454	6.606	7.675
GBC	0.979	3.126	2.212	1.599	1.578	1.474	1.425	1.595	2.263	2.368
LR	5.886	2.078	1.627	0.849	1.170	0.834	0.629	0.459	1.101	2.356
SGD	1.665	0.142	0.143	0.118	0.086	0.090	0.136	0.105	0.082	0.190
LSVM	5.160	0.628	14.244	6.087	4.514	1.040	0.682	***0.461***	16.330	13.157

RFC random forest classifier, *DT* decision tree, *GBC* gradient boosting classifier, *LR* logistic regression, *SGD* stochastic gradient descent, *LSVM* linear support vector machines, *w2v* word2vec, *d2v* doc2vec

Table 4 Accuracies comparison of doc2vec and our modification applied to word2vec (in percentage) for LSVM classifier

	Doc2vec	Proposed approach
Stanford dataset	88.708	**89.01**
Pang Lee dataset	86.75	**87.4**

input documents (here review files). Thus, our approach achieves both better time and space complexity as compared to doc2vec and still achieves better results. The F1 scores for precision and accuracy obtained for our modified approach are 0.88 and 0.896, respectively, for LSVM classifier on skip-n-gram variant combined with polar weights, which gives an F1 score of 0.89. The running times comparison for different techniques and two classifiers is depicted in Fig. 5, which shows considerably less running times for our modified approach as compared to both the variants of doc2vec.

Thus, after predicting the reviews polarity, the producers can figure out whether the movie is reaching out to the people in a positive way or not. Accordingly, producer can change his marketing strategy to improve the perception of people towards the movie and thus can expect better returns on the movie. This can be done in multiple phases before the release of the movie, so that at least producer can be assured that the movie will get a proper opening. Along with this, the distributor can also benefit by ordering more or less number of prints of the movie catering to people's requirements.

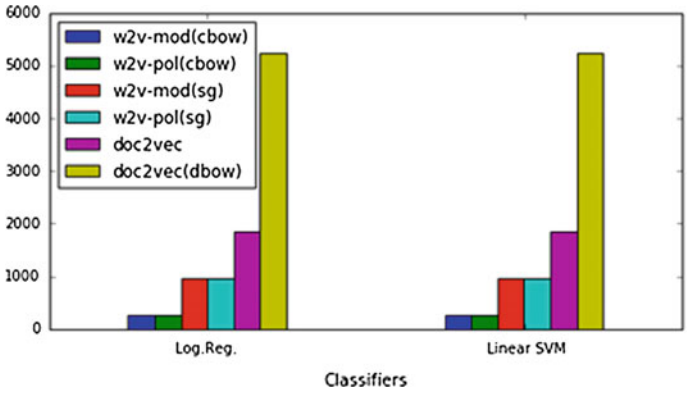


Fig. 5 Running times comparison for our approach and doc2vec for logistic regression and linear SVM classifiers

5 Conclusion and Future Work

The results for predicting the gross for the IMDB data are presented here. The best possible classifier, i.e., random forest classifier, is used for gross prediction. Then sentiment analysis on the user reviews is done and sentiment polarity classification is obtained using various classifiers, along with the accuracies provided for each of them, and their running times. Along with this, our approach is discussed, and accuracies obtained using that approach were comparable to doc2vec approach. Our technique is also tested on Pang and Lee's dataset, and the accuracies are presented. Better accuracy with less time complexity and less space complexity was obtained. It is tentatively concluded that producer can thus change his marketing strategies according to the user reviews and thus increase the profit for the movie and as a result the gross.

The future work includes integrating reviews from all social media possible, like Twitter, YouTube, etc. Also, the model can be extended to distributed systems.

References

1. Maas AL et al (2011) Learning word vectors for sentiment analysis. ACL
2. Rajaraman A, Ullman J (2011) Mining of massive datasets. Cambridge University Press, pp 1–17
3. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. ICLR Workshop
4. Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: Proceedings of the 31st international conference on machine learning, PMLR 32(2):1188–1196
5. Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment classification using machine learning technique. In: Proceedings of EMNLP, pp 79–86
6. Han J, Kamber M, Pei J (2006) Data mining: concepts and techniques (The Morgan Kaufmann Series in Data Management Systems), 2nd edn. Morgan Kaufmann, San Francisco, CA, USA
7. Ng A (2016) Machine learning yearning: draft version 0.5
8. Malouf R (2002) A Comparison of algorithms for maximum entropy parameter estimation. In: Sixth conference on natural language learning (CoNLL), pp 49–55
9. Potts C (2011) On the negativity of negation
10. Sharma V, Rai S, Dev A (2012) Compr Study Artif Neural Netw 2(10)
11. Pang B, Lee L (2004) A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd ACL, pp 271–278

Improving the Utilization of Licensed Spectrum in Cognitive Radio

Soumya Gaddam(✉), Gaurav Verma, and Sudakar Singh Chauhan

Department of ECE, NIT, Kurukshetra, Haryana, India
{gsoumya491, gauravnitkl3}@gmail.com,
sudakar@nitkkr.ac.in

1 Introduction

Increase in the usage of wireless services and applications, limited availability of spectrum, and spectrum scarcity problem increase created demand of spectrum. The reports of federal communications commission (FCC) explained that the allotted spectrum was not utilized efficiently [1, 2]. All situations resulted in the development of cognitive radio technology. The cognitive radio allows the secondary users (SU) to use licensed spectrum when it is not used by the users to which it is actually allocated. The users which are allocated the spectrum have high priority known as primary users (PU)/ licensed users. The users which try to use the spectrum when the spectrum is vacant have low priority and are known as secondary/unlicensed users [3–8]. One of the important functions of CR is finding the presence of PU. For that spectrum, sensing plays a key role [8]. Energy detection scheme technique is used for spectrum sensing because it has fewer complexes and no need of prior information about primary users in this technique. The two important parameters which measure the performance of spectrum sensing are probability of false alarm (P_{fa}) and probability of detection (P_d). P_d is the probability with which PU is correctly detected; it should be as high as possible for sufficient protection of PU. P_{fa} is the probability of falsely detecting primary signal which is actually noise signal; it should be as low as it can be done to improve the efficient usage of licensed spectrum. SU can access licensed spectrum using two methods: opportunistic spectrum access (OSA) [4] and spectrum sharing schemes [6]. In OSA scheme, the SU are allowed to use the licensed spectrum when it is idle, whereas in spectrum sharing scheme, SU allocation is independent of the status whether it is idle or busy [6]. In this case, SU have to adjust the transmission power values not to disturb the PU communication [5].

In the energy detection technique, the average value of energy of the received signal is compared with detection threshold λ for finding the position of target. If λ is high, then P_d decreases that may interfere the primary user communication. If λ is low, then P_{fa} increases that cause the reduction of efficient utilization of idle spectrum which is allotted to some users. The value of λ is calculated by fixing P_{fa} and P_d values in CFAR and CDR principles, respectively. If sufficient protection for primary users is considered as primary concern, then CDR principle is used. But sometimes because of the overprotective nature of CDR principle, PU communication is prohibited even

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_31

in situation where distance between CR location and PU is more [3, 5, 6]. Because of this, CR misses good transmission opportunities which lead to an inefficient utilization of the licensed spectrum. The authors in [3] worked on an approach in which the critical distance is calculated based on the signal-to-interference-plus-noise ratio criterion, while authors in [6] used a criterion in which a given distance of SU from PU is considered as the critical distance at which the P_d in CFAR principle equals to the chosen P_d of CDR principle. The authors [5, 7, 8] further extended the works of [3, 6] to opportunistically use the distance of SU and PU in improving utilization of licensed spectrum. This article efficiently uses the hybrid approach of [3, 6] to enhance the throughput of CR system while sufficiently protecting the licensed users.

Finding the distance between PU and CR is an important task. The proposed approach considers that the position of PU is on the cell boundary. PU is located on the cell boundary where it is very closer to SU transmitter. Critical distance is calculated using critical *SNR* (SNR_C) at particular value of threshold λ where both CDR and CFAR principles are considered. After finding d_c value, minimum power (P_{min}) obtained at a sensing node is calculated when the separation is of $d_{min} = d_c + R$, where R is the range of primary cell. Under the energy detection technique whenever CR detects that the signal power received from the primary transmitter $(P) \geq P_{min})$, it is considered that the separation of secondary transmitter and primary receiver (d) is less or equal to the value of d_{min}, and the reverse also considered. For case $P \geq P_{min}$, mean distance is less then CDR–OSA scheme which should be used. For $P < P_{min}$, there is no interference to PU; so, here spectrum sharing can be used. The CR working under this proposed approach improves the throughput than the other approaches where CDR principle is used blindly.

2 Analysis of CDR, CFAR Principles, and CDR–OSA Scheme

The noise and PU signals are taken as circularly symmetric complex Gaussian (CSCG) signal and phase-shift-keying signal, respectively; as a consequence, P_d and P_{fa} are given as [3]

$$P_{fa} = Q\left(\left(\frac{\lambda}{\sigma_u^2} - 1\right)\sqrt{N}\right) \tag{1}$$

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

In CDR principle, probability of detection is taken as constant, i.e., $P_d\left(= \overline{P_d}\right)$. From Eq. (1) threshold $\lambda(= \lambda_d)$ is as follows:

$$\lambda_d = \sigma_u^2\left(\left(Q^{-1}\left(\overline{P_d}\right)\cdot\sqrt{\frac{2SNR_p+1}{N}}\right)+SNR_p+1\right), \tag{3}$$

whereas in CFAR principle, P_{fa} is taken as constant, i.e., $P_{fa}\left(=\overline{P_{fa}}\right)$. From Eq. (1), threshold $\lambda\left(=\lambda_{fa}\right)$ is calculated as

$$\lambda_{fa} = \sigma_u^2\left(\left(\sqrt{\frac{1}{N}}\cdot Q^{-1}\left(\overline{P_{fa}}\right)\right)+1\right) \tag{4}$$

Transmission rates C_{00} and C_{01} for a CR system using frame structure of [4] are given by [4]

$$C_{00} = \log_2\left(1+\frac{P_{rs}}{\sigma_u^2}\right) = \log_2(1+SNR_s) \tag{5}$$

$$C_{01} = \log_2\left(1+\frac{P_{rs}}{\sigma_u^2+P_{Pr}}\right) = \log_2\left(1+\frac{SNR_s}{1+SNR_p}\right), \tag{6}$$

where C_{00} and C_{01} are transmission rates of CR when target channel is detected as idle under H_0 and H_1, respectively. P_{rs}, SU is the power received at its receiving end, and SNR_s is corresponding SNR. In case of CDR-based OSA scheme, P_{rs} which uses the fix value of transmission power P_{pre} is given by

$$P_{rs} = \mathrm{c}\cdot P_{pre}\cdot\frac{1}{d_{ss}^{\mathrm{r}}} \tag{7}$$

Here, d_{ss} is the separation of secondary transmitter and its receiver, r is the path loss exponent, and c is a constant term. Total throughput of CR system with CDR–OSA approach is given as

$$\begin{aligned} C_{CDR-OSA} = &\left(\frac{T-\tau}{T}\right)P(H_0)\left(1-P_{fa,CDR}\right)C_{00} \\ &+\left(\frac{T-\tau}{T}\right)P(H_1)\left(1-\overline{P_d}\right)C_{01} \end{aligned} \tag{8}$$

Analysis of one fact causes the reduction in efficiency of CR system in CDR principle:

By observing Eq. (3), to maintain the target probability, the threshold λ_d decreases as the *SNR* of the obtained PU at the sensing node $\left(SNR_p\right)$ is reduced for known values of N and σ_u^2. The transmission power loss of the PU can be given by the following equation:

$$P_p = P_t \cdot K \cdot \left(\frac{d_0}{d}\right)^r, \tag{9}$$

where P_t is the transmission power from PU, P_p is the received power of sensing node from the PU, d_0 is the reference distance, d is the separation of PU from sensing node, r is the pathloss exponent, and K is the constant which is dependent of free space and antenna characteristics pathloss up to distance d_0. From Eq. (9), the SNR_p can be written as (Fig. 1)

$$SNR_p = \frac{P_p}{\sigma_u^2} = \frac{1}{\sigma_u^2} \cdot P_t \cdot K \cdot \left(\frac{d_0}{d}\right)^r \tag{10}$$

3 The Proposed Approach

The PU network is assumed in this paper which has one fixed primary transmitter serving as a base station. The service range of primary transmitter is R (Fig. 2) [3].

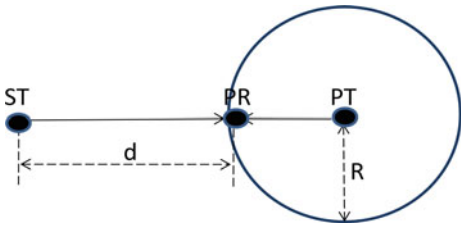


Fig. 1. Assumed system model for proposed approach [3, 5–8]

The proposed approach is given in Fig. 1 in which SU first computes its distance d_{ps} from primary transmitter [3]. d_{ps} is used to calculate the distance d of primary receiver from CR transmitter as

$$d = d_{ps} - R \tag{11}$$

After knowing the value of distance d, it is compared with a critical distance d_c, here, d_c is the distance between receiver and transmitter at SNR_c for particular value of λ.

Consider a threshold value such that $\lambda_d = \lambda_{fa}$ (means CDR and CFAR both the principles are considered), then the number of samples becomes as [2]

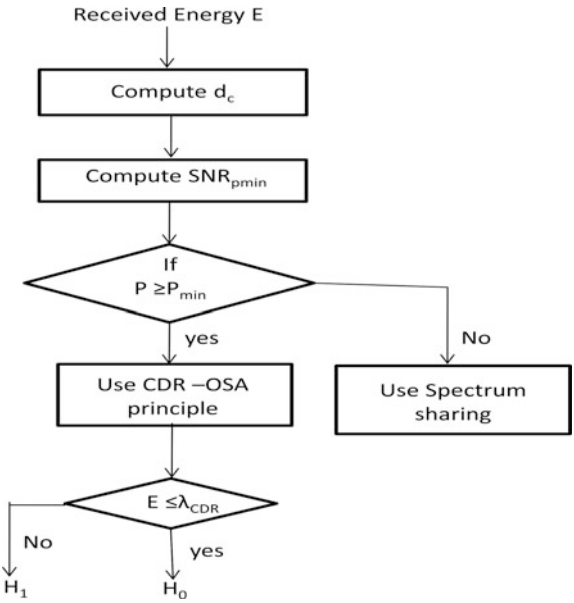


Fig. 2. Flowchart representation of the proposed approach

$$N = \frac{1}{SNR_p^2}\left(Q^{-1}\left(\overline{P_{fa}}\right) - Q^{-1}\left(\overline{P_d}\right)\sqrt{\left(2SNR_p + 1\right)}\right)^2 \tag{12}$$

From the above equation, for the low values of SNR_p, it is assumed $\left(2 \cdot SNR_p + 1\right) \cong 1$, then N can be reduced to

$$N = \frac{1}{SNR_p^2}\left(Q^{-1}\left(\overline{P_{fa}}\right) - Q^{-1}\left(\overline{P_d}\right)\right)^2 \tag{13}$$

Critical value of SNR_p (i.e., SNR_c) is written as [6]

$$SNR_c = \left(\frac{Q^{-1}\left(\overline{P_{fa}}\right) - Q^{-1}\left(\overline{P_d}\right)}{\sqrt{N}}\right) \tag{14}$$

To make the proposed approach simple, it is considered that hidden terminal problems are neglected at sensing node. Using Eqs. (10) and (14), critical distance is calculated as [6]

$$d_c = d_0\left[\frac{\sigma_u^2 \cdot SNR_c}{P_t K}\right]^{\frac{1}{r}} \tag{15}$$

After calculating critical distance d_c, the distance d_{min} and the corresponding power P_{min} are obtained as follows [3–6]:

$$d_{min} = d_c + R \tag{16}$$

$$P_{min} = c \cdot P_p \cdot \frac{1}{d_{min}^r} \tag{17}$$

The SNR corresponding to $P_{min}\left(i.e., SNR_{pmin}\right)$ can be written as

$$SNR_{pmin} = \frac{P_{min}}{\sigma_u^2} \tag{18}$$

The proposed approach can be explained with the help of flowchart in Fig. 3 as follwos: During sensing, where condition of $P \geq P_{min}$ exists, CDR–OSA scheme of spectrum access is used for accessing licensed spectrum. In case of $P < P_{min}$ where $d > d_c$ and probability of sufficient protection for PU communication is more, the CR uses spectrum sharing to access the spectrum. This paper basically uses the approach of [3, 6] in a hybrid manner. Throughput of CR using spectrum sharing scheme is given by

$$C_{CFAR} = \left(\frac{T-\tau}{T}\right) P(H_0) C_{00} + \left(\frac{T-\tau}{T}\right) P(H_1) C_{01} \tag{19}$$

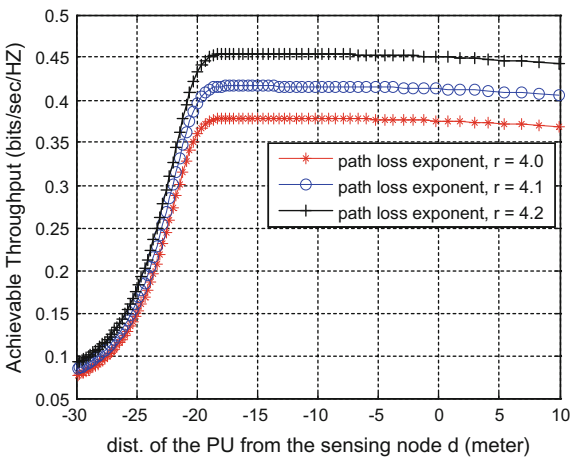


Fig. 3. The throughput of CR in blind usage of CDR principle

4 Results and Analysis

For simulation, the parameters are taken as follows: Total frame duration T = 100 ms, frequency of the target channel is 60 MHz, $P(H_1) = 0.2$, path loss exponent r = 4.2,

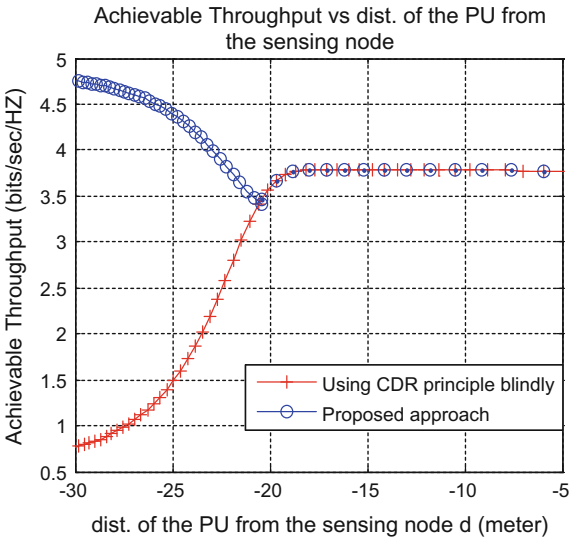


Fig. 4. Comparison of throughput with that of blind usage of CDR principle

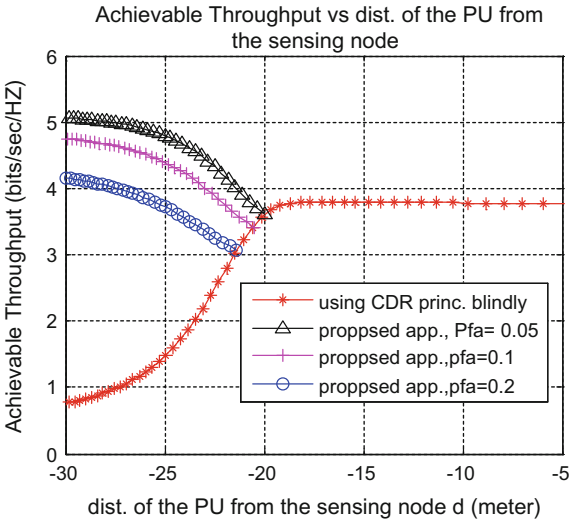


Fig. 5. The throughput using proposed approach for various values of P_{fa}

sensing time $\tau = 14.2$ ms, PU power of transmission P_p is 1 W, SU transmission power $P_t = 0.1$ W, and gains of antenna both sides are taken as unity.

Figure 3 plots the variations in the throughput of CR in the approach where CDR principle is used blindly with respect to d for different values of r. As the value of pathloss r increases, throughput of CR is increased. Figure 4 plots the comparison of throughput of proposed approach with that of approach where CDR principle is used blindly with respect to d. For $d > d_c$, the proposed approach which opportunistically adopts spectrum sharing scheme which gives better throughput. Figure 5 plots the variation of throughput of CR in CDR principle and that of proposed approach for different values of P_{fa}. It shows that as P_{fa} is decreased, the throughput is increased.

5 Conclusion

In the proposed approach, required benefits are taken from both OSA scheme and spectrum sharing scheme. The proposed approach is better than the techniques in which CDR principle is used blindly or in which both CDR and CFAR principles are used in other cases. This approach shows efficient results with respect to the distance between primary receiver and sensing node.

References

1. Federal Communications Commission (2002) Spectrum policy task force report, pp 02–135
2. Haykin S (2005) Cognitive radio: brain-empowered wireless communications. IEEE J Select Areas Commun 23:201–220
3. Verma G, Sahu OP (2016) Intelligent selection of threshold in cognitive radio system. In: Telecommunication systems. Springer, pp 1–10
4. Liang YC, Edward Peh CY (2008) Sensing throughput tradeoff for cognitive radio networks. IEEE Trans Wireless Commun 7(4):1326–1337
5. Verma G, Sahu OP (2016) Efficient use of location of unlicensed users in improving the utilization of licensed spectrum. ETAEERE, Springer, pp 1–8
6. Verma G, Sahu OP (2016) Opportunistic selection of threshold in cognitive radio system. In: A journal in wireless personal communications. Springer, pp 1–16
7. Verma G, Sahu OP (2016) Location based throughput gain in cognitive radio. In: IEEE international conference on communication control and intelligent systems (CCIS), pp 36–40
8. Verma G, Sahu OP (2016) Distance based utilization of licensed spectrum in cognitive radio. In: An IEEE international conference on the recent advances and innovations in engineering (ICRAIE), pp 1–4. Jaipur, Dec 2016

SmarThings: An Utility Service for Wireless Home Automation, Health Monitoring, and Surveillance System

Subhamoy Maitra(✉), Madhumita Hazra, Aindrila Ghosh, Soumik Pattanayak, and Bulbul Mukherjee

Department of Computer Science and Engineering, Bengal Institute of Technology, West Bengal, India
{smaitra94, madhu.hazra14, aindrila95ghosh, soumik.p495, mebulbulmukherjee}@gmail.com

1 Introduction

The advancement of technology has made all-round inroads in the daily life of mankind. One of the recent is the Internet of Things (IoT) [1] which provides connectivity for everything and everyone. In this era of technology, human beings are becoming more and more dependent on Internet for Things in order to maintain their routine household works like running and controlling gadgets like TV, fridge, pump, etc. Automation is now the order of the day and more so because people are often running after hectic work schedule outside their homes and more often than not leaving behind elderly and sometimes physically challenged parents and/or small children. It is often frightening to even think how shall they open or close the doors while felt necessary or switch off microwave oven which the ones who had gone out forgot to do in a hurry. With these problems in mind, this paper illustrates a framework to use wireless technology to design home automation system with sensors by which the household electrical/electronic gadgets can be controlled from any place, inside or outside, through a smartphone app or a website. The bot, with camera to look into every detail of the house, designed for the purpose of live streaming also, eliminates the placement of several CCTV cameras needed at every other corner of the house thereby reducing the cost. This paper also introduced sensors for health monitoring as well as for surveillance purposes in their proposed system. The temperature [2] and LPG [3] sensors are also incorporated to sense any kind of gas leakage and fire. The temperature sensor shall also be used for controlling the room temperature that is suitable for the ailing persons and/or the children at home. In the proposed system, sensors (Body temperature and blood pressure [4]) can also be attached to any patient's body so that his/her health conditions can be monitored at any point of time. Physically challenged person can advantage of this system, they can control entire house and monitor it without movement. Elderly person can monitor their health on their own also when they are living alone at the same time they can do certain jobs efficiently using this system.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_32

2 Related Work

From the available literature, it is noted that the technologies being used are mostly wired which are simple to implement but may cause various problems in future, if not properly executed according to plan and built during the construction; further, due to lots of overheads, the overall cost of the system gets increased. And till now, it is recorded that wireless systems are available which are based on Bluetooth [5]. Bluetooth can give the advantages of wireless systems but it has a major drawback in the application due to a limited range (approximately 10 m); thus using Bluetooth it is not possible to control an entire house or building sitting at the same position or from a distance outside the house. Moreover, all the available systems are dedicated to one single purpose, i.e., either health monitoring [6, 7] or home automation [8–10]. Using different modules and operating them simultaneously naturally adds to more overhead cost.

In [11], the author has proposed a Zigbee-based home automation system which provides a wireless home automation where a common home gateway has been integrated with a Wi-Fi [12] system. This provides a remote control access to the home appliances within a close proximity.

A Bluetooth-based home automation system is proposed in Piyare and Tazil [13] which includes Arduino BT board which is communicating to the cell phone, use of Bluetooth enlarges the range (approximately 10 m) of control.

In the paper [14], the author has proposed a health monitoring system through behavioral pattern.

In the paper [15], the author has proposed a health monitoring system using an incremental diagnosis method (IDM) including wearable sensors, through which any health problem can be diagnosis at an early stage of occurrence.

This paper describes a system which could overcome certain limitations encountered in abovementioned research papers. In [11] and [13], both are wireless systems but Zigbee and Bluetooth can provide a certain range limiting its access control, whereas this paper incorporates the technology of Internet of Things which overcomes the problem of area limitation, as the system can be operated through Internet, anyone can operate this through anywhere, but with complete user ID and password security. In [14], the proposed health monitoring system is based upon human behavioral pattern which sometimes may provide faulty prediction whereas use of sensors in [15] can overcome that, but it can be only monitored by the patient/aged person or someone who is near to them, but if they are living alone their physical condition cannot be monitored by their concerned family member sitting at a remote side; so use of IoT can resolve this issue. This paper proposed a system which has combined both the health monitoring and home automation system which can give an ease of living to the senior citizen or aged person in the home as well as people going outside of the home they can monitor their house using IP camera [16] mounted on Arduino bot, they can get live video streaming through which person can monitor their house from anywhere. So, overall this proposed system is cost-effective, flexible, and easy to use.

3 Proposed Framework

In Fig. 1, the framework has been represented. It has been built upon a four-layer architecture. The four layers, which are created on the basis of their area of operation, are user layer, service layer, link layer, and physical layer. The working process of each layer is described as follows:

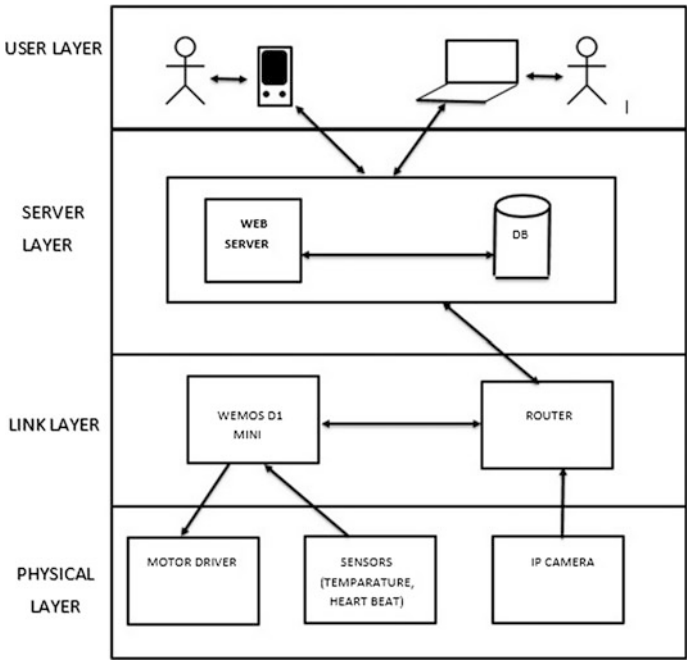


Fig. 1. Layered architecture of proposed framework

User Layer: It is the topmost layer of the architecture where the user sits. Figure 2 shows the flowchart of this layer's functionalities. The app and the website is the interface through which the user interacts with the server. The basic operations of this layer are login and sign up. Once the user is successfully logged in, he can monitor the sensor data and video stream simultaneously and also can send control signals to the server to move the bot or change the status of any switch.

Server Layer: This layer consists of two servers, a webserver and a database server with a centralized database. It is the core part of the framework. It works as a two-way communication channel as shown in the flowchart in Fig. 3. It takes data from both user layer and link layer. It also sends data to these layers. The data coming from the user layer (user's personal information or switch, bot state) is processed by the web-server and then stored inside the database, at the same time the sensor data is also fetched from the database. The data coming from link layer which is mainly sensor data

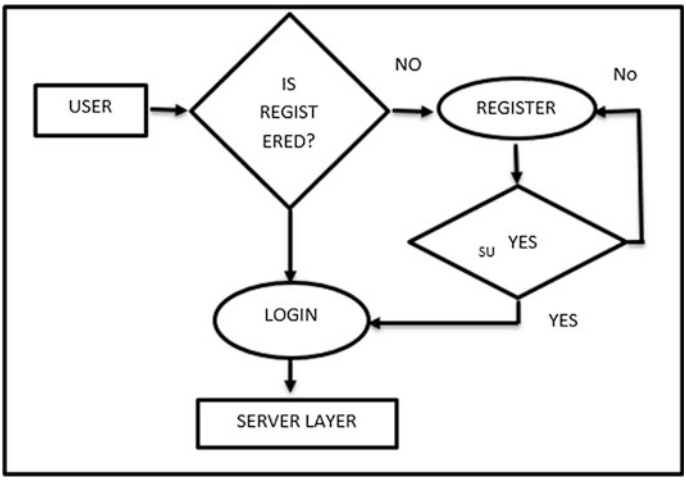


Fig. 2. Flowchart of user level functionalities

is also processed at the webserver and after finding the corresponding user, the data is stored for that particular user. The videos stream is directly passed from the webserver to the user; no database activity is involved here.

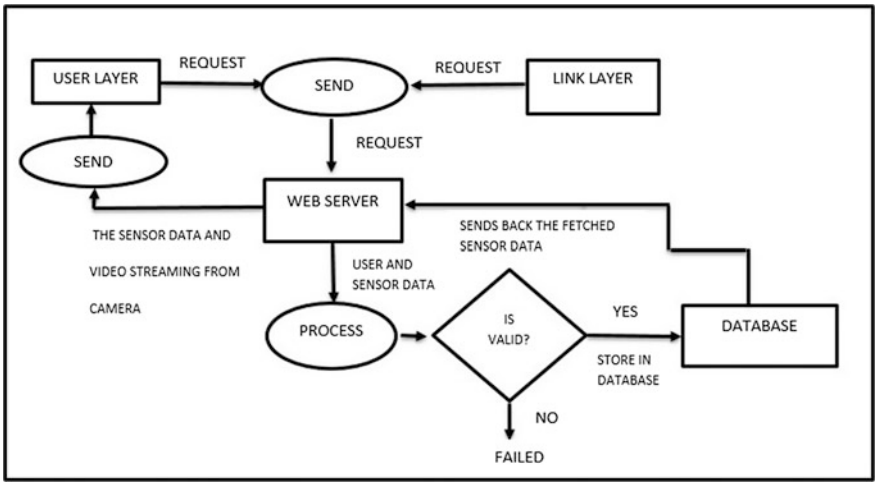


Fig. 3. Flowchart of serer level functionalities

Link Layer: This layer is client-side network. The main components of this layer are the Wemos [17] and the router. The Wemos connects the local router to gain access to the Internet. It reads the sensor data from the physical layer and sends this data to the server layer via Internet (using the router). Simultaneously it also reads the bot and

switch state from the centralized database at server layer for a particular user for which it is programmed for. After reading these states, the Wemos sends control signals to the hardware components at the physical layer accordingly. Functionalities of this layer have been shown in the flowchart in detail in the Fig. 4.

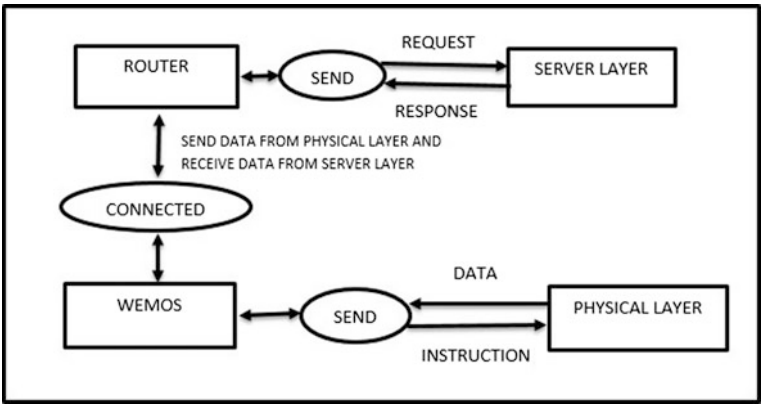


Fig. 4. Flowchart of link level functionalities

Physical Layer: This is the last layer where all the physical hardware components reside. The sensors at this layer send the data to the Wemos at link layer whereas the relay board [18] and the motor driver [19] only receive control signals. The control signals received by the relay board controls the switchboard and the signals received by the motor driver controls the movement of the bot (Fig. 5).

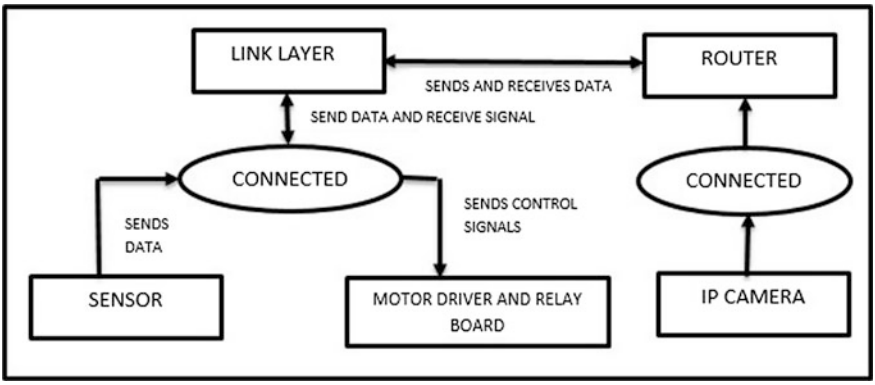


Fig. 5. Flowchart of physical level functionalities

4 Implementation

Figure 6 is the actual image of the bot. The base of this bot has been made using a metal chassis; Chassis is a mechanical assembly for making a four-wheel drive platform. Four motors for four wheels are fixed to the chassis and the wheels are fitted to the DC geared motors. Motor driver (L298 N) is attached on the top surface of chassis they are connected to the motor to control the car directions. There is an attached Wemos which is connected to the motor driver. Temperature and humidity sensor have been attached with the Wemos. The IP camera is mounted at the front of the car to get the live video streaming through it. The client using webpage or android app can send request to the server to move the car on a particular direction or stop the car on desired direction. The server stores the information inside a centralized database. From that database, Wemos reads the information and accordingly sends control signal to the motor driver.

Figure 7 shows the detailed connection inside the home automation kit, it consists of one adapter circuit, two-channel relay boards, and a Wemos. With this, a switchboard has been connected to this home automation kit serving the purpose as main switchboard. A 220 V 50 Hz AC power comes as input to this kit, which is then divided into two halves. One goes to the adapter circuit another to the switchboard through the relay. After getting power from an AC source adapter gives output 5 V DC which activates relay and the Wemos. The client using webpage or android app sends ON/OFF request to the server for switches. Server stores this commanded state in the centralized database for particular switch. The Wemos connects to local router configured in the client side to gain access to the Internet. The Wemos then sends a GET request to the author's server for reading the status of switches or the car from the database. The server checks the ID of the Wemos to validate and if it a match is found corresponding data is sent from the server. Wemos reads this status and sends control signal to the relay. If the status is OFF, then relay breaks the positive half on the input power source and the current flow is blocked and if the status is ON then it allows the current to flow for that particular switch.

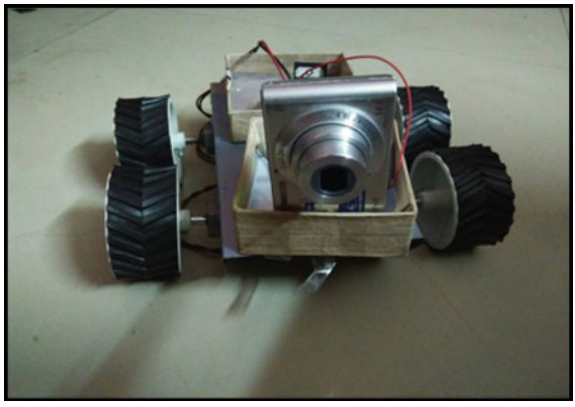


Fig. 6. Bot

Likewise, health monitoring instrument like blood pressure sensor, pulse sensor, etc., are connected with Wemos which sends the updated data from the patient's body to the centralized database, and thus can be monitored through app and website. Thus, all the three functions are assembled in the android app and website.

Figures 6 and 7 show the respective screenshots of android app and the website through which the robot car and the home automation are being accessed. The system is found to be running successfully.

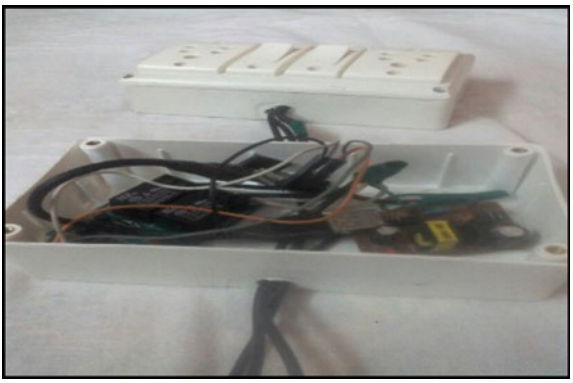


Fig. 7. Connection between switchboard and home automation kit

5 Result and Performance Analysis

After successful connection establishment of all our hardware, connecting Wemos(s) to the network and successful user registration and login phase, there was a test to check the response time to calculate its speed depending upon various bandwidths. In Table 1, the entire result of that experiment has been given and Fig. 8 shows the graphical relationship between network bandwidth and response time.

Table 1. Bandwidth versus response time analysis

Network bandwidth (Kbps)	Response time (s)
22.3	1.01
20.2	1.39
16.1	1.43
12.7	1.82
10.5	2.37
9	2.93

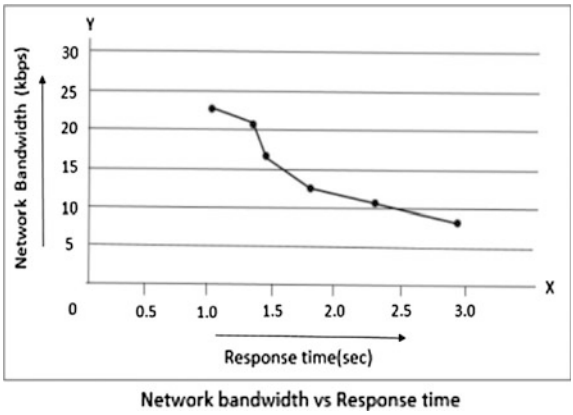


Fig. 8. Network bandwidth versus response time

6 Conclusion and Future Work

6.1 Conclusion

From this experiment, the complete status of the house can be fetched from outside anywhere. The entire experiment is not only limited to smartphone, but inclusion of website is also implemented. Instead of temperature sensor, any other sensor can be also attached in the bot as per user's need like gas leakage sensor, motion sensor, etc. System proposed in this paper provides advantages for large offices or big complexes as it is really a tough job to move floor to floor or block to block to switch on and switch off lights and other appliances and used large manpower. Through this system, it is possible for a single user to control all the lights sitting at one place. In the proposed system, assembling all three in one modules (home automation, health monitoring and surveillance) could be controlled from one single app or website, so it makes the job simpler and cost efficient at the same time.

6.2 Future Work

The proposed system can be extended to by various modifications. The bot can be modified by high-standard wheels and designs. Other kinds of different sensors can also be added in the bot. The system including these changes can be used in earthquake rescue mission. Moreover, attaching the mine detect sensor, the bot can be used in mine detecting operation as the bot can be controlled from a safe location. For the purpose, the bot can be made smaller in size so that it can be used for surveillance for any kind of attack in a closed zone like a hospital or a hotel where police cannot locate the exact position of the attacker.

References

1. Xia F, Yang LT, Wang L, Vinel A (2012) Internet of things. Int J Commun Syst 25(9):1101
2. DHT11 Basic Temperature-Humidity Sensor. https://www.adafruit.com/product/386
3. LPG gas sensor—MQ-6. https://www.sparkfun.com/products/9405
4. Blood Pressure Sensor—Serial Output. http://www.sunrom.com/p/blood-pressure-sensor-serial-output
5. Bluetooth SIG (2001) Bluetooth specification version 1.1. http://www.bluetooth.com
6. Pantelopoulos A, Bourbakis NG (2010) A survey on wearable sensor-based systems for health monitoring and prognosis. IEEE Trans Syst Man Cybern Part C (Applications and Reviews) 40(1):1–12
7. Varshney U (2007) Pervasive healthcare and wireless health monitoring. Mob Netw Appl 12 (2–3):113–127
8. Agaskar APVA, Mithagari A, Mhatre A, Shetty N (2016) Internet of things: home automation and surveillance system. Int Educ Sci Res J 2(4)
9. Alkar AZ, Buhur U (2005) An Internet based wireless home automation system for multifunctional devices. IEEE Trans Consum Electron 51(4):1169–1174
10. Al-Ali AR, Al-Rousan M (2004) Java-based home automation system. IEEE Trans Consum Electron 50(2):498–504
11. Gill K, Yang SH, Yao F, Lu X (2009) A zigbee-based home automation system. IEEE Trans Consum Electron 55(2)
12. Wi-Fi (wireless networking). http://www.webopedia.com/TERM/W/Wi_Fi.html
13. Piyare R, Tazil M (June 2011) Bluetooth based home automation system using cell phone. In: 2011 IEEE 15th international symposium on consumer electronics (ISCE). IEEE, pp 192–195
14. Barger TS, Brown DE, Alwan M (2005) Health-status monitoring through analysis of behavioral patterns. IEEE Trans. Syst. Man Cybern Part A Syst. Hum. 35(1):22–27
15. Wu WH, Bui AAT, Batalin MA, Liu D, Kaiser WJ (2007) Incremental diagnosis method for intelligent wearable sensor systems. IEEE Trans Inf Technol Biomed 11:553–562. ISSN 1089-7771
16. IP Security Cameras & IP Video Servers. https://www.cctvcamerapros.com/IP-Security-Cameras-s/83.htm
17. A Documentation on Wemos. https://www.wemos.cc/
18. Documentation on Relay Board. http://www.nskelectronics.com/relay_boards.html
19. L298 N Dual Full Bridge Motor Driver documentation. http://www.st.com/en/motor-drivers/l298.html

Speech Background Noise Removal Using Different Linear Filtering Techniques

Minajul Haque(✉) and Kaustubh Bhattacharyya

Assam Don Bosco University, Azara, Guwahati, India
minajul.h.electronics@gmail.com,
kaus-tubh.bhattacharyya@dbuniversity.ac.in

1 Introduction

Among the human beings, speech is the simplest and reliable way of communication. A speech signal transmits different information to the listener like type of language being spoken and emotion, gender and identity of the speaker [1]. The individual characteristic like pitch, fundamental frequency, formant frequency can be distinguishing components of human speech. The three main processes by which sound generates—are twisting of nerves, wire beating of membranes or blowing of air through holes but the mechanism of human voice is different as it comes out in different languages and feelings by the control of brain [2]. The range of frequency under which the speech signal falls is the 300–3400 Hz. However, the audible frequency range for human beings are from 20 Hz to 20 KHz [3]. Audio signal processing often suffers from noise trouble [4]. Speech background noise is an undesired signal which mixed with speech signal at the time of generation of speech signal or at the time of transmission [5]. The quality and the intelligibility of the speech signal gets degraded because of the present background noise. Therefore, it becomes important to reduce this background noise from the speech. For removing or reducing the noise, different filtering techniques are there like adaptive filter, Kalman filter, sub-band coding, wavelet transform, etc. The various filters that have been designed and tested are presented in this paper. First, the different kinds of adaptive filtering techniques have been implemented on both the additive white Gaussian noise (AWGN) and the practical noise then the results of both have been compared. After the adaptive filters the optimum filter, i.e., Kalman filter is implemented for both types of the noise. Finally, the results of adaptive filters and the Kalman filter have been analyzed.

2 Adaptive Filtering

In any filter, if the weights of the filter get self-adjusted according to an optimizing algorithm or some predefined rule then the filter is called

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_33

an adaptive filter [6]. According to Premanada B.S et al. the parameters which are used for the processing of the signals change according to some predefined criterion, usually this criterion is chosen as estimated mean squared error (MSE) or correlation. This phenomenon can be considered as an adaptive filtering [7]. A general block diagram of an adaptive filter is shown in Fig. 1. An adaptive filter is a self-regulating system that takes help of recursive algorithm for processing. First input and training are compared and accordingly error signal is generated and that is used to adjust some previously assumed filter parameters under the effect of incoming signal. Filter parameter adjustment continues until steady-state condition. As far as application of noise reduction from speech is concerned, adaptive filters can give best performance. Reason for that noise is somewhat similar to the randomly generates signal and every time its very difficult to measure its statistic. Design of fixed filter is a completely failed phenomena for continuously changing noisy signal with the speech. Some of the signal changes with very fast rate in the context of information in the process of noise cancelation which requires the help of self regularized algorithms with the characteristics to converge rapidly. Some applications of adaptive filtering are system identification, channel equalization, and signal enhancement by noise reduction [8].

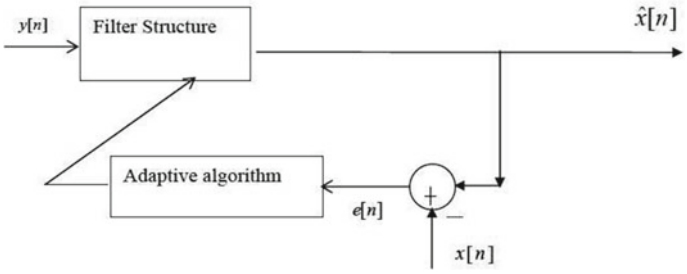


Fig. 1. General block diagram of adaptive filter

2.1 Least Mean Square Algorithm (LMS)

If an adaptive filter is driven by a Least Mean Square algorithm (LMS) then the filter is said to be a LMS filter. LMS filter is class of adaptive filter which is used to mimic a desired signal by finding the coefficient of the filter in terms of least mean square error of the signal. It is a stochastic gradient descent method in that the filter is only adapted based on the error at the current time [9]. The LMS filtering algorithm can be expressed by three expressions as follows:

1. Output of the filter: $Y(n) = \hat{w}(n)u(n)$
2. Estimated error: $e(n) = d(n) - y(n)$
3. Updated weight: $\hat{w}(n+1) = \hat{w}(n) + \mu u(n)e^*(n)$

where $\hat{w}(n)$ is the weight vector, μ is the step size, and $e^*(n)$ is the conjugate of error vector.

Adem Ukte et al. proposed a performance evaluation application of LMS adaptive algorithm for removing the noise and training an adaptive filter is given in [10].

2.2 Normalized Least Mean Square Algorithm (NLMS)

In the standard form of a least mean square filter, weight vector of the filter at iteration $(n + 1)$ gets the necessary adjustment and gives the product of three terms as follows:

1. The step size parameter μ, which subject to design concept
2. The input vector $u(n)$, which is actual input information to be processed
3. The estimated error $e(n)$ for real-valued data, or its complex conjugate $e^*(n)$ for complex-valued data, which is calculated at iteration n.

There is a direct relationship between the adjustment and the input vector $u(n)$. AS a result LMS filter fails because of its gradient noise amplification problem in the case when $u(n)$ is very large. As a solution to that problem of LMS normalized LMS filter can be used [9]. The term normalized can be considered because the adjustment given to the weight vector at iteration $(n+1)$ is "normalized" with respect to the squared Euclidean norm of the input vector $u(n)$. Sivaranjan Goswami et al. proposed an example of NLMS filter for noise filtering from speech, where they have considered the first 0.5 s to be the noise reference and based on that they have trained the adaptive filter to remove the noise from the input noisy speech signal [11].

3 Kalman Filter

The Kalman filter operates through a prediction and correction mechanism, because of this it is also called as a mathematical procedure. Kalman filter combines all the available data, i.e., measured, system knowledge, and the measurement devices to get an estimation of the desired variables in such a manner that the error between the measured and original data gets statistically minimized [12]. Usually, Kalman filter is used for white noise reduction. However, different methods were developed to fit the Kalman approach to reduce the colored noises also. A discrete Kalman filter can be used to estimate a process $d(n)$ from a set of observation $x(n) = d(n) + v(n)$. If this $d(n)$ and $x(n)$ are jointly wide sense stationary process then it can be predicted by using the Wiener filter but as in practical most of the process are nonstationary, therefore Wiener filter cannot be used. A causal Wiener filter for estimating a process $x(n)$ from a noisy measurement can be written as

$$y(n) = x(n) + v(n) \tag{1}$$

Considering a specific problem of estimation an autoregressive $AR(1)$ process of the form

$$\hat{x}(n) = a(1)x(n-1) + w(n) \tag{2}$$

where $w(n)$ and $v(n)$ are uncorrelated white noise process. The optimum estimation of $x(n)$ using all of the measurements, $y(k)$, for $k \leq n$ for k could be computed with a recursion of the form

$$\hat{x}(n) = a(1)\hat{x}(n-1) + K[y(n) - a(1)\hat{x}(n-1)] \tag{3}$$

where K is a constant, referred to as the Kalman gain, that minimizes the mean square error $E\{|x(n) - \hat{x}(n)|^2\}$

An application of Kalman filter to reduce the noise from the old audio recordings in which quality is degraded by aging or limitation of the recording reproduction mechanism is described by Rayan Kutty P et al. [13] such application of Kalman filter is based on the noise variance estimation from the silent region of the signal. It takes the same consideration like spectral subtraction where the initial first one or two seconds of recording are considered to be noise [14]. Orchisama Das et al. in their paper [15], they have used a Kalman filter to filter the noise by tuning its measurement noise covariance factor and distinct Kalman gain for silent and voiced frames.

4 Experimental Details

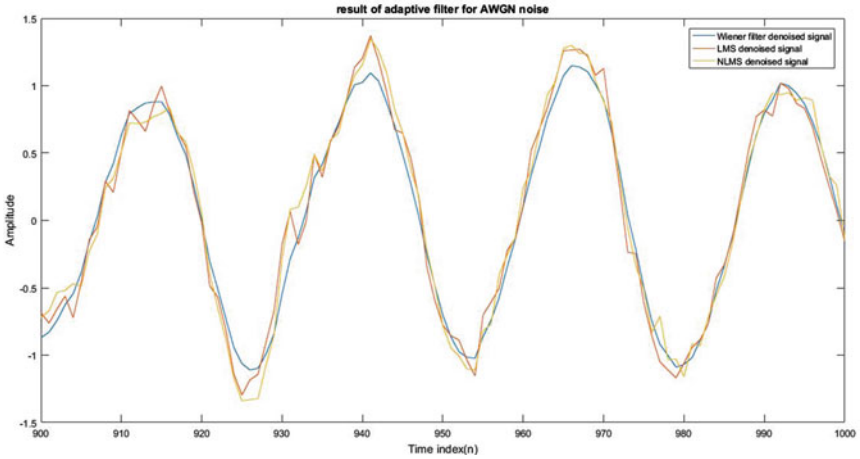
For testing of the various speech background noise removal filtering techniques, three different practical speech signals are recorded, i.e., a clean speech signal without any noise, a noisy version of the same speech signal and a noise signal which is present in the noisy speech signal. As the number of samples in all the signals are very high so only a segment of all the signals is tested on the designed filters. For theoretical noise analysis, the AWGN noise is considered.

As referred in Sect. 2 the various adaptive filters for both AWGN noise and practical noise mixed signals results are given below. The filter length of the filter is 7 and which is kept constant for all the filters.

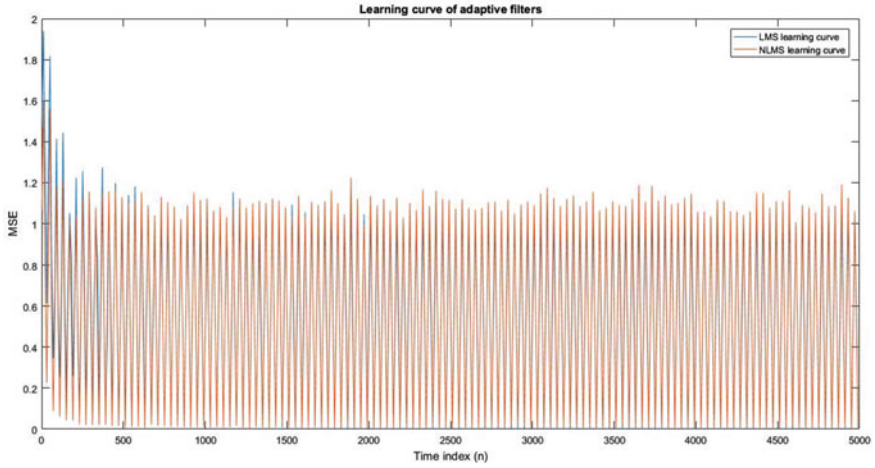
Figure 2a shows the output of the different adaptive filters, as the Wiener filter result was assumed to be the optimum so from here it can be concluded that both LMS and NLMS tried to match the Wiener filter result. From the Fig. 2b it can clearly concluded that the convergence of NLMS is faster than LMS. Now, the same adaptive filters are applied to practical signals and their results are shown below.

Figure 3a, b shows the time domain plot of the desired signal and the input signal respectively. These signals were recorded using stereo mode and then converted into the mono mode.

Figure 4a, b shows the output of both LMS- and NLMS-adaptive filter with desired signal. Here, the Wiener filter output is considered as the optimum solution and based on that it can concluded that the NLMS filter gives the best approximation of the Wiener filter result.



(a) Filtered signal of the LMS and NLMS filter with Wiener filter as optimum



(b) Theoretical learning curve of both LMS and NLMS filters

Fig. 2. Adaptive filter results

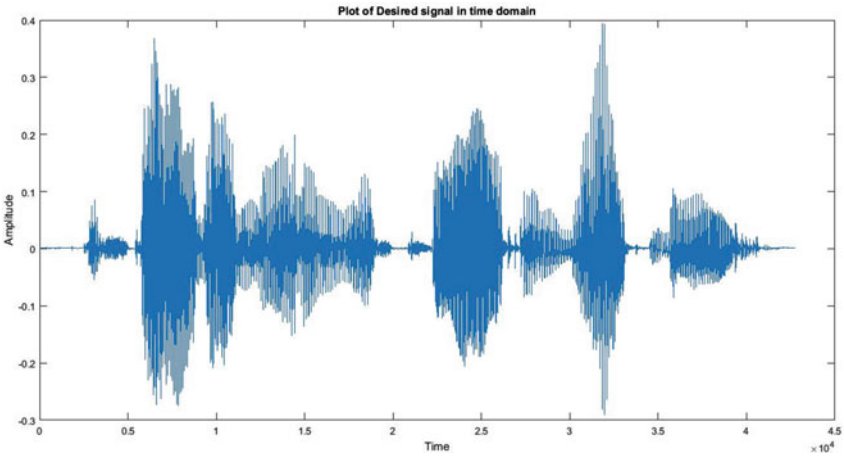
To verify whether the noise is suppressed or not in the filtered signal, the power spectral density of the input signal is calculated before and after filtering.

From the Fig. 4c it is confirmed that the level of SNR is high in the output signal of both LMS and NLMS filter as compared to the input signal.

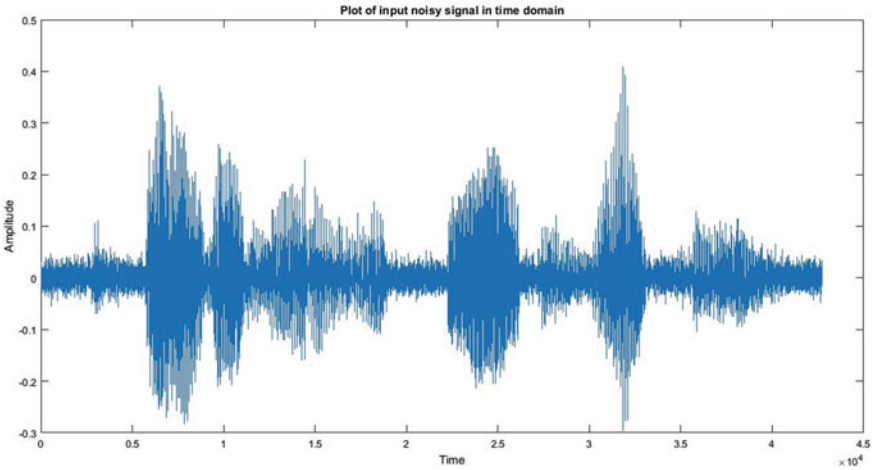
Now, as referred in Sect. 3 a Kalman filter is implemented on the same signals and the results were compared.

Figure 5a shows the time domain plot of both the desired and output of the Kalman filter and it is found that both the signals are quite similar.

Now to analyze the output of the Kalman filter the Mean Squared Eerror (MSE) was calculated and also the power spectral density of the input signal



(a) Desired signal



(b) Noisy signal

Fig. 3. Input signals

and the output signal of Kalman filter is computed to see the effect of noise reduction.

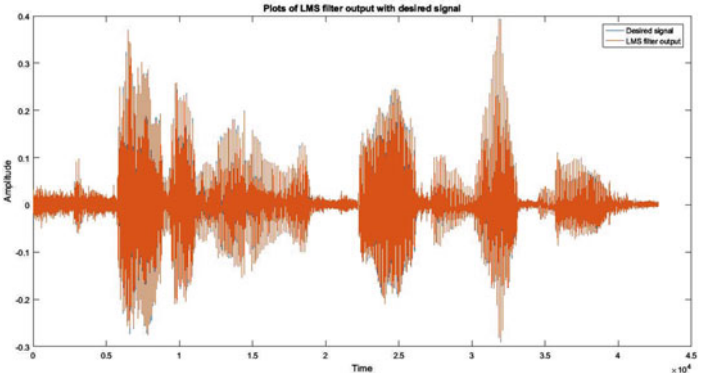
In Fig. 5b, c gives the PSD of the input signal of the Kalman filter and the output signal of the Kalman filter and mean square error of the estimated signal of the Kalman filter.

5 Comparison

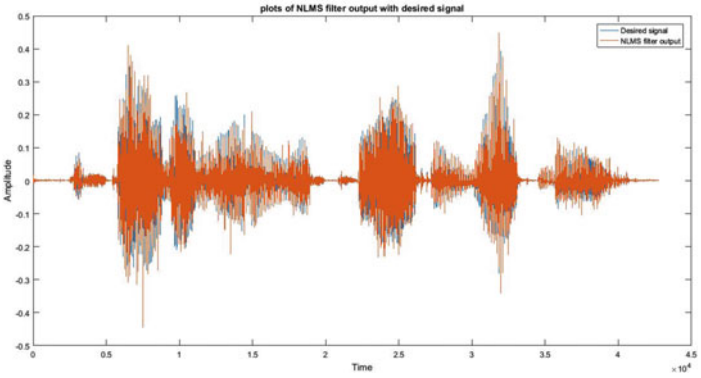
A table of comparison of all the implemented filters with Additive White Gaussian Noise (AWGN) and with practical noise based on the Peak

Table 1. Comparisons of the filter with practical noise

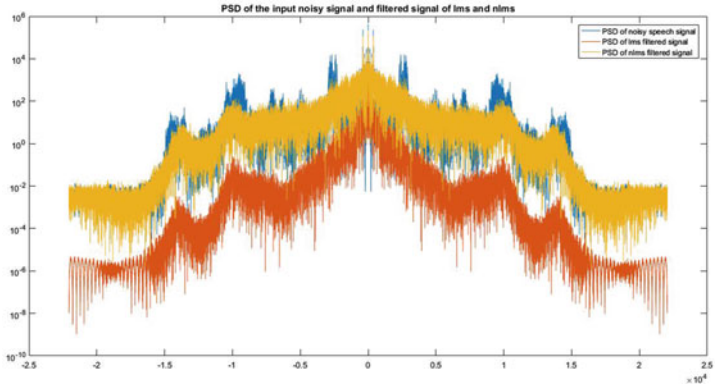
Sl No.	Parameter	LMS	NLMS	Kalman
(a) Comparisons of the filter with awgn noise				
1	PSNR	56.2380	57.8854	69.6564
2	MSE	0.1546	0.1058	0.0017
3	Max error	2.1650	1.4057	1.3563
4	L2Rat	62.87	43.2673	40.4175
(b) Comparisons of the filter with practical noise				
1	PSNR	93.2435	78.7775	75.7824
2	MSE	$3.0813e^{-05}$	$8.6166e^{-04}$	0.0017
3	Max error	0.0366	0.3719	0.4084
4	L2Rat	1.0084	0.9614	0.2973



(a) Filtered Signal with desired signal for LMS

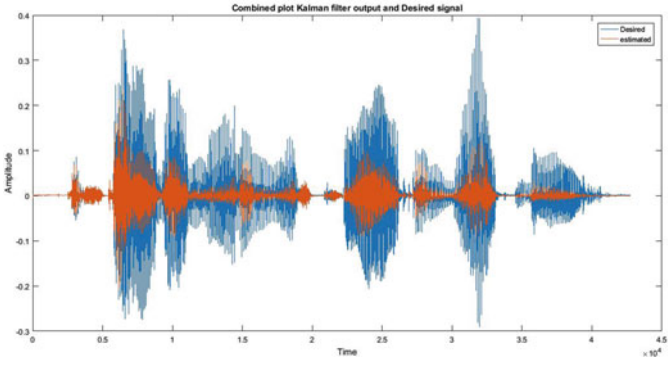


(b) Filtered Signal with desired signal for NLMS

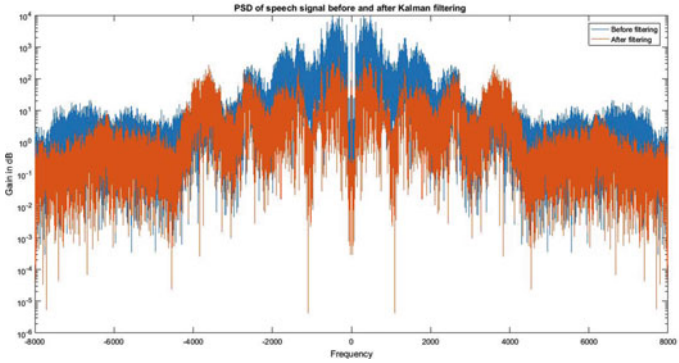


(c) PSD of input signal and output signal of LMS and NLMS filter

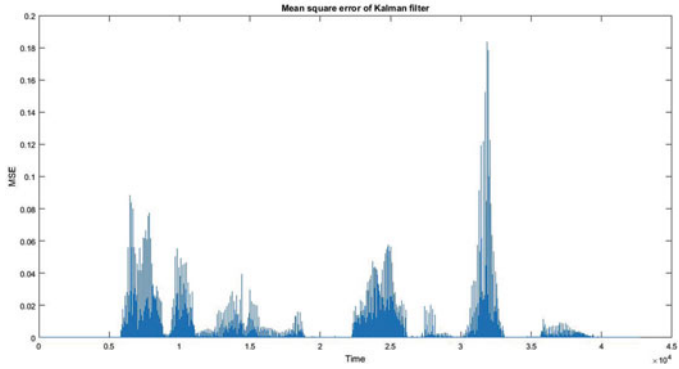
Fig. 4. Filtered signal with desired signal and PSD plot of all the signals



(a) Combine plot of Desired signal and estimated signal of Kalman filter



(b) PSD plot of the signal before and after Kalman filtering



(c) Mean square error of the output signal of Kalman filter

Fig. 5. Kalman filter outputs

Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), Maximum Error (Max error), and Ratio of Squared Norms (L2Rat) is summarized below.

From the above Table 1a, b it is seen that for AWGN noise the performance of Kalman filter is better than the other adaptive filters, but when the noise becomes practical the performance varies.

6 Conclusion

For removing the background noise in speech signal different techniques have been implemented such as adaptive filters (LMS, NLMS, and Wiener) and Kalman filter. The Wiener filter result was assumed to be optimum. In the literature, it is found that the Kalman filter gives better result then other adaptive filter like LMS and NLMS, also from the result we have found that for AWGN noise the performance of Kalman filter is better than LMS, NLMS, and wiener filter but in practical noise the performance varies. Therefore, from the analysis it can be concluded that for AWGN noise the Kalman filter performance is the best among all the filters but for practical noise this is not true.

References

1. Hayes MH. Statistical digital signal processing and modeling, chapter 4th and 9th ed. Wiley
2. Pawar RV, Jalenkar RM (2014) Review on speech production model. Int J Eng Innov. Technol (IJEIT) 3(9)
3. Kalandharan N (2014) Speech enhancement by spectral subtraction method. Int J Comput Appl 96(13):0975–8887
4. Chen Z, Shi D, Yin F (2011) Dynamic noise reduction algorithm based on time-variety filter. In: 2011 3rd International conference on awareness science and technology (iCAST). https://doi.org/10.1109/ICAwST.2011.6163085
5. Proakis, Manolakis. Digital signal processing principle algorithms and application, 3rd ed. ISBN-13:9780133737622
6. Levitt H (2001) Noise reduction in hearing aids: a review. J Rehabil Res Dev 38(1):111–121
7. Premanada BS et al (2013) Speech enhancement algorithm to reduce the effect of background noise in mobile phones. Int J Wirel Mob Netw (IJWMN) 5(1)
8. Fan BW et al (2015) The improvement and realization of speech enhancement algorithm based on Wiener filtering. In: 8th International congress on image and signal processing (CISP 2015)
9. Patil L et al (2014) Efficient algorithm for speech enhancement using adaptive filter. Int J Electr Electron Comput Eng 3(1):98–103
10. Ukte A, Kizikaya A (2016) Comparing the performances of least mean squares based multirate adaptive filters. In: 2016 26th international conference on Radioelectronika. https://doi.org/10.1109/RADIOELEK.2016.7477383
11. Goswami S et al (2014) A novel approach for design of a speech enhancement system using auto-trained NLMS adaptive filter. Int J Inf Commun Technol 6:3–4
12. Mathe M, Nadyala SP (2012) Speech enhancement using kalman filter for white, random and color noise. In: International conference on devices, circuits and systems (ICDCS)

13. Kutty PR, Murthy AS (2011) Kalman filter using quantile based noise estimation for audio restoration. In: Proceedings of ICETCT 2011, 978-1-4244-7926-9/11/$26.00 2011 IEEE
14. Pratapwar SS (2003) Reduction of background noise in alaryngeal speech using spectral subtraction with quantile based noise estimation In: 7th World multiconference on systemics, cybernetics and informatics (SCI 2003), Orlando, USA, 27–30 July 2003
15. Das O, Goswami B, Ghosh R (2016) Application of the tuned kalman filter in speech enhancement. In: 2016 IEEE first international conference on control, measurement and instrumentation (CMI)

Neural Network Classification of EEG Signal for Detection of Brain Abnormalities

Maya V. Karki(✉) and Shaguftha Yasmeen

Ramaiah Institute of Technology, Bangalore, India
mayavkarki@msrit.edu, shagufthay@gmail.com

1 Introduction

The brain is one of the strongest parts in the human body, which consists of many nerves that are connected to different parts of the body. Hence, if any changes take place in any part of the brain, it may cause difficulty in day-to-day activities. Therefore, early detection of brain abnormality is important in clinical diagnosis. Brain abnormalities [1] include seizure, autism, migraine, leukodystrophies, depression, Parkinson's disease, encephalitis, sleep disorder, etc. The proposed algorithm provides a method to detect two very common disorders of the brain, which is seizure and migraine. Both are episodic disorders and have few common symptoms and one is misdiagnosed as other. Here an attempt is made to classify seizure and migraine and enhance clinical diagnosis between seizure and migraine.

There are many noninvasive methods for the acquisition of brain signals such as Electroencephalogram (EEG), functional Magnetic Resonance Imaging (fMRI), Near Infra-Red Spectroscopy (NIRS), and Magneto Encephalography (MEG). Due to its high temporal resolution, safety, and ease of use, EEG is an important and effective tool for measuring the brain activity and understanding complex behavior of the brain. EEG is a dynamic noninvasive and relatively inexpensive method used to monitor the state of brain.

There are two main groups of epilepsy and they are called as generalized seizure and focal epilepsy. Generalized seizure is characterized by sudden loss of consciousness followed by jerky movements. Focal epilepsy is partial seizure which causes one side muscle contractions which leads unusual head or eye movement. Non-epileptic seizure are caused due to the reaction of some drug and in turn irritate the brain [2,3]. Every channel of EEG recording indicates the generalized seizure as it involves the whole brain. Focal epilepsy is detected only a few channels of EEG recording.

Migraine is a common brain abnormality which is characterized by moderate to severe headache [4]. The exact cause of a migraine [4] headache is unknown; it is thought to be due to abnormal brain activity causing a temporary alteration in the nerve signals, chemicals, and blood flow in the brain. It is influenced by certain physiological changes in hunger, sleep, etc. Migraine symptom varies from person to person. Symptoms include feeling uneasy to light and sound, vomiting sensation, giddiness, etc.

EEG shows a continuous oscillation of electrical activity in the brain. The amplitude and pattern determine the reticular activating system of the brain stem. EEG

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_34

signals can be divided into 5 different frequency bands and each band is associated with different brain functions. The five main frequency bands are: Delta (0.5–4 Hz), Theta (4–7 Hz), Alpha (8–15 Hz), Beta (16–31 Hz), and Gamma (>31 Hz).

2 Related Work

Literature review indicates that many have proposed brain abnormality detection using wavelets and its variants. Zandi et al. [5] has proposed a wavelet-based technique for the real-time detection of an epileptic seizure. Decomposition of EEG signal was done using wavelet packet transform and frequency band which represents maximum separation was obtained. The algorithm was experimented on 14 patients with 75.8 h with 63 seizure. The method resulted in 90.5% of sensitivity and false detection rate of 0.51/h. Liu et al. [3] has described an algorithm using wavelet transform and SVM classifier for seizure detection. The data set consists of 2–5 h of seizure data and 24–26 h of non-seizure data. The decomposition of multichannel iEEG (intracranial EEG) was done and frequency bands were extracted from three selected bands. Features were extracted and then sent to SVM for classification purpose. The specificity achieved was 95.26% and sensitivity of 94.46% with a false detection rate of 0.58/h. Batuhan Akben et al. [6] proposed signal processing algorithm to detect a migraine. Preprocessing step was done using Ar Burg method to obtain PSDs. Subasi et al. [7] have described how ANN can give accurate result as a classifier. In his work, he mentioned how the AR model has an advantage over spectral loss problem and also how it gives a better frequency resolution. The results gave an accuracy of 92.3% with a specificity of 96.2% and a sensitivity of 90.3%.

3 Proposed System

Figure 1 represents a block diagram of EEG signal analysis for the detection of brain abnormalities. The major blocks are signal acquisition, preprocessing, feature extraction, and classifier.

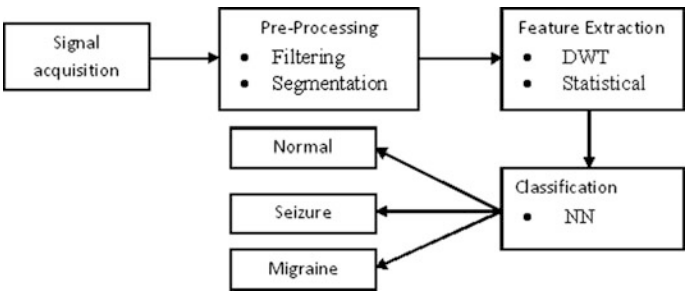


Fig. 1. Block diagram of the proposed system

3.1 Signal Acquisition

The analysis of EEG signal was done using EEG data collected over the time in EEG laboratory. It is an old EEG-recorded dataset which is annotated. It consists of totally 34 patients (10 normal persons, 10 seizure patients, and 14 migraine patients). The sampling rate of signal is 128 Hz. The duration of EEG recording of an individual person holds a minimum of 20 min. Another dataset which is taken from PhysioNet includes EEG recording of 13 persons with 1024 Hz sampling rate. Recording was done using monopolar montage with 28 channels (EMG, EKG, A1, Fp1, Fp2, A2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2, T1, T2, Pg1, Pg2, MK).

3.2 Preprocessing and Segmentation

The raw EEG data is applied with for the preprocessing techniques to remove noise. Raw EEG signal was filtered using a band-pass (0.5–60 Hz) fourth-order Butterworth filter. The removal of artifacts such as eye blinks, muscular movement, power line artifacts, etc., is done by passing the band-passed signal through the notch (50 Hz) filter. After denoising the signal, segmentation was performed by taking a window size of 60 s. From each EEG data, 4 segments were extracted for the feature extraction and detection of abnormalities.

3.3 Feature Extraction

Since EEG is a nonstationary signal, it is advantageous to use time–frequency domain technique such as Discrete Wavelet transform (DWT) to analyze the signal. Scaling and shifting are the two key features of wavelets. Wavelet works on multi-scale basis. This feature of WT allows the decomposition of a signal into several scales [3]. Study on wavelet families shows that smoothing feature of Daubechies-4 wavelet is appropriate for EEG signal detection and hence db4 has been used in this algorithm [6]. Wavelet-decomposed bands give a large number of coefficients and feature vector size increases. To reduce feature vector size, statistical moments are calculated on these coefficients and moment features are used as features. In the proposed algorithm mean, standard deviation, skewness, kurtosis, and magnitude are calculated from each band and used as salient features.

3.4 Classification

Classification of the signal is performed using feed forward neural network. The performance of network completely depends on the number of hidden neurons present in the network. If hidden neurons are less in number, the network will be incapable of differentiating between complex patterns. In contrast, if the network has more number of hidden neurons, then it will lead to the addition of noise within the actual data because of over-parameterization. Performance analysis of a classifier is obtained by calculating parameters such as specificity (true negative ratio), sensitivity (true positive ratio), and accuracy. They are defined by Eqs. (1)–(3).

$$sensitivity = TPR = \frac{TP}{TP + FN} \times 100\% \tag{1}$$

$$specificity = TNR = \frac{TN}{TN + FP} \times 100\% \tag{2}$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{3}$$

4 Simulation Results

Important information in an EEG signal lies between 0 and 60 Hz, hence a band-pass filter has been used to remove signals above 60 Hz. Recording of an EEG signal is 20–22 min duration. Processing such huge data is difficult and hence segmentation of the signal will be carried out. Initially, the complete signal was considered for feature extraction. Keeping 60 s duration, 20 segments are considered. And later one segment at the center of the signal with 60 s duration was considered from individual recording. Similarly, more segments have been considered for feature extraction and for the accurate classification. Further 4 segments each with 60 s duration is selected. In this, center of whole EEG signal with 60 s duration of two segments is selected and either side of this center one segment of 60 s is considered for feature extraction.

Figure 2 shows the decomposition of EEG signal for normal, seizure, and migraine patient. For EEG signal with 128 Hz sampling rate, fourth-level DWT decomposition is applied to extract five bands whereas signal with sampling rates of 1024 Hz, eight-level decomposition has been applied. Different EEG bands: gamma (>32 Hz), beta (15–32 Hz), alpha (7–15 Hz), theta (4–7 Hz), and delta (<4 Hz) are obtained. Figure 3 shows all five bands for a normal person after the fourth level of wavelet decomposition. When coefficients of these five bands are considered as features, the size of the feature vector is 3840.

To reduce feature vector size, moment features are calculated from the five bands and feature vector size becomes 850(34 persons × 25 features). The statistical parameters mean, standard deviation, skewness, kurtosis, and magnitude are calculated for four segments and the average value of each parameter for four segments are considered as features for classification. Table 1 shows the different average values obtained after calculation of statistical moments on different bands of EEG signal. The values are obtained for one normal person (N1), one seizure patient (S1), and one migraine patient (M1).

Classification is done using a single-layer feed forward network. Input layer consists of 25 neurons, since there are totally 25 features for individual EEG signal (five bands with five average features from 4 segments). The hidden layer consists of 10 neurons and output layer consists of 3 neurons to detect normal, seizure, and migraine. The network will be trained with scaled conjugate gradient backpropagation. From the dataset, 70% of data is used for training the network and 15% is used for validation of

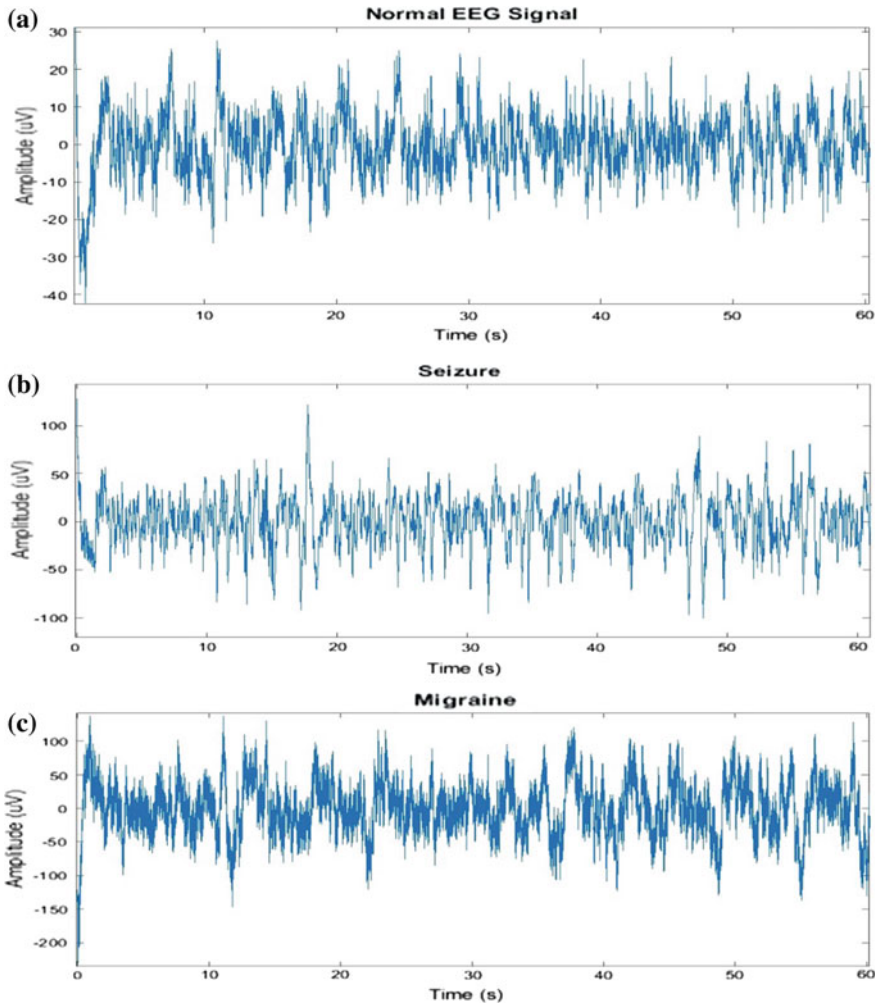


Fig. 2. Segmented EEG signal of **a** normal person, **b** seizure patient, and **c** migraine patient

result and another 15% is used for testing. The proposed work has been tested on 10 normal persons, 14 migraine patients, and 13 seizure patients EEG signal with a sampling rate of 128 Hz.

Figure 4a represents the performance plot of the algorithm which is the plot of error versus epoch for the training, validation, and testing of the network. It shows how the error is reducing as epochs increases. The best validation obtained is at epoch 11, that is, 0.18615. Figure 4b represents the ROC plot. It indicates the receiver operating characteristic for individual output class. The more each curve goes towards of left and top edges of the plot, the better results we can get for the classification. The accuracy of the algorithm can be obtained using confusion matrix.

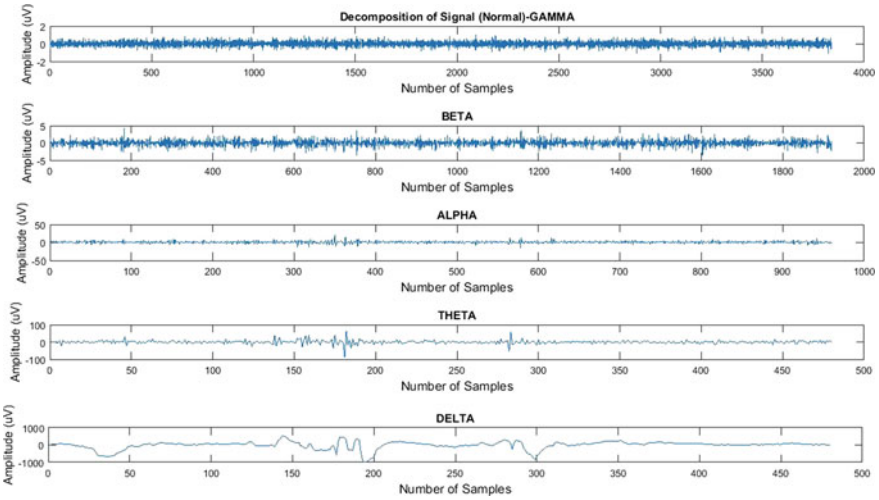


Fig. 3. Decomposition of EEG signal of a normal person

Table 1. Feature values for different bands of normal (N1), seizure (S1), and migraine (M1)

	Peak magnitude			Mean		
	N1	S1	M1	N1	S1	M1
Gamma	886.4586	1.90E+03	3847.562	−2.17E−17	2.41E−17	−2.2E−17
Beta	1.10E+03	2.06E+03	3581.173	4.87E−17	−6.26E−17	−3.2E−17
Alpha	1.83E+03	9.12E+03	6136.559	3.38E−17	−8.10E−18	3.67E−18
Theta	2.17E+03	6.85E+03	8985.86	−1.02E−16	−1.83E−16	−8.3E−17
Delta	7.28E+04	7.29E+04	98103.15	−2.54E−15	−6.65E−14	−1.1E−14
	Standard deviation			Skewness		
	N1	S1	M1	N1	S1	M1
Gamma	1.6509	1.64	2.179935	0.0254	0.2403	−1.76725
Beta	2.2801	3.7914	8.494449	−0.0099	−0.023	4.731707
Alpha	2.9408	7.938	12.35543	0.081	−0.0042	−3.13439
Theta	2.8393	9.763	11.28251	0.0842	−0.1952	1.009319
Delta	21.7924	52.6295	62.42071	−0.4455	0.0257	1.876412
	Kurtosis					
	N1	S1	M1			
Gamma	3.437	37.2684	190.2746			
Beta	3.7995	4.4581	632.8433			
Alpha	4.2053	4.4456	304.4211			
Theta	5.3264	7.4164	37.7896			
Delta	2.8727	9.6291	16.35753			

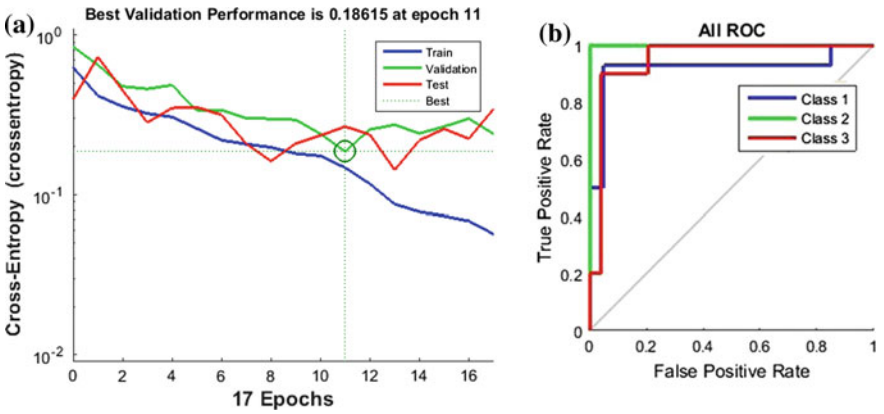


Fig. 4. EEG signal analysis for detection of brain abnormalities, **a** performance plot **b** ROC plot of EEG analysis

Figure 5 represents the classification result for the detection of brain abnormalities. From the figure, it is observed that class M (migraine) has been classified properly with one error. Similarly, class S (seizure) has classified without error and class N (normal) gives 4 misclassifications. Recognition rate for above confusion matrix is 85.29%, sensitivity 80.95%, and specificity of 85.64% for a dataset size of 37.

Fig. 5. Confusion matrixes for test samples, **a** normal, **b** seizure, and **c** migraine

Initially, all 28 channels with different sampling rates are considered for finding the accuracy of the system. Then average features from 4 segments for all 28 channels are considered as input to the NN network and recognition rate is calculated. Figure 6 represents the recognition rate for different channels considering all segments at once. From the figure, it is noticed that A1 gives the best recognition rate for migraine and T3 gives best recognition rate for seizure detection.

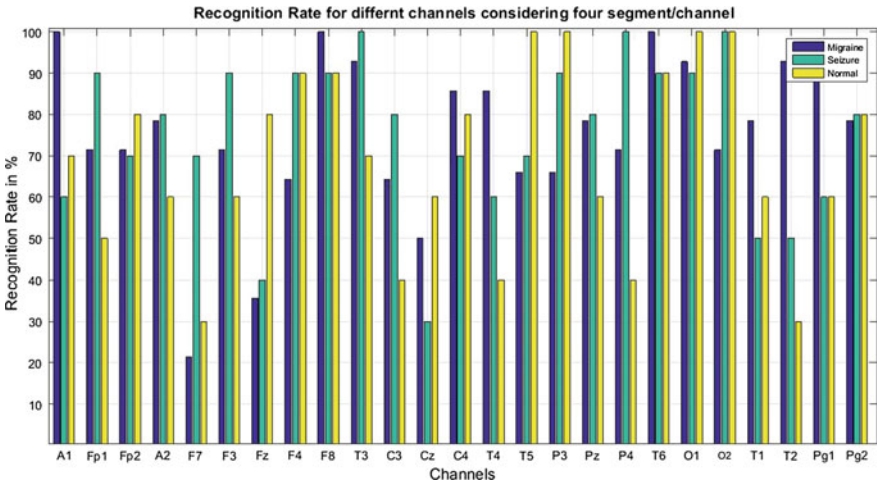


Fig. 6. Recognition rate for different channels of the EEG recording considering all segments

Table 2 shows recognition rate obtained by using six channels. Six out of 28 channels are selected which gives good recognition accuracy. Six channels considered are A1, F8, T3, T6, O1, and O2. From the table, it is noticed that A1 (Channel 2), F8 (Channel 10), T6 (Channel 20) gives 100% accuracy for migraine, and T3 (Channel 11), O2 (Channel 22) gives 100% accuracy for seizure. Classification performance with using only 6 channels gives a sensitivity of 89.23%, specificity of 90.7%, and accuracy of 89.96%.

Table 2. Recognition rate of different channels for detection of brain abnormalities using statistical moments

	A1 (channel 2)	F8 (channel 10)	T3 (channel 11)	T6 (channel 20)	O1 (channel 21)	O2 (channel 22)
Migraine	100	100	92.8	100	92.8	92.8
Seizure	60	90	100	90	90	100

5 Conclusion

The paper proposes an algorithm which can be used for the detection of the most common brain disorder: seizure and migraine. The EEG signal was first filtered with a cutoff frequency of 60 Hz to remove noise and then segmented for easy analysis. After preprocessing of the signal, DWT decomposition was done for the extraction of EEG signal bands. The statistical moment parameters on wavelet coefficients are calculated to reduce features. The moment features are given to the NN classifier to classify the signal as normal, seizure, or migraine. The proposed work has been tested on 10 normal persons, 14 migraine patients, and 13 seizure patients with a sampling rate of

128 Hz, 10 seizure patients with a sampling rate of 1024 Hz. Electrodes A1, F8, T3, T6, O1, and O2 are used for the detection of seizure and migraine instead of 28 channels. The average classification accuracy of 95% is achieved on dataset size of 47. The algorithm can be used to detect brain abnormalities such as autism, sleep disorder, etc., by selecting proper channels and with proper training of neural network.

Acknowledgements. We wish to thank Neurology Department, Ramaiah Memorial Hospital for providing us with old recorded EEG data for the above-conducted experiments.

References

1. Rajya Lakshmi M, Prasad TV, Chandra Prakash V (2014) Survey on EEG signal processing methods. Int J Adv Res Comput Sci Softw Eng 4(1). ISSN: 2277 128
2. Md Mursalin M, Zhang Y, Chen Y, Chawla NV (2017) Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier. Neurocomputing 241:204–214
3. Liu Y, Zhou W, Yuan Q, Chen S (2012) Automatic seizure detection using wavelet transform and SVM in long-term intracranial EEG. IEEE Tran Neural Syst Rehabil Eng 20(6)
4. Cao Z-H, Li-Wei K, Lai K-L, Huang S-B, Wang S-J, Lin C-T (2015) Classification of migraine stages based on resting state EEG power. IEEE. 978-1-4799-1959-8/15
5. Zandi AS, Javidan M, Tafreshi R (2010) Automated real-time Epileptic seizure detection in scalp EEG recordings using an algorithm based on wavelet packet transform. IEEE Trans Biomed Eng 57(7)
6. Batuhan Akben S, Tuncel D, Alkan A (2016) Classification of multi-channel EEG signals for migraine detection. Biomed Res 27(3):743–748
7. Subasi A, Kemal Kiyamik M, Alkan A, Koklukaya E (2005) Neural network classification of EEG signals by using AR with MLE preprocessing for epileptic seizure detection. Math Comput Appl 10(1):57–70

Measurement of Walking Speed from EMG Signal using Kurtosis of Approximate Coefficients

Habib Masum[1(✉)], Surajit Chattopadhyay[2], Ranjit Ray[3], and Subhasis Bhaumik[1]

[1] Aerospace Engineering and Applied Mechanics Department, IIEST, Shibpur, WB, India
kingmasum@gmail.com
[2] Department of Electrical Engineering, GKCIET, Malda, WB, India
[3] Robotics and Automation Group, CSIR-CMERI, Durgapur, WB, India

1 Introduction

Revealing of EMG signals and its analysis are becoming a very significant advancement in biomedical engineering not only for diagnosis of neurological and neuromuscular problems, also to monitor health of muscle, determine the activation timing, estimate joint angle, measure force and torque produced by it, hardware implementations, and control of multifunctional prostheses [1–3]. Therefore, a continuous advancement has been observed in this field through several decades. In 1849, Dubios–Raymond revealed the electrical activity of contraction of a voluntary muscle, but in 1890, the maiden recording was made by Marey and he introduced the word electromyography [4]. Initially, rough information could only be found from the EMG signals due to its stochastic behavior. In early 1970s, Hardyck and his group successfully utilized clinically surface EMG in more organized form for the treatment. The intramuscular electrodes provide a higher electromyogram but uses are restricted to deep muscles only [5]. The processing of EMG signals is mainly concerned with rectification of signals, its integration over a specified period and subsequently developing a time series of the integrated values and analysis in the domain of time–frequency. Fang et al. developed a technique using spectrum matching based wavelet transform for the segmentation of the EMG signal and to classify their single motor unit (SMU) potentials [6]. The classification algorithms has been applied in EMG signal denoising where the process was consisting of decomposition of signals, followed by applying proper thresholds to the detail coefficients, zeroing all coefficients below their associated thresholds, and then reforming the denoised signals considering modified detail coefficients [7]. Kurtosis values and Negative entropy have been utilized for computing the non-Gaussianity in Surface Electromyogram signal from three different fingers and wrist [8]. The decomposition of electromyographic (EMG) signals has also been done using communication technical interpretation which carries the information about the firing of different muscle fiber groups [9]. Discrete cosine transform (DCT)-based feature extraction technique from the EMG signals has been proposed for classification of normal event and a

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_35

neuromuscular disease [10]. The discrete wavelet transform and artificial neural network (ANN) have been proposed in combination for pattern recognition of EMG signal to operate the prosthetic hand [11].

However, fewer attempts have been made on the lower limb EMG signal analysis compared to upper limb. Therefore, still there are vast unattempted areas of EMG signal analysis. This has motivated authors to deal with EMG signals of leg muscles by applying the tool of discrete wavelet transform to assess walking speeds. In the author's previous work [12], the skewness-based discrete wavelet transform tool has been applied successfully on force data of human gait. At the end, an effective comparison has been made leading to specific outcome with sensor-based optimization.

2 EMG Signal Acquisition of VM, TA, GnL

Many of body muscles of human are responsible for locomotion like walking, running, sitting, standing, stair ascending–descending, etc. [13]. The electromyography refers the technique of monitoring electrical activity of muscles. The EMG signals are basically electrical currents produced in muscles at the time of its contraction or relaxation. It is not always an easy task to choose most suitable muscle(s) for capturing EMG signal(s) which is responsible for particular mode of locomotion for individual as the signals are influenced by anatomical and physiological construction of muscles and there are numbers of muscles overlaps each other's [14]. The surface EMG electrodes fit on skin, collect signal from different motor units simultaneously, which increases the noise level in captured signals. Apart from that, the placement of electrodes and holding it in place during movement, is also a complicated job. Therefore, keeping all these critical issues associated with EMG extraction, the following three muscles of lower leg have been selected. These are Vastus Medialis (VM), Tibialis Anterior (TA), and Gastrocnemius Lateral Head (GnL) as shown in Fig. 1. These muscles respond well with comparatively low level of noise.

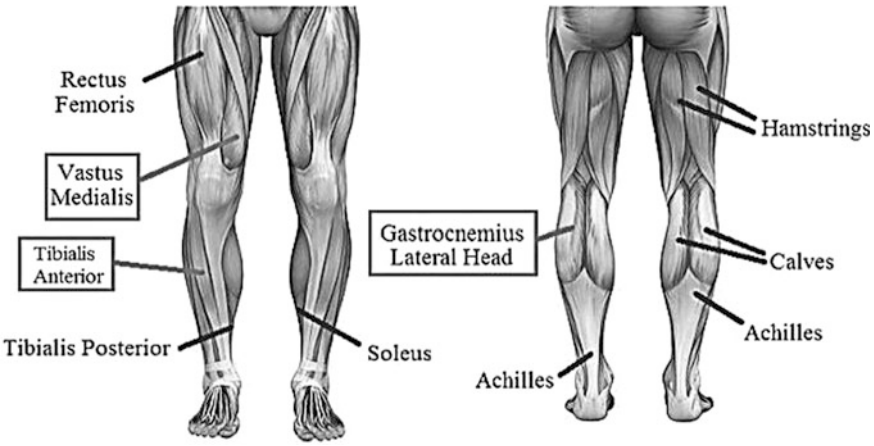


Fig. 1. Muscles of lower extremity [14]

The signals have been captured through Laboratory EMG Systems of Biometrics Ltd. which is a powerful data acquisition tool with real time display connected through DataLINK. The sEMG (surface electromyography) sensors and systems are used for movement analysis and biofeedback. It consists of bipolar, differential EMG Amplifier (SX230FW) having variable electrode distance that uses disposable electrodes and an online programmable Data Acquisition System, DataLINK (DLK900), which allows user to collect both analog and digital data. Signals from six different walking speeds (2–7 kmph with interval of 1 kmph) of 17 normal objects (age of 23–35 years and height of 161–173 cm) are collected as shown in Fig. 2. All walks were level walking on motorized treadmill.

Fig. 2. Laboratory trial of EMG signal acquisition

3 Assessment of Walking Speeds using EMG Signal in Time Domain

The EMG signal is biomedical signal, typically a function of time and can vary in its amplitude. Few samples collected from three muscles of left leg, namely Vastus Medialis muscle, Tibialis Anterior muscle, and Gastrocnemius Lateral Head muscle are shown in Fig. 3. All of them are of periodic and non-stationary in nature.

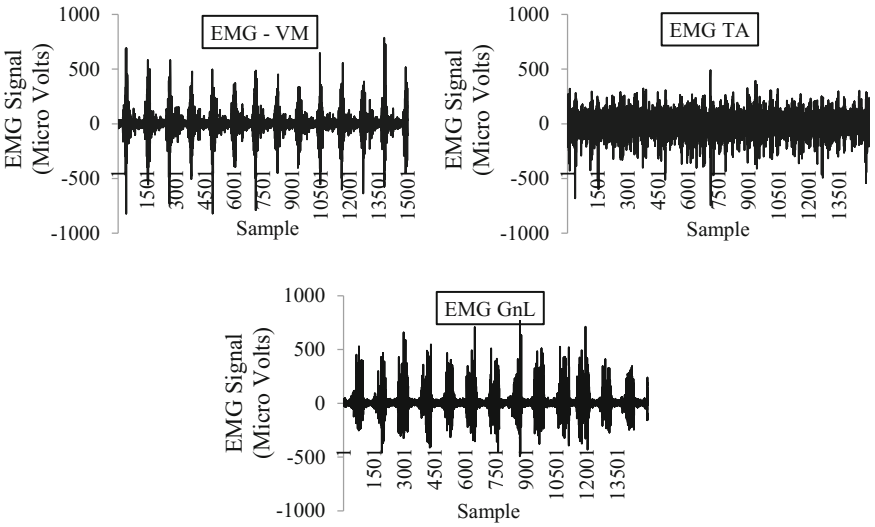


Fig. 3. Raw EMG signals captured from VM, TA, and GnL

4 Assessment of Speeds using Kurtosis of Approximate Coefficient from EMG Signal

Kurtosis is a measure of the "tailedness" of the probability distribution of univariate data. The Kurtosis is defined as

$$\text{Kurtosis} = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})^4 / N}{S^4}$$

where Y_1, Y_2, …, Y_N, are the univariate data [15]

As the biomedical signals are nonstationary (samples presented in Fig. 3) and captured in digitized form, therefore, Kurtosis of DWT based approximated coefficients can have an option for analysis of the signals. The EMG signal acquisition and analysis process are presented by blockdiagram as shown in Fig. 4.

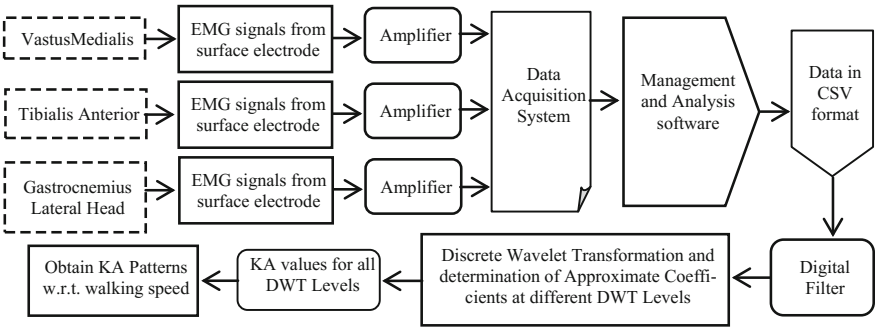


Fig. 4. Block diagram for gait data capture and analysis

4.1 Determination of Kurtosis of Approximate Coefficients (KA)

The Kurtosis values of approximate coefficients (KA) related to different walking speeds have been determined up to DWT level 9 for all three muscles. The KA values with respect to DWT levels have been plotted to observe its variation and it has been seen that the magnitudes, in initial DWT levels, are almost unchanged whereas, in later levels, change immensely. The variations are depicted in Fig. 5. Therefore, to find relation of KA value with walking speeds, the values corresponding to each DWT level have been plotted again against walking speeds as shown in Fig. 6.

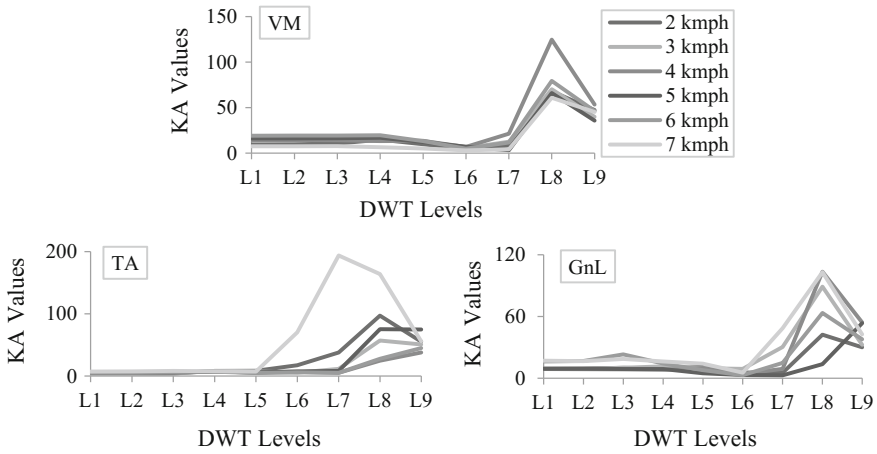


Fig. 5. Variations of KA values with DWT levels

4.2 Optimization of Level of Decomposition for Speed Measurement

Figure 6 shows that every DWT levels has definite pattern of variation. The patterns correspond to few levels are identical to each other and remaining have unique in

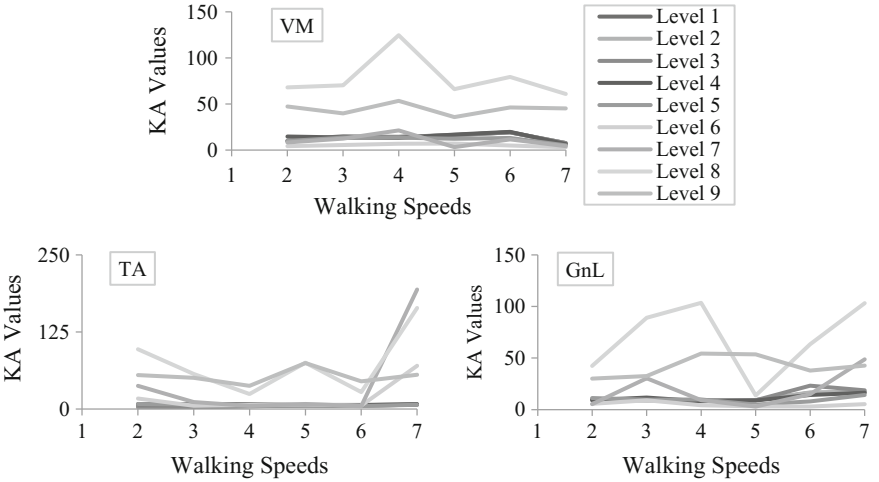


Fig. 6. Variations of KA values with respect to walking speeds

nature. Another observation has been made that the patterns are either zigzag or regular in nature. The gradually varying regular patterns are following approximate linear natures which are useful for assessment. Thus, three numbers of most regular patterns, as shown in Fig. 7, corresponds to each muscle have been selected for scaling in unknown walking speed measurement.

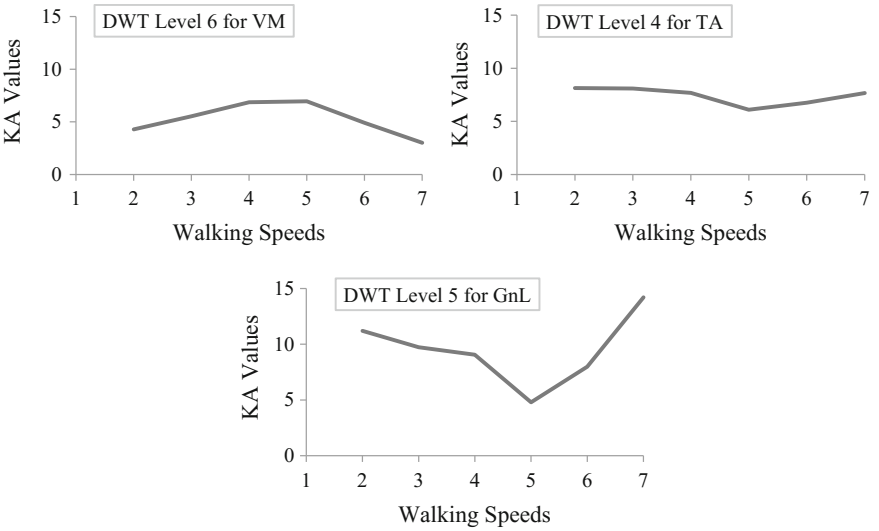


Fig. 7. Selected patterns correspond to each muscle

The selected patterns indicate steady variations of coefficient values with the change of walking speed. It is very significant that in all three cases, the values are changing gradually up to 5 kmph and then following reverse slope. For Vastus Medialis muscle, the KA values increasing with the walking speeds up to 5 kmph and the decreasing whereas for Tibialis Anterior muscle and Gastrocnemius Lateral Head muscle, nature is exactly reverse. These outcomes indicate that the natures of signals of two muscles, i.e., Tibialis Anterior and Gastrocnemius Lateral Head, are similar from the point of KA values and opposite to the nature of signal of Vastus Medialis. This finding may also be used to identify muscle from its signal.

5 Algorithm

For measurement of unknown walking speeds, an algorithm has been suggested based on nature came from optimized levels of decomposition, as shown in Fig. 7:

(a) Acquire EMG signals from vastus medialis muscle, tibialis anterior muscle, and gastrocnemius lateral head muscle, as shown in Fig. 1, using standard data acquisition kit.
(b) Digitize the signals.
(c) Pass the noisy signals through digital filter for denoising.
(d) Decompose the denoised signals up to level 9 and find approximate coefficients for each level.
(e) Determine kurtosis value of each approximate coefficients.
(f) Make pattern of kurtosis values w.r.t. walking speed for every DWT level.
(g) Select the most regular pattern to scale and measure the unknown walking speed using the curve.

6 Validation of Algorithm based on Optimized Level

The validation of the proposed algorithm is necessary for further utilization. This has been done by measuring unknown speeds of the individual. For the same, the EMG signals from the same muscles have been captured from treadmill walking and the speedometer reading has been noted for further comparison. The Kurtosis of approximate coefficient has been calculated applying the proposed algorithm and compared it with the reading of speedometer. All three curves produce satisfactory results with very close accuracy among which the curve correspond to vastus medialis muscle give result more close to the actual. The calculated speeds of unknown walking speeds of different objects are presented in Fig. 8.

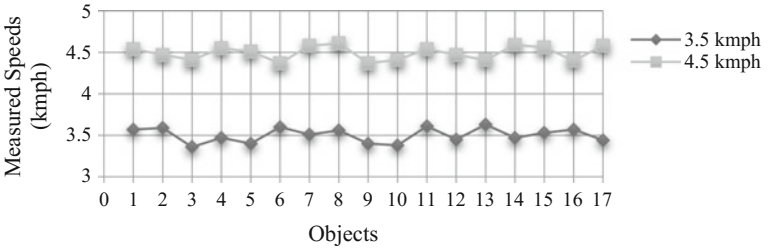


Fig. 8. Calculated speeds of unknown walking speeds of different objects

7 Conclusions

In this paper, the speed of human walking has been assessed from EMG signals using Kurtosis of approximate coefficients-based wavelet transform. The signals have been captured from three muscles of lower extremity of normal individuals during level walking at different speeds. The collected signals have been denoised as the signals were heavily noised. The denoised EMG data have been decomposed at various DWT levels for feature extraction and subsequently Kurtosis values have been determined. The patterns of change of Kurtosis values with respect to walking speeds corresponding to individual DWT levels have been studied and compared. Results demonstrated many definite relations between walking speeds and Kurtosis values. Most regular patterns have been utilized for scaling. Finally, an algorithm has developed for assessment of unknown walking speed data and subsequently validated. EMG sensor-based different scheme for monitoring health parameters of disable persons like muscle strength, response, etc., are becoming familiar in biomedical image processing. The method proposed in this work is adding another effective means of monitoring walking speed using same type of sensors which incorporates both time–frequency domain analysis. Therefore, the method seems a little bit complex; however, it can be used in a compact and robust way with other existing schemes for monitoring health parameters. The work can be extended to inter-link the walking speed features with those parameters and can be utilized in closed loop prosthetic control having EMG sensory feedback.

Acknowledgements. The authors acknowledge the participation and support extended by the subjects during the laboratory trials at GKCIET, Malda.

References

1. Liu YH, Huang HP (2009) Towards a high-stability EMG recognition system for prosthesis control: a one-class classification based non-target EMG pattern filtering scheme. In: Proceedings of IEEE international conference on systems, man, and cybernetics, pp 4752–4757. IEEE, San Antonio, TX, USA

2. Stefano AD, Burridge JH, Yule VT, Allen R (2004) Effect of gait cycle selection on EMG analysis during walking in adults and children with gait pathology. J. Gait Posture 20(1): 92–101
3. Saito H, Watanabe T, Arifin A (2009) Ankle and knee joint angle measurements during gait with wearable sensor system for rehabilitation. In: Dössel O, Schlegel WC (eds) Proceedings of IFMBE WC 2009, vol 25(9), pp 506–509
4. Cram JR, Kasman GS, Holtz J (1998) Introduction to surface electromyography. Aspen Publishers Inc., Gaithersburg, Maryland
5. Kleissen R, Buurke J, Harlaar J, Zilvold G (1998) Electromyography in the biomechanical analysis of human movement and its clinical application. J Gait Posture 8(2):143–158
6. Fang J, Agarwal GC, Shahani BT (1997) Decomposition of EMG signal by wavelet spectrum matching. In: Proceedings of the 19th international conference of the IEEE, engineering in medicine and biology society, IEEE/EMBS, vol 3, pp 1253–1256. Chicago, IL, USA
7. Zhang X, Wang Y, Han RPS (2010) Wavelet transform theory and its application in EMG signal processing. In: Seventh international conference on fuzzy systems and knowledge discovery (FSKD), pp. 2238–2238. IEEE, Yantai, China
8. Naik GR, Kumar DK, Arjunan SP (2011) Kurtosis and Negentropy investigation of Myo electric signals during different MVCs. In: Proceedings of IEEE ISSNIP biosignals and biorobotics conference biosignals and robotics for better and safer living, (BRC), pp 40–43. Vitoria, Brazil
9. Gut R, Moschytz GS (2000) High precision EMG signal decomposition using communication techniques. IEEE Trans Signal Process 48(9):2487–2494
10. Doulah AB, Fattah SA, Zhu WP, Ahmad MO (2014) DCT domain feature extraction scheme based on motor unit action potential of EMG signal for neuromuscular disease classification. IET Healthcare Technol Lett 1(1):26–31
11. Arozi M, Putri FT, Ariyanto M, Caesarendra W, Widyotriatmo A, Munadi Setiawan JD (2016) Electromyography (EMG) signal recognition using combined discrete wavelet transform based on artificial neural network (ANN). In: International conference of industrial, mechanical, electrical, and chemical engineering (ICIMECE), pp. 95–99. IEEE, Yogyakarta, Indonesia
12. Masum H, Chattopadhyay S, Bhaumik S, Ray R (2016) Utilisation of skewness of wavelet-based approximate coefficient in walking speed assessment. IET Sci Meas Technol 10(8):977–982
13. Knudson D (2007) Fundamentals of biomechanics, 2nd edn. Springer
14. Drake R, Vogl AW, Mitchell A, Tibbitts R, Richardson P (2014) Gray's Atlas of anatomy, 2nd edn. Churchill Livingstone, Elsevier
15. Chattopadhyay S, Mitra M, Sengupta S (2011) Electric power quality, Springer

GAE: A Novel Approach for Software Workflow Improvement by Unhidding Hidden Transactions

Shashank Sharma(✉) and Sumit Srivastava

Manipal University Jaipur, Jaipur, India
shashanksharmaaa@gmail.com

1 Introduction

Workflow mining a.k.a. process mining is a considerably new and emerging area of academic research within data analytics. The key objective here is to deploy workflow-related data in the direction to obtain pertinent info and knowledge by employing data analytic algorithms and determining a workflow model. This section discusses the concept of how an event log is the foundation of exploration along with other main building blocks of process mining. Work process mining articles to course the hole between huge information examination and traditional business work process/process administration. This field can fundamentally be categorized into (1) workflow revelation, (2) conformance checking, and (3) upgrade [1]. This permits the extraction of experiences about the by and large and internal conduct contained in any given procedure. Work process disclosure strategies underline on utilizing the occasion information in order to decide work process models. Conformance checking procedures underline on supporting the occasion information on a work process model to confirm how well the model fits the information and the other way around [2]. Despite the fact that increase strategies utilize occasion information and work process models to repair or expand the work process show. Thus, work process mining gives the course the crevice between information mining and machine learning hones and the business procedure administration teach.

1.1 The Event Log as Main Focus of Analysis

Data is a crucial building block in various discovery domains. Process Mining uses data that is accounted by event logs. A sample log is shown in Table 1. An event log in terms of this situation can be defined as the process of recording of an action instance on the system. Action or activity instances are units of work that are registered by the system when work is piloted in the situation of an assured process. Statuses of activity or action are specified to a set of languages that are fixed to the workflow modeling hypothesis. Various workflow modeling hypothesis have unique implementation standards that are governed by their state evolution diagrams. A simple state evolution diagram has been explained in Fig. 1. More elaborate diagrams have been described in process modeling literature. Consider an example that the Object Management Group

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_36

(OMG) characterizes the life cycle of an action in its Business Process Modeling Notation (BPMN) particular [3]. Adding to this, different process displaying ideal models, for example, YAWL (Yet Another Workflow Language), case taking care of [4] and revelatory methodologies (EM—BrA2CE [5], Declare [6]) which proposes positive however similar to base representations for their relating delineation semantics.

Table 1. An example of typical event log

Case ID	Activity name	Event type	Originator	Timestamp	Extra data
011	Make order form	Start	Employee A1	10-19-1955 15:15:02	….
012	Make order form	Complete	Employee A2	10-19-1955 05:14:01	….
….	….	….	….	….	….
….	….	….	….	….	….

It can get fascinating to experience different establishment semantics and state change charts, for work process mining and work process examination; notwithstanding, the real information is what is significant. Process data is generally pooled from various vaults that are gotten from CRM, ERP, WFM, and other different data frameworks. This outcome is in trouble while recognizing hypothetical establishment of an occasion as far as state progress charts and genuine information found practically speaking. In various sectors such as CRM (customer relationship management), product development, financial services, etc., business workflow depends heavily on legacy information systems or low workflow-oriented information systems. Further, the registered business process data is well defined such that only a certain type of state transition (e.g. accomplishment of an activity occurrence) can be mined via. actual data blocks. Workflow mining can be considered most fruitful in modular conditions where business information systems have broader options of behavior, here; the transition of accessible data into an event log is most often a significant task.

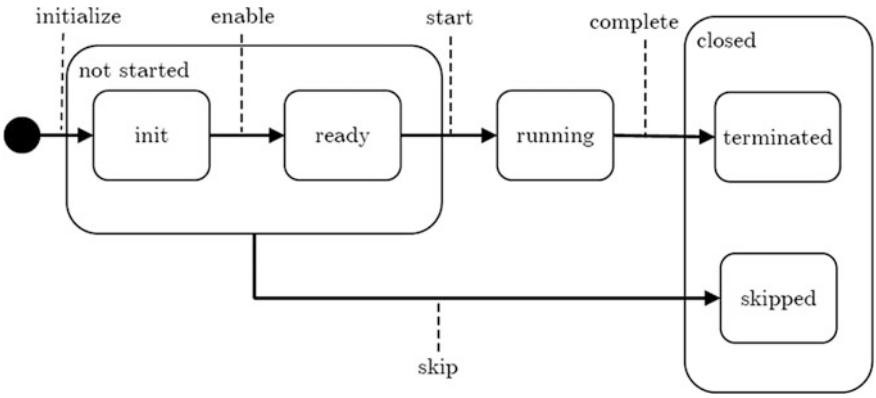


Fig. 1. A basic state transition diagram

Filtration and extraction of event logs through business information systems is usually carried out by text-based data scripts. In reality, process data is more often distributed over various data sources and it is painstaking to define the precise scope of process that is studied. Further, a comprehensive ETL-phase is required before a concrete analysis is initiated. Adding to which, the data must be in an event log storage format.

One of the underlying ways to deal with putting away log in view of occasion performed/executed is in the MXML design (Mining Extensible Markup Language). Since 2003, the MXML organize has been utilized as the reality-based (de facto) standard since it is exceedingly inter-ground with the ProM-structure which is a system utilized for scholarly reason for work process change. It was not up to this point IEEE team was favored rather than MXML for work process mining. This new configuration, XES (eXtensible Event Stream) enhances the first standard, since it is less prohibitive. The meta-model of XES is portrayed in [7].

As occasion logs are the establishment of the approach of workflow mining, it is significant to express the variable necessities to which an occasion log must approve. After taking these three suppositions are required and basic:

- Activity instance must be well defined for the workflow instance of an event which specified by a unique activity name
- Unique process instances or ID which are used by cases must be referred by an events
- Ordered Timestamp must be recorded by the events.

2 Literature Review

As Agrawal et al. [8] and Pinter and Golani [9] are the fundamental ones that don't explicitly get the possibility of the split/join centers in the mined models. The reason is that they concentrate on a model for the Flow check work structure [10] and each point in this system has an OR-part/join semantics. In fact, each organized round segment in the model has a Boolean limit that surveys to real or false after an errand is executed. The appraisal of the Boolean conditions sets what number of branches are established after an endeavor is executed. Cook et al. [11] have the primary approach that does not concentrate on a whole mined model. Their approach looks for the most progressive cases in the model. In actuality, all over they do mine a whole procedure appear, however that is not their essential point. The build ups that can't be mined by all frameworks are circles, sans non-choice, imperceptible assignments moreover, duplicate endeavors. Grecco et al. [12] can't mine any kind of circles. The reason is that they show that the models their computations delve think about as small extra lead (that is not in the event log) as could be normal the situation being what it is. They do as such by posting each one of the takes after that the mined model can deliver and differentiating them and the follows in the event log. Models with circles would make this endeavor outlandish. Some different procedures can't mine self-assertive loops in light of the fact that their model documentation (or portrayal) does not bolster this sort of

loops. The fundamental motivation behind why most strategies can't mine non-neighborhood sans non-decision is that the majority of their mining calculations depend on neighborhood data in the logs. The systems that don't mine nearby without non-decision can't do as such on the grounds that their portrayal does not bolster such a develop. Generally the system depends on a piece organized documentation, as Herbst et al. also, Schimm. Skip errands are not mined in light of depiction limitations as well. Split/join imperceptible endeavors are not mined by various techniques, except for Schimm and Herbst et al. Actually, we similarly don't center at finding such kind of assignments. Nevertheless, it is routinely the case that it is possible to make an exhibit with no split/join imperceptible errands that communicates a comparable direct as in the model with the split/join subtle assignments. Duplicate endeavors are certainly not mined in light of the way that various methods expect that the mapping between the endeavors what's more, their imprints is injective. By the day's end, the names are exceptional per task. The principal strategies that mine duplicate endeavors are Cook et al. [11] for progressive structures just, and Herbst et al. [13] for both back-to-back and parallel techniques. We don't consider Schimm [4] to mine technique models with duplicate errands since his approach expect that the revelation of the duplicate endeavors is done in a preplanning step. This movement identifies every one of the duplicates and guarantees that they have unique identifiers when the event log is given as commitment to the mining calculation. As a matter of fact, every procedure that we survey here would handle copy undertakings if this same pre-preparing step would be done before the sign in and is given as contribution to them.

3 Proposed Algorithm with Procedure with a Sample Case Study

With a specific end goal to apply a hereditary calculation, we have to speak to people. Every individual compares to a conceivable procedure model and its portrayal ought to be anything but difficult to deal with. Our underlying thought was to speak to forms straightforwardly by Petri nets. Tragically, Petri nets end up being a less helpful way to speak to forms in this specific circumstance. So to overcome this problem we use causal matrix [3, 14].

Table 2. Concise and concrete encoding format of an individual in Table 2

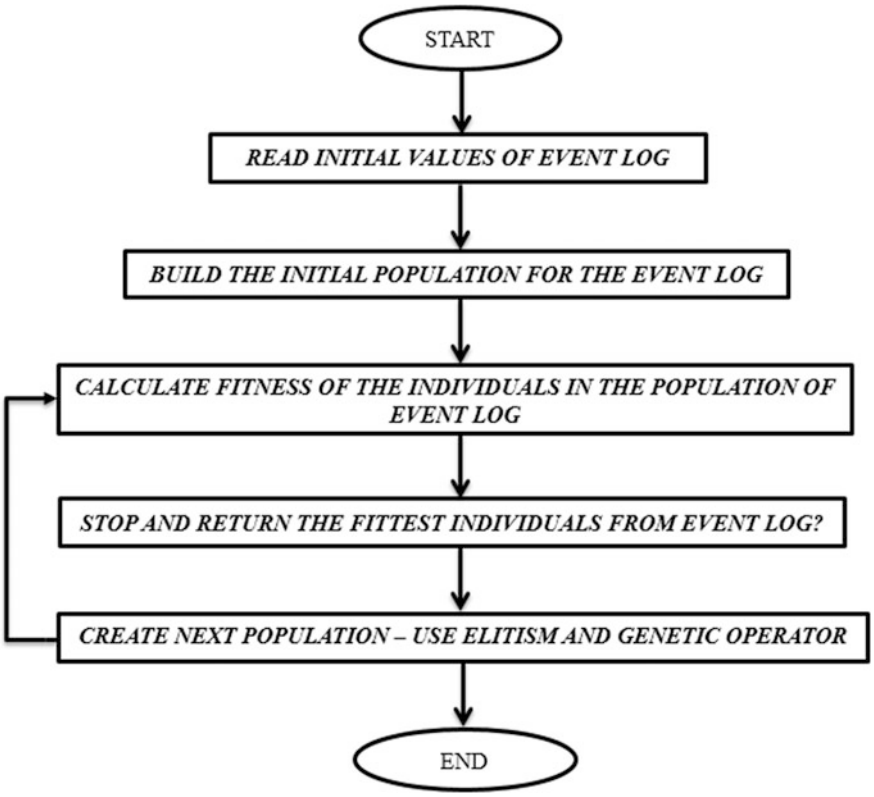
Activity	Input	Output
A	{}	{{*B*, *C*, *D*}}
B	{{*A*}}	{{*H*}}
C	{{*A*}}	{{*H*}}
D	{{*A*}}	{{*E*}}, {{*F*}}
E	{{*D*}}	{{*G*}}
F	{{*D*}}	{{*G*}}
G	{{*E*}}, {{*F*}}	{{*H*}}
H	{{*B*, *C*, *G*}}	{}

A matrix of causality is a row, X with elements named as (M, N, O, P) where

- M comprises of activities with type of finite sets,
- $N \subseteq M \times M$ is the relation of causality,
- $O \in A \rightarrow P(P(M))$ is the function of input condition type, 3
- $O \in A \rightarrow P(P(M))$ is the function of output condition type,

such that

- $N = \{(m1, m2) \in M \times M \mid m1 \in O(m2)\}$, 4
- $N = \{(m1, m2) \in M \times M \mid m2 \in P(m1)\}$,
- $\forall m \in M \; \forall \; QQsR \in O(m) \; Q \cap QL = \varnothing \Rightarrow Q = QL$,
- $\forall m \in M \; \forall \; Q, QR \in P(m) \; Q \cap QL = \varnothing \Rightarrow Q = QL$,
- $N \cup \{(mo, mi) \in M \times M \mid mo \; N\bullet = \varnothing \wedge N\bullet \; mi = \varnothing\}$ is a connected graph of strong type. The proposed process efficient GAE (Genetic Algorithm for Events) is based on GA and process mining of event logs. In this, fitness function is computed



Flowchart 1: Proposed algorithm for GAE

by the quality of an individual. The quality of an individual is basically set by its replaying of the log traces. This semantics permits us additionally to characterize an idea of wellness required for the hereditary calculations. Details regarding the calculations of the various parameters are beyond the scope of this paper. Utilizing the Petri net portrayal, we can play the "token amusement" to perceive how every occasion follow in the log fits the individual spoke to by a causal network. The entire flow of GAE is shown in Flowchart 1. The steps of the proposed algorithm are as follows:

In this, first we take the event log of hospital case of healthcare information system. In this particular event log we have 42 events or events flow. Our aim is to classify the events on the basis of type of events. We start with event log then we select MXML legacy classifier and process discovery algorithm for the extraction of Petri net. Along with this, we apply some user-specified constraints to get expected result (Fig. 2, Table 3).

Table 3. Causal matrix made and utilized for portrayal

Input									Output
	true	A	A	A	D	D	E ∧ F	B ∨ C ∨ C	
→	A	B	C	D	E	F	G	H	
A	0	1	1	1	0	0	0	0	B ∨ C ∨ D
B	0	0	0	0	0	0	0	1	H
C	0	0	0	0	0	0	0	1	H
D	0	0	0	0	1	1	0	0	E ∧ F
E	0	0	0	0	0	0	1	0	G
F	0	0	0	0	0	0	1	0	G
G	0	0	0	0	0	0	0	1	H
H	0	0	0	0	0	0	0	0	true

4 Comparison Analysis

For verifying and validating the effectiveness of GAE, we used the standard algorithm which are used for workflow management as well as analyze and compare it with proposed GAE. In figure number 5, you find black boxes (dark or complete black), which referred to hidden transactions, which shows that other or previous algorithms are unable to extract or locate hidden transactions although algorithms are able to classify the event log (Figs. 3, 4, 5 and 6, Table 4).

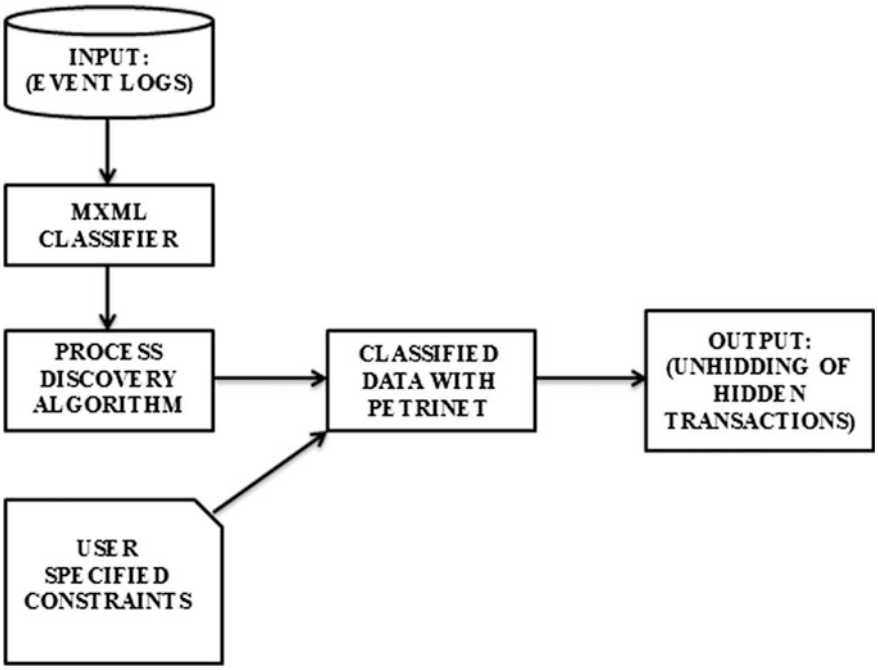


Fig. 2. Procedure for GAE

Table 4. Comparison of result for various algorithms

Name of algorithm	Input format	Output format	Intermediate output	Able to unhide hidden transactions
Alpha	MXML log file	Petri net	No	No
Alpha++	MXML log file	Petri net	No	No
Tsinghua-Alpha	MXML log file	Petri net	No	No
GAE	MXML log file	Petri net	Heuristic Net	Yes

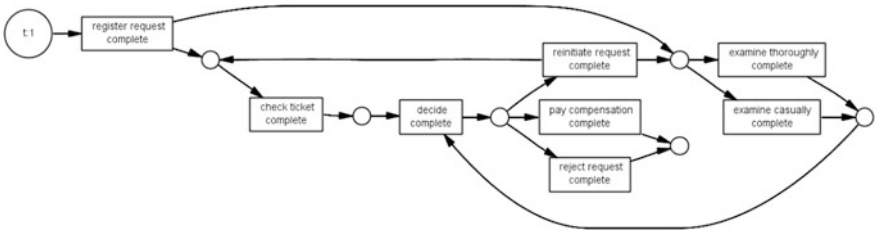


Fig. 3. Alpha algorithm

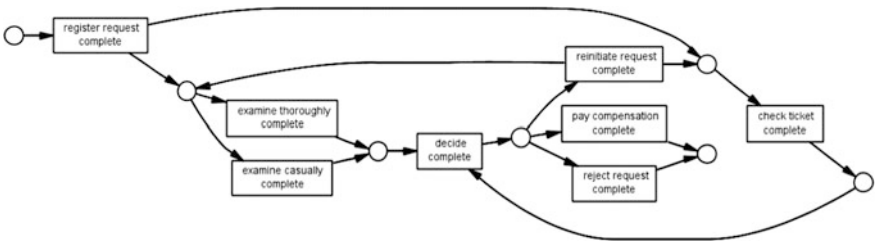


Fig. 4. Alpha++ algorithm

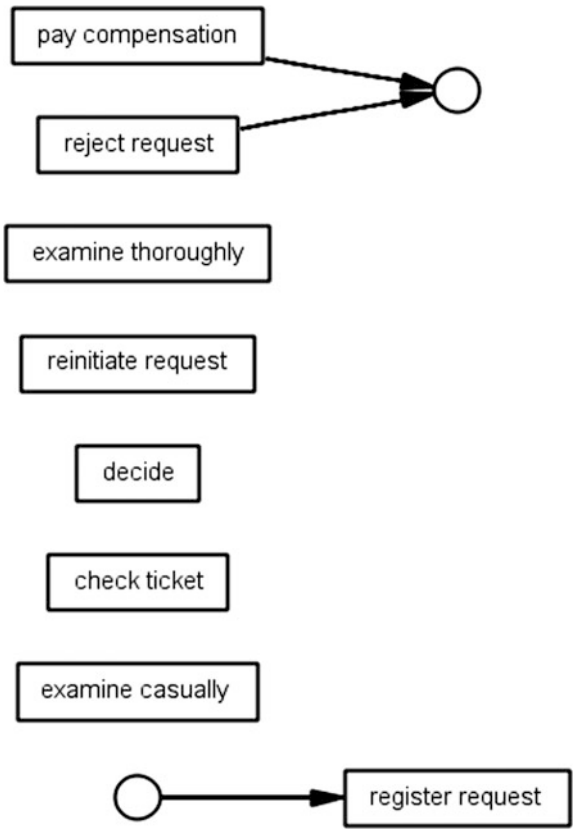


Fig. 5. Tsinghua–Alpha algorithm

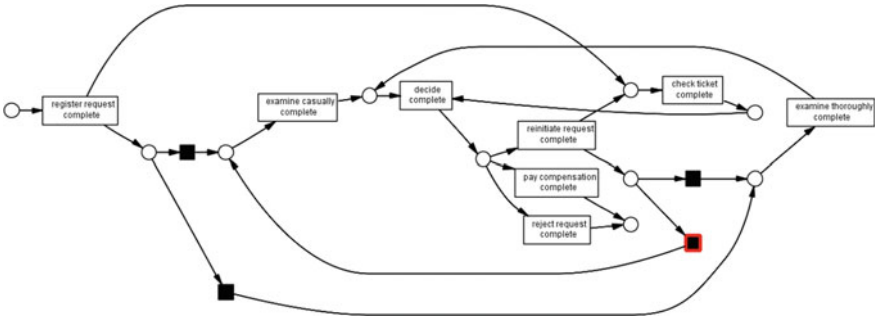


Fig. 6. GAE

5 Results and Outcomes

In this, we received classified result for that particular hospital case. We got 8 classified event classes. For this classification, we used 2 criteria, i.e., on the basis of event functionality and other one on the basis of event type. Along with this, complete classification is further subdivided into 3 parts, i.e., start event (count is 1), originators (count is 6) and end event (count is 2). After interpreting the Petri net, we find that which workflow model we need to work upon to improve the process of information system. Please find below the outcomes (Table 5).

Table 5. Classified event log data

Model element	Event type	Occurrences (absolute)	Occurrences (relative) (%)
Check ticket	Complete	9	21.429
Decide	Complete	9	21.429
Register request	Complete	6	14.286
Examine casually	Complete	6	14.286
Reinitiate request	Complete	3	7.143
Examine thoroughly	Complete	3	7.143
Pay compensation	Complete	3	7.143
Reject request	Complete	3	7.143

6 Conclusion and Future Work

"Workflow scientist" desires to possess particular/exact to initiate innovation in a progressively digitalized ecosphere. In this paper, we just only conceptualized our idea with a small case study. This allows us to analyze the operational process workflow of healthcare information systems under real-life scenarios, and use extraction procedures for processes to acquire specific and recognized software improvement models. This paper oriented on associating the organized mining approach for the event classes from event data for the software process improvement by using the Petri nets flow model

approach. In the future, we aim at conducting additional experiments using different variety of event log data sets. A reasonable succeeding phase is to progress with tool support for domain-based information management systems.

References

1. van der Aalst WMP (2011) Process mining: discovery, conformance and enhancement of business processes. Springer, Berlin
2. Adriansyah A, van Dongen BF, van der Aalst WMP (2011) Towards robust conformance checking. In: Business process management workshops. Lecture notes in business information processing, vol. 66, pp. 122–133. Springer, Berlin, Heidelberg
3. Dehnert J, van der Aalst WMP (2004) Bridging the gap between business models and workflow specifications. Int J Coop Inf Syst 13(3):289–332
4. Schimm G (2004) Mining exact models of concurrent workflows. Comput Ind 53(3):265–281
5. Goedertier S (2008) Declarative techniques for modeling and mining business processes. Phd thesis. Katholieke Universiteit Leuven, Faculty of Business and Economics, Leuven, Sept
6. Pesic M, van der Aalst WMP (2006) A declarative approach for flexible business processes management. In: Eder J, Dustdar S (eds), Business process management workshops. Lecture notes in computer science, vol 4103, pp 169–180. Springer
7. Günther CW XES standard definition. http://www.xes-standard.org
8. Agrawal R, Gunopulos D, Leymann F (1998) Mining process models from work logs. In: Ramos I, Alonso G, Schek H-J, Saltor F (eds), Advances in database technology—EDBT'98: sixth international conference on extending database technology. Lecture notes in computer science, vol 1377, pp 469–483
9. Pinter SS, Golani M (2004) Discovering workflow models from activities lifespans. Comput Ind 53(3):283–296
10. IBM (1999) IBM MQSeries workow—getting started with buildtime. IBM Deutschland Entwicklung GmbH, Boeblingen, Germany
11. Cook JE, Du Z, Liu C, Wolf AL (2004) Discovering models of behavior for concurrent workflows. Comput Ind 53(3):297–319
12. Greco G, Guzzo A, Pontieri L, Sacca D (2004) Mining expressive process models by clustering workflow traces. In: Dai H, Srikant R, Zhang C (eds), BIBLIOGRAPHY 365 PAKDD. Lecture notes in computer science, vol. 3056, pp 52–62. Springer
13. Herbst J, Karagiannis D (2004) Workow mining with InWoLvE. Comput Ind 53(3):245–264
14. Regev G, Soffer P, Schmidt R (2006) Taxonomy of flexibility in business processes, Input BPMDS'06 workshop

Lung Cancer Detection in CT Scans of Patients Using Image Processing and Machine Learning Technique

Karan Sharma[1(✉)], Harshil Soni[1], and Kushika Agarwal[2]

[1] System Level Solutions (I) Pvt. Ltd., V.U. Nagar 388121, Gujarat, India
{karan.sharma,hsoni}@slscorp.com
[2] Department of CSE, Sikkim Manipal Institute of Technology, Sikkim, India
kushikaagarwal@gmail.com

1 Introduction

Lung cancer strikes more than 1.8 million people every year and accounts for billions of dollars in healthcare costs [1]. Early detection is critical to give patients the best chance of recovery and survival. All accepted contemporary works and studies have suggested methods to improve detection of cancerous lesions if any present in lungs. Many of the existing concepts have described about feature extraction methods for training of machines [2], but a huge void exists in terms of processing and analysis of such voluminous data by using simple neural network and other machine learning algorithms like random forest, decision tree, etc. Also, to note that images occupy comparatively large space and again to use any data stored in very low level format like JPEG will result in loss of crucial information that is required for machine learning algorithms. Therefore, it is wise to make use of patient's CT data stored as DICOM (Digital Imaging and Communication in Medicine) file. DICOM is the standard to store and communicate datasets in medical science. One can have better access to CT scan images in uncompressed version and get access to important information regarding patient's demographics from DICOM [3]. Again, with all precautions about data format, an inevitable element about CT images still exist in the form of unwanted additional information of surrounding body organs and external noises due to the different scanning devices used. Such irregularities have been recognized as a possible source of error for tissue classification [4]. To overcome these challenges and to make lung cancer detection easy, the entire work is divided into two subgroups as: image preprocessing and machine learning.

Set of sequential methods implemented for data normalization and segmentation composes image preprocessing. Input data normalization is the first step aimed to bring images from various sources to a regular standard by resampling pixel data contained in DICOM file of patients. It is followed by segmentation required to depict the lung abnormal regions and boundaries of the lung from surrounding thoracic tissues. Threshold-based segmentation is the most basic, well understood, and effective

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_37

technique for obtaining segmentation from images with a well-defined contrast difference among the regions. This method segments the image by creating binary partitions that are based on image attenuation values and grouping together all the image elements to a region that satisfies the threshold interval. It is efficient as it takes only a few seconds to yield complete reproducible segmentation [5]. The coherency of the segmented image is not guaranteed and still may have holes, extraneous pixels, and noises. Region-growing segmentation is slightly advanced and effective method which serves as an efficient tool for extracting homogeneous regions with more precise lung segmentation results without false values. In this method, one pixel is compared with its neighboring pixel and if the predefined criterion (homogeneity) is met, then the pixel is said to belong to the same class as one or more of its neighbor. This method is useful for their efficiency and robustness in dealing with attenuation variations by reinforcing spatial neighborhood information and region criteria. Therefore, it is considered for segmentation and extracting pulmonary lung lesions in this paper. Other methods include shape-based segmentation which is modeled to give high accuracy, but are computationally expensive and performance highly depends on the feature set and training data. Two more segmentation methods being considered for clinical practice are neighboring anatomy-guided methods and machine learning-based methods. Both works well for specific conditions like when attenuation-matrices fails and to classify ground-glass opacity nodules, aided they require high computational power. To give a detailed idea of all methods is beyond the scope of this paper, but one can refer literature survey for more information on them [5, 6]. Finally, all preprocessed data is recorded along with a label set which contains labels for each input patient having cancer or not. Convolutional Neural Network (CNN) based on deep learning technique is used for training patients dataset. CNN having the advantage of both classification and prediction of input serves a robust and reliable algorithm to train against large datasets as CNN has N number of parameters to be manipulated for accurate training.

2 Materials and Methods

2.1 Image Preprocessing

2.1.1 DICOM and Hounsfield Units Conversion of Input Pixel Data

DICOM basically is a software standardization, which is used for image diagnosis and communication of image data using standard network protocol. The DICOM dataset of patients who undergo medical CT scanning serves as abundant amount of information related to his/her body organ which can effectively be utilized for image processing for diagnosis purpose. The biggest advantage to use this standard is the raw pixel image data along with other metadata about the patients and pixel values is present which further helps reduce a lot of computational complexity. Figure 1 shows the file view of a DICOM slice of a patient's lung. The crucial information about pixel dimension and Hounsfield conversion factors, that is, rescale slope and rescale intercept are some important attributes of DICOM.

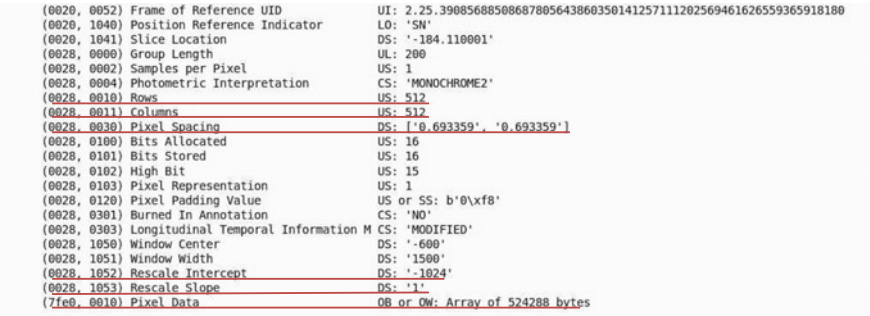


Fig. 1. A DICOM file representing a slice of patient's lungs

Attenuation coefficient of material describes the fraction of incident X-ray beam being scattered or absorbed by the material as the beam penetrates inside. This helps in determining the material thickness and the material itself [7].

The Hounsfield Unit (HU) scale is a linear transformation of the original linear attenuation coefficient measurement into one in which the radio density of distilled water at Standard Temperature and Pressure (STP) is defined as zero Hounsfield Units (HU), while the radio density of air at STP is defined as 1000 HU. This is a universally accepted standard for all the CT scan tomography scanners which determines the effect of attenuation coefficient on the input intensity [8]. In short, it is a scale for identifying material whose values are recorded as attenuation coefficient. The relation between HU values and the attenuation coefficient is a linear equation described as below

$$y = m * x + c \tag{1}$$

where y is Hounsfield Unit value of pixel attenuation coefficient x, m, and c represents rescale slope and rescale intercept shown in Fig. 1, respectively.

Thus, conversion of input attenuation coefficients to HU values is evident to understand the nature of what body composition is present in the scanned data and also to know how scattered is the region of interest, lung nodules in this case. The theoretically prepared Hounsfield units scales which apply to medical-grade CT scans is used as reference to plot the histogram plot for the HU values of input scan dataset of a patient.

2.1.2 Resampling of HU-Converted Values

HU-converted input values are to be resampled as because of difference in scanning method and devices used, input data of different patients show different pixel spacing values along the 3 axes. Thus, bringing these values to a uniform value assures that the dataset of various patients are uniform structures free from errors generated because of scanning methods. When resampling is done the density of pixel in a patient's data changes, thus bringing a change to attenuation values and corresponding HU values. Now, by choosing a standard new spacing value we can easily get uniformity within our datasets. In general, the standard value of new spacing along the 3 axes is chosen as

(1, 1, 1) and existing values are converted to this, the algorithm for n patients data is discussed as follows.

```
Begin
Declare two list variables Slices, Images and initialize them with list of DICOM slices of patients and their pixel data respectively
Declare 1D arrays pixel_spacing, new_pixel_spacing, resample_factor, new_shape and resize_factor
Initialize new_pixel_spacing with [1, 1, 1]
for (i = 0, slice = Slices[i], image = Images[i] to i = n) do
pixel_spacing = Pixel Spacing of slice
resample_factor = pixel_spacing/new_pixel_spacing
new_shape = Shape of image
resize_factor = new_pixel_spacing/Shape of Image
Resample all pixel data in image by resizing them with resize_factor
Until termination condition is false
End
```

2.1.3 Segmentation of Resampled Pixel Data

The pixel data array which gets constructed after scanning of patients and the images that we plot from such data are seen to have many unwanted portion of pixel values which is of no use for diagnosis, that is, those are because of the images being square and scanners being oval, therefore this unwanted portions like air, bones, hard tissue, etc., are of no interest in the diagnosis of patient lungs. Therefore, segmenting necessary lung portions reduces the complexity of our algorithm and makes us have the region of interest necessary for analysis. Segmentation comprises of thresholding, connected component analysis [9], and removal of unwanted region. Figure 2 illustrates the segmented pulmonary nodules of patient's lungs. Algorithm for region-based segmentation [10] is discussed as follows.

```
Begin
Declare and Initialize seed variable SEED with position of seed (x, y)
Declare counter RCOUNT, stack PX, BP to keep track of current region, store pixels to grow and boundary pixels of region grown respectively
Declare array REGION and CP to store labels of grown region and 8-neighbours of CP respectively
R, C = size of the image
SEED = (R/2, C/2)
RCOUNT = 1, i = 1, j = 1 and PX(i) = SEED
Set threshold value THR as -320
while(PG not empty) do
CP = PG(i) and i = i - 1
```

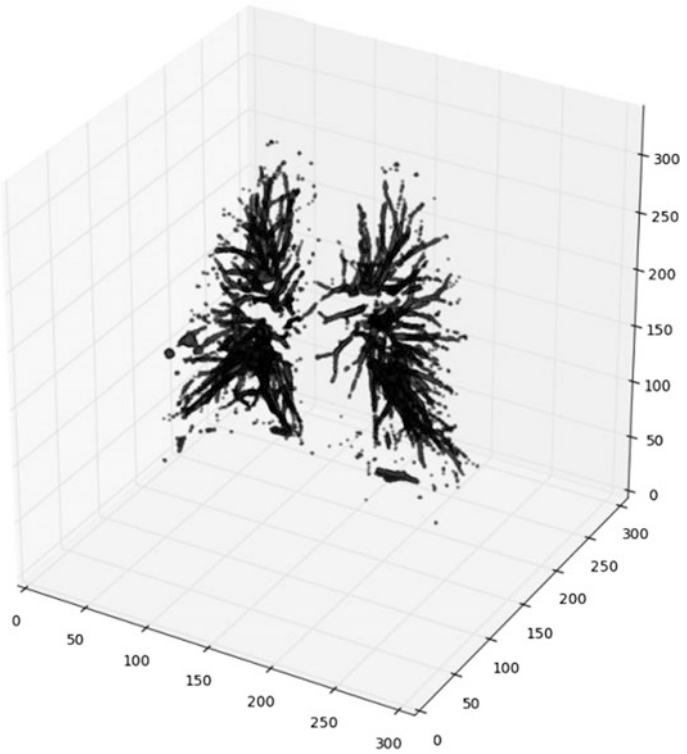


Fig. 2. Segmented pulmonary lung nodules

```
for (8-nb of CP, k = 1 to 8) do
If (REGION (CP(k)) is not labeled) then
Calculate the distance DIST of CP(k) from SEED
If (DIST > THR) then
REGION (CP(k)) = RCOUNT and i = i + 1
Else j = j + 1 and BP(j) = CP(k)
End for
End While
While (BP not empty) do
SEED = BP(j), j = j - 1, RCOUNT = RCOUNT - 1
i = 1 and PX(i) = SEED
Go to step 9
End while
End
```

2.1.4 Resizing of Images

The input dataset of each patient contains slices of resolution 512 × 512 and the number of slices for each one of them varies from 100 to 400 because of different

calibration of the instruments from which the scan was taken, this nonuniformity is highly undesirable as the construction of neural network to perform operation of feature extraction on the input data depends on the size of the matrix passed as the input. Otherwise, every patient will require constructing a new neural network, which contradicts the training of machine. Thus, we conclude that this nonuniformity is highly undesirable and can be handled by downsampling or upsampling the input data to a certain fixed value across the depth.

2.1.5 Standardization

The standardization or feature scaling is important to step up the processing of machine learning algorithm and enhance the time taken for classification of datasets. It basically helps us reach values which bring easy path of convergence of weight values in training the machine using neural networks, thus reducing the time to predict and classify the data. This can be easily achieved by using following equations

$$\rho = \omega - \mu \tag{2}$$

where ρ is the result of subtraction of segmented pixel values ω from their mean μ.

$$\rho = \rho \div \sigma \tag{3}$$

where σ is the standard deviation of segmented pixel values ω.

3 Machine Learning

To train a Convolutional Neural Network (CNN) for extracting features from the input images by implementing deep learning technique is the basis of image classification and prediction of data. The design of CNN is illustrated in Fig. 3 for a better understanding.

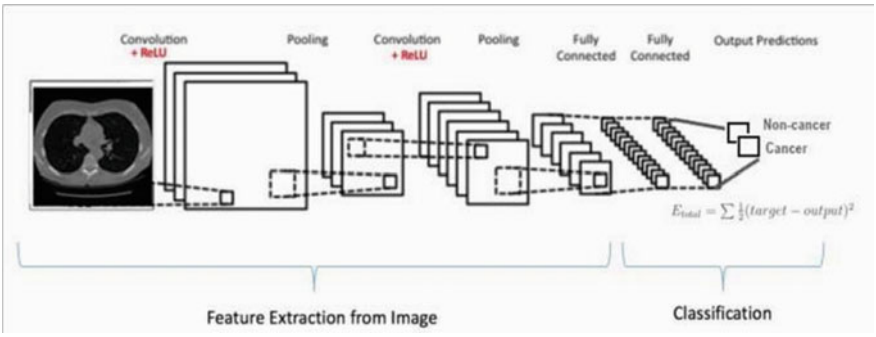


Fig. 3. Design of convolutional neural network [11]

The features extracted from convolutional part of CNN are used for prediction by passing them through a fully connected network and finally giving output with respective number of classes in the output layer of the network; two in this case. The reduction of loss function by changing weights using backpropagation algorithm supported with Adam optimizer for better convergence of weights is the real goal of training the network with correct weight values. The detailed understanding of CNN is necessary to bring it in use, and the study of CNN is very much limited in this paper as because, N number of parameters involved for tuning the constructed network to give accurate results. One can see the literature in references for further in-depth knowledge on CNN. The activation function is one key parameter of CNN which gives network nonlinearity aspect. The choice of activation function is independent, but results obtained by using Rectified Linear Unit (ReLU) are more promising. ReLU overcomes the limitation of traditional sigmoid and hyperbolic function, that is, the point of saturation which stops further change in weight values. Figure 4 shows ReLU curve.

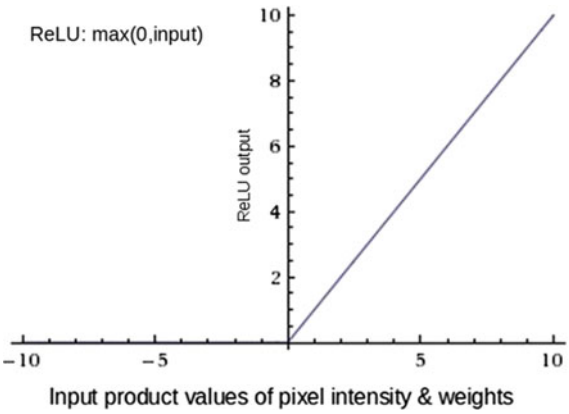


Fig. 4. Rectified linear unit curve

4 Experimental Results and Discussions

The proposed machine learning algorithm was tested on available dataset of 20 patients having cancer and no-cancer. The experimental results from training depict a gradual decrease in the loss function as shown in Fig. 5.

The performance evaluation of the proposed method resulted in detection accuracy of approximately 65%. The occurrence of lower accuracy is mainly due to the use of limited hidden layers in CNN computation. These layers are responsible for deep feature extraction depending upon the number of layers used, thus limiting to only 2 hidden layers might have failed to detect deep features accounting to cancer in the lung tissues. Also to mention that region-growing algorithm is good at detecting nodules in the lung region, but accurate extraction of nodules attached to lung walls, detection of

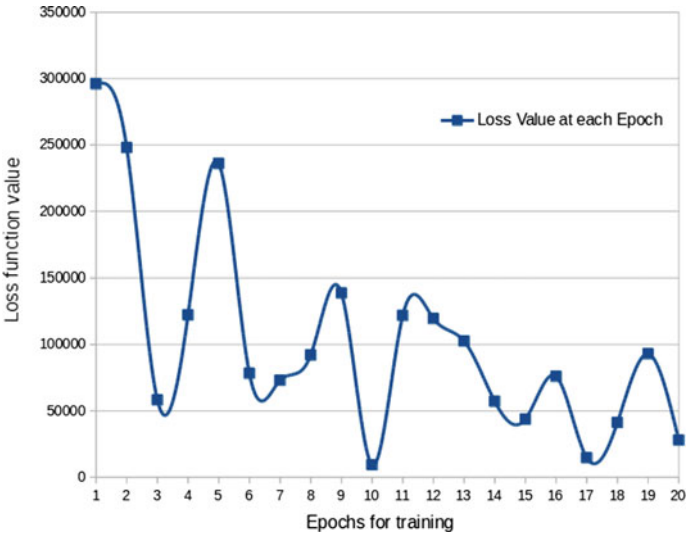


Fig. 5. Showing loss curve with respect epoch

ground-glass opacity nodules and to distinguish closely attached nodules and vessels are limitations of region-growing algorithm.

5 Future Scope

The image preprocessing can be enhanced to classify nodules better by detecting ground glass opacity nodules (which are deformed in shape but malignant), nodules attached to lung walls, anisotropic nodules, and also the ones which are closely attached with blood vessels. To attain this mean shift analysis based on the dot filter method of segmentation and Hessian matrix can be implemented.

The loss function values can result in a smooth curve with a gradual decrease in loss value by increasing the epoch size and once again by increasing the number of hidden layers in convolutional step more hidden features can be extracted from the input for better classification of patient's data.

6 Conclusion

Image classification and feature extraction from images to make accurate prediction is challenging and beneficial at the same time, it gives the ability to define objects, predict variation in images accurately at very deep dimension and also by leveraging the advanced methods of Neural Network (CNN), it is now possible to build an AI which

can help reduce the burden on radiologist, analyst, hospitals, and clinical practitioners. Image networks is a special area of study for training machines to learn image data to classify and predict correct results, by bringing in advance additions to existing networks and also by making use of medically accurate segmentation methods it will definitely be possible to attain an accuracy of 95% and above for a wide variety of patients dataset.

Acknowledgments. We would like to thank Big Data team of System Level Solutions (SLS) Pvt. Ltd. Gujarat, India (www.slscorp.com), for their fruitful support, guidance, and mentorship towards the completion of this problem. We also would like to heartily thank SLS for giving us a noble opportunity to work on such a good problem statement.

We would like to thank Mr. Tejas Vaghela, General Manager, System Level Solutions Pvt. Ltd. Gujarat, India for his throughout support and guidance in this project.

We would like to thank Ms. Srijana Pradhan, Assistant Professor-I, CSE Department, Sikkim Manipal Institute of Technology for fruitful discussion.

References

1. Statistical Data of Lung Cancer. http://www.wcrf.org/int/cancer-facts-figures/data-specific-cancers/lung-cancer-statistics. Accessed 5 May 2017
2. Miah Md. BA, Yousuf Md. A (2015) Detection of lung cancer from CT image using image processing and neural network. In: ICEEICT, May 2015
3. The DICOM Standard (2017) Medical imaging and technology alliance—a division of NEMA
4. Horwood AC, Hogan SJ, Goddard PR, Rossiter J (2001) Image normalization, a basic requirement for computer-based automatic diagnostic applications. University of Bristol, UK
5. Mansoor A, Bagci U, Foster B, Xu Z, Papadakis GZ, Folio LR, Udupa FK, Mollura DF (2015) Segmentation and image analysis of abnormal lungs at CT: current approaches, challenges, and future trends. Radio Graph 35:1056–1076
6. Mesanovic N, Grgic M, Huseinagic H, Males M, Skejic E, Smajlovic M (2011) Automatic CT image segmentation of the lungs with region growing algorithm
7. Transmitted Intensity and Linear Attenuation Coefficient. https://www.nde-ed.org/EducationResources/CommunityCollege/Radiography. Accessed 13 May 2017
8. Hounsfield Scale. https://en.wikipedia.org/wiki/Hounsfield_scale. Accessed 15 May 2017
9. Panpaliya N, Tadas N, Bobade S, Aglawe R, Gudadhe A (2015) A survey on early detection and prediction of lung cancer. Int J Comput Sci Mobile Comput IJCSMC 4(1):175–184
10. Verma OP, Hanmandlu M, Susan S, Kulkarni M, Jain PK (2011) A simple single seeded region growing algorithm for color image segmentation using adaptive thresholding. IEEE
11. Karpathy A, Johnson J, Li F-F (2016) Convolution neural networks for visual recognition. Stanford University

Implementation of Lifting Scheme Discrete Wavelet Transform Using Modified Multiplier

G. Kishore Kumar[1(✉)] and N. Balaji[2]

[1] Department of ECE, VRSEC, Vijayawada 520007, Andhra Pradesh, India
kishore.chiya@gmail.com
[2] Department of ECE, JNTUK UCEN, Narasaraopet 522603, Andhra Pradesh, India
prof.balaji.ece@gmail.com

1 Introduction

There has been an extensive use of discrete wavelet transform (DWT) in many applications of image processing and signal processing [1–3]. The present designs for the implementation of the DWT are mostly categorized into two types: 9/7 filter and lifting scheme. For the second category, designs have advantages when compared to the first category architectures in area requirement and computing complexity.

The digital signal processing (DSP) applications such as discrete wavelet transform, finite impulse response (FIR) filters and infinite impulse response (IIR) filters involves multiplication and addition as fundamental operations. In the real-time environment having hardware restrictions, delay, area and power metrics plays a key role in analyzing the system performance. The multiplier plays key role as hardware point of view in discrete wavelet transform, as result multiplier consumes larger area, power, and delay.

A shift adder-based multiplier is implemented in [4]. This work involved with less logic resources but with marginally higher delay. An SOPOT-based multiplier is proposed in [5]. This work involved with large logic resources and more power consumption. The authors in [5] implemented the SOPOT architecture which consists of scalable shifter and variable shifter. We found that some redundant operations are involved in scalable shifter. The main contribution in this paper, scalable shifter is replaced with one shifter and one multiplexer.

In this research work, a modified CSD multiplier is designed and the same is used for developing an efficient lifting-based DWT. The ASIC results show the efficiency of the modified multiplier.

2 Lifting Scheme

The construction of bi-orthogonal wavelets for lifting scheme was developed by Sweldens [6]. The constructions are obtained in the spatial domain, which is the main feature of this lifting scheme. It also does not need complex mathematical calculations

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_38

which are necessary in traditional methods. The lifting scheme is simplest and competent algorithm to compute wavelet transforms. It does not rely on Fourier transforms.

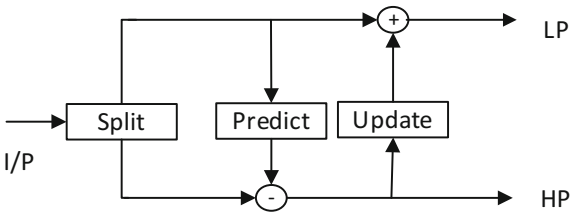


Fig. 1. Forward lifting scheme

Figure 1 shows the block diagram of the lifting-based structure. Usually, the lifting scheme consists of three steps: (1) split, (2) predict, and (3) update [4].

1. Split Step:
   To find the utmost correlation between adjacent pixels for using them in during the next predict step, the signal is split into even and odd points in this step. Each pair of given input samples x(n) is split into odd coefficients x(2n + 1) and even x(2n).
2. Predict Step:
   To generate the complete coefficients (dj), the product of even samples and the predict factor is done and the output is added to the odd samples, in this step. Since the detailed coefficients are taken into account, this results in high-pass filtering.
3. Update Step:
   Here, the product of the detailed coefficients calculated in the previous predict step and the update factors is done in this step. The output is added to the even samples to obtain the coarse coefficients (sj). This results in a low pass filtering. The block diagram of lifting scheme DWT is shown in Fig. 2 [6].

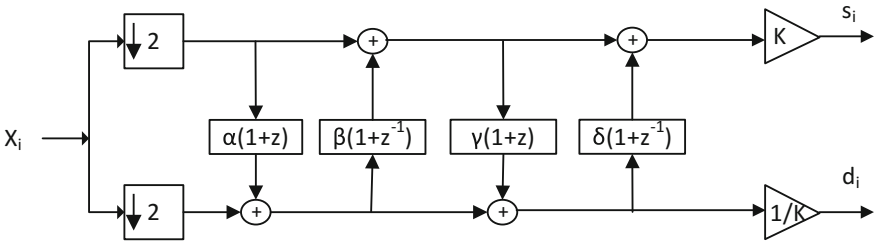


Fig. 2. Block diagram of lifting scheme DWT

The lifting coefficients δ, β, α, γ, and K are $\delta \approx 0.88701370409$, $\gamma \approx 1.7658221510$, $\alpha \approx -3.172268684$, $\beta \approx -0.105960237$, and $K \approx 0.812893066$, respectively.

The complete lifting procedure consists of four steps, in which the input sequence X_i, where i = 0, 1, 2, … N − 1. Here N is the input data range.

(1) Splitting step:

$$\text{Odd part} \quad d_i^{(0)} = x_{2i+1} \tag{1}$$

$$\text{Even part} \quad s_i^{(0)} = x_{2i} \tag{2}$$

2) First lifting step:

$$\text{Predictor} \quad d_i^{(1)} = d_i^{(0)} + \alpha \times \left(s_i^{(0)} + s_{i+1}^{(0)} \right) \tag{3}$$

$$\text{Updater} \quad s_i^{(1)} = s_i^{(0)} + \beta \times \left(d_{i-1}^{(1)} + d_i^{(1)} \right) \tag{4}$$

3) Second lifting step:

$$\text{Predictor} \quad d_i^{(2)} = d_i^{(1)} + \gamma \times \left(s_i^{(1)} + s_{i+1}^{(1)} \right) \tag{5}$$

$$\text{Updater} \quad s_i^{(2)} = s_i^{(1)} + \delta \times \left(d_{i-1}^{(2)} + d_i^{(2)} \right) \tag{6}$$

4) Scaling step:

$$d_i = d_i^{(2)} / K \tag{7}$$

$$s_i = K \times s_i^{(2)} \tag{8}$$

3 Proposed Multiplier

The DWT architecture consists of adders and multiplies. The conventional DWT architecture has multiplied with constant coefficients, which consume more area and power. Several multiplication operations are performed in DSP algorithm. So constant multipliers are replaced with canonical-signed (CSD) [7]. The multipliers are replaced with shift adders, which lead to multiplier-less implementation. The same method is applied to all the coefficients of the lifting structure and the values are tabulated in Table 1. Table 1 shows the power-of-two with negative and positive components. The positive exponent and negative exponents represent left shift and right shift respectively.

Table 1. CSD representation of lifting coefficients

Coefficient	Coefficient value	CSD representation
α	−3.17226868408289	$-2^{2} + 2^{0-2-2} + 2^{-4} + 2^{-6} - 2^{-11} + 2^{-13}$
β	−0.10596023714940	$-2^{-3} + 2^{-6} + 2^{-8} - 2^{-11}$
γ	1.76582215102264	$2^{1-2-2} + 2^{-6} + 2^{-12-2-14}$
δ	0.88701370409649	$2^{0-2-3} + 2^{-6} - 2^{-8} + 2^{-12} + 2^{-14} - 2^{-16}$
K	0.81289306611600	$2^{0-2-2} + 2^{-4} + 2^{-11} + 2^{-13}$
1/K	0.61508705245638	$2^{-1} + 2^{-3} - 2^{-7} - 2^{-9} - 2^{-13} - 2^{-15}$

Authors in [5] proposed architecture have scalable shifter and variable shifter. By observing Table 1, there is no higher power-of-two difference (i.e., greater than 7 shifts). So scalable shifter is optimized without disturbing functionality and accuracy of the multiplier. The modified multiplier architecture is shown in Fig. 3.

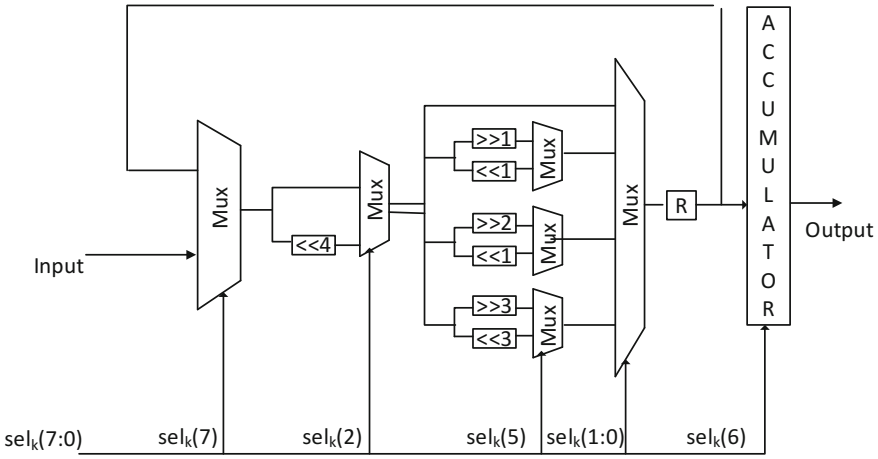


Fig. 3. Proposed architecture for multiplier

The proposed architecture defines SOPOT term of shifters and accumulator unit. The size of control register is 8 bit. The selection of input data is not mentioned in existing architecture. The sign of CSD term is controlled by $sel_k(6)$ and based on this control bit accumulator performs addition and subtraction. If the control bit $sel_k(6)$ is positive accumulator executes addition and negative accumulator executes subtraction. The sign of exponent of CSD term is controlled by $sel_k(5)$ and based on this control bit variable shifter performs right shift and left shift. If the control bit $sel_k(5)$ is positive variable shifter executes left shift and negative variable shifter executes right shift. The control bit $sel_k(2)$ selects scalable shift operation. The control bits $sel_k(1:0)$ selects variable shift operation. The control bit $sel_k(7)$ selects input data/register data. The configuration table for α lifting coefficient is given in Table 2.

Table 2. Configurations of the lifting coefficient α

CSD term	Control register	Configuration
$+2^{0}$	Sel_0	10000000
-2^{2}	Sel_1	01000010
-2^{-2}	Sel_2	01100010
$+2^{-4}$	Sel_3	00100010
$+2^{-6}$	Sel_4	00100010
-2^{-11}	Sel_5	01100101
$+2^{-13}$	Sel_6	00100010

4 ASIC Implementation Results

The proposed work is designed in Verilog HDL and the synthesis is done using Cadence Genus Synthesis of 90 nm technology library [8]. The ASIC implementation results of proposed, shift adders and existing SOPOT architectures are shown in Table 3. Table gives delay, area, power values and excess area delay product (EADP), excess power delay product (EPDP) results compared to other works for dissimilar bit-widths respectively. The proposed design has 21.75 and 12.35% less ADP; 21.54 and 22.07% less PDP when compared to shift adder and SOPOT architectures, respectively, for dissimilar bit-widths.

Table 3. ASIC implementation results of proposed and existing designs

Design	Width (n)	Delay (ps)	Area (cells)	Power (pW)	ADP	PDP	EADP %	EPDP %
Shift adder [4]	10	1380	1065	236.60	1469700	326509.4	2.34	2.42
	12	1571	1672	363.55	2626712	571137.1	32.11	34.62
	16	1981	2319	425.40	4393939	842731.3	30.80	27.59
Existing CSD [5]	10	815	1976	471.96	1610440	384647.4	12.14	20.66
	12	949	2327	529.40	2208323	502404.4	11.07	18.42
	16	1222	3272	687.17	3998384	839727.9	13.84	27.14
Proposed CSD	10	815	1762	391.13	1436030	318778.3	–	–
	12	949	2095	447.03	1988155	424238.1	–	–
	16	1222	2874	540.47	3512028	660464.1	–	–

5 Conclusion

In this work, a modified CSD architecture is proposed targeting area and power optimization. The logic operations involved in SOPOT-based multiplier is analyzed, we found that architecture having some redundant operations. The redundant logic operations that are present in SOPOT architecture are eliminated in the proposed novel multiplier using CSD representation. The implementation results show that the

proposed architecture has 21.75 and 12.35% less ADP compared to shift adder and SOPOT architectures, on average, respectively, for dissimilar bit-widths. The implementation results show that the proposed architecture has 21.54 and 22.07% less PDP compared to shift adder and SOPOT architectures, on average, respectively, for dissimilar bit-widths. The implementation results show that the modified CSD architecture occupies low area, consumes less power and is efficient for the VLSI hardware implementation.

References

1. Barsanti RJ, Athanason A (2013) Signal compression using the discrete wavelet transform and the discrete cosine transform. IEEE southeast conference, pp 1–5
2. Urriza I, Artigas JI, Garcia JI, Barragan LA, Navarro D (1998) VLSI architecture for lossless compression of medical images using the discrete wavelet transform. In: Proceedings of the Design Automation Test Europe, pp 196–201
3. Vaithiyanathan D, Seshasayanan R (2013) High speed low power DWT structure with log based FPU in FPGAs. In: International conference on green computing, communication and conservation of energy (ICGCE 2013). Chennai, India, pp 308–313
4. Shi G-M, Liu W-F, Zhang L, Li F (2009) An efficient folded architecture for lifting-based discrete wavelet transform. IEEE Trans Circuits Syst II Exp Briefs 56(4):290–294
5. Wu JF (2015) Efficient implementation and design of a new single-channel electrooculography-based human machine interface system. IEEE Trans Circuits Syst II Express Briefs 62(2):179–183
6. Sweldens W (1996) The lifting scheme: a custom-design construction of biorthogonal wavelets. Appl Comput Harmon Anal 3(15):186–200
7. Lim YC, Parker SR (1983) FIR filter design over a discrete power of-two coefficient space. IEEE Trans Acoust Speech Signal Process ASSP-31:583–591
8. Cadence genus synthesis solution (2015)

Evolutionary Algorithms to Minimize Interaction Energy Between Drug Molecule and Target Protein in Streptococcus

Ayan Chatterjee[1(✉)], Uttam Kumar Roy[1], and Dinesh Haldar[2]

[1] Department of IT, Jadavpur University, Kolkata, India
ayanl.c2@gmail.com, u_roy@it.jusl.ac.in
[2] Kalyani University, Nadia, India
Haldar.dinesh85@gmail.com

1 Introduction

The main aim of drug designing is to develop or search a stable drug that cures disease with minimal cost and time. Different docking strategies are taking place to make it more efficient with evolutionary algorithms. The problems can be classified as optimization problems with NP-completeness. Here, we have used GA (fixed and variable length), swarm intelligence-based algorithm (PSO) to find out the probable orientation or geometry of a drug molecule that can bind to a particular active site of a target protein to produce a stable drug. Active sites are some area on the surface of a protein molecule and that area becomes active when it comes in contact with an appropriate drug. As a result, the normal function of the protein is hampered. Computational drug design is important as it reduces conventional research deadline and cost. Here, we have considered structure-based drug design where we have identified a protein target (active site) and going ahead we have to find a molecule or ligand that nicely fits there. In this paper, we have used pdb file of Streptococcus and developed an algorithm to predict protein and ligand structure for docking energy minimization. At the output, we prepare a set of molecules having energy less than a threshold value. All these molecules are the potential drugs for a particular protein. These set of molecules will help the biologists or pharmacists to choose the best drug for a particular disease with very less effort [1–5].

2 Bacteria Protein

The study has been conducted on Streptococcus bacteria's protein structure 5a35.pdb and its corresponding drug structure is 1PE.pdb. The structure of 5a35.pdb is shown in Fig. 1.

Streptococcus pyogenes, a Group A Gram-positive bacteria responsible for pharyngitis, cellulitis, tonsillitis, rheumatic fever, scarlet fever, etc. Based on the statistics, 700 million infections occur worldwide each year with a mortality rate of 25%.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_39

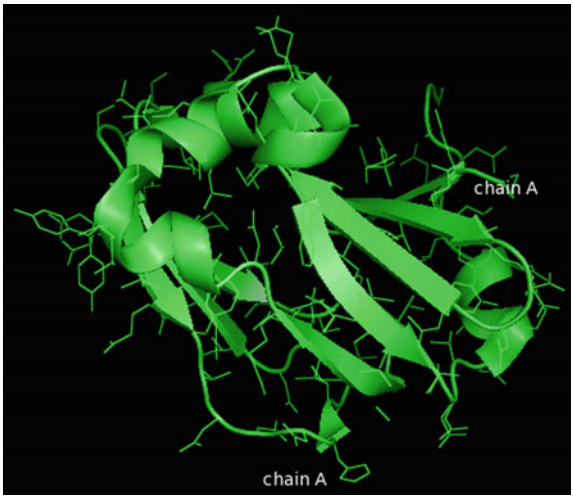


Fig. 1. 5a35.pdb structure with PyMol [Chatterjee et al.]

3 Proposed Algorithms

3.1 For Docking Energy Minimization with Evolutionary Algorithm—GA

Step#1 Coordinates of active site
Step#2 Valencies and length of functional groups.
Step#2 van der Waal constants and electrostatic constant.

Output: Desired ligand structure L for receptor target protein P.
Begin
Call GA(active_site_P, Functional group List);
End.
Procedure GA(active_site_P, Functional group List)
Begin
While generation_count < X do/* X = max. number of generations. */
Begin

Select parent randomly → Crossover → Mutation
Use "Correct function" to adjust parameters based on the assumption for Ligand Tree
Use "adjustPosition function" to adjust the coordinate of added functional group (must be in search space)
Increment generation_count
end
Output the best individual found

Step#3 Ends

3.2 For Docking Energy Minimization with Evolutionary Algorithm—PSO

Step#1 Coordinates of active site
Step#2 Valencies and length of functional groups.
Step#3 van der Waal constants and electrostatic constant.

Output: Desired ligand structure L for receptor target protein P.
Begin
Call PSO(active_site_P, Functional group List);
End.
Procedure PSO(active_site_P, Functional group List)

Begin

A. Initialize all particles Xi with each particle having D dimensions.
B. Assign initial coordinates to each particle and call calculate_energy (Xi) to find respective Fitness
C. Initialize Particle pBest of each particle and find Particle gBest
D. Calculate velocity and position of PSO parameters
E. Check the length of newly formed particle by calling check_length()
F. Update the coordinates of each particle
G. Update Particle pBest and Particle gBest as necessary
H. Use "Correct function" to adjust parameters.
I. Use "adjustPosition function" to adjust coordinate of added functional group (must be in search space)
J. Repeat steps C to I until a convergence criteria is satisfied or maximum iteration ends.

Step#4 Ends.

4 Ligand Tree Structure

We have used GA and PSO-based algorithm technique to predict the ligand orientation, shape and minimize the interaction energy and compare our results with previous results. We also have used a tree-like structure whose length cannot be determined in advance. To do this, we have introduced a group which is called absent group. The absent group can be present only in leaf [6, 7]. Two extra functions namely "correct" and "adjust" are used to check the length of the newly formed ligand molecule with adjusted new position inside search space. Some constraints should be maintained as follows: [16, 19, 30]—A. The first, third, and sixth position of right tree and first and third position of left tree cannot contain any type of polar group unless they are present at the terminals of the tree. B. If any of the parent positions contain any child group then parent position cannot contain 8. C. Crossover, mutation operators are implemented similarly as described in this paper. We can look for a simple example: say

right tree chromosomes for two parents are: [5 5 2 8| 8 1 5 |1 4 6]; [2 7 18 3 2 |5 8 6| 1] Then corresponding two children will be: [5 5 2 8 5 8 6 1 4 6]; [2 7 1 8 3 2 8 1 5 1]. Now, the right tree of the first children contains 8 at sixth position though its eighth, ninth, and tenth places are not empty. Thus, by using "Correct" function we place a 2 at sixth position so that the two children will be: [5 5 2 8 5 2 6 1 4 6]; [2 7 1 8 3 2 8 1 5 1]. Now two chromosomes are valid and accepted. **N.B**: the distance between a residue of the target protein receptor and its closest functional group should be at most 2.7 Å for the molecules to interact and should not be closer than 0.65 Å to avoid steric contact (Figs. 2 and 3).

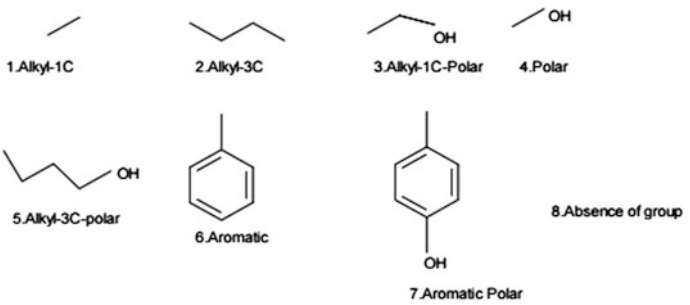


Fig. 2. Functional groups used in algorithm [1]

Functional group	Bond length along x axis (Å)	Code
Alkyl-1C	0.65	1
Alkyl-3C	1.75	2
Alkyl-1C-Polar	1.1	3
Alkyl-3C-Polar	2.2	4
Polar	0.01	5
Aromatic	1.9	6
Aromatic-Polar	2.7	7
NUL (no group)	-	8

Fig. 3. Bond lengths of functional groups [1]

In the ligand molecule, we have used above necessary functional groups including double (=C=O) and triple bonds (–C≡N). The size of representing tree is made variable. The length of the chromosome can vary between certain range denoted by (lmax, lmin), where these values are defined as: Lmax = (Max distance in between two CA)/2.7 and Lmin = (Min distance in between two CA)/0.65.

4.1 Necessary Setup

We have used http://zhanglab.ccmb.med.umich.edu, a Michigan University Lab portal for "Department of Computational Medicine and Bioinformatics" to determine binding site residues for 5A35.pdb and in 2D diagram, we have found active site that resembles

the active site as shown below with 1PE ligand binding and populated by RASMOL using X-ray diffraction (Fig. 4).

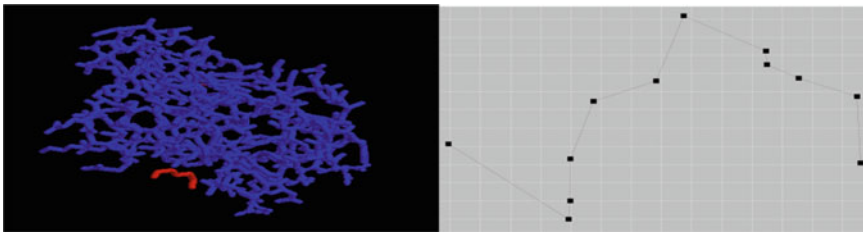


Fig. 4. Docking structure comparison [Chatterjee et al.]

Using evolutionary algorithms, we will generate tree on both the sides of the pharmacophore maintaining Lmax, Lmin, and a boundary size of search space will be defined and that search space is divided into two parts on either side of the pharmacophore.

Active site size	12
L max	7
L min	3
South boundary of polygon	−31.584, −4.099 | −31.485, −7.72 | −29.547, −8.727 | −24.808, −8.597 | −22.741, −7.462 | −21.977, −4.321 | −21.972, −1.988
North boundary of polygon	−30, −1 | −27, −0.8 | −24, −0.5
Pharmacophore left side coordinate	(−25.15, −6)
Pharmacophore right side coordinate	(−24.5, −5)
Distance in between residue and functional group	0.65–2.7

N.B: All distances are measured in Angstrom and South boundary points are coordinates of active sites residues in X-Y. **Assumptions are**—A. Random integer is generated to produce the size of tree and it must fall under Lmax, Lmin. Then group selection is random out of 8 groups to fill gene position. B. If Ligand Tree is avariable-length tree, and if randomly generated tree value is < Lmax then add padding (Lmax-Lrand) by '#' to keep the ligand tree length correct. Example: lmax = 8, the size of the right tree, i.e., li is five, then a chromosome may look like 01252### and here we have used group 8 to fill it. C. To complete the polygon structure of search space, we have taken 3 assumed points for the north boundary.

4.2 Experiment

Total Nodes except Pharmacophore: 11 (Left size: 4 and right size: 7). We have used our functional group index from 0 to 7 in our program.

Genetic Algorithm Parameters: Population: 30 Exhaustiveness: 100 Mutation Rate: 0.01 Elitism: 10% Crossover: Uniform (Figs. 5 and 6).

Side	Length (Fixed)	Energy	Sequence	Execution time (ms)	Fitness
Left tree	4	0.92	5 2 5 6	2764	−0.296
Right tree	7	0.7	1 2 1 2 3 5 2	1995	13.22

Side	Length (Variable)	Energy	Sequence	Execution time (ms)	Fitness
Left tree	4	0.35	5 3 5 3	1375	−0.07
Right tree	7	1.16	1 4 1 3 2 7 6	3149	13.22

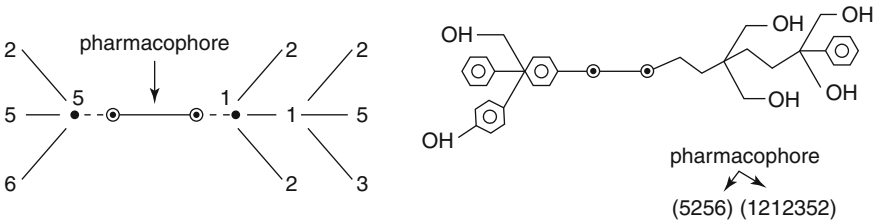


Fig. 5. Drug structure with GA (Fixed length) [Chatterjee et al.]

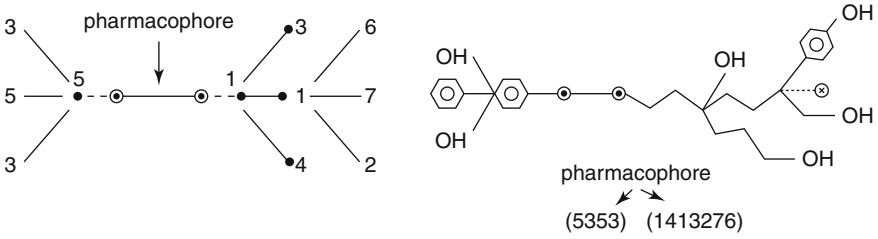


Fig. 6. Drug structure with GA (Variable length) [Chatterjee et al.]

PSO Algorithm Parameters: SWARM_SIZE = 30 MAX_ITERATION = 100 PROBLEM_DIMENSION = 2 C1 = 2.0 C2 = 2.0 W_UPPERBOUND = 1.0 W_LOWERBOUND = 0.0 (Fig. 7).

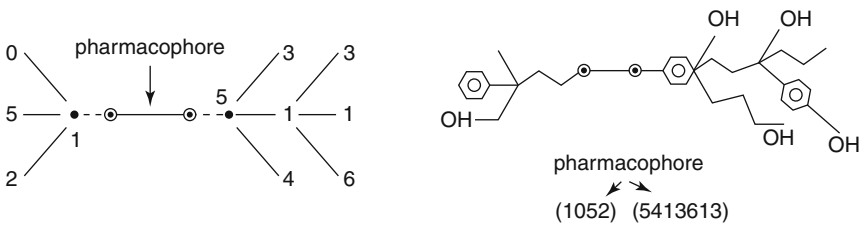


Fig. 7. Drug structure with PSO [Chatterjee et al.]

Side	Length	Energy	Sequence	Execution time (ms)	Fitness
Left tree	4	0.00004	1 0 5 2	289	167.5
Right tree	7	0.008	5 4 1 3 6 1 3	381	96.7

Fitness Evaluation: We have considered here fitness value must be dependent on below two things [1]: A. Distance and Polarity → Fitness F1 and B. Interaction energy (Van Der Waal and electrostatic) → Fitness F2

$$V_{non-bond} = \sum_{non-bond} \left[\varepsilon_{i,j} \left[\left(\frac{R_{min,i,j}}{r} \right)^{12} - 2 \left(\frac{R_{min,i,j}}{r} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon r} \right]$$

$$V(r) = \left[\left(\frac{c_n}{r^6} \right) - \left(\frac{c_m}{r^{12}} \right) \right]$$

Finally, fitness is computed as

$$F = \frac{k}{E}$$

where 'k' is a constant (let it be 100) and 'E' is the total interaction energy in Kcal/mol. Hence, maximizing fitness leads to minimizing interaction energy. A penalty will be given if any functional group crosses the boundary, if a functional group possesses opposite polarity to its closest residue or if functional group crosses the minimum constraint boundary. So, we will choose that drug that will have overall minimum interaction energy and minimum penalty value. Let P is the value used for penalty calculation and D is the distance in between functional group and the closest residue and the relationship in between them is P = (2.7 − D)/2.7. Penalty = −|P| if a functional group possess opposite polarity to its closest residue Penalty = 10 * (−|P|) if the position of a functional group of the tree exceeds the boundaries. Fitness of a Functional group → **If** the polarity of the function group is not the same as that of the residue or D < 0.65 then Fitness = Penalty **Else if** P < 0.0 then Fitness = 0.0 **Else if** polarity of the functional group is approximately the same as the polarity of the residue and P > 0.0 then Fitness = D.

4.3 Result

We have found the below result after several trials run over 30 populations with exhaustive value 100: (Fig. 8).

Methods	Total interaction energy (VDW + ES)	Total execution time (ms)	Total fitness
GA—Fixed length	1.62	4759	12.924
GA—Variable length	1.51	4524	13.15
PSO	0.00804	670	264.2

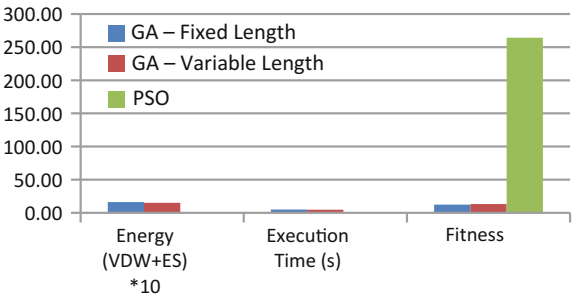


Fig. 8. Comparison of different methods [Chatterjee et al.]

5 Conclusion

It has been shown how PSO method gives better performance than GA (both fixed and variable length). In this paper, we have taken input from pdb file, analyzed it, created 2D and 3D architecture, and used evolutionary algorithms to find out probable ligand structure. This work has helped in digging very basic domain knowledge of computational biology and evolutionary algorithms. A similar case study can also be applied to other drug design problems to have optimized data based on different evolutionary algorithms.

References

1. You Z-H, Lei Y-K, Zhu L, Xia J, Wang B (2013) Prediction of protein-protein interactions from amino acid sequences with ensemble extreme learning machines and principal component analysis
2. Raman K (2010) Construction and analysis of protein–protein interaction networks
3. Pasupuleti S (2008) Detection of protein complexes in protein interaction networks using n-clubs

4. Jothi R et al (2006) Co-evolutionary analysis of domains in interacting proteins reveals insights into domain-domain interactions mediating protein-protein interactions. J Mol Biol 362:861–875
5. Iqbal M, Freitas AA, Johnson CG (2008) Protein interaction inference using particle swarm optimization algorithm
6. Hesselberth JR et al (2006) Comparative analysis of Saccharomyces Cerevisiae WW domains and their interacting proteins. Genome Biol 7(4):R30
7. Yu H et al (2008) High-quality binary protein interaction map of the yeast interactome network. Science 322(5898):104–110

Assessment of the Fetal Health Using NI-aECG

Niyan Marchon[1(✉)] and Gourish Naik[2]

[1] Padre Conceicao College of Engineering, Goa, India
niyanmarchon@gmail.com
[2] Electronics Department, Goa University, Goa, India
gmnaik@unigoa.ac.in

1 Introduction

Globally, it is reported that many stillbirths occur at labor due to various congenital anomalies [1]. This growing concern of fetal well-being during pregnancy has raised a need for fetal monitoring during pregnancy, especially during the third trimester or labor. It was first suggested by Winkler that an FHR of less than 120 beats per minute (bpm) or greater than 160 bpm were signs of fetal distress [2]. It was observed that decelerations of the fetal heart rate-predicted fetal hypoxia. Hence, if the basic parameters of fetal health are monitored by the mother then the status of the fetus can be electronically communicated to the clinician to save the fetus from going through irreversible neurological brain damage [3].

Fetal distress, now commonly known as non-reassuring fetal status is a complication during pregnancy or labor. It occurs mostly due to placenta abruption, umbilical cord prolapse, breathing problems, etc. The various non-reassuring patterns seen on a Cardiotocography (CTG) to mark fetal distress are increased or decreased fetal heart rates especially during and after a contraction, decreased variability in FHR and late decelerations among other signs [4]. FHR monitoring nowadays during pregnancy is, therefore, a very important exercise by mothers to keep their unborn babies safe and secure. Presently, ultra sound transducers are the common methods used by doctors to check the well-being of the fetus. It is realized in many cases that the FHR observed by using the aECG signals are more accurate than using the ultrasound method [5].

Noninvasive abdominal ECG (NI-aECG) signals extracted from the abdominally placed ECG electrodes are very low-voltage amplitude signals buried in maternal complexes, baseline wander, power line interferences, maternal electromyogram, and other artifacts [6]. Invasive FECG obtained using the fetal scalp electrode directly applied to the skin allows easy detection of fetal QRS complexes. These signals have much larger amplitudes than abdominally obtained signals, however with the risk of infection to the fetus [1]. Detection of Non–Invasive fetal ECG (NIFECG) signal and its analysis is becoming a very powerful and advanced method in obtaining valuable QRS parameters from the FECG so as to assist the doctors and biomedical researchers to make appropriate decisions during pregnancy and labor. However, there exists a limitation for perfectly monitoring the fetal heart rates improving the signal to noise ratio (SNR) of the FECG [7]. The latest technologies in digital signal processing have been developed over the years towards advanced FECG detection, extraction, and analysis methods.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_40

1.1 Noninvasive Fetal ECG Extraction Techniques

Researchers in this biomedical field have continuously strengthened the extraction algorithms and improved the detection techniques which have led to further reduce noise and thus acquire reliable FECG signals assuring fetal health during pregnancy and labor. They have put in extensive work to effectively and efficiently separate FECG from maternal ECG. Some of the methods are Kalman filtering [8], LMS (least mean square) algorithm [9], ANN (artificial neural networks) [10], ANFIS (adaptive-network-based fuzzy inference system) [11], methods of BSS (blind source separation)—ICA [12], PCA (principal component analysis) [13], and SVD (singular value decomposition) [14]. Other methods include auto-correction methods [15], subtraction [16], averaging techniques [17], finite impulse response (FIR) and infinite impulse response (IIR) filtering [18] and wavelet transform based techniques [19].

Extraction of FECG is challenging because the FECG is a very weak signal and has low SNR as compared to the MECG and the fetal signal overlaps with the MECG in the frequency domain [1]. Table 1 gives us the MECG and FECG relationship calculated as mean values for QRS amplitude and width for various online aECG databases. It is observed from the table, that the mean amplitude value of the MQRS is nearly 4–5 times larger than the FQRS and the MQRS width is almost double that of FQRS [20]. In this paper we compare fetal extraction using the ICA technique and compare it with our method using correlations and synthesized QRS template matching (see Table 2). The proposed methodology can be evaluated by using the Physionet online abdominal and direct fetal ECG database (adfecgdb) [21, 22].

Table 1. MQRS and FQRS amplitude and width calculations as mean values

ECG signal	MQRS amplitude (μv)	MQRS width (ms)	FQRS amplitude (μv)	FQRS width (ms)
Matonia et al. [20]	150	100	30	50
adfecgdb [22]	211	115	192	58
nifecgdb [23]	10000	132	2600	46
Phy C 2013 [24]	23	93	21	47

Table 2. Comparison between ICA and correlation [17]

Method	Complexity	Computational time (s)	Progress in SNR	Accuracy (%)
ICA	Complex	Medium	Medium	Medium
Correlation	Low	Low	Low	Medium

2 Method I: Extraction of FQRS Using Independent Component Analysis

ICA is a statistical method for breaking down complex combined signals into independent parts. It looks for components that are both statistically independent and non-Gaussian. ICA can be easily expressed as $Y(n) = A\ S(n) + N$, where Y(n), S(n) and N are vectors of sensor signals, source signal, and N noise respectively [25] and A is the unknown mixing matrix and n is the discrete time index. In a noisy environment the source signal may be obtained as, $S(n) = A^{-1}\ [Y(n) - N]$, if $N \neq 0$. Therefore, pre-processing before applying ICA may improve the performance of the ICA [12]. In this technique the raw aECG signals are preprocessing followed by a two-stage ICA fetal and maternal separation. As the pre-filtering step is crucial for this method, a composite filter of high pass and low pass filter was used. The filter eliminated the low-frequency baseline wander (due to patients breathing) which is in the range of 1 Hz and the high-frequency noise of the EMG and the 50 Hz PLI. After the filtering sequence, the ICA was applied to the four filtered abdominal channels to separate maternal ECG from the other components. The following conditions must be met for ICA to function correctly. They are (i) the number of measured signals should be equal to or greater than the number of input sources (ii) instantaneous linear time-invariant mixing matrix (iii) statistically independent input sources and (iv) non-Gaussian and/or auto-correlated sources [26]. In our application of maternal and fetal QRS separations, the first two do not fully satisfy because artifacts increase the number of sources and fetal movement leads to a non-invariant mixing matrix. The maternal ECG is the strongest independent source of the four measured abdominal signals. Hence, it would result in at least one independent component. For records, where clean ECG signals and an invariant mixing matrix exist, ICA would also be able to separate the fetal ECG.

3 Proposed Method II: Synthesized QRS Template

One of the simplest and best known methods of recognizing and matching the QRS patterns is based on template matching [27] where the synthesized template is compared with the test signals of well-known databases. The algorithm is based on determining the similarity between the two signals. The synthesized template is generated based on the average values of the test signal. The QRS detector can be developed based on concepts from the automata theory [28], template subtraction technique [16] or template cross-correlation [29]. Using this method, we first filtered the raw aECG for baseline wandering with a low pass first order Butterworth filter and then performed the convolution of one of the filtered abdominal channels with the proposed synthesized maternal and fetal QRS complexes.

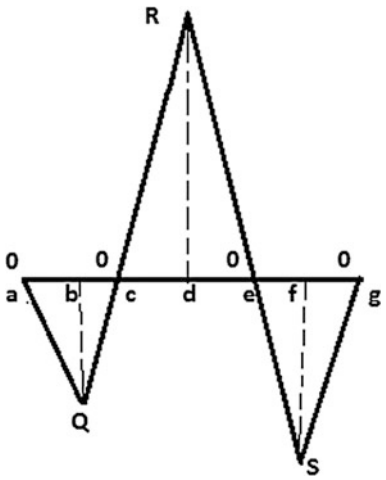


Fig. 1. Proposed QRS synthesized template format for MQRS and FQRS

3.1 Generation of the Synthesized QRS Waveforms

The synthesized QRS template waveforms for both maternal and fetal ECG signals were computed for the three available Physionet databases. The following six slopes for the QRS waveform were generated: 0-Q, Q-0, 0-R, R-0, 0-S and S-0 (See Fig. 1).

The synthesized MQRS waveform for adfecgdb is computed from Table 1: QRS amplitude = 211 μv; QRS width = 115 ms and Q = −37 μv, R = 211 μv, S = −163 μv and no = 1.

MQRS	= −1.21(n − no)	no < n < no + 33	(point a)
	= −37 + 5.28 (n – no − 33)	no + 33 < n < no + 40	(point b)
	= 11.72 (n – no − 40)	no + 40 < n < no + 58	(point c)
	= 211 − (11.72) (n − no − 58)	no + 58 < n < no + 76	(point d)
	= −9.58 (n – no − 76)	no + 76 < n < no + 93	(point e)
	= −163 + 7.76(n − no − 93)	no + 93 < n < no + 114	(point f)

Using the above method, the synthesized FQRS can also be computed with QRS amplitude = 192 μv; QRS width = 58 ms. To determine the correlations and convolution between the original maternal and synthesized MQRS, the following two operations were simulated using Matlab 2013a. A pulse train of synthesized signals $x_1(k)$ can be multiplied with the filtered aECG signal x_2 (n − k) as per Eq. (1).

$$y(n) = \frac{1}{N}\sum_{n=0}^{N-1} x1(k)\, x2(n-k) \tag{1}$$

4 Results and Discussions

The two methods, namely ICA and our proposed synthesized template matching method were evaluated and compared to obtain mainly the R peak of the FQRS using the Physionet afecgdb records [22]. The R peaks extracted from both the techniques were subjected to a QRS detector which is based on the QRS slope, width and R peak amplitude detection method [30, 31]. The QRS detector algorithm uses an adaptive threshold which helps compute the fetal R-R interval. Using the R-R intervals, the fetal heart rate variability (FHRV) was plotted and compared with the reference fetal QRS annotations (direct fetal scalp ECG).

4.1 ICA Method Results

After the first stage application of ICA, channel 1 contains a strong MECG, channel 2 displays a mixture of FECG and MECG signals, while channel 3 and 4 contain the noise components shown in Fig. 2. After zeroing the noisy channels 3 and 4, four new channels are generated from the two valid channels 1 and 2. The second stage ICA is applied to these four generated channels, thus obtaining stronger MECG on channel 1, a prominent fetal signal on a channel with smaller amplitude traces of maternal ECG while channel 3 and 4 contain no noise signals as shown in Fig. 3. EEGLAB, an interactive Matlab toolbox was used for processing continuous and event-related EEG, and other electrophysiological data incorporating independent component analysis (ICA).

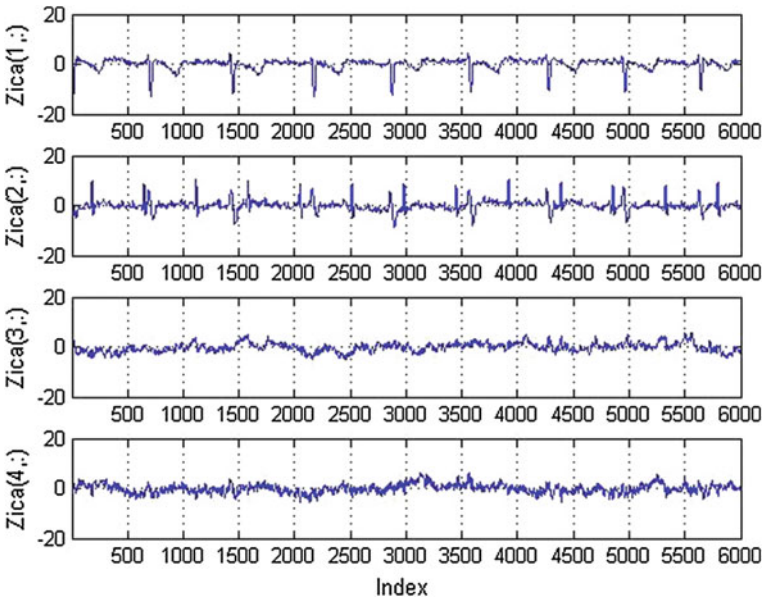


Fig. 2. Stage 1: Application of ICA to the four preprocessed abdominal channels

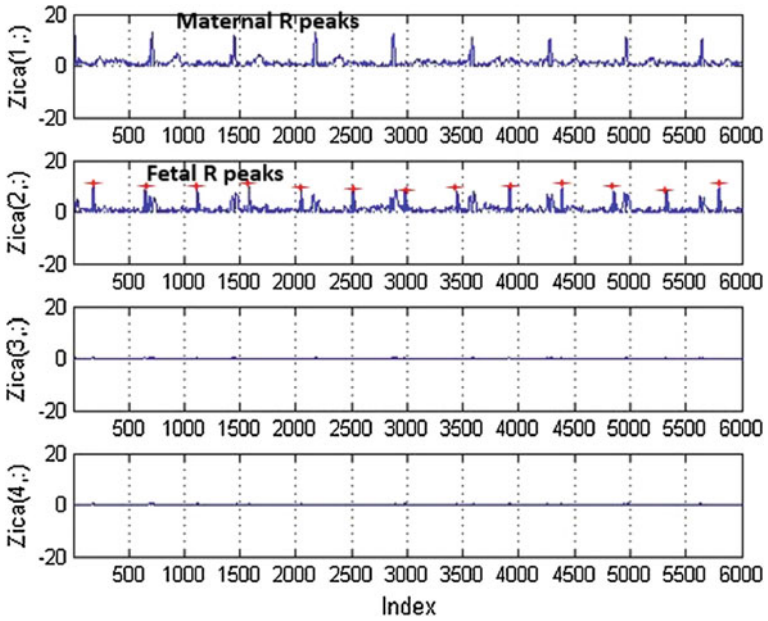


Fig. 3. Stage 2: Application of ICA to the four preprocessed abdominal channels (Red markings indicate the correct and stronger fetal R peaks)

4.2 Results of the Proposed Synthesized QRS Template and Pulse Matching

The QRS synthesized template and pulse matching was used to extract MQRS and FQRS signals. As in the case to extract MQRS and FQRS a rectangular pulse (tau) of

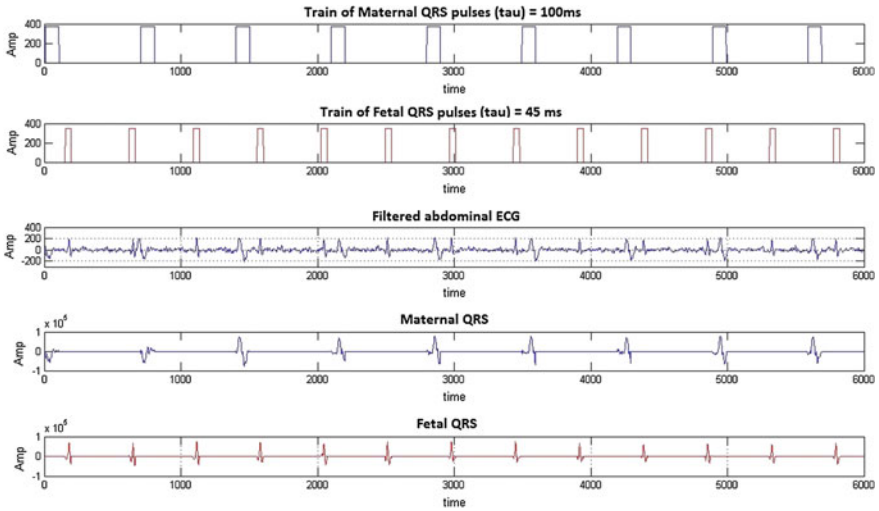


Fig. 4. Multiplication of synthesized pulses with filtered abdominal ECG to obtain MQRS and FQRS signals

100 ms and 45 ms was used respectively. The preprocessed filtered aECG was multiplied with the train of fixed k bpm and for a fixed rectangular width (tau). A clean MQRS and FQRS can be obtained if both the QRS are periodic (see Fig. 4).

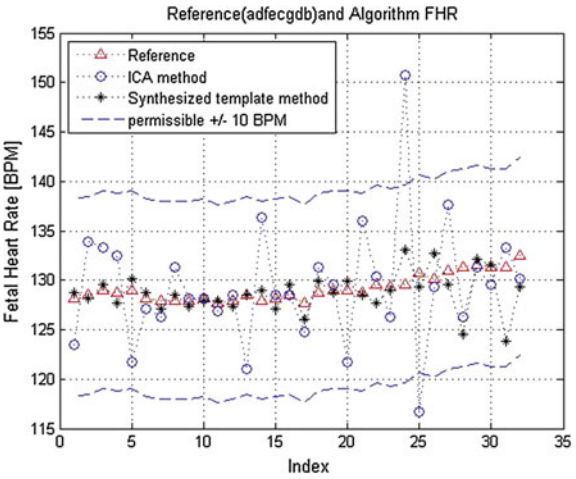


Fig. 5. Comparison of the FHRV using ICA and synthesized template method with the true reference annotations

5 Conclusion

The aECG channels were preprocessed by filtering the baseline wandering frequencies for both the methods, i.e., ICA and synthesized template method. The synthesized QRS signals for maternal and fetal ECGs were computed after carefully comparing the real time abdominal ECG signals. The filtered aECG was multiplied with the respective synthesized pulse trains having QRS (tau) = 100 ms for maternal and QRS (tau) = 45 ms for fetal. It was seen that, by synchronizing synthesized pulses alternatively with MQRS and FQRS present in the aECG, a noise-free individual maternal and fetal QRS were obtained. Further the fetal heart rate was computed by using the QRS detector giving us accurate heart rates for most records. In the ICA method it was observed that some of the FHR bpm values differed with the true reference range (direct fetal scalp annotation) and crossed the assumed ± 10 bpm limits. Additionally, using the synthesized template method the FHR values were closer to the true FHR values (see Fig. 5).

References

1. Sameni R, Clifford GD (2010) A review of fetal ECG signal processing; issues and promising directions. Open Pacing, Electrophysiol Therapy J 3
2. Winkler CL et al (1991) Neonatal complications at term as related to the degree of umbilical artery acidemia. Am J Obstet Gynecol 164(2):637–641
3. Hasan MA, Ibrahimy MI, Reaz MBI (2007) Techniques of FECG signal analysis: detection and processing for fetal monitoring. WIT Trans Biomed Health 12:295–305
4. Corton MM et al (2014) Williams obstetrics 24/E. McGraw Hill Professional
5. Cohen WR et al (2012) Accuracy and reliability of fetal heart rate monitoring using maternal abdominal surface electrodes. Acta Obstet Gynecol Scand 91(11):1306–1313
6. Clifford GD et al (2014) Non-invasive fetal ECG analysis. Physiol Meas 35(8):1521
7. Behar J et al (2016) A practical guide to non-invasive foetal electrocardiogram extraction and analysis. Physiol Meas 37(5):R1
8. Vullings R, De Vries B, Bergmans JWM (2011) An adaptive kalman filter for ECG signal enhancement. IEEE Trans Biomed Eng 58(4):1094–1103
9. Widrow B, McCool JM, Larimore MG, Johnson CR (1976) Stationary and non-stationary learning characteristics of the LMS adaptive filter. IEEE Proc. 64:1151–1162
10. Marques de Sa JP, Reis LP, Lau JN et al (1994) Estimation and classification of fetal heart rate baselines using artificial neural networks. Comput Cardiol 541–544
11. Kezi Selva Vijila C, Kanagasabapathy P, Johnson S (2005) Adaptive neuro fuzzy inference system for elicitation of fECG. In: Annual IEEE India conference-INDICON, pp 224–227
12. Ye Y, Zhang ZL, Zeng J, Peng L (2008) A fast and adaptive ICA algorithm with its application to fetal electrocardiogram elicitation. Appl Math Comput 205(2):799–806
13. Romero I (2010) PCA-based noise reduction in ambulatory ECGs. Comput Cardiol 37: 677–680
14. De Lathauwer L, De Moor B, Vandewalle J (2000) SVD-based methodologies for fetal electrocardiogram elicitation. In: Proceedings of IEEE on Acoustics Speech and Signal Processing, vol 6, pp 3771–3774
15. Van Bemmel JH (1968) Detection of weak foetal electro-cardiograms by autocorrelation and cross correlation of envelopes, IEEE Trans Biomed Eng, BME-15(1):17–23
16. Bergveld P, Meijer WHJ (1981) A new Technique for the suppression of the MECG. IEEE Trans Biomed Eng 28(4):348–354
17. Jagannath DJ, Immanuel Selvakumar A (2014) Issues and research on foetal electrocardiogram signal elicitation. Biomed Signal Process Control 10:224–244
18. Kam A, Cohen A (1999) Detection of fetal ECG with IIR adaptive filtering and genetic algorithms. In: IEEE international conference on acoustics, speech, and signal processing (ICASSP'99) Phoenix, AZ, USA, vol 4, pp 2335–2338
19. Echeverria JC et al (1996) Fetal QRS extraction based on wavelet analysis and pattern matching. In: Engineering in medicine and biology society. Bridging disciplines for biomedicine. Proceedings of the 18th Annual International Conference of the IEEE, vol 4
20. Matonia ADAM et al (2005) Modelling of non-invasively recorded maternal and fetal electrocardiographic signals. Biocybern Biomed Eng 25(2):27–39
21. Goldberger AL et al (2000) PhysioBank physiotoolkit, and physionet: components of a new research resource for complex physiologic signals, Circulation 12 (101) e215–e220. http://physionet.org/physiobank/database/
22. Abdominal and direct fetal electrocardiogram database. www.physionet.org/physiobank/database/adfecgdb. Accessed 15 June 2017

23. Non-invasive fetal electrocardiogram database. https://physionet.org/physiobank/database/nifecgdb. Accessed 15 June 2017
24. Physionet challenge 2013. http://physionet.org/physiobank/database/challenge/2013/set-a. Accessed 15 June 2017
25. Algunaidi M, Sheikh M et al (2009) Fetal heart rate monitoring based on adaptive noise cancellation and maternal QRS removal window. Eur J Sci Res 27(4):565–575
26. Varanini M et al (2014) An efficient unsupervised fetal QRS complex detection from abdominal maternal ECG. Physiol Meas 35(8):1607
27. Jain AK et al (2000) Statistical pattern recognition: a review. IEEE Trans Pattern Anal Mach Intell 22(1):4–37
28. Furno GS (1982) QRS detection using automata theory in a battery-powered microprocessor system. University of Wisconsin–Madison
29. Dobbs et al (1984) QRS detection by template matching using real-time correlation on a microcomputer. J Clin Eng 9(3):197–212
30. Pan J, Tompkins WJ (1985) A real-time QRS detection algorithm. IEEE Trans Biomed Eng 3:230–236
31. Marchon N, Naik G (2016) QRS detector for maternal abdominal ECG. In: Presented at the IEEE International Conference on Signal and Information Processing (IConSIP-2016), Nanded, India, Oct 6–8

Cost-Effective Vertical Handoff Strategies in Heterogeneous Vehicular Networks

Praneet Dhingra(✉) and Prem Chand Jain

Electrical Engineering Department, School of Engineering, Shiv Nadar University, G. Noida 201314, UP, India
{pd410, premchand.jain}@snu.edu.in

1 Introduction

Vehicular network aims to connect vehicle-to-vehicle (V2V) or vehicle-to-the roadside unit (RSU) using wireless communication technologies. IEEE has released 802.11p standard for vehicular ad hoc network (VANET) that connects V2V or V2I using dedicated spectrum. Currently vehicles are able to access Internet via cellular networks GPRS/UMTS/LTE. Other wireless technologies like Wi-Fi, Zigbee, Bluetooth, RFID are also being used as an alternative for short distances. Wi-Fi technology is an attractive solution for increase in data demand in mobile network because of high data rate achieved with low cost. However, limited coverage of Wi-Fi access points and user mobility poses difficulty to optimize offloading performance. Cellular and Wi-Fi radio access technologies are involved in data offloading. An interworking between cellular and Wi-Fi network may help to offload cellular data. Better offloading performance can be achieved if more number of Wi-Fi APs available in certain geographical area due to their short range. However, large-scale dense Wi-Fi APs deployment may incur high capital cost. Hence tradeoff between cost and offloading performance should be examined. A good strategy is to consider population density, user mobility, mobile data usage patterns, and communication environment while deploying Wi-Fi APs.

Internet applications are infotainment such as entertainment, web browsing, mobile commerce, etc., through Internet. Many Internet applications require continuous Internet connectivity. This connectivity becomes more challenging when vehicle moves across overlapping heterogeneous wireless networks. In such a case frequent switching from serving network to target network may occur, which often degrades the performance of the network. High vehicle mobility results in very short connection time with Wi-Fi APs (e.g., several tens of seconds) which limits volume of data transferred because one has to consider time spent in Wi-Fi association, authentication, and IP configuration before data transfer which is not negligible. Communication in vehicle environment suffers from high packet loss rate due to channel fading, shadowing, and Doppler shift because Wi-Fi protocol stack is not designed for high mobility environments [1].

We focus on cellular networks interworking with Wi-Fi in HVN. WLAN offers higher data rates with lower cost for a short range while cellular network offers low data rates with high cost but with large range. In the presence of both above, vehicle on-board unit (OBU) can choose either of two for data transfer. However, when OBU

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_41

leaves the Wi-Fi due to short range, cellular network will be connected to OBU, if V2V communication is not feasible. The availability of base stations (BS) and APs, and their data transfer prices are represented by cc ($) and cw ($), respectively, broadcasted periodically by all BSs in cellular and all APs in Wi-Fi network [2]. Next, we consider APs on the roadside on a model according to more practical situation like urban, semi-urban, and highways. Further, we include V2V communication where vehicles use multi-hop data transfer approach to relay the data to the APs. In this paper Sects. 2 and 3 discusses about efficient VHO decision with fixed inter-distance APs, and statistical inter-distance APs in urban, semi-urban, and rural areas. Section 4 includes V2V in addition to cellular and Wi-Fi network for VHO decision, Sect. 5 discusses performance evolution, and Sect. 6 concludes the paper.

2 VHO Decision with Inter-distance APs

For minimizing the cost of data transfer Wi-Fi is always preferred over cellular network. The data transfer cost c1 ($) per bit for cellular and c2 ($) per bit for Wi-Fi is assumed in VHO decision algorithm.

2.1 Cost Minimization Approach

Let us assume that all vehicle's OBU consists of both cellular and Wi-Fi interfaces. Let *bt* be the number of bits transmitted, *rc* and *rw* be the data rates (bits/sec.) of cellular and Wi-Fi, respectively, at the time of decision-making, and let *W* meters be the range of Wi-Fi AP. All APs are equidistant and remaining area between two consecutive APs is *A* meter as shown in Fig. 1. Let *v* (m/s) is velocity of the vehicle again at the time of decision-making, then the cost *c1* for cellular and *c2* for Wi-Fi can be obtained as

$$c1 = bt \times cc \tag{1}$$

$$c2 = Nw \times \frac{W}{v} \times rw \times cw + \left(Tw - Nw \times \frac{W}{v}\right) rc \times cc \tag{2}$$

In Eq. (2) *Nw* is number of APs and *Tw* is time taken by Wi-Fi to transfer data. Again in Eq. (2), first term represents cost of data bits to be transmitted by Wi-Fi

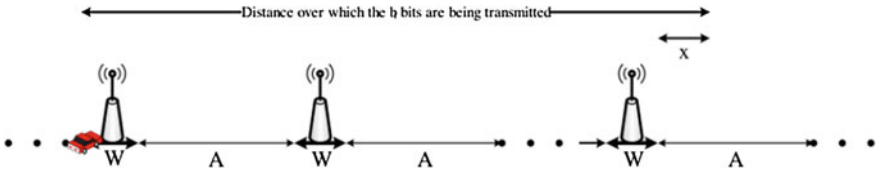


Fig. 1. Vehicular network with inter-AP distance [3]

network and second term represents cost of signaling bits and data transmitted by cellular network [3].

If b_{VHO} represents signaling bits and Nw the maximum integer for VHO, then

$$Nw \times \frac{W}{v} \times rw + (Nw - 1) \times \frac{A}{v} \times rc - 2Nw \times b_{VHO} \leq bt \tag{3}$$

$$Tw = \frac{[x + Nw \times W + (Nw - 1)A]}{v} \tag{4}$$

and x as shown in Fig. 1 is obtained by

$$bt - \left[Nw \times \frac{W}{v} \times rw + (Nw - 1) \times \frac{A}{v} \times rc - 2Nw \times b_{VHO} \right] = \frac{x}{v} \times rc \tag{5}$$

In Eq. (3) cellular network signaling is comparable with overhead of VHO signaling, and $2Nw \cdot b_{VHO}$ bits for signaling are needed irrespective of initial location of vehicle user. As velocity increases, number of VHO also increases which may cause more VHO signaling bits compared to data bits which result in high cost for data transmission.

2.2 Transmit Time Minimization Approach

User has no control over Wi-Fi and cellular network data rates, but appropriate access network at any instant can be chosen to minimize the total data transmit time. Let *Tc* be the total data transmission time by cellular network and *Tw* when Wi-Fi together cellular network is used, then

$$Tc = \frac{bt}{rc} \tag{6}$$

and *Tw* is given by Eq. (4). The vehicle will select the network to minimize the transmission time. It may be noted that the decision-making process is event based, and thus parameters will be computed whenever any significant changes in the network like new data rates, or new access network, or change in the velocity or direction of vehicle, or advertisement of new costs [4].

3 VHO Decision Based on Statistical Inter-distance APs

The distance between consecutive APs follows a pattern designed according to the area urban, semi-urban, or rural/highway chosen. The distance between APs and velocity of vehicle have been chosen arbitrarily as shown in Table 1. Urban areas will have maximum vehicle density, hence vehicle speed will be quite low compared to semi-urban and rural areas. Also due to higher traffic and higher Internet requirements,

APs inter-distance is quite low to ensure better connectivity to all the vehicles. In rural areas, density of vehicles is quite low and velocity is quite high. APs inter-distance is large enough. This lowers the installation cost of APs as well. The cost using Eq. (2) and transmission time using Eqs. (6) and (4) for the above three areas are calculated and handover is performed accordingly.

Table 1. Distribution of APs

Velocity		Inter-AP distance	Area
Min. (km/h)	Max. (km/h)		
20	45	0.3 km	Urban
47	65	0.6 m	Semi-urban
67	90	1 km	Rural

4 VHO Decision Including V2V Mode

VHO decision algorithm is generalized by including multi-hop V2V communication between vehicles. In multi-hop V2V communication intermediate vehicles relay data to next vehicle using ad hoc network which can act as an alternative to costlier network to some extent. The V2V delays include multi-hop relaying delays and delays when no next hop vehicle is available. These delays are quite large compared to communication delays in core network. Above delays should be investigated thoroughly particularly for delay-sensitive applications (voice/video).

4.1 Ad hoc Network Delay

Vehicles should be equipped with GPS receiver to know their accurate geographical positions. Vehicles broadcast beacon messages periodically to inform their own position to the neighbor vehicles and based on the above information vehicles update neighbors list in LUT (look-up table) [5]. Consider inter-distance between APs fixed then distance between two adjacent APs d = A + W from Fig. 1. Let ad hoc communication delay *dAH* be the time taken by a vehicle to forward the data to adjacent vehicles. This delay depends on average vehicle speed *vh* (m/s), average hop delay d_{hop} (s), average vehicle population density ρ (cars/km) at different points of highway [6]. If arrival of vehicles on highway behaves like *Poisson,* then inter-distances between vehicles on highway is exponential distribution with an average $1/\rho$ where ρ is *a*verage density of vehicles. Then,

$$dAH = (1 - e^{-\rho r})\left(\frac{A/2}{r}\right)d_{hop} + e^{-\rho r}\left(\left(\frac{A/2}{v_h}\right)\right) \tag{7}$$

where $e^{-\rho r}$ is the probability of vehicle to carry packet itself when next hop vehicle is not available within its range, and $(1 - e^{-\rho r})$ is the probability of packet forwarded in

wireless communications mode. If the packet is carried by the vehicle along a distance of $A/2$ then $\{(A/2)/vh\}$ is the packet delay while if packet goes through V2V hop, then $\{(A/2)/r\}$ is the number of hops needed to cover a distance of $A/2$, which constitutes the forwarding delay when multiplied by d_{hop}.

4.2 Wi-Fi and Ad hoc Network

The combined cost c3 of Wi-Fi and ad hoc network can be represented by

$$c3 = T_u r_w c_w \tag{8}$$

where Tu is the usage time (total time for which connection to APs remains established) represented by

$$Tu = (Nw - 1)\left(\frac{W + A}{v} - d_{AH}\right) + \frac{W}{v} + \left(\frac{x}{v} - d_{AH}\right) \tag{9}$$

and Nw maximum integer value can be found out by

$$\left[(Nw - 1)\left(\frac{W + A}{v} - d_{AH}\right) + \frac{W}{v}\right] rw \leq bt \tag{10}$$

and x is obtained from

$$bt - \left[(Nw - 1)\left(\frac{W + A}{v} - d_{AH}\right) + \frac{W}{v}\right] rw = \left(\frac{x}{v} - d_{AH}\right) rw \tag{11}$$

Note that the transmission time ($\mathbf{T_{w+AH}}$) is different from of Tu when Wi-Fi and ad hoc network is used and it is given by

$$T_{w+AH} = \frac{[(Nw \times W) + (Nw - 1)A + x]}{v} + d_{AH} \tag{12}$$

4.3 Combination of Cellular, Wi-Fi, and Ad hoc Network

The combination of Wi-Fi, cellular, and ad hoc network is beneficial when dAH does not satisfy delay requirements and the vehicle is willing to have the most economical set of access networks. In above network number of VHOs is lesser or equal to the Wi-Fi and cellular network. Thus, if c2 < c1, the cost (c4) incurred by using combination of Wi-Fi, cellular, and ad hoc network will also be c4 < c2 < c1. Similarly, if Tw < Tc, the transmission time in the case of combination of Wi-Fi, cellular, and ad hoc network will be $T_{c+w+AH} < Tw < Tc$.

5 Performance Evolution

Simulations using MATLAB 2014R were conducted for methodologies discussed above in Sects. 2–4. The latest Wi-Fi and cellular standard taken for simulation are Gigabit IEEE802.11ac Wi-Fi and 4G LTE cellular respectively. The parameters used for simulation are shown in Table 2.

Table 2. Simulation Parameters

Cost	Parameter
150 m	W (AP range)
300 m	A (Inter-AP distance)
18–90 km/h	v (Decision-making vehicle's velocity)
2–10 veh/km	ρ (Average vehicle density)
1.3 Gb/s	r_w (Data rate of WLANs)
1 Gb/s	r_c (Data rate of cellular)
8.8 Mb	b_{VHO} (Handover signaling bits)
Rs. 4.025 * 10^{-8} = 1 unit	c_w (Cost of transferring one bit in WLAN)
Rs. 8.05 * 10^{-8} = 2 units	c_c (Cost of transferring one bit in cellular)
300 Gb	b_t (Total data to be transferred)
100 m	r (Transmission range)

The other factors like *Nw* (Number of APs on the roadside), total distance traveled by the car, the preference, the area in which the node is, etc., are required from the user and the cost and time functions are calculated based on them. We have investigated performance of VHO decision algorithm for fixed and statistical inter-AP distances and then analyzed both transmission cost and transmission time versus velocity of decision-making vehicle.

5.1 VHO Decision with Fixed Inter-AP Distance

User Parameters: Total Distance the car travels: 2 km, number of APs on the road: 8.

The cost of cellular is fixed at Rs. 600 while Wi-Fi cost varies from Rs. 350 to Rs. 540 as velocity goes up from 20 to 90 kmph. User will prefer to be in Wi-Fi network because Wi-Fi frequency band is unlicensed and is less costly than the licensed cellular, which is also what we have considered in the fixed parameters in Table 2 the cost of Wi-Fi is half of the cost of cellular. If the user is currently using Cellular, it will switch to Wi-Fi by performing the Vertical Handover. The Wi-Fi data rate is faster than cellular. If the preference of the user is to minimize the transmission time of data bits or maximize the Quality of Service, it will at all times stay in Wi-Fi and not switch to cellular. Despite the wide connectivity of cellular, the node tries to stay in Wi-Fi network in fixed inter-AP distance case due to cellular higher cost and lower data rate.

5.2 VHO Decision with Statistical Inter-AP Distance

- **Urban area**. Total distance covered is 2 kms.

Considering transmission cost, the crossover takes between Wi-Fi and cellular at 28.6 kmph, the cost of Wi-Fi system was higher at lower velocity and lesser at higher velocity. Also transmission time followed similar pattern but the crossover is near 33 kmph. There is higher traffic in urban areas i.e. higher vehicle density and hence a significant amount of vehicles would be connected to a single AP which slows down the data rate and thus require higher transmission time when connected to an AP. The cost is also proportional to time because more the time it is connected to an AP, more is the cost. While at higher speeds, the vehicle can effectively switch APs at a much faster rate and expect an average overall load to be distributed among rest of the APs, hence the transmission time is less and likewise the cost.

- **Semi-urban area**. Total distance covered is 3 kms.

Due to higher velocities and greater inter-AP distances, the vehicle tends to move faster and hence the switchover of the two networks comes at a comparatively early stage compared to urban area. The vehicle tries to stay maximum in Wi-Fi network after a certain speed. The cost is low and data rate of the Wi-Fi network is usually higher than Cellular which is one of the other reasons why it tries not to switch back to cellular. However, for larger total distance of say 5 km the cost and time are at all velocities are less in cellular compared to Wi-Fi which could be due to high installation costs of APs in semi-urban areas and due to the greater inter-AP distances, the vehicle is mostly connected to cellular. So, if a node is in Wi-Fi, it will always tend to vertical handover to cellular network.

- **Rural area**. Total distance covered is 8 kms.

The interpretation of this scenario is comparable to the one in semi-urban but with greater distances. The inter-AP distance is pretty high leaving the node connected to cellular for most of the time. Also, the installation cost of APs plays a key role in rural areas, hence the vehicle tries to stay in Cellular network and will hand over from Wi-Fi at all velocities.

5.3 VHO Decision Including V2V Mode

We have included V2V mode in addition to Cellular and Wi-Fi network. The ad hoc delay calculated in Sect. 4 depends on the vehicle density, hopping delay, etc., and then the cost and transmission time is accounted for. Since, the global knowledge of the roads is not considered here, hence the user provides the vehicular density in a certain area. The hopping delay is considered to be 30 ms.

User inputs: Total Distance: 2 km, number of APs on the road: 8, vehicular Density: 5

It is expected to have the ad hoc network the minimum amount of transmission cost when compared to the other two networks because of an additive feature of car hoping

to transfer the data. The vehicles in a network communicate through DSRC and hence the cost inculcated is minimal. Hence, ad hoc network plays the best role to minimize the cost. It is observed that as long as the speed is too high, the combination of Wi-Fi, cellular, and ad hoc network yields remarkable performance improvements over the above cases. This could be because at higher speeds the inter-AP distance is covered faster compared to the case of lower speeds. Let us consider the worst case scenario with a minimal amount of vehicular traffic on the roads, and the data to be deployed by the vehicle has to be carried by itself until the next AP arrives. Thus, at lower velocities, the transmission time will automatically shoot up because the vehicle takes more time to cover the same distance while at higher speeds the adjacent AP arrival rate increases and transmission time gradually falls.

6 Conclusions

Optimal choice of access technology is needed in a vehicular network particularly in a heterogeneous wireless environment. The cost-effective VHO decision depends on capacity available in each access network, cost of transmitting data traffic in the network, and the speed of the vehicle. In this paper we have discussed a heterogeneous vehicle network using Wi-Fi and cellular network. The results show that VHO can be avoided and stay in Wi-Fi network at higher speeds while VHO is an appropriate choice in lower speeds to minimize the communication cost and time in fixed inter-AP distance. However, in the statistical inter-AP distance to stay in cellular networks is better choice at higher speeds in semi-urban and rural areas due to the extensive installation costs of APs. By including V2V communication in Wi-Fi and cellular network, the combination of Wi-Fi, cellular, and ad hoc networking outperforms in terms of transmission time and cost at higher speeds compared to the above two networking strategies.

Considering future aspects, the analysis can be extended by lifting off the assumptions made in this paper to further extent. Also, the case where the algorithm is location specific, it can be extended by adding digital maps of the area which are readily available on the Internet. This will provide more accurate vehicular density and road data of cities and states, and by incorporating them one can get more accurate results.

References

1. Kumaran U (2010) Vertical handover in vehicular ad-hoc networks—a survey. Int J Latest Trends Eng Technol (IJLTET). 3:132–138
2. Hong K, Lee S, Kim L, Song P (2009) Cost based vertical handoff decision algorithm for WWAN/WLAN integrated networks. EURASIP J Wirel Commun Netw, 1–11
3. Shafiee K, Attar A, Victor CML (2011) Optimal distributed vertical handoff strategies in vehicular heterogeneous networks. IEEE J Sel Areas Commun 29(3):534–544

4. Lee S, Sriram K, Kim K, Kim YH, Golmie N (2009) Vertical handoff decision algorithms for providing optimized performance in heterogeneous wireless networks. IEEE Trans Veh Technol 58(2):865–881
5. Zhang J, Chan H, Leung V (2006) A location-based vertical handoff decision algorithm for heterogeneous mobile networks. In: Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM 06), pp 1–5
6. Huang J, Huang Y, Wang J (2014) Vehicle density based forwarding protocol for safety message broadcast in VANET. Sci World J, 1–9

Energy-Efficient Optimum Design for Massive MIMO

Ankita Sahu(✉), Manish Panchal, and Rekha Jain

Department of Electronics and Telecommunication Engineering, Shri Govindram Seksaria Institute of Technology and Science, Indore 452003, M.P., India
{ankitasahu21, hello.rjain.me}@gmail.com,
mpanchal@sgsits.ac.in

1 Introduction

Over the past few decades, wireless access has been increased due to advancement of applications of cellular applications and internet of things. It leads to exponential growth in network traffic. Environmental and economic concerns demand to improve energy efficiency therefore research has been diverted towards optimizing energy efficiency. Future wireless communication system, e.g., 5G network which will be deployed up to year 2020 expected to achieve energy efficiency through various techniques, e.g., use of more number of base station in the smaller area, use of unused mm wave spectrum and massive MIMO [3, 4, 6, 13, 14].

In channel estimation and precoding, the circuit power consumption was very high [4]. Looking toward energy-efficienct operations in calculating channel estimation and precoding has been changed in the proposed work. A single cell of downlink massive MIMO channel is considered. A new refine model highlights the total number of complex operation performed to estimate channel information and ZF precoding. Results of the proposed work shows that by using ZF precoding and MRT technique at base station in massive MIMO regime gives area throughput that changes with increasing number of antenna at base station and also optimum value of energy efficiency with respect to optimum number of antennas at the base station and optimum number of user terminals.

The paper is organized as follows: The system model is discussed in Sect. 2. Power consumption model is described in Sect. 3. These are then used to compute optimal number of base station antennas, the optimal number of user terminals and transmit power under the assumption of perfect channel state information by using zero forcing precoding. In Sect. 4, numerical results are used to confirm the theoretical analysis. Finally, conclusion and future scopes are described in Sect. 5.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_42

2 System Model

Consider the downlink of a massive MIMO system of single cell, where a BS having M antennas transfer information to K user terminals (UTs) having single antenna. The channel is Rayleigh fading channel with zero mean and λ_k variance (Fig. 1).

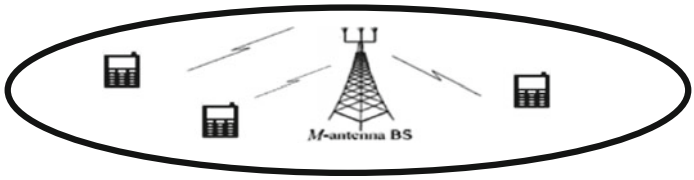


Fig. 1. A massive MIMO scenario: a circular cell with M-antenna BS and K single-antenna UTs

Let 'x' is Mx1 precoded vector of the complex information symbol transmitted from antenna of base station. The signal received by the user antenna $y \in C^{Kx1}$ is then given as [13]

$$y = \sqrt{\rho} H^H x + n \tag{1}$$

where H represents the M × K channel matrix between the M antenna at BS and the K user terminals. $n \in C^{KX1}$ is the additive white gaussian noise (AWGN) having zero mean and variance $\sigma^2 = N_0 B$. Here channel bandwidth and power spectral density of the AWGN denotes by B (Hz) and N_0 (W/Hz) respectively.

Channels are static within time–frequency coherence blocks of $T = B_c T_c$ symbols (channel uses) where B_c is coherence bandwidth and T_c is coherence time. The assumption is taken that BS and UTs are perfectly synchronized and using the time-division duplex (TDD) protocol to acquire the channel state information (CSI). The pilot signaling occupies τK symbols in uplink. Pilot signals length $\tau K < T$ for each transmission where $\tau \geq 1$ to enable orthogonal pilot sequences among the UT [2, 8, 13]. The case undertaken when the BS has perfect CSI, i.e., the BS has full knowledge of the instantaneous channel realization and possibly of the interferences statistics at the UT.

3 Generic Modal of Energy Efficiency (EE)

EE (bits/joule) is the ratio of the total number of bits transferred and total power consumed. EE might not increase only by reducing transmit power because additional power consumed from digital signal processing and analog filters used for RF and baseband processing. EE is given as

$$EE = \frac{R}{\frac{\rho KBA_{\lambda}}{\eta} + P_{CKT}} \tag{2}$$

where information rate (R) for ZF precoding for all K user is given below as [3]

$$\mathrm{R} = BK\left(1 - \frac{\tau \mathrm{K}}{T}\right)\log_2(1 + \rho(M - K)) \tag{3}$$

where $\left(1 - \frac{\tau \mathrm{K}}{T}\right)$ is pre-log factor accounts for pilot overhead. This term is included due to the fact that in training period τK time slots are used to estimate CSI during one coherence block of T. T − τK shows the number of data transmission slots and obtained information rate is to be averaged over T channel uses. Information rate can be increased by using precoding technique. For zero forcing precoding $\mathrm{V} = H(H^H \mathrm{H})^{-1}$ and for MRT V = H.

Power amplifier (PA) power is $\frac{\rho KBA_{\lambda}}{\eta}$ [3] where ρ is radiated power and $0 < \eta \leq 1$ is PA efficiency. The user distribution and propagation environment is indicated by A_{λ} as remark 1 from [3]. P_{CKT} is power consumed by other circuit systems which have described in power consumption model.

3.1 Circuit Power Consumption Model

A massive MIMO transceiver is consumed power M P_{tx} + K P_{rx} + P_{syn} watt [4]. P_{tx} is power consumed by component of each antenna attached to BS (as converters, mixers, and filters). P_{rx} power is need by circuit component of each single-antenna UT (as amplifiers, oscillator, mixer, and filters). P_{syn} is power consumed by single local oscillator which is used for all BS antenna. power is required for channel coding and decoding RK(P_{cod} + P_{decd}) watt [10], where P_{cod} and P_{decd} are the power needed for coding and decoding respectively. Fixed power P_{fix} is required for control signaling, site-cooling and by backhaul infrastructure that is independent of load power and baseband processors [4]. P_{bh} is power required for load dependent backhaul infrastructure [3]. The complete complex operations performed in the channel estimation and ZF multiuser detection, i.e., (2MKT + 4MK^2 + (8 K^3/3)) [11] that are less computed operation than assumed in [3, 4].

(1) Total number of operations calculated 2$MK^2\tau$ by multiplying M × τK matrix with τK × K matrix for computing the channel matrix ($\widehat{H}$) [11].

(2) To calculate the pseudo-inverse of $\widehat{H}$ i.e.$V = \widehat{H}\left(\widehat{H}^H\widehat{H}\right)^{-1}$, {4$MK^2$ + (8K^3/3)} total number of operations needed that is explained below:

- 2MK^2 number of operations need to calculate $\left(\widehat{H}^H\widehat{H}\right)$ by multiplying K × M matrix with M × K matrix.
- 8K^3/3 number of operations need to calculate $\left(\widehat{H}^H\widehat{H}\right)^{-1}$ by taking inversion of K × K matrix [5].

- $2MK^2$ number of operations need to calculate $\widehat{H}\left(\widehat{H}^H\widehat{H}\right)^{-1}$ by multiplying $M \times K$ with $K \times K$ matrix.

(3) $2MK\,(T - K\tau)$ number of operations need for data phase($T - K\tau$ channel uses) ZF multiuser detection by multiplying $K \times M$ matrix with $M \times 1$ vector [11].

Let a single complex operation is need $L_{0\,J}$ energy. Total number of operations described above are computed in T_c s. Hence, the average power need to compute these many operations is described below as [11].

$$P_c = 2MKL_0B_c + 4MK^2\frac{L_0}{T_c} + \frac{8K^3}{3}\frac{L_0}{T_c} \tag{4}$$

Total power needs to estimate channel information and to detect multiuser by MRT technique is $2MKL_0B_c$ [3, 12].

3.2 Energy Efficiency Optimization with ZF Processing

This section is designed to elaborate the EE model in more compact form for better understanding and comprehensibility using ZF precoding. Analytic processing to get optimize value of M, K, and ρ is proposed.

$$EE = \frac{BK\left(1 - \frac{\tau K}{T}\right)\log_2(1 + \rho(M - K))}{\frac{\rho KBA_\lambda}{\eta} + \sum_{i=0}^{3} C_iK^i + M\sum_{i=0}^{2} D_iK^i + EKB\left(1 - \frac{\tau K}{T}\right)\log_2(1 + \rho(M - K))} \tag{5}$$

Björnson et al. [4] proposed EE is given in (5) and proposed circuit power consumption coefficients C_i, D_i and E for ZF are $C_0 = P_{syn} + P_{fix}$, $C_1 = P_{rx}$, $C_2 = 0$, $C_3 = \frac{8L_0}{3T_C}$, $E = P_{cod} + P_{dec} + P_{bt}$, $D_0 = P_{tr}$, $D_1 = 2L_0B_c$, $D_2 = 4\frac{L_0}{T_c}$ and for MRT are $C_0 = P_{syn} + P_{fix}$, $C_1 = P_{rx}$, $C_2 = 0$, $C_3 = 0$, $E = P_{cod} + P_{dec} + P_{bt}$, $D_0 = P_{tr}$, $D_1 = 2L_0B_c$, $D_2 = 0$.

Lemma 3 from [4] gives the solution of the EE optimization problem explained in (5) and the behavior of z^{opt} determines by lemma 4 from [4] need to study the behavior of optimized solution of M, K, and ρ.

Optimal Transmit Power

Transmit power ρ is directly proportional to SINR and also to power amplifier (PA) transmit power under ZF processing.

Proposition 1 Optimized value of ρ is given by (6) for maximizing EE explained in (5).

$$\rho^{opt} = \frac{e^{W\left(\frac{\eta}{BA_\lambda}\frac{(M-K)(C\prime + MD\prime)}{e} - \frac{1}{e}\right) + 1} - 1}{M - K} \tag{6}$$

$$C' = \frac{\sum_{i=0}^{3} C_i K^i}{K} \quad \text{and} \quad D' = \frac{\sum_{i=0}^{2} D_i K^i}{K} \tag{7}$$

$C' > 0$ and $D' > 0$ are defined above. The proof of Proposition 1 is very similar to Theorem 3 from [4]. Only circuit power coefficients have changed. W(x) is a lambert W function of x [7].

Optimal Number of Base Station Antennas

Proposition 2 We find optimum number of base station antenna for given value of ρ *and* K as $M^{opt} = \lfloor M \rceil$. M is given by Eq. (8)

$$M = \frac{e^{\mathcal{W}\left(\frac{\rho\left(\frac{\rho BA_\lambda}{\eta} + C_l\right)}{Dle} + \frac{\rho K - 1}{e}\right) + 1} + \rho K - 1}{\rho} \tag{8}$$

where $C' > 0$ and $D' > 0$ are defined in (7). The proof of proposition 2 is very similar to theorem 2 from [4]. Only circuit power coefficients have changed. Optimum value of M is noninteger value got in a solution of objective function, but quasiconcavity tells that the M^{opt} is attained at one of the two closest integers.

Optimal Numbers of User Terminals

Proposition 3 We look forward to find optimal number of user terminals when M and ρ are given. We assume $\rho K = \bar{\rho}$ *and* $\frac{M}{K} = \bar{\beta}$ as constant with $\bar{\rho} > 0$ and $\beta > 1$. We found

$$K^{opt} = \max_i \lfloor K_i \rceil \tag{9}$$

K_i is real positive roots of the quartic equation given below

$$K^4 - \frac{2T}{\tau} K^3 - \mu_1 K^2 - 2\mu_0 K + \frac{T\mu_0}{\tau} = 0 \tag{10}$$

where

$$\mu_1 = \frac{\frac{T}{\tau}\left(C_2 + \bar{\beta} D_1\right) + C_1 + \overline{\beta} D_0}{C_3 + \bar{\beta} D_2} \quad \text{and} \quad \mu_0 = \frac{C_0 + \frac{A_\lambda B \bar{\rho}}{\eta}}{C_3 + \bar{\beta} D_2} \tag{11}$$

The notation $\lfloor . \rceil$ in (9) tells that optimal value K^{opt} is either the closest smaller or closest larger integer to K_i, that can be find out with the comparison of corresponding EE. The proof of proposition 3 is very similar to theorem 1 from [4]. Only circuit power coefficients changed.

Table 1 Simulation Parameters

Cell radius (single cell): d_{max}	250 m	P_{fix}	18 W
Minimum distance: d_{min}	35 m	P_{syn}	2 W
Path loss at distance d: λ	$\frac{10^{-3.53}}{d^{3.76}}$	P_{rx}	1 W
B	20 MHz	P_{tx}	1 W
B_c	180 kHz	P_{cod}	0.1 W/(Gbit/s)
T_c	10 ms	P_{dec}	0.8 W/(Gbit/s)
L_0 (Joule/operation)	10^{-9}	P_{bt}	0.25 W/(Gbit/s)
Pilot length τ	2	M	280
		K	180

4 Numerical Results

Simulation parameter values are taken from [1, 9] and summarized in Table 1. We made an assumption that user is uniformly distributed in a 250 m radius circular cell. The propagation parameter $\boldsymbol{A}_{\lambda}$ is calculated as in remark 1 from [3].

Figure 2 depicts that the obtained EE for various value of M and K under ZF precoding (note that $M \geq K$ due to ZF). Each value of M and K used to find the value of ρ from (6) where EE is maximized. The figure illustrates that there is a global optimal point at M = 231 and K = 180 with ρ = 1.1466 and EE = 76.25. Standard alternating optimization algorithm is taken from [4] to search the combined global optimum. Figure 2 is concave and smooth. EE is improved two times the results of Bj*ö*rnson et al. [4]. The algorithm initiation point is M = 3, K = 1, ρ = 1 taken. The algorithm converged after 8 iterations to a suboptimal solution.

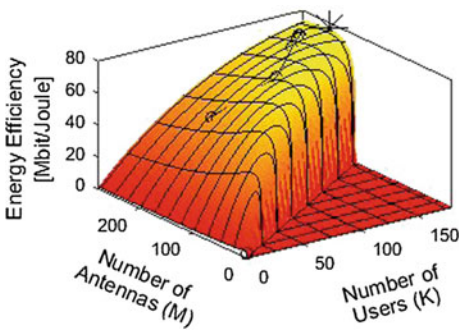


Fig. 2. Under the consideration of single-cell scenario, the variation of energy efficiency (in Mbit/J) for different combinations of M and K with ZF precoding. The global optimum is indicated by star while the circles shows the obtained convergence point of alternating optimization algorithm

For comparisons Fig. 3 shows energy efficiency for MRT precoding. Its result was generated by Monte Carlo simulations while Fig. 2 was computed using our analytic results. EE optimal value is achieved at M = 123, K = 99. EE in MRT case is improved 6.4% than the result of Bj*ö*rnson et al. [4].

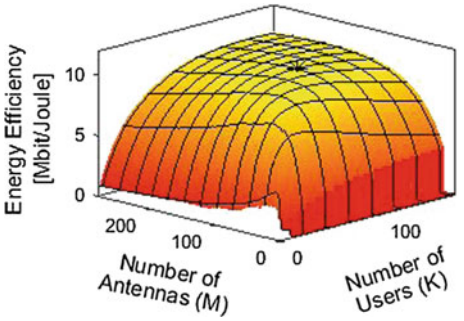


Fig. 3. Under the consideration of single-cell scenario, the variation of energy efficiency (in Mbit/J) for different combinations of M and K with MRT. The global optimum is marked with a star

By comparing Fig. 2, 3 it is deduce that EE is better in ZF than MRT under perfect CSI in single-cell scenario. Figure 4 shows the area throughput at maximizing point of the EE for different M and at that point area throughput is increased than result of Bj*ö*rnson et al. [4]. Area throughput is very high for ZF than MRT processing. Area throughput at optimizing point of EE is 181.2 Gbits/s. for ZF case and 9.99 Gbits/s. for MRT case which shows three and two times improvement than the result of Bj*ö*rnson et al. [4] respectively.

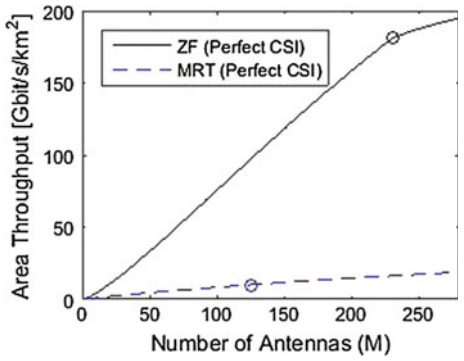


Fig. 4. Under the consideration of single–cell scenario, the variation of Area Throughput at the EE-maximizing solution for different number of BS antenna M

A brief summary of work done and proposed methodology in massive MIMO for ZF and MRT described in Table 2. From the analysis we deduce that optimal point of EE is improved with increasing M and K up to a particular limit.

Table 2. Comparison on various parameter employed in different literature

Properties	Björnson et al. [3]		Björnson et al. [4]		Proposed work	
	ZF	MRC	ZF	MRC	ZF	MRC
Energy efficiency	7	1.5	30.7	9.86	76.25	10.49
M^{opt}	165	4	165	81	231	123
K^{opt}	85	1	104	77	180	99

5 Conclusion

In this paper, circular single cell equipped with M number of antenna at base station and K user terminals with single antenna. Information rate is used under ZF precoding. A new refine model for power consumption is acquired hence the EE is increased with the increment of optimum number of antenna at base station and user terminals. Linear precoding scheme only useful to mitigate intracell interference. In multi-cell scenario, pilot contamination diminishes the linear precoder advantage. To mitigate inter-user interference complex nonlinear precoding technique like dirty paper coding and maximal likelihood detection technique should be used. Multi-cell scenario considered as future work.

References

1. Auer G et al D2.3: Energy efficiency analysis of the reference systems, areas of improvements and target breakdown. In: INFSO-ICT-247733 EARTH, ver. 2.0 (2012) [Online]. http://www.ict-earth.eu/
2. Björnson E, Hoydis J, Kountouris M, Debbah M (2014) Massive MIMO systems with non-ideal hardware: energy efficiency, estimation, and capacity limits. IEEE Trans Inf Theor 60(11):7112–7139
3. Björnson E, Sanguinetti L, Hoydis J, Debbah M (2014) Designing multi-user MIMO for energy efficiency: When is massive MIMO the answer? In: Proceedings of the IEEE wireless communications and networking conference (WCNC), vol 3, pp 242–247, Apr 2014
4. Björnson E, Sanguinetti L, Hoydis J, Debbah M (2015) Optimal design of energy-efficient multi-user MIMO systems: is massive MIMO the answer? IEEE Trans Wirel Commun 14 (6):3059–3075
5. Boyd S, vandenberghe L Numerical linear algebra background. http://www.ee.ucla.edu/ee236b/lectures/num-lin-alg.pdf
6. Gupta A, Jha RK (2015) A survey of 5G network: architecture and emerging technologies. IEEE J 3:1206–1232
7. Hoorfar A, Hassani M (2008) Inequalities on the Lambert W function and hyperpower function. J. Inequalities Pure Appl Math 9(2):1–5

8. Hoydis J, Ten Brink S, Debbah M (2013) Massive MIMO in the UL/DL of cellular networks: How many antennas do we need? IEEE J Sel Areas Commun 31(2):160–171
9. Kumar R, Gurugubelli J (2011) How green the LTE technology can be? In: Proceedings of wireless VITAE
10. Mezghani A, Nossek JA (2011) Power efficiency in communication systems from a circuit perspective. In: Proceedings of IEEE international symposium on circuits and systems (ISCAS), pp 1896–1899
11. Mohammed SK (2014) Impact of transceiver power consumption on the energy efficiency of zero-forcing detector in massive MIMO systems. IEEE Trans Commun 62(11):3874–3890
12. Mukherjee S, Mohammed S (2014) On the energy-spectral efficiency trade-off of the MRC receiver in massive MIMO systems with transceiver power consumption. http://arxiv.org/abs/1404.3010
13. Ngo H, Larsson E, Marzetta T (2013) Energy and spectral efficiency of very large multiuser MIMO systems. IEEE Trans Commun 61(4):1436–1449
14. Yang H, Marzetta T (2013) Total energy efficiency of cellular large scale antenna system multiple access mobile networks. In: Proceedings of IEEE online green communication, pp 27–32

Reckoning of Music Rhythm Density and Complexity through Mathematical Measures

Sudipta Chakrabarty[1(✉)], Gobinda Karmakar[1], Md. Ruhul Islam[2], and Debashis De[3]

[1] Department of MCA, Techno India, Salt Lake, Kolkata, West Bengal, India
chakrabarty.sudipta@gmail.com
[2] Department of CSE, SMIT, Majhitar, Rangpo, East Sikkim, India
[3] Department of CSE, MAKAUT, Kolkata, West Bengal, India

1 Introduction

Rhythm can be considered as the most important feature of music. Furthermore, Density and Complexity are the two primary features of Rhythm in Indian music. Therefore, measures of the rhythm density and rhythm complexity of a particular rhythm are very useful in the field of Music Information Retrieval (MIR). Rhythm density of a particular rhythmic phrase is rate of stroke beats present in the total number of beats or meter of the rhythm. Rhythm Complexity is the sum total of the degree of syncopation of each rhythmic phrase in the particular rhythm divide by the total meter length. In Indian music, Syncopation is the unexpected element to a predictable rhythmic pattern. The paper proposes a mathematical concept for computing the density and complexity of a particular rhythm structure by using rank percentage, relative rank percentage, indispensability vectors, and degree of syncopation of the fundamental frequencies of a particular rhythm structure as follows:

$$\text{Rank}\,\% = \frac{\text{Total Frequency}}{\text{Sum of Total Frequencies}} \tag{1}$$

$$\text{Relative Rank}\,\% = \frac{\text{Individual Rank}\,\%}{\text{Individual Highest Rank}\,\%} \tag{2}$$

$$\text{Indispensability Vector} = 1 - \text{Relative Rank}\,\% \tag{3}$$

$$\text{Degree of Syncopation} = \text{Indispensability Vector} \times \text{Binary Encoding value} \tag{4}$$

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_43

2 Related Works

Versatile rhythms generation is one of the common events of most of the music composers. In paper [1] introduces a machine recognition technique for generation of musical patterns and proposed several measures of rhythm complexity and a novel technique to resolve a restricted tonal context. Some other papers [2, 3] deal with the implementation in the field of Pattern Recognition of music by mathematical expressions. Again consider some other papers that are based on the creativity of music using Genetic Algorithm technique. The contributions describe the study of usefulness of Genetic Algorithm [4, 5] in the field of music composition. The paper [6] introduces a offspring rhythm from a predefined set of rhythms applied to initial population using Genetic Algorithm operator. Another paper proposes the concept of generating realistic drum-set rhythm automatically using Genetic Algorithm [7, 8]. Some researchers present a system in paper [9] in the field of pattern recognition in Indian music by the help of Median Filter and it also proposed three measurements of generating rhythm complexity. Some researchers have been described the musical pattern recognition and rhythmic features retrieval by object-oriented concept [10, 11]. Some papers describe the classification of different music features using Petri Nets and also describe the implementations of musical percussion-based tempo [12, 13].

The genetic algorithm is a search procedure to find the fittest rhythm structure to improve the quality of the Music. The primary objective of the paper [14] is to create a pervasive teaching-learning tool for music rhythm learner in context awareness. This contribution is a music rhythm developing tool to create a versatile music rhythm structure from a set of rhythms that contain more melody and tempo. Computational Musicology is an emerging field that fully depends on Computer Science. A few works have taken place in this area in Indian Music. Raga or music origin is the backbone of Indian Classical Music. Automatic Raga Recognition is identified by finding the note structures by mapping the fundamental frequencies of each note and Pitch Contour data values associated with that particular song and again matching these notes with Raga Knowledgebase [15]. Another very good approach has been incorporated in a paper that song origin and song can be represented by the Unified Modeling Language (UML) [16]. Music recommendation system is one of the most intelligent tools that classifies different songs at different time slots of a particular day [17] and proposes another time-based raga recommender using Neural Network [18]. The paper [19] proposes a statistical method that is classifying the similar song patterns depends on coefficient of variance.

3 Proposed Work

Music rhythm density and complexity are the two primary features of Rhythm in Indian music. Therefore, measures of the rhythm density and rhythm complexity of a particular rhythm are very useful in the field of Music Information Retrieval (MIR). Rhythm density of a particular rhythmic phrase is the rate of stroke beats present in the total number of beats or meter of the rhythm. Rhythm complexity is the sum total of the

degree of syncopation of each rhythmic phrase in the particular rhythm divide by the total meter length. The workflows of the proposed work are given below:
Step 1: Take all the rhythms.
Step 2: Click on the wave surfer button to open the software. To Convert the song in f0 format and extract the frequencies the steps are follows:
Click on File → Open and choose a song
Click on Transform → Convert → sample rate = 22050 and channel = Mono
Right click on the black line → Create Pane → Pitch Contour
Right click on the black dots → Properties → Pitch Contour → Pitch Method → AMDF (Average Magnitude Difference Formula) → Apply → Ok
Right click on the black dots → Save the data file → Save the file with .xlsx extension
Step 3: Accepted all those frequency values ranges from 1 to 500 of the rhythms from their corresponding .f0 file.
Step 4: Then the number of occurrences has been calculated of all the fundamental frequencies of each rhythms.
Step 5: Fix the number frequencies that is equal the number of meter of the rhythm which has the highest occurrence. respectively, from the list of frequencies of the .fo file of each of the rhythms.
Step 6: Calculate total frequency by the following formula-

$$\text{Total frequency} = \text{Frequency} \times \text{Occurrence} \tag{5}$$

Step 7:

$$\text{Calculate Rank}\,\% = \frac{\text{Total Frequency}}{\text{Sum of Total Frequencies}} \tag{6}$$

Step 8: In Value Encoding, taking numeric value 1 for single stroke beat and 2 for double stroke beat.
Step 9: Then this individual beat values are converted into equivalent binary digits in Binary Encoding Mechanism.
Step 10:

$$\text{Calculate Rhythm Density} = \frac{\text{The Meter Length of the Rhythm}}{\text{Total number of beats in the Binary Encoding}} \tag{7}$$

Step 11:

$$\text{Calculate Relative Rank}\,\% = \frac{\text{Individual Rank}\,\%}{\text{Individual Highest Rank}\,\%} \tag{8}$$

Step 12:

$$\text{Compute Indispensability Vector} = 1 - \text{Relative Rank}\,\% \tag{9}$$

Step 13:

$$\text{Compute Degree of Syncopation} = \text{Indispensability Vector} \times \text{Binary Encoding value} \tag{10}$$

Step 14:

$$\text{Calculate Rhythm Complexity} = \frac{\sum \text{Degree of Syncopation}}{\text{The Meter Length of the Rhythm}} \tag{11}$$

4 Result Set Analysis

To establish the proposed work, there are a lot of Rhythm Structures of Meter 16 have been taken and find their similarity mapping using our proposed workflow. In this paper, consider three rhythm structures and compute their Densities and Complexities. Table 1 depicts the rhythm structure and their value and binary encoding, Fig. 1. represents the bar diagram of frequency versus occurrence, and Table 2 depicts the density and complexity of rhythm 1. Table 3 depicts the rhythm structure and their value and binary encoding, Fig. 2. represents the bar diagram of frequency versus occurrence, and Table 4 depicts the density and complexity of rhythm 2. Table 5 depicts the rhythm structure and their value and binary encoding, Fig. 3. represents the bar diagram of frequency versus occurrence, and Table 6 depicts the density and complexity of rhythm 3.

Test Case 1:

See Tables 1, 2 and Fig. 1.

Table 1. Rhythm 1 and its value and binary encoding

Rhythm 1	dha	dha	ti	te	dha	dha	tu	na	ta	ta	ti	te	dha	dha	tu	na
Value encoding	2	2	1	1	2	2	1	1	1	1	1	1	2	2	1	1
Binary encoding	10	10	1	1	10	10	1	1	1	1	1	1	10	10	1	1

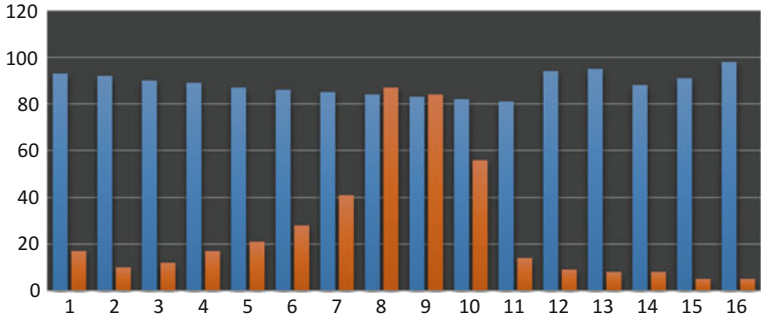


Fig. 1. Bar diagram of frequency value and their corresponding occurrences of rhythm 1

Table 2. Measuring rhythm density and complexity of rhythm 1

Frequency	Occurrence	Total frequency	Rank %	Binary encoding	Rhythm density	Relative rank %	Indispensability vector	Degree of syncopation	Rhythm complexity
93	17	1581	0.043825	10	0.72727	0.216338	0.783661781	7.836617805	3.43531933
92	10	920	0.025502	10		0.125889	0.874110587	8.74110587	
90	12	1080	0.029938	1		0.147783	0.852216776	0.852216776	
89	17	1513	0.04194	1		0.207033	0.79296665	0.79296665	
87	21	1827	0.050644	10		0.25	0.750000046	7.500000462	
86	28	2408	0.06675	10		0.329502	0.670498145	6.704981452	
85	41	3485	0.096604	1		0.476875	0.52312543	0.52312543	
84	87	7308	0.202578	1		1	0	0	
83	84	6972	0.193264	1		0.954023	0.045977188	0.045977188	
82	56	4592	0.12729	1		0.628352	0.371647626	0.371647626	
81	14	1134	0.031435	1		0.155172	0.844827615	0.844827615	
94	9	846	0.023451	1		0.115764	0.884236475	0.884236475	
95	8	760	0.021067	10		0.103996	0.896004398	8.96004398	
88	8	704	0.019515	10		0.096333	0.903667232	9.036672318	
91	5	455	0.012613	1		0.062261	0.937739475	0.937739475	
98	5	490	0.013583	1		0.06705	0.932950204	0.932950204	

Test Case 2:

See Tables 3, 4 and Fig. 2.

Table 3. Rhythm 2 and its value and binary encoding

Rhythm 2	dha	din	din	ta	dha	din	din	ta	na	tin	tin	na	ta	din	din	ta
Value encoding	2	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1
Binary encoding	10	1	1	1	10	1	1	1	1	1	1	1	1	1	1	1

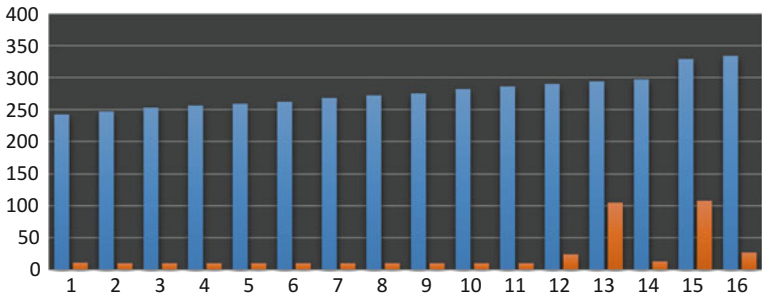


Fig. 2. Bar diagram of frequency value and their corresponding occurrences of rhythm 2

Table 4. Measuring rhythm density and complexity of rhythm 2

Frequency	Occurrence	Total frequency	Rank %	Binary encoding	Rhythm density	Relative rank %	Indispensability vector	Degree of syncopation	Rhythm complexity
242	11	2662	0.023047	10	0.88889	0.074918	0.925081526	9.250815257	1.83868915
247	10	2470	0.021385	1		0.069515	0.930485112	0.930485112	
253	10	2530	0.021904	1		0.071204	0.928796491	0.928796491	
256	10	2560	0.022164	1		0.072048	0.927952181	0.927952181	
259	10	2590	0.022424	10		0.072892	0.927107871	9.271078706	
262	10	2620	0.022683	1		0.073736	0.92626356	0.92626356	
268	10	2680	0.023203	1		0.075425	0.924574939	0.924574939	
272	10	2720	0.023549	1		0.076551	0.923449192	0.923449192	
275	10	2750	0.023809	1		0.077395	0.922604882	0.922604882	
282	10	2820	0.024415	1		0.079365	0.920634824	0.920634824	
286	10	2860	0.024761	1		0.080491	0.919509077	0.919509077	
290	24	6960	0.060258	1		0.19588	0.804119992	0.804119992	
294	105	30870	0.267266	1		0.868795	0.13120462	0.13120462	
297	13	3861	0.033428	1		0.108663	0.891337254	0.891337254	
329	108	35532	0.307628	1		1.000001	0	0	
334	27	9018	0.078076	1		0.2538	0.7462003	0.7462003	

Test Case 3:

See Tables 5, 6 and Fig. 3.

Table 5. Rhythm 3 and its value and binary encoding

Rhythm 3	dha	dha	ti	te	dha	dhin	dhin	dha	ta	ta	ti	te	ta	dhin	dhin	dha
Value encoding	2	2	1	1	2	2	2	2	1	1	1	1	1	2	2	2
Binary encoding	10	10	1	1	10	10	10	10	1	1	1	1	1	10	10	10

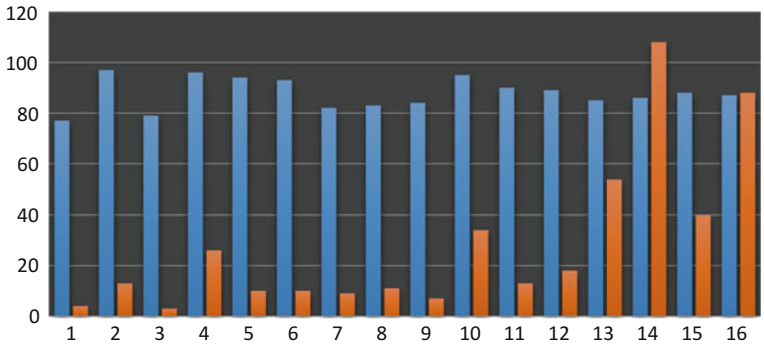


Fig. 3. Bar diagram of frequency value and their corresponding occurrences of rhythm 3

Table 6. Measuring rhythm density and complexity of rhythm 3

Frequency	Occurrence	Total frequency	Rank %	Binary encoding	Rhythm density	Relative rank %	Indispensability vector	Degree of syncopation	Rhythm complexity
77	4	308	0.007804	10	0.64	0.033161	0.966838953	9.668389531	4.24932112
97	13	1261	0.031951	10		0.135766	0.864233506	8.642335059	
79	3	237	0.006005	1		0.025517	0.97448322	0.97448322	
96	26	2496	0.063243	1		0.268734	0.731266321	0.731266321	
94	10	940	0.023817	10		0.101206	0.898794207	8.987942074	
93	10	930	0.023564	10		0.100129	0.899870865	8.998708648	
82	9	738	0.018699	10		0.079457	0.920542686	9.205426862	
83	11	913	0.023133	10		0.098299	0.901701182	9.017011823	
84	7	588	0.014899	1		0.063307	0.936692547	0.936692547	
95	34	3230	0.081841	1		0.34776	0.65223967	0.65223967	
90	13	1170	0.029645	1		0.125969	0.874031088	0.874031088	
89	18	1602	0.040591	1		0.172481	0.82751949	0.82751949	
85	54	4590	0.1163	1		0.494186	0.505814268	0.505814268	
86	108	9288	0.235336	10		0.999999	0	0	
88	40	3520	0.089188	10		0.378983	0.621016606	6.210166064	
87	88	7656	0.193985	10		0.824289	0.175711119	1.75711119	

5 Conclusion

In this work, an intelligent prototype has been proposed to find the rhythmic density and rhythmic complexity of any particular rhythm structure. The measures of rhythm density and complexity of a particular rhythm is very useful in the field of Music Information Retrieval (MIR) as versatile rhythms generation is one common event for most of the music composers. Rhythm density of a particular rhythmic phrase is the rate of stroke beats present in the total number of beats or meter of the rhythm and rhythm Complexity is the sum total of degree of syncopation of each rhythmic phrase in the particular rhythm divide by the total meter length.

The main objective of this contribution is to explore the nature of any rhythm structure from a set of rhythms to compute the density and complexity for practical implementation in Indian music as well as world music. Performance evaluation approaches for this system help to evaluate their systems and have conducted a comprehensive testing of near about hundred rhythm structures to find their complexity and density. The work has been implemented by some mathematical methods like rank percentage, relative rank percentage, indispensability vectors, and degree of syncopation. This contribution is implemented as a step towards developing tools to the music composer as well as music listeners to evaluate the density and complexity of a particular rhythm that has more versatile features with more tempos.

References

1. Shmulevich I, Yli-Harja O, Coyle EJ, Povel D, Lemstrm K (2001) Perceptual issues in music pattern recognition complexity of rhythm and key finding. In: Computers and the humanities. Kluwer Academic Publishers, pp 23–35
2. Chakraborty S, De D (2012) Pattern classification of Indian classical ragas based on object oriented concepts. Int J Adv Comput Eng Archit 2:285–294

3. Chakraborty S, De D (2012) Object oriented classification and pattern recognition of Indian classical ragas. In: Proceedings of the 1st international conference on recent advances in information technology (RAIT). IEEE, pp 505–510
4. Gartland-Jones A, Copley P (2003) The suitability of genetic algorithms for musical composition. Contemp Music Rev 22(3):43–55
5. Matic D (2010) A genetic algorithm for composing music. Proc Yugosl J Oper Res 20 (1):157–177
6. Dostal M (2005) Genetic algorithms as a model of musical creativity—on generating of a human-like rhythimic accompainment. Comput Inform 22:321–340
7. Alfonceca M, Cebrian M, Ortega A (2006) A fitness function for computer-generated music using genetic algorithms. WSEAS Trans Inf Sci Appl 3(3):518–525
8. Chakrabarty S, De D (2012) Quality measure model of music rhythm using genetic algorithm. In: Proceedings of the international conference on RADAR, communications and computing (ICRCC). IEEE, pp 203–208
9. Bhattacharyya M, De D (2012) An approach to identify thhat of Indian classical music. In: Proceedings of international conference of communications, devices and intelligence system (CODIS). IEEE, pp 592–595
10. De D, Roy S (2012) Polymorphism in Indian classical music: a pattern recognition approach. In: Proceedings of international conference on communications, devices and intelligence system (CODIS). IEEE, pp 632–635
11. De D, Roy S (2012) Inheritance in Indian classical music: an object-oriented analysis and pattern recognition approach. In: Proceedings of international conference on RADAR, communications and computing (ICRCC). IEEE, pp 296–301
12. Roy S, Chakrabarty S, Bhakta P, De D (2013) Modelling high performing music computing using Petri Nets. In: Proceedings of international conference on control, instrumentation, energy and communication. IEEE, pp 757–761
13. Roy S, Chakrabarty S, De D (2013) A framework of musical pattern recognition using Petri Nets. In: Emerging trends in computing and communication 2014. Springer-Link Digital Library, Submission No-60
14. Chakrabarty S, Roy S, De D (2014) Pervasive diary in music rhythm education: a context-aware learning tool using genetic algorithm. In: Advanced computing, networking and informatics, vol 1. Springer International Publishing, pp 669–677
15. Chakrabarty S, Roy S, De D (2013) Automatic raga recognition using fundamental frequency range of extracted musical notes. In: Proceedings of the eight international multiconference on image and signal processing (ICISP 2014). Elsevier
16. Chakrabarty S, Gupta P, De D (2015) Behavioural modelling of ragas of Indian classical music using unified modelling language. In: Proceedings of the second international conference on perception and machine intelligence (PerMIn '15). ACM Digital Library, pp 151–160
17. Chakrabarty S, Roy S, De D (2016) Time-slot based intelligent music recommender in Indian music, International Book Chapter, ISBN13: 9781522504986, ISBN10: 1522504982. IGI Global, USA
18. Roy S, Chakrabarty S, De D (2017) Time-based raga recommendation and information retrieval of musical patterns in Indian classical music using neural network. IAES Int J Artif Intell (IJ-AI), ISSN: 2252-8938, pp 33–48
19. Chakrabarty S, Islam MdR, De D (2017) Modelling of song pattern similarity using coefficient of variance. Int J Comput Sci Inf Secur, ISSN 1947-5500, pp 388–394

Medical Diagnostic Models an Implementation of Machine Learning Techniques for Diagnosis in Breast Cancer Patients

Rupam Borah(✉), Sunil Dhimal, and Kalpana Sharma

Sikkim Manipal Institute of Technology, Majitar, Rangpo 737136, Sikkim, India
{rupamborah663, sunildhimal, headcs. smit}@gmail. com

1 Introduction

1.1 General Overview

Data generated in the healthcare industry has exceeded in volume compared to the techniques used in its analysis for over a long period of time. However, the procedures for inferring and standardizing meaning from it are still in an active development phase.

With the advancement in computational capability of problem-solving, there is a need for it in the field of medical applications, where machine learning techniques are to be tested for their ability in behaving as diagnostic tools. This project aims to demonstrate the working and the accuracy of a few machine learning models on a given set of data on breast cancer and also, making a comparison between them to determine the best model suitable of a particular paradigm.

1.2 Literature Survey

Prerequisite Activity for Data Collection Details Finite Needle Aspirate (FNA): Fine-needle Aspiration (FNA) is an efficient procedure used to investigate cancerous lumps or masses in the breast area. In this method, a thin, hollow needle is inserted into the mass to collect a small sample of cells, which are examined under a microscope to analyze their similarity with cancer symptoms.

FNA biopsies are very safe, minor surgical procedures. Often, a major surgical procedure which is unnecessary can be avoided by performing an FNA instead. Currently, this procedure is widely used in the diagnosis of cancer.

Concepts of Individual Machine Learning Models:
Decision Trees. Belonging to the paradigm of divide and conquer, decision tree learners are powerful classifiers, which utilize a tree structure to model the relationships among the features and the potential outcomes. One of the most well-known implementations is the C5.0 algorithm. A few characteristics of decision trees taken into consideration are as follows:

- Entropy: The degree to which a subset of examples contains only a single class is known as purity, and any subset composed of only a single class is called pure.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_44

$$\text{Entropy}\,(S) = \sum_{i=1}^{c} -p_i \log_2(p_i) \tag{1}$$

Typically, entropy is measured in bits. If there are only two possible classes, entropy values can range from 0 to 1. For n classes, entropy ranges from 0 to $\log_2(n)$ [1].

- Information Gain: To use entropy to determine the optimal feature to split upon, the algorithm calculates the change in homogeneity that would result from a split on each possible feature, which is a measure known as information gain. Higher the information gain, the more efficiently the feature can be classified into well-known groups [1]. The information gain for a feature F is calculated as the difference between the entropy in the segment before the split (S1) and the groups resulting from the split (S2):

$$\text{InforGain}\,(F) = \text{Entropy}\,(S_1) - \text{Entropy}\,(S_2) \tag{2}$$

K-Nearest Neighbors. The k-NN algorithm belongs to the lazy learner paradigm [1]. Classification algorithms based on the nearest neighbor methods are considered lazy learning algorithms. It merely stores the training data verbatim. A few characteristics of k-NN algorithm taken into consideration are as follows:

- Measuring similarity of new data points with distance: The new incoming data is compared to existing data points to fit them into a certain class.
- Choosing an appropriate k: The optimal value of k and hence, the best model are chosen using repeated cross-validation by the virtue of accuracy.

Support Vector Machines. Support vector machines fall under a paradigm of black box methods. An SVM conceptualizes a plane surface that creates a separation between input data samples of varying types that represent examples and their feature values. A few characteristics of support vector machines are as follows:

- Ability to separate nonlinearly separable data: This is done with the use of a slack variable, which casually allows some input data to fall on the wrong side of the plane surface. A restriction to this feature called cost value (denoted as C) is applied to all input data that determines the diagnosis, and then, algorithm attempts to minimize the cost of slack.
- Kernel Transformation: SVMs are able to map the input data of certain dimensions into a mapping result, which is simpler to understand (differentiating data points is easier) using a process known as the kernel trick [1]. In doing so, a nonlinear relationship may be converted into a linear relationship.

Neural Networks. Neural networks also fall under the paradigm of black box methods [1]. A neural network has the following key components:

- A mapping known as activation function which takes input data and produces a single output *would* determine the diagnosis.
- The topology of the network graphically denotes the way the neurons are connected, the number of *layers*, *and* their connections.
- The algorithm to train the network that specifies how much weightage is given to each data sample in order to produce the desired output.

Resampling Techniques for Model Evaluation and Performance:
K-fold Cross-validation. In k-fold cross-validation, the original dataset is randomly divided into k number equal-sized subsamples. A single subsample is selected as the testing data for validating the model, and the remaining k − 1 subsamples are used as training data. Usually k is set to 10 [1], but in general, k remains as an unfixed parameter.
Bootstrapping. The original dataset of n examples is used by this procedure to create one or more new training datasets that will also contain n examples (some of the datasets may be redundant). The remaining datasets which were not used as training datasets are then used as testing datasets.

2 Design Strategy

2.1 Elicitation of Outcomes and Findings from Patient Data

The datasets for breast cancer were available from a website known as the UCI Machine Learning Repository. Characteristics of the cell nuclei present in a breast mass from a Finite Aspirate Needle test recorded used as features. 569 samples in total were used for analysis.

Ten real-valued features computed for each cell nucleus [2] are as follows:

- Radius of the cell.
- Texture of the cell.
- Perimeter of the cell border.
- Surface area of the cell.
- Smoothness (Change in radius of neighbor cells).
- Compactness of collective cells = $(\text{Perimeter}^2)/(\text{Area}-1)$.
- Numerical value to estimate concavity of the cells.
- Number of concave regions of the center of the cell.
- Symmetry of the cell.
- Fractal dimension.

The mean, standard error, and worst (mean of the largest values) were calculated for each feature, resulting in 30 features. The outcome to be predicted is in the form of response, “Benign” or “Malignant”, to define if the breast mass is benign or malignant. The total amount of samples used for training the models is 469, while the remaining 100 were used as test cases.

2.2 Workflow

Initially, we have two sets of data, one for medical diagnosis of the nature of the breast mass cells, and the other for text mining out of clinical notes. The FNA measures data is split into training and testing parts for the purpose of model learning and model validation. The training data is used as input prior to the training of the model. This data also has the outcomes along with it, so supervised learning is implemented. Each model is fit with the training data. The models are resampled with respect to the training data itself using either k-fold cross-validation or bootstrapping. Each algorithm produces its best model based on accuracy. The predictions are made based on the testing data by each model and their individual accuracies are determined. The best model of a particular algorithm is chosen by comparing the accuracies of all the models (Fig. 1).

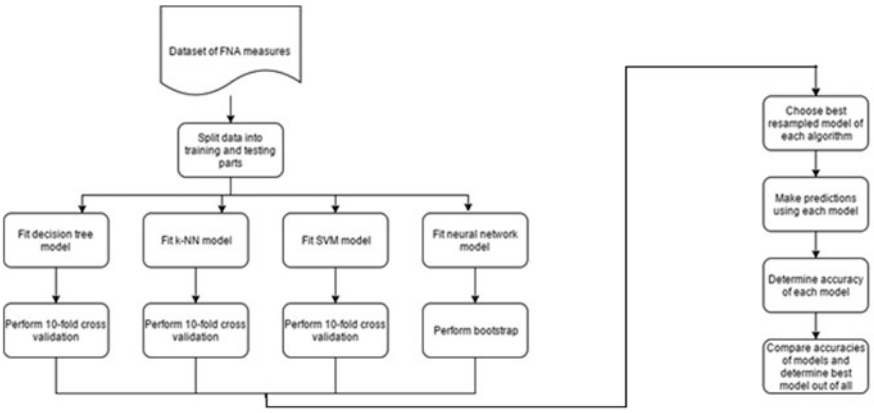


Fig. 1. Detailed workflow of procedures involved in implementation

3 Implementation Details

3.1 Attribute Correlation Analysis

Correlation analysis for features was performed and the results were summarized (Fig. 2).

3.2 Preliminary Tasks Prior to Training

Before training the models, the parameters for training have to be specified in the form of a training controller and/or tuning grid, which would look over the entire process of the model's learning phase. No explicit parameters are required for k-NN and SVM algorithms For C5.0 decision tree, the parameters are as follows:

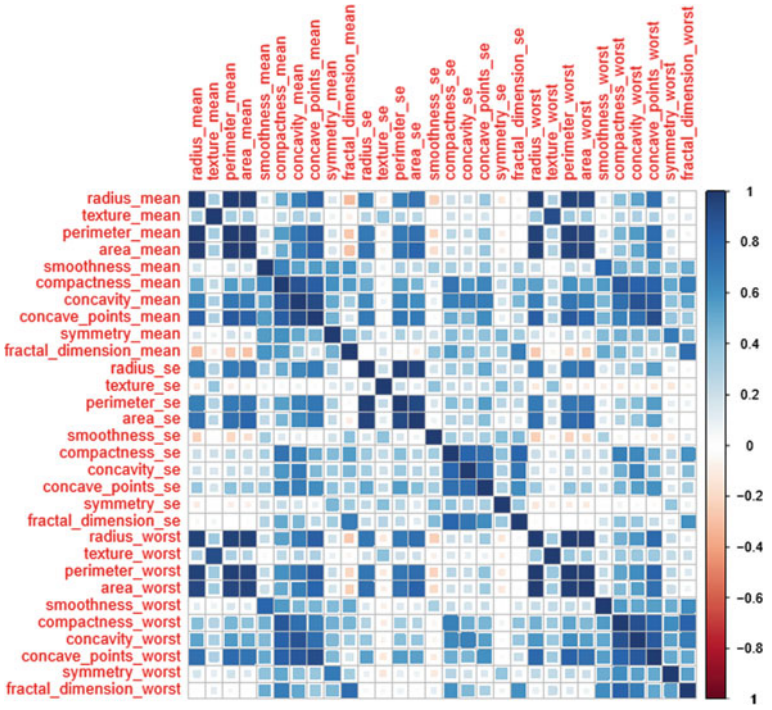


Fig. 2. Correlation matrix for feature values. Dark blue indicates strong positive correlation while dark red indicates strong negative correlation

- *Boosting factor*: In boosting, several classifiers are generated, i.e., decision trees or rule sets, rather than just one. When a new case is to be classified, each classifier votes for its predicted classes and the votes are counted to determine the final class.
- *Winnowing factor*: In winnowing, the most relevant features are selected for prediction, in case there are many numbers of attributes.

For neural networks, the parameters are as follows:

- *Weight decay*: Used in backpropagation method of training, gradual decay of weights of input variables as machine inclines toward an optimal decision.
- *No. of hidden layer nodes*: Number of intermediate nodes required for the process.

No explicit parameters were needed to be specified in respect of k-NN algorithm and linear SVM algorithm (Table 1).

The decision tree models and the neural network models were shown in a graphical layout as they provide a better understanding of how the features were used for the diagnosis procedure. In Fig. 3, we have the modeled decision tree, which was later used for the prediction of the outcomes. In Fig. 4, the designed neural network with three levels (input, hidden, and outcome) is shown.

Table 1. Testing and tuning grid specifications

Model Name	Training controller used	Tuning Grid			Training/Testing data size
C5.0 Decision tree	10-cross fold validation for 10 iterations	**Iterations**	**Parameter 1**	**Parameter 2**	Training Data 469 samples Testing Data 100 samples
		10	Winnow(T/F)	Boosting Factor(5,10,15,20)	
k-NN	10-cross fold validation for 10 iterations	No tuning grid required			
Linear SVM	10-cross fold validation for 10 iterations	No tuning grid required			
Neural network	Bootstrapping for 25 iterations	10	Weight decay (0.1, 0.2, 0.5)	Hidden layer nodes (5, 6 ,7)	

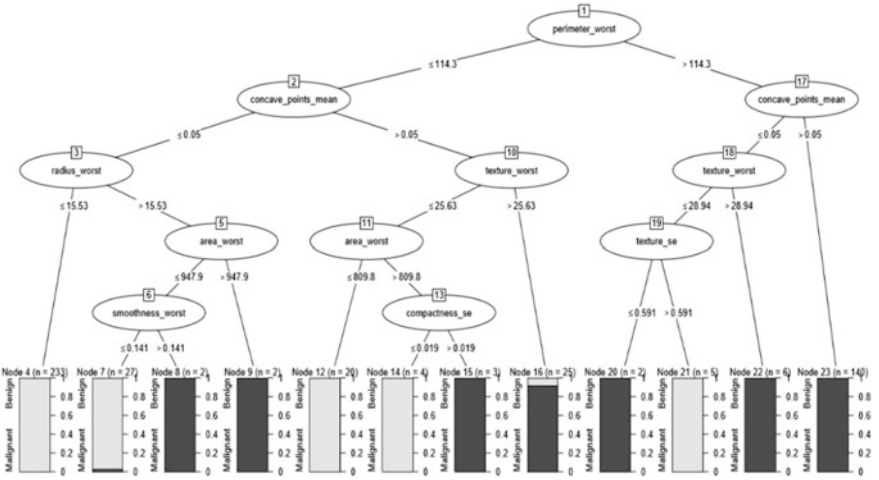


Fig. 3. Modeled decision tree for diagnosis

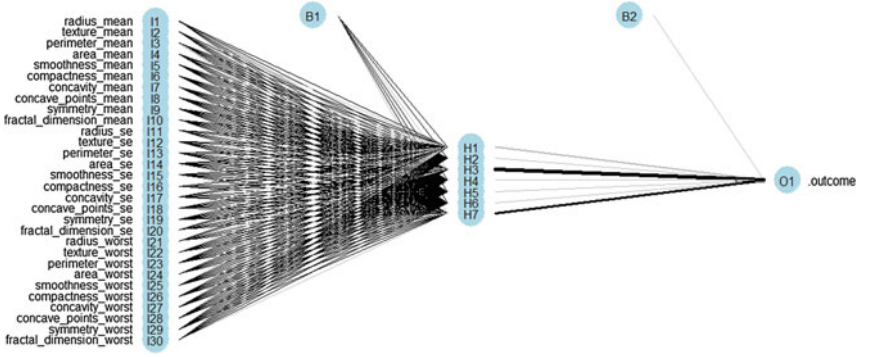


Fig. 4. Layout of neural network

4 Results and Discussion

The predictions of the models and their accuracies are shown in the following table and charts. The green and red values indicate number of correct and incorrect diagnoses (Table 2 and Fig. 5).

Inference of Chart Values:
For C5.0, the number of voting models was chosen as 15 and no winnowing option to make the result optimal. For k-NN algorithm, nine neighbors have been chosen as optimal, where this value is the nature of the dataset. For optimal accuracy by linear SVM, the cost factor is at 0.1, which indicates that there is a very tight restriction on values falling on the wrong side of the plane. Finally for neural networks, the weight decay is at 0.1 and no. of hidden layer nodes show that slower and more gradual decay in weight values ensures better learning.

Table 2. Confusion matrix for the implemented models

	Actual diagnosis/Predicted diagnosis	Benign	Malignant	Total
C5.0 predictions	**Benign**	68	0	68
	Malignant	3	29	32
k-NN predictions	**Benign**	71	6	77
	Malignant	1	22	23
SVM predictions	**Benign**	68	0	68
	Malignant	2	30	32
Neural network predictions	**Benign**	71	6	77
	Malignant	1	22	23

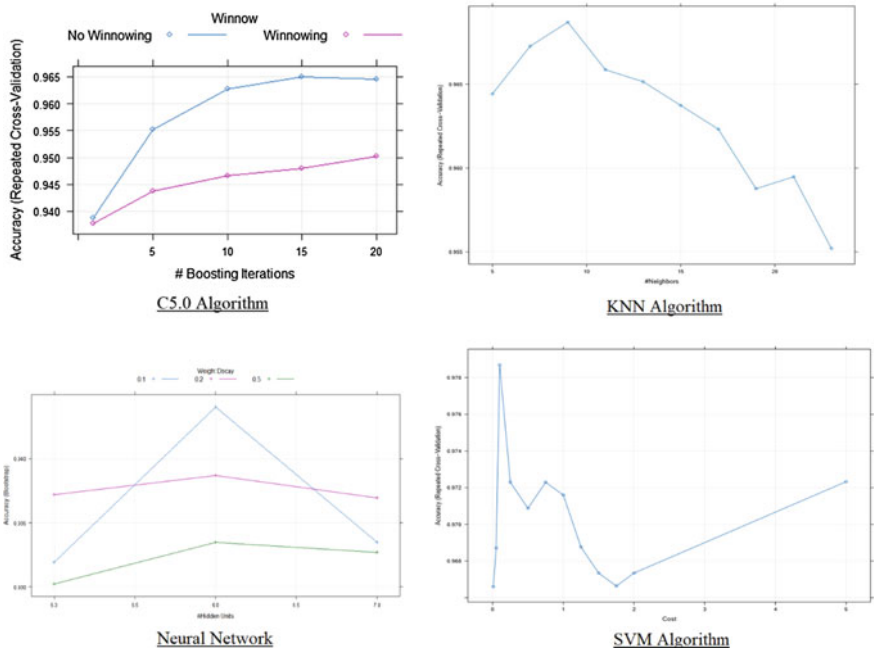


Fig. 5. Accuracy charts for various models

4.1 Comparison of Accuracies of Individual Models

Comparing the models to determine the best model for diagnostics is needed, so we have compared each of them according to their accuracy. From the bar chart below, we can see that the model with the best accuracy is the linear SVM model with an accuracy of 98.1% (Fig. 6).

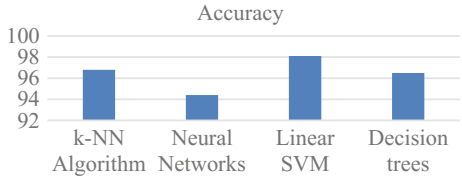


Fig. 6. Comparison of accuracies of individual models

4.2 Comparison with Existing Results

The following graph results of a similar analysis done on β-thalassemia patients. Here, the machine learning models used were k-NN classifier, multilayer perceptron, naïve Bayes classifier, and logistic regression model. From the graph, we see that maximum

accuracy for two different variations of outcomes for measuring the severity of the disease is highest for multilayer perceptron with a value 86.6142% [3] (Fig. 7).

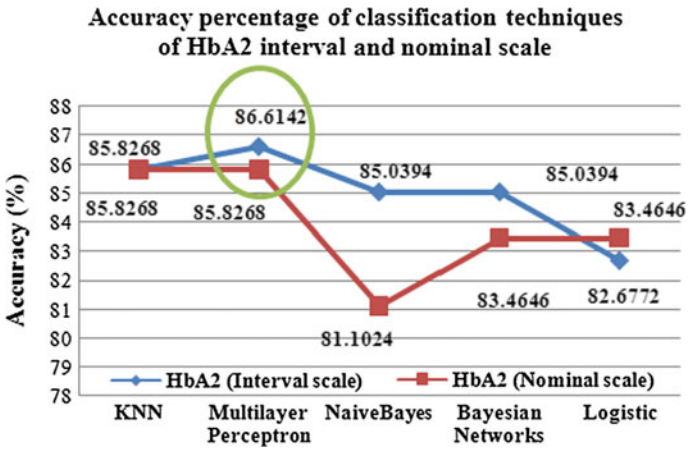


Fig. 7. Accuracy chart on predictions on β-thalassemia patients [3]

5 Conclusion

The trained models are able to diagnose a cancer condition with an accuracy of at least 94%, which is quite reliable. However, there are certain limitations. The machine learning models fail to suggest appropriate treatment for conditions after its diagnosis. Wrong data fed into the models may result in incorrect experience and hence, incorrect diagnosis. Also, from the comparisons above and the reference made in [4] we can see that different machine learning models are portrayed to be the most suitable for diagnosis, which indicates that a lot of factors come into play such as the nature of datasets and correlation values between the features used in predictions. But the main motive is to achieve a prediction which is closest to 100% accuracy so that its reliability increases.

References

1. Lantz B (2004) Machine learning with R, 66, 240, 340–343, 220–221, Packt Publications, Birmingham, UK, pp 121–122
2. Street N, Wolberg, WH, Mangasarian OL (1993) Nuclear feature extraction for breast tumor diagnosis. In: International symposium on electronic imaging: science and technology, San Jose, CA, pp 861–870
3. Paokanta P (2012) β-Thalassemia knowledge elicitation using data engineering: PCA, Pearson's Chi square and machine learning. Int J Comput Theor Eng

4. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B (1998) Support vector machines. IEEE Intell Syst Appl
5. Fatih A, Mehmet (2009) Support vector machines combined with feature selection for breast cancer diagnosis. Exp Syst Appl 3240–3247
6. Kishore R, Kaur T (2012) Backpropagation algorithm: an artificial neural network approach for pattern recognition. Int J Sci Eng Res
7. Tang L, Refaeilzadeh P Cross Validation, Arizona State University
8. Zoubir M, Abdefihak I, Robert D (2007) Bootstrap methods and applications. IEEE Signal Process Mag 10–19
9. Kohli S, Singal H (2015) Data analysis with R. Utility Cloud Comput
10. Linna L, Zhang X (2010) Study of data mining algorithm based on decision tree. Comput Des Appl

Breast Blood Perfusion (BBP) Model and Its Application in Differentiation of Malignant and Benign Breast

Sourav Pramanik[1(✉)], Debapriya Banik[1], Debotosh Bhattacharjee[1], Mita Nasipuri[1], and Mrinal Kanti Bhowmik[2]

[1] Department of Computer Science and Engineering, Jadavpur University, Kolkata, India
{srv.pramanik03327, debu.cse88}@gmail.com, {debotosh, mnasipuri}@cse.jdvu.ac.in
[2] Department of Computer Science and Engineering, Tripura University (A Central University), Agartala, Tripura, India
mrinalkantibhowmik@tripurauniv.in

1 Introduction

Nowadays, medical infrared breast thermography has emerged as an essential tool for early detection, better diagnosis, and effective treatment of breast cancer [1]. It generally records the variation of temperature on the breast surface using IR radiation discharged from the breast surface [3]. Some researches on breast cancer reveal that the abnormal growth rate of a tumor is proportional to its temperature [2]. The changes in thermal patterns occur due to various causes such as inflammation, the presence of a tumor, or angiogenesis. The thermal infrared camera can record these changes very well and thus, breast thermal image can be used to detect breast abnormality in its early stage [4]. However, from the perception of thermal characteristics, it has been seen that a higher level of blood perfusion from any source of abnormality causes heat to be conveyed from the core of the breast to the breast surface with a much greater efficiency. These underlying facts motivated us to derive a breast blood perfusion (BBP) image model from the original gray-level breast thermogram. However, it is very hard to find any related work that uses BBP image for the identification of the breast abnormalities. Over the past few decades, a lot of works have been reported in the literature for the identification of breast abnormalities by directly using the originally captured breast thermograms [5–9]. But unfortunately, they have not achieved any acceptable accuracy so far. Pramanik et al. [5] have proposed a new block variance measure to convert grayscale thermal breast image (TBI) into the corresponding texture features image to predict breast abnormality. Gaber et al. [6] have extracted 30 features from each breast of an individual to detect an abnormality in the breast. Sathish et al. [7] have extracted first- and second-order statistical features by computing histogram and GLCM matrices in four directions. Krawczyk et al. [8] performed asymmetry analysis by taking a set of 38 statistical features from each breast. Gogoi et al. [9] have extracted a total number of six first-order statistical features for asymmetry analysis to differentiate between healthy and unhealthy breasts.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_45

However, in contrast to the TBI processing, blood perfusion information has widely been used in the thermal face recognition [10]. They have shown that the blood perfusion information is more powerful and consistent than the raw thermograms for the identification of the face. Therefore, in this work, we have proposed a blood perfusion image model, called BBP model, based on an existing model [10] to convert the raw breast thermogram into the blood perfusion image. The fundamental difference between the proposed BBP model and the existing model [10] is that the BBP model has paid a special attention to the pathological condition, which is ignored in [10]. After transformation of the raw breast thermogram into the BBP image, a new combination of texture features is extracted for the analysis of the breast thermograms and fed to the feed-forward artificial neural network (FANN) for classification. Our experimental evaluation on a dataset of benign and malignant breast thermograms outperforms in terms of classification accuracy, sensitivity, specificity, PPV, and NPV compared to the raw gray-level TBI and existing blood perfusion image model [10].

The rest of the paper is arranged as follows. Section 2 demonstrates the proposed system in detail. Section 3 describes the experiment conducted along with performance analysis of the proposed system, and the conclusion is drawn in Sect. 4.

2 Proposed System

Figure 1 shows the flow diagram of the proposed breast abnormality prediction system using TBI. Each of the steps in Fig. 1 will be described in the following subsections.

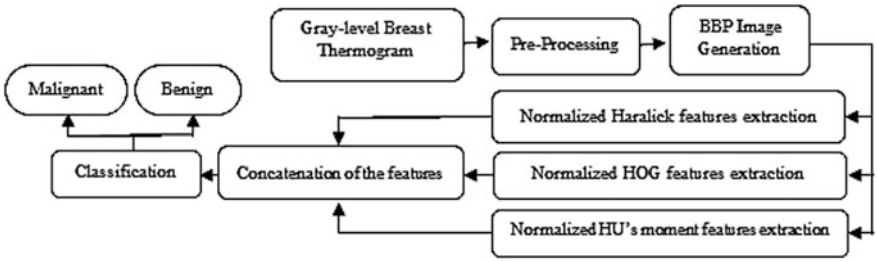


Fig. 1. Flow diagram of the proposed system

2.1 Preprocessing

Usually, in the original thermograms, some extra body parts (arms, neck, and abdomen) also appear beside the breast region as shown in Fig. 2a. Due to the difficulties in the automatic segmentation of the breast region, we have manually segmented the breast region into left and right breast by taking expert's suggestions, as shown in Fig. 2b.

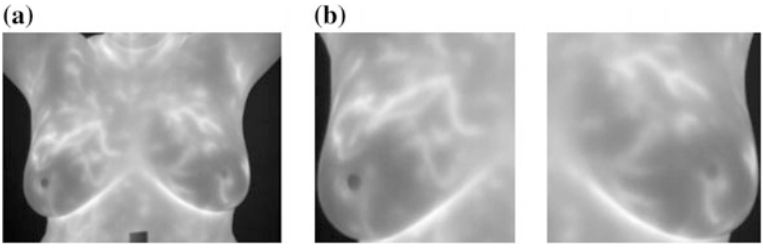


Fig. 2. **a** Original breast thermogram **b** Manually segmented left and right breast region

2.2 Breast Blood Perfusion (BBP) Image Generation

The breast thermogram usually indicates the surface temperature of the breast. The active rise in the temperature of the breast surface is due to the transport of blood through the veins, which are superficial to the skin surface. Hence, the venous patterns on the skin surface play a crucial role in the assessment of the physiological state of the breast thermograms. In this work, we thus converted the breast thermogram into the BBP image to further analyze it.

Wu et al. [10] have proposed a skin heat transfer (SHT) model, which is also called the blood perfusion image model to convert thermal face images into the corresponding blood perfusion face images. This model has made the following assumptions: (1) the body temperature is supposed to be constant with no thermal regulation, (2) the surrounding temperature is generally assumed to be lower than the body temperature, and (3) psychological and pathological conditions are ignored. However, the SHT model has some useful properties (proof is given in [10]) that motivated us to use it in our work. The properties are it preserves the exact shape and location of the objects; it increases the contrast on the high-temperature area and decreases the contrast on the low-temperature area. Although it works well for the face images, we cannot directly apply it to the TBIs because the capturing conditions and pathological conditions of the breast are completely different from the face. Thus, we have modified the SHT model for the application to the TBI.

In a thermally neutral condition, heat equilibrium equations can be defined at the skin surface [11]:

$$T_r + T_e + T_f = T_c + T_m + T_b, \tag{1}$$

where T_c, T_m, *and* T_b are the body heat production elements and stand for deep body conduction, metabolic heat, and blood flow convected heat, respectively. T_r, T_e, and T_f are the outflows terms and stands for radiation, basic evaporation, and convection, respectively.

In Wu et al. [10] model, they have ignored evaporative heat loss T_e because under the equilibrium state evaporative heat loss is very low. Similarly, the metabolic heat component T_m is considered as a constant term in the SHT model (Assumption 3). In our work, TBIs are used to identify the presence of abnormality in the breast. Hence, metabolic heat cannot be a constant term. Based on the above argument, we have proposed a new SHT model, called BBP model, which incorporated the SHT model

along with a new metabolic heat term. Theoretical justification of the proposed BBP model is similar to the SHT model [10]. The following parameters T_r, T_f, T_c, and T_b in (1) are directly obtained from the SHT model and depicted in (2), (3), (4), and (5). T_m is the metabolic heat defined in (6) [11].

$$T_r = \varepsilon\sigma\left(t^4 - t_e^4\right) \tag{2}$$

$$T_f = Ak_f d^{3M-1}\left(P_r g\beta/\nu^2\right)^M (t - t_e)^{M+1} \tag{3}$$

$$T_c = \frac{k_s(t_c - t)}{D} \tag{4}$$

$$T_b = \alpha c_b \omega_b (t_a - t) \tag{5}$$

$$T_m = m_0 * 2^{(t-t_a)/10} \tag{6}$$

where t is the apparent temperature of the skin, t_e is the environmental temperature, $\varepsilon = 0.98$ is the emissivity of the body, σ is the Stefan–Boltzmann constant, A and M are constants, d is the length of the object, P_r is the Prandtl number, g is the local gravitational acceleration, β is the air thermal expansion coefficient, v is the kinematic viscosity of air, k_f is the air thermal conductivity, k_s is the skin thermal conductivity, t_c is the core body temperature, D signifies the distance from the skin to the body core, α is the skin countercurrent exchange ratio, c_b signifies the specific heat of the blood, ω_b is the blood perfusion rate, t_a is the arterial or core body temperature, and m_0 is the metabolic constant, respectively.

From the above equations, it can be seen that there are various controlling parameters associated with the equations. The appropriate estimation of these parameters plays a vital role during the transformation. Thus, by several trial and errors, we have selected the value of the parameter, which is best suited for our domain of application. Table 1 depicts the parameters' values used for our case. Finally, based on the above equations, the BBP model can be obtained as follows:

Table 1. Parameters and their values used in Eq. (7)

Parameter	Value	Parameter	Value
σ	$5.67 * 10^{-8}$ W/m^{2}k^4	β	3.354×10^{-3} K^{-1}
t_e	292.15 K	v	1.56×10^{-5} m^2/s
A	0.29	k_s	0.5 W/mK
M	0.25	t_c	310.15 K
k_f	0.024 W/mK	α	0.8
d	0.170	ρ_b	1060 kg/m^3
D	0.085	c_b	0.92 cal/ml.K
P_r	0.72	t_a	310.15 K
g	9.8 m^2/s	m_0	0.0144

$$I^{(BP)} = \omega_b \frac{\varepsilon\sigma\left(t^4 - t_e^4\right) + Ak_f d^{3M-1}\left(\frac{P_r g\beta}{\nu^2}\right)^M (t - t_e)^{M+1} - \frac{k_s(t_c - t)}{D} - m_0 * 2^{(t-t_a)/10}}{\alpha c_b(t_a - t)} \quad (7)$$

From the perspective of breast thermography, it can be easily visually demonstrated by naked eye that the BBP image enhances the contrast of the high-temperature region and suppresses the contrast of the low-temperature region. As a result, the pathological variations can be easily diagnosed. Figure 3a shows the gray-level left and right breast thermogram, whereas Fig. 3b demonstrates the left and right BBP image.

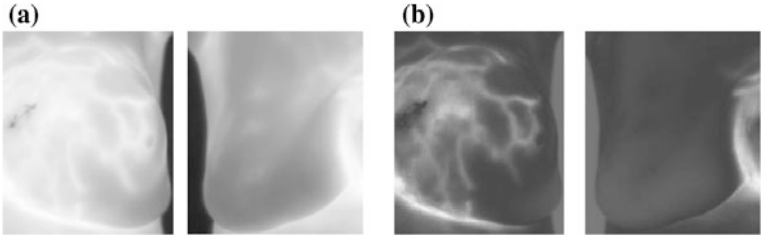


Fig. 3. **a** Gray-level breast thermograms **b** Corresponding BBP images

2.3 Feature Extraction

After transformation of the TBIs into the blood perfusion images, the following three types of features are extracted: texture features (Haralick features), gradient-based features (HOG features), and Hu's moment invariants.

The basis of Haralick texture features is the gray-level co-occurrence matrix (GLCM) [12]. Basically, computation of Haralick features involves two steps. The GLCM is computed in the first step, which describes the texture of the image followed by the extraction of texture features from the GLCM. A set of 14 Haralick texture features are extracted from the co-occurrence matrix. The feature sets are then normalized using z-score [13]. Histogram of oriented gradients (HOG) captures local edge or gradient structure from an image [14]. It is significant for breast thermograms with weak or blur edges and independent of local edge orientation. In this study, HOG is computed by dividing the image into smaller nine rectangular blocks, which are followed by nine bins histogram generation per block. The nine histograms with nine bins were concatenated to form an 81-dimensional feature vector. The gradients were locally normalized over each cell. Hu's moment invariants are computed based on the information provided by both the shape boundary and its interior region. The features extracted from Hu's moment are invariant to rotation, scaling, and translation and so the breast thermograms are independent of any geometric transformation [15]. In this work, a set of seven normalized Hu's moment invariants are calculated based on the central moments, which are invariant to image translation.

Each of the normalized feature sets is then concatenated to form 102-elements feature vector for each of the left and right breasts. Finally, the asymmetry features are calculated for each patient's breast thermogram by taking the absolute difference

between $\left[f_v^{(L)}\right]_{1\times102}$ and $\left[f_v^{(R)}\right]_{1\times102}$, where these are concatenated feature vectors of left and right breast thermogram of each patient. Let $[F]_{1\times102}$ be the asymmetry feature vector of a patient's breast thermogram and is defined as $F = \left|f_v^{(L)} - f_v^{(R)}\right|$.

3 Experimental Results

The breast thermograms used in this research work are collected from the existing Database for Mastology Research (DMR) database [16]. DMR is an available online database that includes 287 breast thermograms of normal, benign, and malignant patients. From the dataset, we have randomly chosen 150 frontal view TBIs (45 malignant and 105 benign) to validate the proposed system.

Here, we have shown the effectiveness of our proposed BBP image for the differentiation of the malignant patients from the benign patients. The 102-elements feature vector is formed for each patient's breast thermogram, which is then fed to the three-layer FANN for the classification [5]. 102 neurons in the input layer of the network are considered to fit the 102-elements feature vector. Since our problem of classification is a binary problem; we have considered one neuron in the output layer. One hidden layer with 51 neurons is considered for this network. For the training of the network, the Levenberg–Marquardt backpropagation algorithm (with learning rate = 0.1) is used. 75 (35 malignant and 40 benign) BBP images are randomly chosen from a set of 150 BBP images (45 malignant and 105 benign) to train the network. Remaining TBIs are used to test the network. We have performed some classical performance measurements such as accuracy, sensitivity, specificity, NPV, and PPV based on the classification results [4]. In this work, the obtained accuracy, sensitivity, specificity, NPV, and PPV are 98%, 93.5%, 100%, 97.1%, and 100%, respectively. In order to verify the superiority of the proposed BBP image over the existing blood perfusion image (EBPI) model [10] and raw grayscale image (RGI), the same set of features are extracted and fed to the FANN classifier. Table 2 shows the experimental results for three types of TBIs. From Table 2, it can be seen that the proposed BBP image-based breast abnormality prediction system achieved significantly better results compared to other two different image-based methods.

Table 2. Comparison of performance measures

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUC
RGI-based method	89	70.9	97.1	91.6	88.1	0.929
EBPI-based method [10]	91	77.4	97.1	92.3	90.5	0.941
Proposed method	**98**	**93.5**	**100**	**100**	**97.1**	**0.976**

We have also quantified our classification results using receiver operating characteristic (ROC) curve. The area under the curve (AUC) is very much informative for the analysis of the system results, which typically lies between 0.5 and 1 [17]. Figure 4 shows the comparison between the ROC curves of the proposed BBP image-based method, raw grayscale image-based method, and existing BPI-based method. The obtained AUC value for our proposed model is 0.976 and attained 100% true positive recognition at a value less than 0.1 false positive rate.

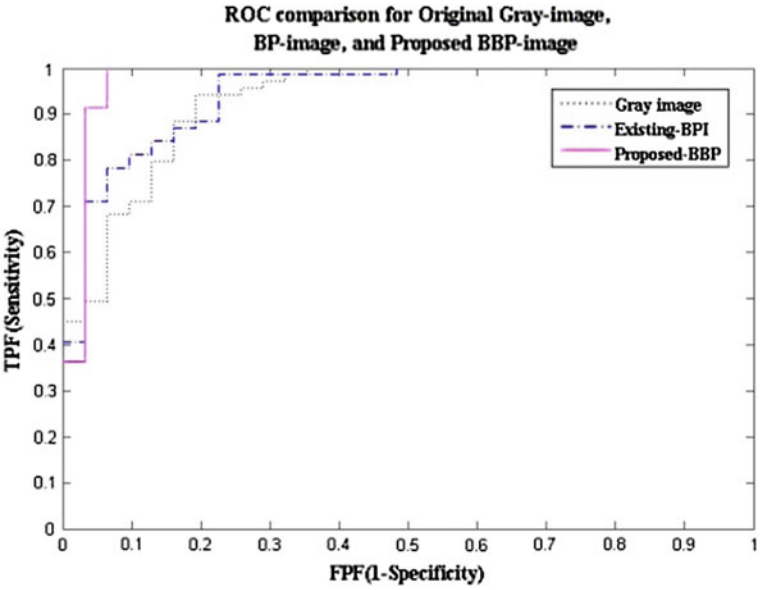


Fig. 4. Comparison of the ROC curves

4 Conclusion

Preprocessing of the TBI plays a crucial role in the computer-assisted diagnosis of breast cancer using TBIs. The raw breast thermogram often suffers from the intra-class scatter problem that makes it very difficult to differentiate a malignant breast thermogram from the benign one. Hence, in this paper, we have proposed a BBP model, which is also the primary contribution of this paper, to convert the raw gray-level breast thermogram into the blood perfusion image. The BBP model used thermal physiology of the breast and reduced the intra-class variability of the breast thermograms. The BBP model can be used as a preprocessing step for the diagnosis of the breast using breast thermogram. The experiment results have shown that the blood perfusion image of the breast is more powerful than the raw gray-level breast thermogram for the prediction of the breast abnormalities. Additionally, we have introduced a new combination of texture features for the analysis of the breast thermograms. In future, we will try to address the problem of suspicious region segmentation using BBP image, which is comparatively very hard for the gray-level breast thermal image.

Acknowledgements. Authors are thankful to DBT, Govt. of India for funding a project with Grant no. BT/533/NE/TBP/2014. Sourav Pramanik is also thankful to Department of Electronics and Information Technology (DeitY), Government of India, for providing him PhD Fellowship under Visvesvaraya PhD Scheme.

References

1. Prabha S, Anandh KR, Sujatha C M, Ramakrishnan S (2014) Total variation based edge enhancement for level set segmentation and asymmetry analysis in breast thermograms. IEEE Int Conf Eng. Med Biol Soc
2. Usuki H, Onoda Y, Kawasaki S, Misumi T, Murakami M, Komatsubara S, Teramoto S (1900) Relationship between thermographic observations of breast tumors and the DNA indices obtained by flow cytometry. Biomed Thermol 10(4):282–285
3. Sobti P, Sobti L, Keith G (2005) Screening and diagnostic mammography: why the gold standard does not shine more brightly? Int J Fert Women Med 50:199–206
4. Borchartt TB, Conci A, Lima RCF, Resmini R, Sanchez A (2013) Breast thermography from an image processing viewpoint: a survey. Sig Proc 93:2785–2803
5. Pramanik S, Bhattacharjee D, Nasipuri M (2016) Texture analysis of breast thermogram for differentiation of malignant and benign breast. IEEE Int Conf Advanc Comput Commun Informat
6. Gaber T, Ismail G, Anter A, Soliman M, Ali M, Semary N, Hassanien AE, Snasel V (2015) Thermogram breast cancer prediction approach based on Neutrosophic sets and fuzzy c-means algorithm. IEEE Int Conf Eng Med Bio Soc
7. Sathish D, Surekha K, Keerthana P, Rajagopal K, Roshan JM (2016) Asymmetry analysis of breast thermograms using automated segmentation and texture features. J Signal Imag Video Process 1–8
8. Krawczyk B, Gerald S (2014) A hybrid classifier committee for analysing asymmetry features in breast thermograms. J Appl Soft Comput 112–118
9. Gogoi US, Majumdar G, Bhowmik MK, Ghosh AK, Bhattacharjee D (2015) Breast abnormality detection through statistical features analysis using infrared thermograms. IEEE Int Sympos Advanc Comput Commun 258–265
10. Wu S, Lin W, Xie S (2008) Skin heat transfer model of facial thermograms and its application in face recognition. Patt Recogn 41:2718–2729
11. Houdas Y, Ring EFJ (1982) Human body temperature: its measurement and regulation. Plenum Press, New York
12. Haralick RM (1979) Statistical and structural approaches to texture. Proc IEEE 67:786–804
13. Jain A, Nandakumara K (2005) A Ross score normalization in multimodal biometric systems. Patt Recogn 38:2270–2285
14. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. IEEE Comput Soc Conf Comput Vision Patt Recog 1–8
15. Ming-Kuei H (1962) Visual pattern recognition by moment invariants. Informat Theor IRE Trans 8:179–187
16. Silva LF, Saade DCM, Sequeiros-Olivera GO, Silva AC, Paiva AC, Bravo RS, Conci A (2014) A new database for breast research with infrared image. J Med Imag Health Informat 4(1):91–100(9)
17. Tilaki KH (2013) Receiver Operating Characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med 4:627–635

Development and Feasibility Studies of a Device for Early Prediction of Asthma Attack

Shreya Das, Supriyo Sadhya(✉), Megha Dasgupta, Pratyush Anand, Manoswita Biswas, Soham Chatterjee, and Surajit Bagchi

Heritage Institute of Technology, Kolkata, India
{dasneha02081994, supriyosadhya}@gmail.com

1 Introduction

Asthma is a disease of the bronchial tubes in the lungs (the "airways") [1]. The commonly known symptoms of asthma attack are coughing, wheezing, chest tightness, and shortness of breath [2]. There are three major features of asthma: (1) Airway obstruction, (2) Inflammation, (3) Airway Irritability. Asthma is a serious threat to mankind. According to WHO (World Health Organization), 9 people die every day across the globe because of asthma and an estimated 300 million people worldwide suffer from asthma. It is estimated that the number of people with asthma will grow by 100 million by the year 2025 [1]. Thus, the diagnosis and control of asthma are of utmost importance.

Paredi et al. (2002) have proposed a temperature sensor to measure rate of the exhaled breath temperature. They have shown that the faster rise of exhaled breath temperature is an indicative of inflammation in conducting airways [3].

Popov et al. (2010) have developed an exhaled temperature measuring device for assessment of inflammatory processes in the conducting airways [4].

Dieffenderfer et al. (2016) have proposed a multimodal sensing platform using an array of sensors for an asthma predictor to monitor environmental pollutants, heart rate, breath rate, wheezing sound, expiratory airflow, three-axis acceleration, and many more [5].

However, the above devices are bulky, operationally complex, and have incorporated multiple sensors, and therefore, the common people are reluctant to use them.

Siemens have manufactured a device that detects the early signs of airway inflammation by measuring the concentration of nitric oxide (NO) in the patient's exhaled breath [6]. Star et al. (2007) have developed a NO sensor using carbon nanotube that can detect an asthma attack before its onset [7]. However, the presence of NO in exhaled breath for NO detection is not at all full proof as its concentration may change due to other factors also. High expense is an additional disadvantageous factor.

Hence, the objective of the present work is the development and performance characterization of an asthma predictor that would be simple, relatively inexpensive, and easy to use to meet the needs of mankind. It will serve asthmatic patients all around the globe, enabling them to take early precautionary measure.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_46

The proposed single sensor-based device measures primarily the respiration temperature of the subject. Simple programs were embedded in an Arduino controller to derive clinically significant asthmatic indices to foresee possible asthma attack.

2 Theoretical Background

Olson (1958) reported the electrical analog of the ventilatory system [8]. Schmidt et al. (1998) have described the respiratory system using the electrical RC-model. The respiratory tract is analogous to electrical resistor (R) and alveolus represents the electrical capacitor (C) [9].

An oxygen deficiency in the blood stream of an asthmatic patient occurs due to the formation of mucus in the respiratory tract. The presence of the mucus reduces the cross-sectional area of the tract causing a restricted ventilatory flow leading to higher breath rates to meet the required oxygen demand. Moreover, during expiration air leaving the lungs is cooled as it passes through the respiratory tract through direct transfer of heat to the airway mucosa [10]. The excessive mucous layer in the respiratory tract of the asthmatic patients reduces this heat transfer efficiency resulting in higher exhaled breath temperature. The slope of time–temperature response curve is higher for asthmatic patients.

During exhalation, breath temperature rises and during inhalation, breath temperature falls. For mathematical reasons, it was determined that the temperature slope from the beginning of exhalation to 63% of the plateau temperature best characterized the rate of increase, i.e., was the maximum [11] (Fig. 1).

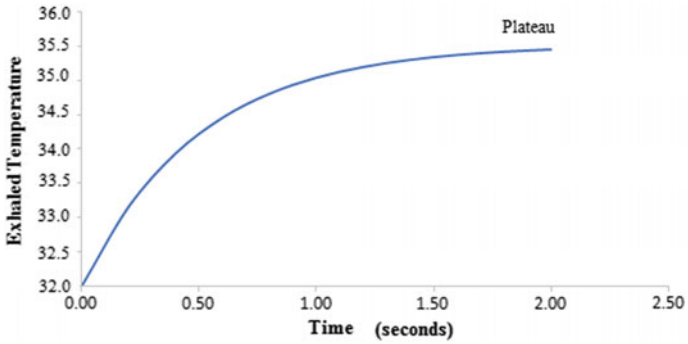


Fig. 1. Exhaled temperature versus time curve

The exhaled breath temperature changes exponentially with time. The shape of the curve depends upon the time constant (τ). In one time constant, the response reaches 63% of its final change. The rate of temperature increase ($\Delta T(t)/\Delta t$) calculated between the beginning of exhalation and 63% of the total temperature increase proved to be the more reproducible parameter to characterize the curves. The response is of the form

$$T(t) = To\left(1 - e^{-\frac{t}{\tau}}\right) \tag{1}$$

To = temperature at steady state

$$\frac{dT(t)}{dt} = \frac{1}{\tau} \times To \times e^{-t/\tau} \tag{2}$$

t = time after exhalation starts, τ (time constant) = R × C

As the value of t in Eq. (2) will increase, the slope will decrease hence we can say that with increase in time slope is decreasing.

$$\frac{dT(t)}{dt}(max) = To/\tau, \text{ at } t = 0 \tag{3}$$

$$\frac{dT(t)}{dt}(min) = 0, \text{ at } t = \infty \tag{4}$$

3 Materials

The components used in the proposed hardware system are: (a) LM35 (Temperature Sensor), (b) OP-AMP (IC-741), (c) Arduino (UNO), (d) 16 × 2 LCD, (e) breathing mask, (f) resistors, and capacitors (passive components).

4 Methodology

4.1 Overview of the Method

The proposed system uses LM35 as the temperature sensor that provides electrical signal (mV) replicating the respiration temperature. A conditioning circuit using an amplifier and a low-pass filter processes the signal and feeds to the Arduino Uno controller. The Arduino (using embedded algorithms) evaluates the desired parameters and displays them on the LCD. The block diagram and prototype of the system are shown in Fig. 2 and Fig. 3, respectively.

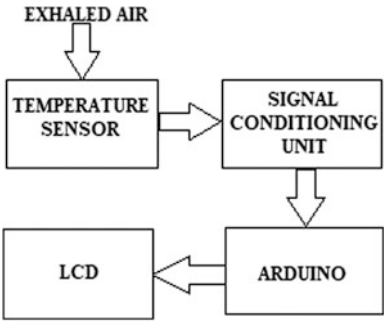


Fig. 2. Block diagram

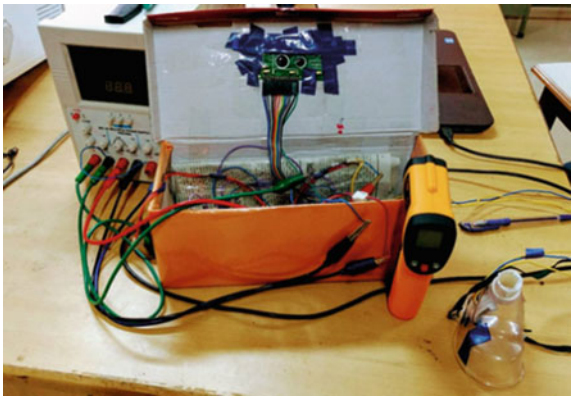


Fig. 3. Prototype of the system

4.2 Principle

The device works on the basic knowledge that the exhaled air has a higher temperature than that of the inhaled air. The temperature sensor follows the breath temperature. Arduino executes programs to evaluate (a) respiratory rate of the concerned subject. The rate of respiration of a normal adult person is 12 to 25 breaths per minute; the same for asthmatic patients is greater than 25 breaths per minute. (b) temperature and (c) rate of rise of the exhaled breath. The differences in parameters have been exploited in predicting asthma [3, 12].

4.3 Procedural Steps

The procedure for evaluating asthma-specific parameters is as follows:

1. The electrical output from LM35 is amplified with a gain of 11, $V_o = \{1 + (R_f/R_i)\} * V_i$, where V_o is output voltage, V_i is input voltage, R_f is feedback resistor, R_i is input resistor. R_f is 100 kΩ and R_i is 10 kΩ.

2. The amplified signal is fed to an active first-order low-pass RC filter of unity gain in order to remove the noise from the final output. The cutoff frequency of the filter is around 5 Hz (frequency = 1/(2ΠRC)). Value of R is 3.3 KΩ and value of C is 10 μF.
3. The final output obtained from the low-pass filter goes to the A4 pin of Arduino Uno. This input received is converted by an ADC in the range of 0–1023 which is further mapped on to a scale of 0–5 V using a conversion factor of 0.0048828 (= 5/1024).
4. The Arduino executes the program to evaluate (a) respiration rate, (b) temperature, and (c) slope (rate of rise of exhaled breath temperature) and the results are displayed on a 16 × 2 LCD screen.

4.4 Program Algorithm

The output of the low-pass filter is fed to the Arduino Uno which executes a preloaded code to determine the asthma-specific parameters. The pseudo-code of the algorithm processed by the microcontroller is given.

```
1: INITIALIZE slope, rate and count1 to zero.
2: WHILE time is less than 60 seconds do
3:      Read data from pin A4 of Arduino into variable a
4:      Convert a into temperature and store in variable c using regression equation
5:      Delay for 300 milliseconds
6:      Read data from A4 pin of Arduino into variable b
7:      Convert b into temperature and store in variable d using regression equation
8:      IF d is greater than c then
9:              count=1; sp=(d-c)/300; slope=maximum of (slope, sp)
10:     END IF
11:     IF c is greater than or equal to d then
12:             count1=count+1
13:     END IF
14:     IF count1=2 then
15:             rate=rate+1; count=0; count1=0
16:     END IF
17: END WHILE
14: PRINT rate, slope and temperature
```

5 Results and Discussion

5.1 Calibration of Temperature Sensor Along with Signal Conditioning Unit

The electrical signal from the overall temperature sensing system is calibrated against a set of predetermined temperatures using an infrared thermometer. The data pairs are plotted to get a calibration curve (Fig. 4) and a regression equation is established as Y = (0.1048 * X) + 0.4441 using Microsoft Excel, 2016, where Y is voltage in Volts and X is temperature in °C.

5.2 Performance Study of the Device

136 healthy male subjects and 53 male asthmatic subjects were studied who belonged to the age group of 20 to 25 years. The device measures breath rate, breath temperature, and rate of rise of exhaled breath temperature of all the subjects. The exhalation and inhalation breath pattern of the subjects were also observed using the Arduino serial plotter. The different temperature values in the serial plotter are actually digital values of the input voltage (converted into temperature) converted by the 10-bit ADC of the Arduino Uno. The breath pattern is a temperature v/s time graph. One complete breath cycle consists of exhalation and inhalation. Temperature rises during exhalation and temperature falls during inhalation. A sample breath pattern is shown in Fig. 5. The Table 1 shows the data obtained from only ten subjects due to space constraint.

The subjects under consideration in Table 1 labelled as ASTH were asthmatic. The breath rates obtained were all in the range of (29–33) breaths per minute. This satisfies the theoretical criteria that breath rate of asthmatic patients should be greater than 25 breaths per minute [12]. The subjects under consideration in Table 1 labelled as HLTHY have no medical history of respiratory diseases. The breath rates obtained are found to be in the range (16–23) breaths per minute. This also satisfies the theoretical criteria that the breath rate should lie between 12 and 25 breaths per minute [12]. However, the breath rate might change with different geographical locations and different age groups.

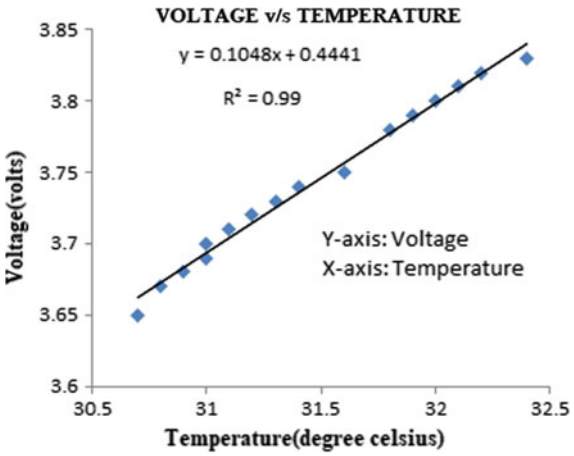


Fig. 4. Calibration curve

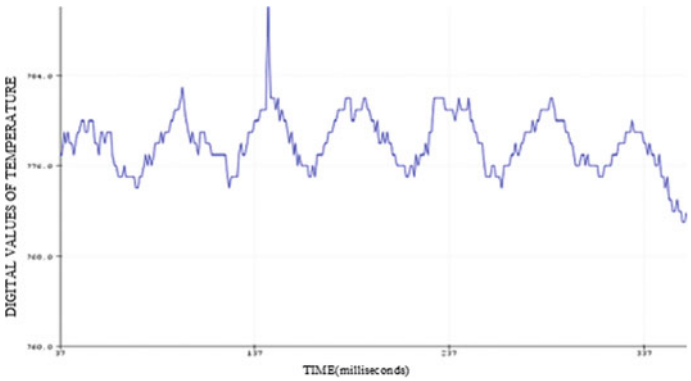


Fig. 5. Exhalation and inhalation breath pattern

Table 1. Values of the asthma-specific parameters of Asthmatic and healthy subjects

Age		Breath rate		Breath temperature (°C)		Rate of rise of exhaled temperature (°C/s)	
ASTH	HLTHY	ASTH	HLTHY	ASTH	HLTHY	ASTH	HLTHY
22	21	34	20	35.71	33.94	1.0310	0.4361
22	20	31	18	35.32	34.56	0.9136	0.4355
23	24	29	16	35.55	34.45	0.9197	0.5052
22	22	30	21	36.11	34.21	0.9256	0.5120
22	20	35	17	35.96	33.96	1.0520	0.4361

The results obtained by performance study of the device as shown in Table 1 indicate that the device is clinically fit for asthma prediction. The device is found to consume 410 mW of power which is close to the power consumed, 500 mW and 800 mW, by some available asthma predicting devices [4, 13].

6 Conclusion

A single sensor-based, inexpensive, and easy-to-adopt asthma attack predictor has been presented. The sensor, fitted inside a breathing mask, tracks the respiration temperature. The temperature pattern is processed by a set of simple algorithms, embedded in an Arduino-UNO controller, to calculate the breath rate, exhaled breath temperature, and its rate of rise of a subject to warn the patient. The device consumes only 410 mW of power. However, the use of Arduino Pro Mini will reduce the power consumption drastically to 0.115 mW. Its predictive performance is clinically satisfactory.

References

1. The Global Asthma Report (2014) In: Globalasthmareport.org. http://www.globalasthmareport.org/burden/burden.php. Accessed 12 Jun 2017
2. The Symptoms of Asthma. In: WebMD. http://www.webmd.com/asthma/guide/asthma-symptoms1. Accessed 12 Jun 2017
3. Paredi P, Kharitonov S, Barnes P (2002) Faster rise of exhaled breath temperature in Asthma. Am J Respir Crit Care Med 165:181–184. https://doi.org/10.1164/ajrccm.165.2.2103053
4. Popov T, Kralimarkova T, Tzachev C et al (2010) Development of an individual device for exhaled breath temperature measurement. IEEE Sens J 10:110–113. https://doi.org/10.1109/jsen.2009.2035732
5. Dieffenderfer J, Goodell H, Mills S et al (2016) Low-power wearable systems for continuous monitoring of environment and health for chronic respiratory disease. IEEE J Biomed Health Informat 20:1251–1264. https://doi.org/10.1109/jbhi.2016.2573286
6. Westly E (2011) Early warning for asthma sufferers. In: MIT technology review. https://www.technologyreview.com/s/422832/early-warning-for-asthma-sufferers/. Accessed 12 Jun 2017
7. (2007) New nano-sensor can detect Asthma symptoms. In: Phys.org. https://phys.org/news/2007–08-nano-sensor-asthma-symptoms.html. Accessed 12 Jun 2017
8. Olson H (1958) Dynamical analogies, 2nd edn. Van Nostrand, New York
9. Schmidt M, Foitzik B, Hochmuth O, Schmalisch G (2017) Computer simulation of the measured respiratory impedance in newborn infants and the effect of the measurement equipment. Med Eng Phys 20:220–228
10. Taylor M, Shores S (1981) Heat exchange in the nose. Laryngoscope 91:1606–1613. https://doi.org/10.1288/00005537-198110000-00004
11. Johnston R (2017) Exhaled air temperature and Asthma. In: PFTBlog. https://www.pftforum.com/blog/exhaled-air-temperature-and-asthma/. Accessed 7 Aug 2017
12. Vital signs | Cleveland clinic. In: Cleveland clinic. https://my.clevelandclinic.org/health/articles/vital-signs. Accessed 7 Aug 2017
13. Malik S, Singh O, Nurifhan A, Malarvili M (2017) Portable respiratory CO_2 monitoring device for early screening of Asthma. In: International conference on advances in computing, electronics and communication pp 90–94. https://doi.org/10.15224/978-1-63248-113-9-61

Game Theory for Vertical Handoff Decisions in Heterogeneous Wireless Networks: A Tutorial

Pramod Goyal(✉), D. K. Lobiyal, and C. P. Katti

School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, India
{pramodgoyal73,lobiyal}@gmail.com, cpkatti@mail.jnu.ac.in

1 Introduction

Heterogeneous wireless networks are an overlay structure of different types of wireless networks as shown in Fig. 1 [1]. The participating multiple networks generally differ with each other in terms of Radio Access Technologies, Network Architecture, Protocols, Type of Service requests from Mobile Users, and Network Operators in the market [2, 3]. In such a heterogeneous environment a mobile user, equipped with multi-mode or multi-home terminals [4], may experience availability of more than one wireless networks at any moment of time, which are better than the current wireless network as per the user preferences. Similarly, a wireless networks may experience more than one mobile users at any point in time, which can provide better revenue without degrading network performance. Such situations lead to the competition between wireless networks or between mobile users or between wireless networks and mobile users to acquire the best user or best networks or best resources respectively. The game theory is an important mathematical tool to solve such problems.

The handoff is an event where a mobile user leaves current network and connects to another network for continuation of its current application/service. A handoff from one cell to the other cell of the same network is termed as horizontal handoff (HHO) while the handoff between two different networks is termed as vertical handoff (VHO). The VHO decision process [5] comprises three important stages: 1. Handoff Initiation and information gathering, 2. Network selection and 3. Handoff decision.

The game theory is vastly studied and applied to solve the problems in economics [6]. However, nowadays, game theory is also increasingly used in solving the problems of VHO decisions in HWNs. Various types of games are used to solve various types of problems related to VHO [7–9], but most of the works are either network centric or user centric and does not considers the both user aspects as well as network aspects simultaneously to perform an efficient VHO in HWN's. In network-centric VHO, the handoff decision factors are related to network aspects like bandwidth, received signal strength, delay, jitter and bit-error rate, etc., while in user-centric VHO, the handoff decision factors are related to the user aspects like battery power, terminal velocity, and usage cost along with their respective relative user preferences in addition to the

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_47

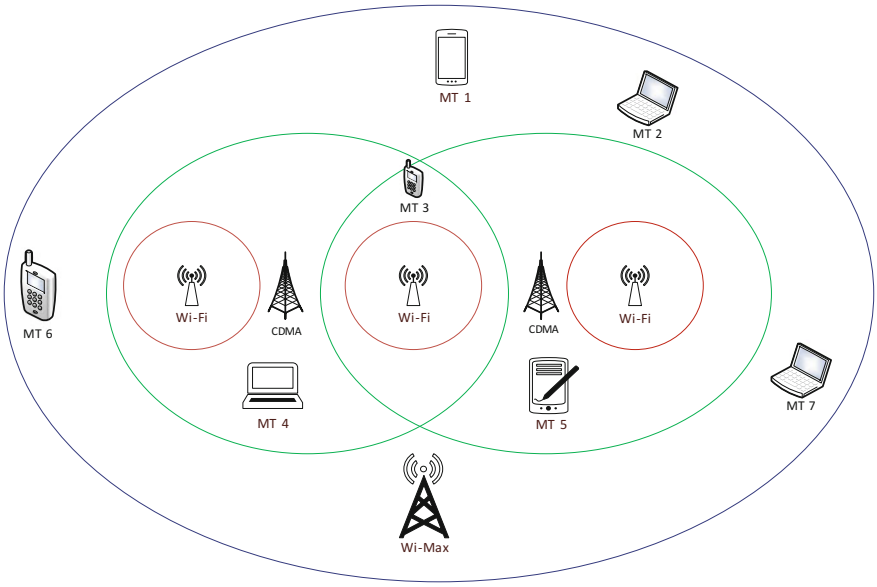


Fig. 1. Structure of heterogeneous wireless networks [1]

network-related aspects. The user preferences are expressed as user preference weights such that the sum of all such user preference weights is equal to one.

Several survey works [10–12] present the applications of game theory for VHO, but lacks in providing all the relevant information regarding Game theory at one place. This paper is an effort in this direction to fulfill this requirement. This paper presents in a tutorial form all the basic concepts and elements of game theory, game types, and game solutions. This paper, in specifically, presents the noncooperative game formulations for modeling the competition between mobile users, between wireless networks and, between mobile users and wireless networks which may become the ready reference for the researchers in this area.

The rest of the paper is organized as follows; Sect. 2 presents game theory overview, Sect. 3 presents three different types of noncooperative game formulation, Sect. 4 presents Game solution using NASH equilibrium and Sect. 5 presents the conclusions and future prospects of application of game theory.

2 Game Theory: An Overview

Game theory is an analytical tool to get the optimum results from the situations where specific actions of decision makers lead to mutually conflicting consequences to each other [6, 13]. The essential elements of a game, as shown in Fig. 2, are as follows:

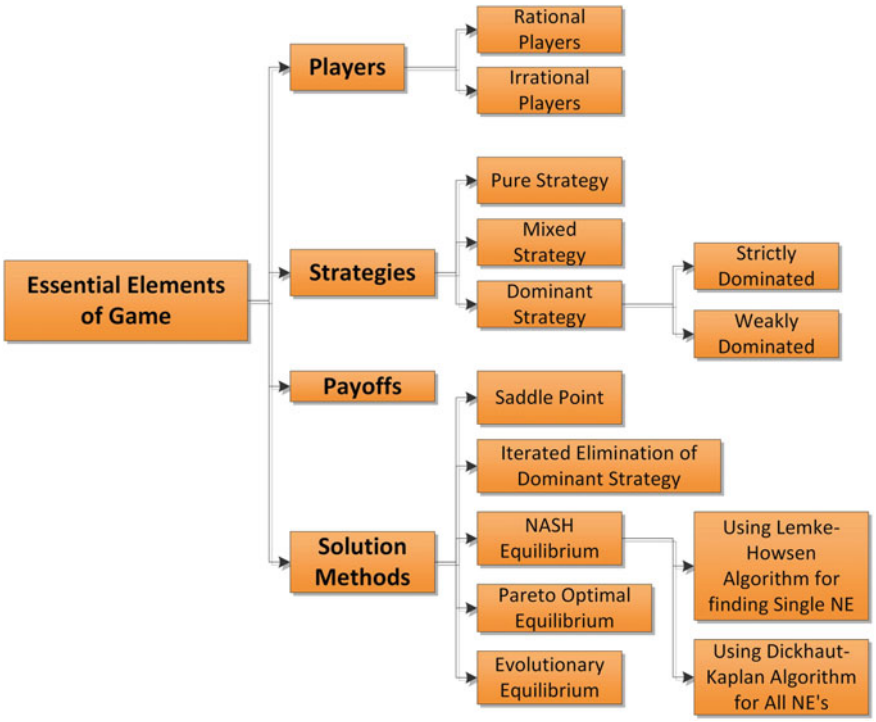


Fig. 2. Essential elements of a game

Players: They are the individual stakeholder who makes the decision. The aim of each stakeholder is to maximize his/her own payoff/profit by adopting a suitable strategy. The players can be of two types—Rational and Irrational. A player who always strives to maximize its payoff is termed as rational player.

Strategy: It is an action by a player which results in a corresponding payoff to the player. The strategies can be classified into three types—Pure, Mixed, and Dominant Strategy. The dominant strategy can be strictly dominated or weakly dominated.

Payoffs: They are the profits receivable by a player when it adopts a certain strategy while the strategies of all the other players are decided. The profits are generally expressed using utility functions and sometimes termed as utility also.

Solution Methods/Equilibrium: A solution of the game, at an equilibrium, gives the best combination of strategies of all the players, which result in the best payoff to each player of the game.

The games can be classified in any of four ways [10, 12, 14]:

1. Noncooperative Games versus Cooperative Games
2. Static versus Dynamic Games
3. Complete Information Games versus Incomplete Information Games
4. Perfect Information Games versus Imperfect Information Games.

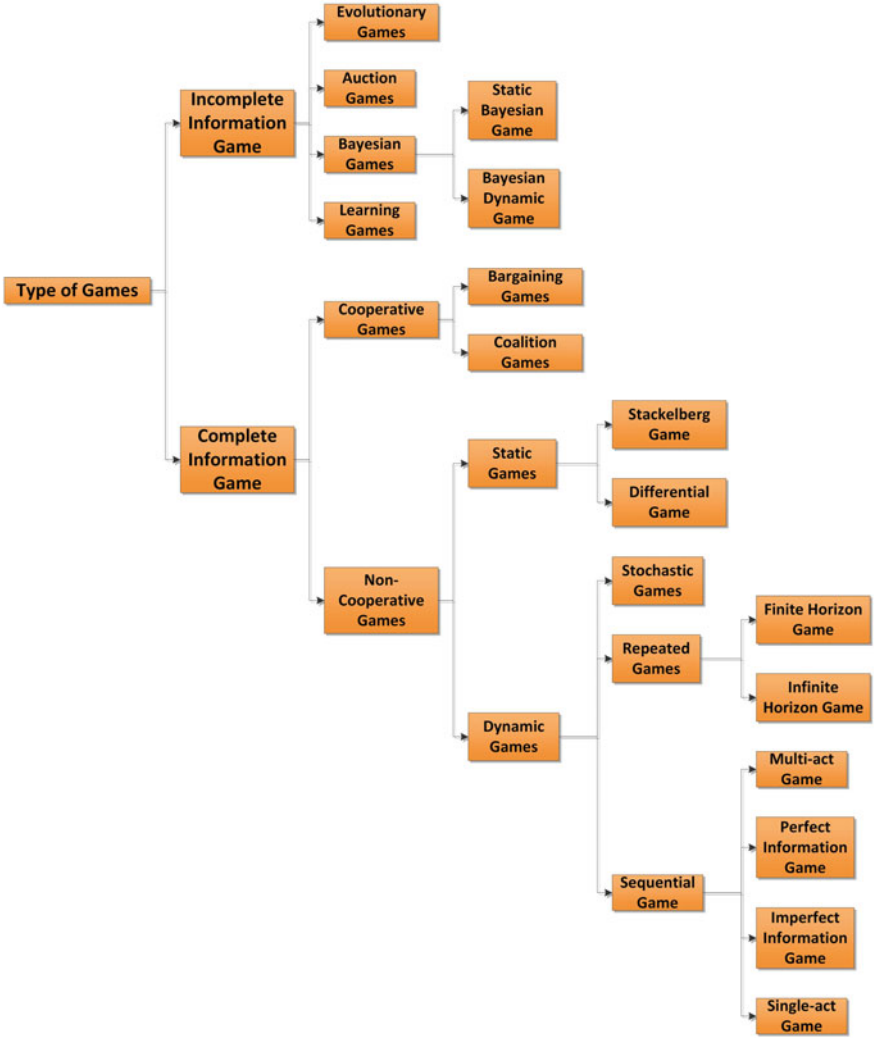


Fig. 3. Classification of games

Figure 3 shows the relationship between different types of the games in a hierarchical structure. A game is said to be Complete Information Game, if all elements of the game are common knowledge. Otherwise, the game is said to be one with incomplete information, or an Incomplete Information Game. In noncooperative games, the strategies/moves/actions of players are independent while in Cooperative games, the strategy/move of one player depends on move of other players. In Cooperative games, the players share the information to decide their moves so that all players can get best payoffs. In Static games, players take their actions only once, independently of each other. While in Dynamic games, time has a central role in the decision making. The static games are represented in strategic form/matrix form due to

which they are sometimes termed as normal form game also, while dynamic games are generally represented in extensive form using tree structure. A game is said to be Perfect information game, if only one player moves at a time and if each player knows every action of the players that moved before it at every point of the game. Otherwise, it is said to be an Imperfect information game where some of the players do not know all the previous choices of the other players.

3 Game Formulation for VHO

Techniques of Game theory are now increasingly applied in network selection for VHO in HWNs. The noncooperative games [15] are always of special interest in heterogeneous wireless networks, since the different WNs controlled by different network operators always try to maximize their revenue without cooperating with any other WN, until and unless it becomes highly necessary to do so. Hence, here we present the concept of game theory through noncooperative games. A noncooperative game in strategic form (or Normal Form), is a triplet

$$G = \left(P, (S_i)_{i \in P}, (U_i)_{i \in P}\right) \tag{1}$$

where

P: A finite set of rational player's, i.e., $P = \{1, 2, 3, \ldots, n\}$
S_i: Set of available strategies for player $\boldsymbol{i}$

$$U_i : S \rightarrow R, \text{ represents a utility(Payoff)function for player } \boldsymbol{i}, \text{with} \tag{2}$$

$$S = S_1 \times S_2 \times \cdots \times S_i \times \cdots \times S_n. \tag{3}$$

Here, the strategy space S is the strategy set defined as the Cartesian product of the individual strategy set of all the players.

In the wireless networking context, the players are the *mobile users* and/or *wireless networks*. The players are assumed to be *rational*, i.e., they always try to maximize their *payoffs* (or utilities). The game theoretical network selection schemes [16] for vertical handoff decisions in heterogeneous wireless networks can be classified into three groups.

3.1 Game Between Mobile Users

A noncooperative game is played between mobile users to select a wireless network that can maximize their payoffs. The payoffs of mobile users are the Quality of Service or Quality of Experience receivable from a wireless network. A noncooperative game between mobile users to select a particular wireless networks can be defined by the Eqs. (1)–(3) with players P defined as a finite set of mobile user's i.e. $P = \{1, 2, 3, \ldots, n\}$.

3.2 Game Between Wireless Networks

A noncooperative game is played between wireless networks to select a mobile user who can maximize their payoffs. The payoffs for the wireless networks are, generally, the revenue receivable from a mobile user for providing the requested service. The noncooperative game between wireless networks can be defined by Eqs. (1)–(3) with players P defined as a finite set of wireless networks, i.e., $P = \{1, 2\,3, \ldots, n\}$.

3.3 Game Between Mobile Users and Wireless Networks

The game between mobile users and wireless networks can be modeled as a non-cooperative game, to mutually select each other in such a way that maximizes the payoffs of each player. The game can be modeled as a two-player strategic game, where First Player is the set of "*Mobile Users*" and second player is the set of "*Wireless Networks*". The game is defined as-

$$\mathrm{G} = (\{\mathrm{U}, \mathrm{N}\}, \{\mathrm{S_U}, \mathrm{S_N}\}, \{\pi_\mathrm{U}, \pi_\mathrm{N}\})$$

where

U: The first player which is specified as a "set of *mobile users*", i.e.,

$$\mathrm{U} = \{\mathrm{U}_1, \mathrm{U}_2 \ldots, \mathrm{U}_\mathrm{n}\}$$

N: the second player which is specified as a "set of *wireless networks*", i.e.,

$$\mathrm{N} = \{\mathrm{N}_1, \mathrm{N}_2, \ldots, \mathrm{N}_\mathrm{m}\}$$

$S_U = \{S_i\}_{i \in U}$ Set of strategies for *mobile users* i.e.,

$$S_U = \{S_1, S_2, \ldots S_i, \ldots, S_n\}.$$

$S_N = \{S_j\}_{j \in N}$ Set of strategies for *wireless networks* i.e.,

$$S_N = \{S_1, S_2, \ldots S_j, \ldots, S_m\}.$$

π_i: $S_U \rightarrow R$, represents a utility (Payoff) function for **ith** *mobile user*

π_j: $S_N \rightarrow R$, represents a utility (Payoff) function for ***jth*** *wireless network* with the strategy space S given as

$$S = \{S_U \times S_N\}$$

4 Game Solution Using NASH Equilibrium

Let i and j are the index values of strategies of Player 1 (s_i) and Player 2 (s_j), respectively, for a two-player noncooperative game, and i^* *and* j^* are the index values of the strategies at the NASH equilibrium [17]. Then the pure strategy pair (s_{i^*}, s_{j^*}) is in NASH equilibrium if

$$\pi_1(s_{i^*}, s_{j^*}) \geq \pi_1(s_{i'}, s_{j^*}), \quad \forall i' \neq i^* and\ i^*, i' \in n$$

and

$$\pi_2(s_{i^*}, s_{j^*}) \geq \pi_2(s_{i'}, s_{j^*}), \quad \forall j' \neq j^* and\ j^*, j' \in m$$

The NASH equilibrium strategies s_{i^*} and s_{j^*} is obtained using Best Response (BR) function as given below.

$$(s_{i^*}) = BR_1(s_{j^*}) = \text{argmax}_{s_i} \pi_1(s_i, s_j), \forall i = 1...m \text{ and } \forall j = 1...n$$

and

$$(s_{j^*}) = BR_2(s_{i^*}) = \text{argmax}_{s_j} \pi_2(s_i, s_j), \forall i = 1...m \text{ and } \forall j = 1...n$$

NASH equilibrium, if exists, gives the pair of strategies (s_{i^*}, s_{j^*}) which can provide best payoffs to both the players. In case NASH equilibrium does not exist, we may opt for suboptimal solutions which can converge the VHO decision algorithms.

4.1 Performance Measures of a Game

Price of Anarchy and *Price of Stability* [4] are two performance measures which are originally proposed for the analysis of economic model/policies using game theory. Here, we are proposing them to measure the performance of game theory models for heterogeneous wireless networks, by remodeling as shown below, where the payoff of the game players is represented as utility.

Price of Anarchy:

$$\gamma = \frac{U_{cs}}{U_{NE}} \tag{4}$$

where

$U_{cs} = Max_{s \in S}(\sum_{i \in N} u_i(s))$ is the maximum utility achieved by any network among all the considered networks and $U_{NE} = Min_{s \in S^{NE}}(\sum_{i \in N} u_i(s))$ is the utility at the worst case NE

Price of Stability:

$$\rho = \frac{U_{cs}}{U_{NE}} \tag{5}$$

where

$U_{cs} = Max_{s \in S}(\sum_{i \in N} u_i(s))$ is the maximum utility achieved and
$U_{NE} = Max_{s \in S^{NE}}(\sum_{i \in N} u_i(s))$ is the utility at the Best Case NE.

5 Conclusion

Game theory is an important mathematical optimization tool to solve the problems involving the contradicting goals of the constituent entities. The VHO decision problems in HWNs are some of such problems which may require to model the game between mobile users, game between wireless networks and, game between mobile users and wireless networks. In this paper, we have presented the three simple non-cooperative game formulations for the same along with the procedure to obtain the solution of the noncooperative games using NASH equilibrium. The concept of Price of Anarchy and Price of Stability is presented to measure the performanc of Game models. This work may be beneficial to the researchers who wish to apply game theory to solve the network selection and vertical handoff decision related problems in HWNs.

References

1. Goyal P, Lobiyal DK, Katti CP (2017) Vertical handoff in heterogeneous wireless networks: a tutorial. In: International conference on computing, communication and automation (ICCCA2017). IEEE, pp 1–6
2. Márquez-Barja J, Calafate CT, Cano JC, Manzoni P (2011) An overview of vertical handover techniques: algorithms, protocols and tools. Comput Commun 34:985–997
3. Kassar M, Kervella B, Pujolle G (2008) An overview of vertical handover decision strategies in heterogeneous wireless networks. Comput Commun 31:2607–2620
4. Wang L, Kuo G-SGS (2013) Mathematical modeling for network selection in heterogeneous wireless networks—a tutorial. IEEE Commun Surv Tutorials 15:271–292
5. Goyal P, Saxena S (2008) A dynamic decision model for vertical handoffs across heterogeneous wireless networks. Int J Comput Electr Autom Control Inf Eng 2
6. Trestian R, Ormond O, Muntean GM (2012) Game theory-based network selection: solutions and challenges. IEEE Commun Surv Tutorials 14:1212–1231
7. Huang W (2012) Application of game theory in wireless communication networks

8. Fu S, Li J, Li R, Ji Y (2014) A game theory based vertical handoff scheme for wireless heterogeneous networks. In: 10th international conference on mobile ad-hoc and sensor networks, pp 220–227
9. Chen M, Li F, Wang J (2011) A game theoretical approach of network selection algorithm in heterogeneous wireless networks. In: IET international communication conference on wireless mobile and computing (CCWMC 2011). IET, pp 148–153
10. Mark Felegyhazi J-PH (2007) Game theory in wireless networks: a tutorial
11. Chen Q-B, Zhou W-G, Chai R, Tang L (2011) Game-theoretic approach for pricing strategy and network selection in heterogeneous wireless networks. IET Commun 5:676
12. Bendaoud F, Abdennebi M, Didi F (2015) Network selection using game theory. In: 3rd international conference on control, engineering and information technology (CEIT). IEEE, pp 1–6
13. Charilas DE, Panagopoulos AD (2010) A survey on game theory applications in wireless networks. Comput Netw 54:3421–3430
14. Meenakshi, Singh NP (2016) A comparative study of cooperative and non-cooperative game theory in network selection. IEEE Int Conf Comput Tech Inf Commun Technol 612–617
15. Cesana M, Malanchini I, Capone A (2008) Modelling network selection and resource allocation in wireless access networks with non-cooperative games. In: 5th IEEE international conference on mobile ad-hoc and sensor systems. IEEE, pp 404–409
16. Antoniou J, Pitsillides A (2007) 4G converged environment: modeling network selection as a game. In: 16th IST mobile and wireless communications summit. IEEE, pp 1–5
17. Buttler J, Akchurina N (2013) Nash equilibria in normal games via optimization methods. In: 2013 European control conference (ECC)

Downlink Spectral Efficiency of ZF Precoding Based Multi-user MIMO System Over Weibull Fading Channel

Keerti Tiwari(✉) and Prabhat Thakur

Department of Electronics and Communication Engineering, Jaypee University of Information Technology, Waknaghat, Solan 173234, India
krt.tiwari@gmail.com, prabhat.thakur@mail.juit.ac.in

1 Introduction

Future cellular networks require a very high data throughput due to continuously increasing worldwide demand for wireless data traffic. This target can be accomplished with the help of spatially multiplexed multiple-input multiple-output (MIMO) systems without requiring more bandwidth or additional base stations (BSs) and with improved spectral efficiency (SE). MIMO is a potential technology in the field of wireless communication which outperforms the single-input single-output (SISO) in terms of high-data rate, low bit-error rate (BER) and enhanced signal-to-noise plus interference ratio (SINR) [1]. The effective process to increase the SE of a cellular network is to yield multiple parallel transmissions. Therefore, multi-user (MU) MIMO is the most scalable and appealing solution [2]. The canonical MU MIMO system contains a BS having N antennas that assists K single-antenna terminals or number of users. The BS practices its antennas for directing each signal on the way to its desired receiver in the downlink, and for separating the multiple signals received in the uplink. The terminal having multiple antennas is advantageous due to extra antennas for alleviating interference and improving the signal-to-noise ratio (SNR) instead of transmitting multiple data streams [3].

The researchers are still working to improve the spectral efficiency of MU MIMO with multi-antenna BSs. Consequently, some prominent works are included in [4–6]. Many non-linear precoding techniques have been investigated to achieve uplink and downlink capacity. However, to achieve the optimal performance, inter-user interference suppression is needed by interference-attentive transmit processing or iterative interference-attentive receive processing. In addition, non-linear techniques demand large mathematical computation and correct channel state information (CSI) [7]. Thus, the performance analysis and optimization of linear processing techniques have gained huge attention. The linear precoding e.g., zero forcing (ZF) is simple and easy to implement in MIMO system [4, 8].

The two popular approaches to implement the downlink and uplink transmission over a specified frequency band are frequency division duplex (FDD) and time division duplex (TDD) [9]. In FDD mode, the bandwidth comprises two parts: one for the uplink and other for the downlink. Pilot sequences are needed in both the downlink and

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_48

the uplink due to the frequency selective fading, giving an average pilot length of $(N+K)/2$ per sub channel. On the other hand, in TDD mode, the entire bandwidth is exploited for both downlink and uplink transmission, however at different time slots. If the system switches between downlink and uplink faster than the changing of channels, then it is sufficient to learn the channels in only one of the directions which leads to the average pilot length of $\min(N, K)$ per sub channel by sending pilots only in the most efficient direction. In the preferable operating regime of $N \gg K$, where TDD systems should send pilots only in the uplink, the pilot length becomes $\min(N, K) = K$ [10].

In [10], the SE for MU MIMO is analyzed over the mostly considered Rayleigh fading scenario. However, we have analyzed the SE for MU MIMO system in Weibull fading scenario which considers the multipath fading, indoor, outdoor and radar clutters [11]. In addition, we have analyzed the performance of the system by considering TDD and FDD mode with slightly high and low SNR regime.

The reminder of paper is organized as follows. Section 2 includes precoded MU MIMO system and Weibull channel model. Section 3 presents the simulation results and related discussion. Finally, the paper is concluded in Sect. 4.

2 System and Channel Model

The proposed system model comprises a base station (BS) having N number of antenna which aids the K number of users. For the sake of simplicity, we assume the imperfect CSI at the transmitter and the channels are modeled as the uncorrelated Weibull fading. The BS multiplexes the data stream per user for downlink and receives one stream per user in the uplink. The BS exploits its antennas to transmit each signal towards the desired receiver in the downlink, and to separate the multiple signals received in the uplink. The multiple antennas at the terminal allows to suppress interference as well as to improve SNR. In this case, $\min(N, K)$ represents the maximal number of data streams that can be simultaneously transmitted in the cell, while still being separable in the spatial user MIMO system. The precoding technique is considered to be linear one and known as zero forcing (ZF).

For an arbitrary BS l, the $x_l \in C^N$ represents the transmitted signal vector for its K users. Since we have considered the linear precoding, therefore, the transmitted vector is evaluated as

$$x_l = \sum_{t=1}^{K} \sqrt{\rho_{l,t}}\, W_{l,t}\, S_{l,t} \tag{1}$$

where the payload symbol $S_{l,t}$ is transmitted for user t in cell l and has unit transmit power $E\left\{\left|S_{l,t}\right|^2\right\} = 1$, $\rho_{l,t}$ denotes the assigned transmission power to the specified user, and $W_{l,t} \in C^N$, (for $t = 1, \ldots, K$) are conforming linear precoding vectors. The received signal $y_{l,k} \in C$ at user K in the cell l is defined as

$$y_{l,k} = \sum_{i=1}^{L} (h_{l,K}^{i})^{H} x_i + n_{l,k} \tag{2}$$

where $n_{l,k}$ and $h_{l,K}^{i}$ denotes the additive white noise with variance σ_{DL}^2, and channel gain from base station to the Kth user in the ith cell respectively.

Adding more number of BS antennas, both the high capacity achieving non-linear precoding and a simple linear ZF precoding come up to the instance without interference. Moreover, it is an important property called promising propagation. For the promising propagation, channels should be orthogonal to each other and hence, we considered the orthogonal channels. There is negligible correlation among the channels. As channel h is modeled as the Weibull channel model, it is represented in terms of Gaussian in-phase elements U and quadrature phase elements V of the multipath components [12]

$$h = (U + jV)^{2/\beta} \tag{3}$$

The amplitude of channel coefficient is $R = |h|$. Probability density function of Weibull random variable r is represented in [12] as

$$p(r) = \frac{\beta}{\Omega} r^{\beta-1} \exp\left(-\frac{r^{\beta}}{\Omega}\right) \tag{4}$$

where β is shape parameter, Ω is average fading power $(\beta > 0, \Omega \geq 0, r \geq 0)$. For $\beta = 1$ and $\beta = 2$, Weibull distribution is approached to exponential and Rayleigh distribution respectively. Correspondingly, with high β, fading severity reduces.

3 Simulation Results and Discussion

In this paper, we have evaluated the average downlink spectral efficiency of MU MIMO system with respect to the number of BS antennas using ZF precoding, and different types of CSI available at the BS over Weibull fading channel. Figure 1a, b illustrate the SE for downlink transmission with K = 10 users having SNR of 8 dB and 20 users having SNR of 1 dB respectively, in an uncorrelated Weibull fading environment for $\beta = 2$. However, Fig. 2a, b display the SE in the comparatively light fading environment ($\beta = 5$). The first case of SE for $\beta = 2$ shows the Rayleigh fading scenario which is comparable to [10], and validate the simulation results of this paper. The simulation results compare the SE achieved when having perfect CSI with the performance when having CSI estimated with pilot sequences of length t_p. The SE is shown as a function of the number of BS antennas, N, and Fig. 1a compares TDD mode for $\beta = 2$ using t_p = K = 10 with FDD mode using either t_p = 10, t_p = N, or

t_p = min(N, 40), where the last-mentioned selects an arbitrarily maximum pilot length of 40. In addition, Fig. 1b shows the same comparison while selecting $t_p = K = 20$ with FDD mode using either $t_p = 20$, $t_p = N$, or t_p = min(N, 40), keeping the pilot length and β same as Fig. 1a, however, reducing SNR from 8 to 1 dB. We observe that increased t_p also increases SE even at a very low SNR.

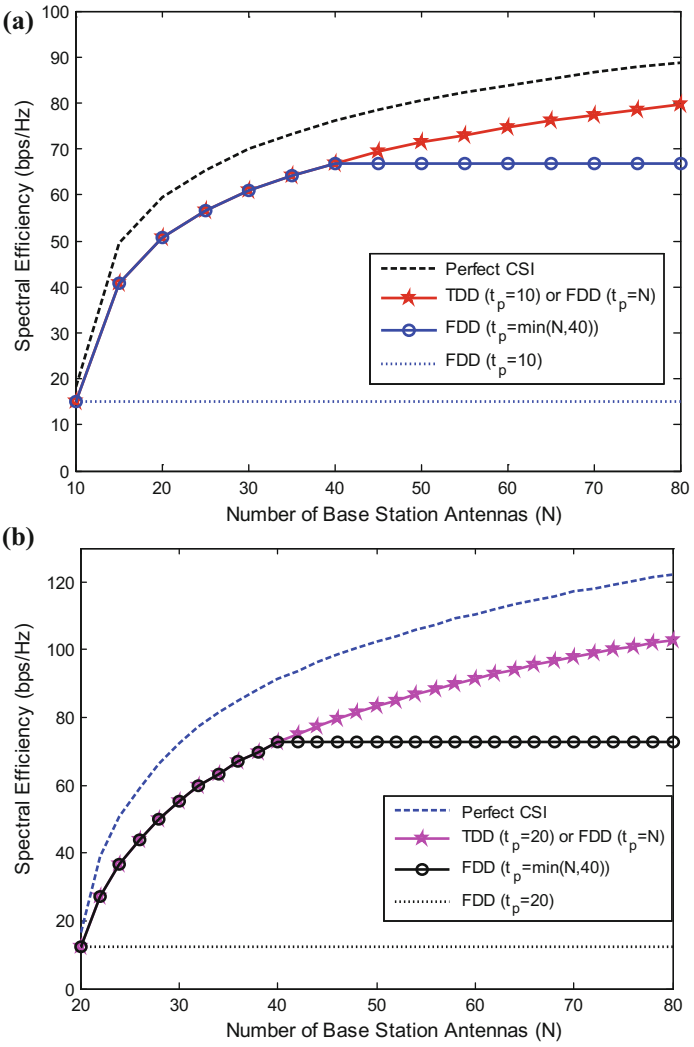


Fig. 1. Average downlink spectral efficiency using zero-forcing precoding with different forms of CSI available at the BS in Weibull fading for $\beta = 2$ **a** SNR = 8 dB **b** SNR = 1 dB

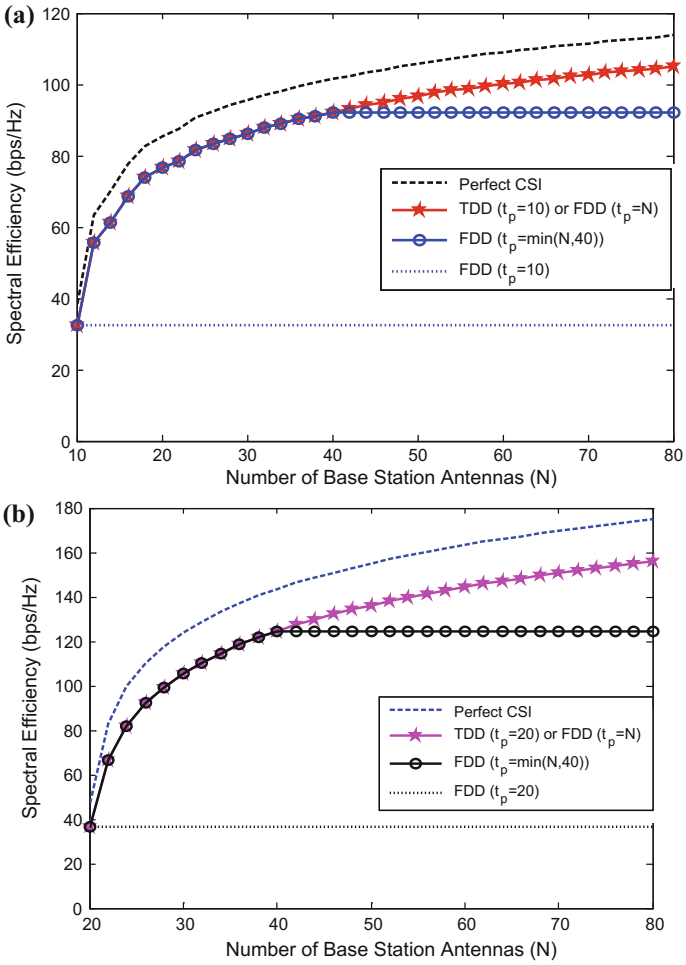


Fig. 2. Average downlink spectral efficiency using zero-forcing precoding with different forms of CSI available at the BS in Weibull fading for $\beta = 5$ **a** SNR = 8 dB **b** SNR = 1 dB

Figures 1 and 2 depict a noticeable performance degradation in TDD mode as compared to having perfect CSI. It is observed that the performance falls are significantly constant regardless of the number of BS antennas, thus TDD systems are always beneficial with added large number of antennas. Consequently, FDD systems are only beneficial by adding more antennas while keeping the pilot sequences larger when $t_p = N$. With $t_p = 10$ (Figs. 1a and 2a) and $t_p = 20$ (Figs. 1b and 2b), we cannot achieve the significant gain in SE by adding more than 10 and 20 antennas for $t_p = \min(N,40)$ respectively, while the performance steeps at 40 antennas. TDD operation shows complete scalability with regard to the number of BS antennas, however, FDD operation can employ more number of antennas with increased pilot overhead. It is practically feasible to arrange FDD systems with many antennas, mainly for slowly

varying channels where we can accept a large pilot overhead, but TDD proofs the superior choice in this context. Different values of β, e.g., 2 and 5 demonstrate the severe and non-severe fading scenario. In which, our results verify the results of [10] by changing the parameters under special case of Rayleigh fading. Much improvement in the SE can be seen by selecting high β value or less fading scenario. Thus, our results are valid in both the scenarios. We see the maximum improvement in SE with perfect CSI by changing β from 2 to 5.

4 Conclusion

This paper concludes that TDD is the superior mode since it not only involves shorter pilots than FDD, but is also extremely scalable due to the independency of pilot length from the number of BS antennas. The improved SE is achieved with MU MIMO ZF processing using TDD systems than FDD, however, we get a perceptible performance loss in TDD mode than having perfect CSI, and performance reduction is significantly constant irrespective of the number of BS antennas. Thus, TDD systems are more advantageous by adding more number of antennas than FDD. Consequently, FDD systems are only beneficial by adding more antennas while taking large pilot sequences. We achieve a significant performance gain in spectral efficiency even after reducing SNR at a very low value with increased pilots under Weibull fading. Consequently, SE is increased in less fading environment. The uplink works in the similar means in both TDD and FDD modes, whereas the advantage of scalability in TDD occurs in the downlink.

References

1. Mietzner J, Schober R, Lampe L, Gerstacker WH, Hoeher PA (2009) Multiple-antenna techniques for wireless communications—a comprehensive literature survey. IEEE Surv Tutorials 11(2):87–105
2. Gesbert D, Kountouris M, Heath R, Chae CB, Sälzer T (2007) Shifting the MIMO paradigm. IEEE Signal Process Mag 24(5):36–46
3. Björnson E, Kountouris M, Bengtsson M, Ottersten B (2013) Receive combining vs. multistream multiplexing in downlink systems with multi-antenna users. IEEE Trans Signal Process 61(13):3431–3446
4. Almelah HB, Hamdi KA (2016) On the downlink spectral efficiency of ZF beamforming over MIMO channels in multi-cell environments. In: 2016 IEEE global communications conference (GLOBECOM), Washington, DC, pp 1–6
5. Finn D, Ahmadi H, Cattoni AF, DaSilva LA (2016) Improved spectral efficiency through multiuser MIMO across small cells. IEEE Trans Veh Technol 65(9):7764–7768
6. Teklu MB, Meng W, Lei C, Gelmecha DJ, Li C (2017) SLNR-based precoding for single cell full-duplex MU-MIMO systems. IEEE Trans Veh Technol 66(9):7877–7887
7. Liao CH, Morikawa H (2015) Latticing the interference: non-linear pre-coding for non-orthogonal multiple access in downlink multi-user MIMO system. In: 2015 IEEE

26th annual international symposium on personal, indoor, and mobile radio communications (PIMRC), Hong Kong, pp 710–714
8. Thakur P, Tiwari K (2018) Error Rate Analysis of Precoded-OSTBC MIMO system over generalized-K fading channel. In: Konkani A, Bera R, Paul S. (eds) Advances in systems, control and automation, Lecture notes in electrical engineering, Springer, Singapore 442, pp 299–307
9. Casta MH, Nossek JA (2013) FDD and TDD single-user capacity bounds. In: 9th international ITG conference on systems, communication and coding (SCC) VDE, pp 1–6
10. Van Chien T, Björnson E (2017) Massive MIMO communications. In: 5G mobile communications. Springer International Publishing, pp 77–116
11. Tiwari K, Saini DS, Bhooshan SV (2016) On the capacity of MIMO Weibull-Gamma fading channels in low SNR regime. J Electr Comput Eng pp 1–7
12. Sagias NC, Karagiannidis GK (2005) Gaussian class multivariate Weibull distributions: theory and applications in fading channels. IEEE Trans Inf Theory 51:3608–3619

Product Recommendation System Using Support Vector Machine

K. Yesodha(✉), R. Anitha, T. Mala, and Shri Vindhya

Department of IST, Anna University, Chennai, India
yesodha@auist.net, {anitharagupathi,space.safia}
@gmail.com, malanehru@annauniv.edu

1 Introduction

The growth of online shopping in today's world is exceptional. The pleasure of shopping things online is remarkable. Shopping in the comfort of the house is the main feature which made online shopping a very big success. There are many shopping sites like Flipkart.com, Amazon.in, ebay.in, shopclues.com, and so on. There are certain limitations in online shopping. The main drawback about this online shopping is that the touch and feels factor of a general shopping is missing. Here, the user cannot touch the product or test the product before buying it. The customer needs some factor to come to a conclusion about the product, this is exactly where customer reviews get into the picture. Customer reviews play a vital role in this sector. There may be thousands of reviews on a product. It is highly impossible for the customer to read all the reviews, to decide the product, make this easy for the customer, this system was initiated. Though shopping from home is comfort, the customer cannot purchase by just looking at the images displayed. A good product with all negative reviews may not hit the higher sales rate.

2 Literature Survey

Most of the Product Recommendation Systems use the Reviews given by the Reviewers to provide the reliable useful information for the end user to take the decision to purchase the particular product.

Mostly reviews are presented with either useful or un-useful reviews. These reviews are mixed with both positive and negative reviews which make the online shopper to get a confused, and it is needed for a system to generate the aggregation of overall product reviews and provide the consolidated ratings from the multiple customer views. In existing works, the ratings are done only by considering the reviews in single sites. It usually provides a wrong view on particular product, whereas the same product has a high rating in another site.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_49

The methodology proposed by Garg et al. [1] has providing simple aggregating evaluation techniques for product reviews.

$$V(p) = \frac{\mathring{a}_{i=1}^{n}(u(T_i(P))\,\text{Polarity of reviews}\,(T_i(P)))}{\mathring{a}_{i=1}^{n}u(\text{Ti}(P))} \tag{1}$$

In the given Eq. (1), the evaluation for product P is considered as V(p). In this, the usefulness score u(Ti(P)) is used as a ranking function. This given ranking functions helps to eliminate unwanted information which creates overload for users who gone through the reviews.

Most of the Natural Language Analysis System uses statistical approaches to learn the given text corpora. The work discussed by Wiebe et al. [2], tested the similarity distribution between words based on their frequencies, collocations, and tag generated for the words. In addition, they have proved that the density of subjectivity clues in the surrounding context strongly affects how likely it is that a word is subjective. Finally, they used the subjectivity clues to identify opinionated chunks in text. The system proposed by Yu and Hatzivassiloglou [3] discussed about extracting the opinions from the fact reviews. The analyses are carried out both in context and document level. They used naive Bayes classifier, which is standard classifications algorithm for extracting the opinions from the context. They also considered and developed a polarity classification based on context-average approaches. These techniques used to provide an opinion of the review by considering all words in the context instead classifying it as positive or negative by considering a single word.

Some of the early works carried out by Turney [4], classifies the reviews as recommended or not recommended by the symbols (thumbs-up and thumbs-down) provided in the review. This work does not consider the exact opinion of the review that is given in the sentence level. The work is proposed by Pang et al. [5], for classifying the movie reviews. In this, the documents are simply classified by considering the probability of positive and negative words. If the document has higher probability for positive words, then it considered as a positive review document, otherwise it takes it as a negative review document.

Another work by Pang et al. [5], considers the multipoint scale approaches which extend the symbol from thumbs-up or thumbs-down to stars (1–5). By this way, they can provide ranking of the review documents from positive to negative. Popescu and Etzioni [6], discussed unsupervised information extraction system, in this system switching from classification to extraction. It can be used to determine the product quality, for customers to make the decision easily.

The framework proposed in Liu et al. [7], for analyzing and contrasts the customer reviews of competing products. This framework provides functionality of visualization to the customer to identify the accurate rate of that product clearly. Each and every product in e-commerce sites, it has the strength and weaknesses based on the customer opinions or reviews. Soo-Min Kim and Eduard Hovy discussed a system that

automatically extracted pro and con reasons from the opinion text. Shifting from classification to extraction, Ana-Maria Popescu and Oren Etzioni introduced Opine, an unsupervised information extraction system (Popescu and Etzioni [6]). Opine mines reviews to build a model of important product features, their evaluation by reviewers, and their relative quality across products (Kim and Hovy [8]).

3 System Design: Proposed System

The main aim of the system is to recommend the online shopper, if the product specified is worth buying. Reviews on that particular product from the shopping websites will be crawled and data preprocessing will be done. Gold Standard Usefulness score is calculated. Usefulness features are extracted from the data. With the above feature the Squared Correlation Coefficient and Mean Squared Error is calculated. Regression Analysis of two standard metrics is used to evaluate the quality namely, the Squared Correlation Coefficient and Mean Squared Error. Then, Review aggregation is done. Three types of review aggregation are considered and compared, they are, SimAvg, PredAvg, and GsAvg. This Review Aggregation is provided as the final output.

3.1 Detailed Architecture

The following Fig. 1 is the block diagram of the proposed system and Fig. 2 is the detailed architecture of the proposed system respectively.

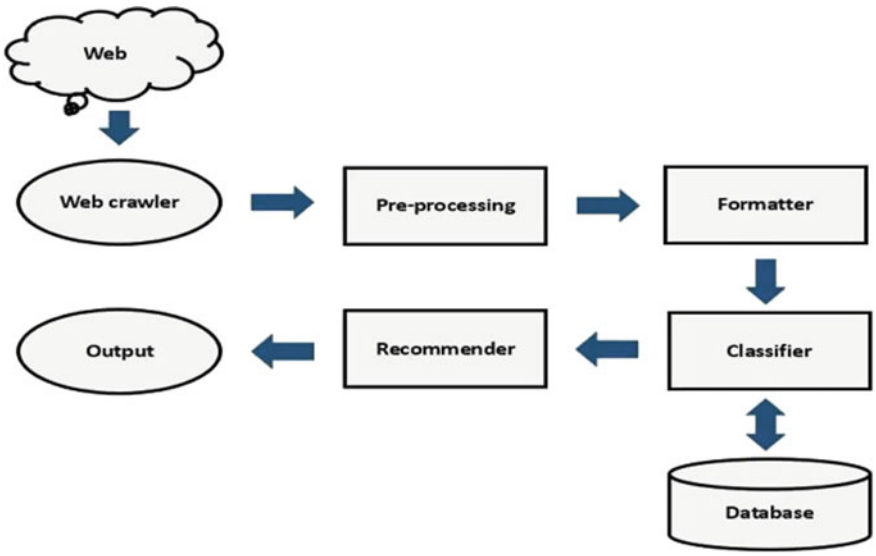


Fig. 1. Block diagram

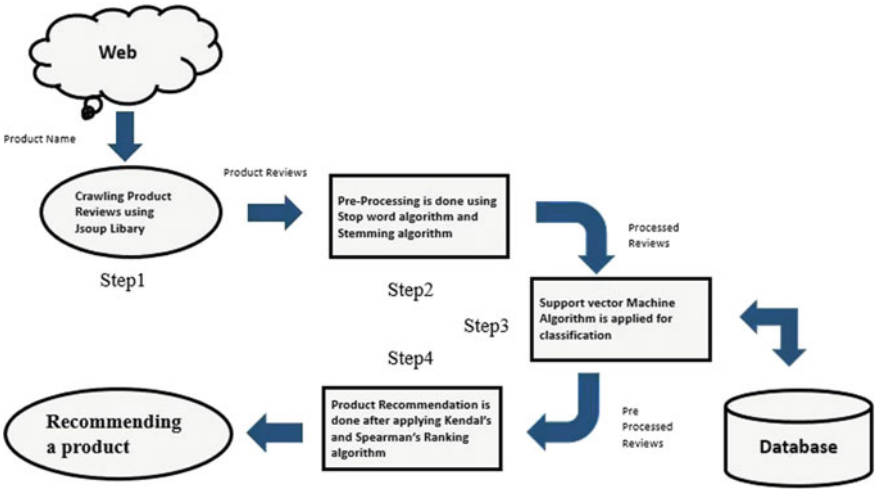


Fig. 2. Detailed architecture

The detailed system architecture components are,

A. Review Crawling

The online shopping site chosen to collect reviews is Flipkart. Product name is given as input to the program. JSoup Html Parser, a Java library is used to parse the web page. Product id is fetched from the web page of the product. URL of the review page framed with the product name and id. Then, the reviews are crawled.

B. Preprocessing

Customer reviews are extracted from Flipkart.com. The extracted reviews will have tags that need to be removed. Punctuations should also be removed. Regex is used to remove all the punctuations from the text. Methods in JSoupHTMLParser are used to extract only the required content from the HTML page fetched.

C. Lexical Similarity

Lexical similarity feature of a review is calculated. Few product reviews are mere copy of the Specifications. Such reviews are to be identified. Tf (Term frequency), idf (inverse document frequency) and Tf-idf (Term frequency—inverse document frequency) are calculated for estimating Cosine Similarity.

Lexsim is a feature which is used to measure the usefulness of the review. A customer review which is a loyal rephrase of the product specification will not work. At the same time a review without any impact of product specification may not be appreciated. Tf-idf for each term in the review and the product specification is calculated and cosine similarity of each product review against product specification is calculated. Then the obtained output is added to the .csv file. The output looks like Table 1 Cosine value is always between −1 and 1: the cosine of a small angle is near 1, and the cosine of a large angle near 180° is close to −1. This is good, because small angles should map to high similarity, near 1, and large angles should map near to −1.

D. Shallow Syntactic Feature

Good reviews will have certain writing style. It is assumed that all the good reviews tend to exhibit certain kind of similarity. Parts of speech tagger has been used to tag the most rated reviews and check if other reviews are similar to it. Again tf, idf, tf-idf and Cosine Similarity is calculated. It is believed that useful reviews hold certain Syntactic Features. Cosine similarity between all the user reviews and the useful review is calculated. For which, top ten useful reviews are handled as single review and its average is taken. There are about 33 parts of speech. For each term Tf-idf is calculated and its cosine similarity is estimated. During cosine similarity each and every review is compared with the aggregated review. Obtained cosine similarity is added to the .csv file, Table 1. This .csv file will be used to find the Useful regression. This will be very much helpful in finding the Squared Correlation Coefficient and the Mean Squared Error.

Table 1. Shallow syntactic feature, lexical similarity score, and calculated usefulness score

Product	Author	Real_Usefulness	Lexsim	Inherent polarity	SSF
Samsung Galaxy On5	Akash Sharma	0.97	0.02202	1	0.153463715
Samsung Galaxy On5	Nitin Kumar	0.95	0.01191	1	0.018647898
Samsung Galaxy On5	Karthik Iyer	0.93	0.15519	1	0.411112955
Samsung Galaxy On5	Akshay Singh	0.94	0.06279	0.8	0.123338421
Samsung Galaxy On5	Bajrang Sewag	0.91	0.04817	1	0.082809204
Samsung Galaxy On5	Adish Jain	0.89	0.07733	1	0.15587808
Samsung Galaxy On5	Ashok Kumar R	0.87	0.02499	0.2	0.050425057
Samsung Galaxy On5	Abhishek Singh	0.89	0.12395	1	0.370998695
Samsung Galaxy On5	Jimmy Kadakia	0.78	0.05505	0.6	0.092267496
Samsung Galaxy On5	Himanshu Kumar	0.81	0.16954	0.8	0.423765535
Samsung Galaxy On5	Sunil	0.79	0.23611	0.8	0.646471335
Samsung Galaxy On5	Sathish Kumar	1	0.02973	1	0.040440377
Samsung Galaxy On5	Pawan Kumar	0.79	0.13091	0.8	0.32697262
Samsung Galaxy On5	Tanmay Srivastav	1	0.13091	0.4	0.310121351
Samsung Galaxy On5	Rakesh Kumar	1	0.01689	1	0.02580486
Samsung Galaxy On5	Balaji	1	0.01191	1	0.018647898
Samsung Galaxy On5	JigarVira	1	0.08617	1	0.270993234
Samsung Galaxy On5	Kavish kumar	1	0.12503	1	0.300824885
Samsung Galaxy On5	Virkumar Chougul	0.8333	0.04188	1	0.054811478

E. Usefulness Regression

Regression performance of all the features is calculated against the real usefulness score. Two support vector regression is used, namely, e-Support vector regression and n-Support vector regression, for which Squared Correlation Coefficient and Mean Squared Error is calculated. Squared Correlation Coefficient and Mean Squared Error is calculated using weka tool. The input given to the weka tool should contain attributes such as Product Name, Author, Gold Standard Usefulness Score, LexSim Score, and SSF Score.

Table 2 represents the statistics for the attribute LexSim and Fig. 3 represents the Lexical Similarity graph.

Table 2. Statistics of lexical similarity

Statistics	Value
Minimum	0.012
Maximum	0.236
Mean	0.086
StdDev	0.06

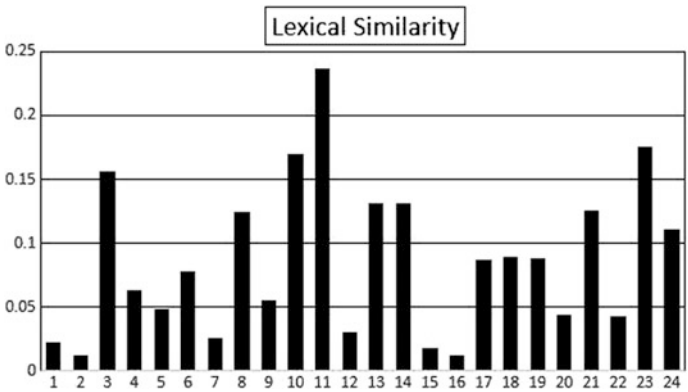


Fig. 3. Lexical similarity graph

Gold Standard Usefulness Score is very much important in this process. All the reviews in the Flipkart come with its usefulness rating. Once, the calculating Gold Standard Usefulness score is done, the content is written to the .csv file. Data such as product Name, Author, Gold Standard Usefulness Score is written to the file, it is depicted in the Table 1. Calculating usefulness feature is the next step after determining the real usefulness score. Table 3 displays the Squared Correlation Coefficient and Mean Squared Error value obtained using e-SVR for Gold Standard Usefulness Score, Lexsim and Ssf.

Table 3. e-SVR

	Gold standard usefulness score	Lexsim	Ssf
Squared correlation coefficient	0.02244	0.06584	0.1330
Mean squared error	0.007089	0.004761	0.02839

Table 4. n-SVR

	Gold standard usefulness score	Lexsim	Ssf
Squared correlation coefficient	0.003844	0.8284	0.15634
Mean squared error	0.006905	0.001962	0.028695

Table 4 displays the Squared Correlation Coefficient and Mean Squared Error value obtained using n-SVR for Gold Standard Usefulness Score, Lexsim, and Ssf. Figure 4 shows the Performance Evaluation of Squared Correlation Coefficient and Mean Squared Error.

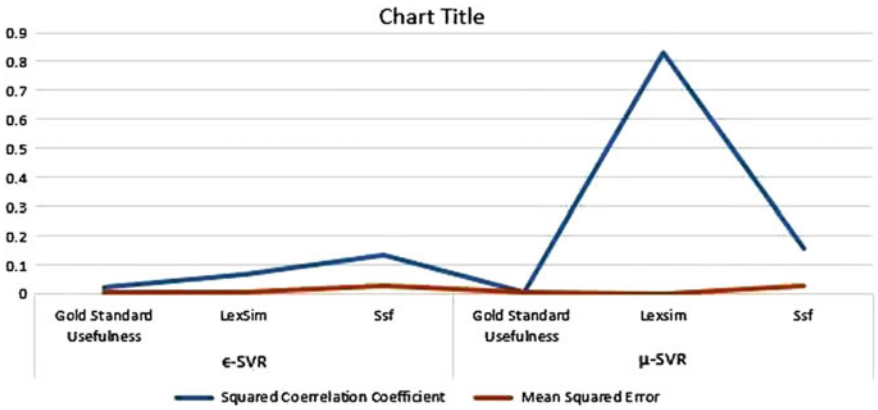


Fig. 4. Performance evaluation

F. Review Aggregation

Aggregated review polarity is calculated. Polarity of the review is found using the star rating given to the review by the shopping site. Three types of Review Aggregation are done; they are SimAvg, PredAvg, and GsAvg. Based on these averages calculated, the product can be recommended to the user.

4 Implementation and Results

Implementation details are discussed here. Reviews are examined, Lexical similarity and shallow syntactic features are extracted, cosine similarity is calculated, a weighted average of each system is calculated and ranking is done through Kendall's ranking correlation. Every Recommendation System should have accuracy, efficiency, and

scalability. So that Customers will get the true value of the items which they need for. In this Product Recommendation System giving the highest accuracy of review results than Compared with the previous works.

4.1 Review Aggregation

This section deals with review aggregation process, there are three types of review aggregation, SimAvg, PredAvg, and GsAvg.

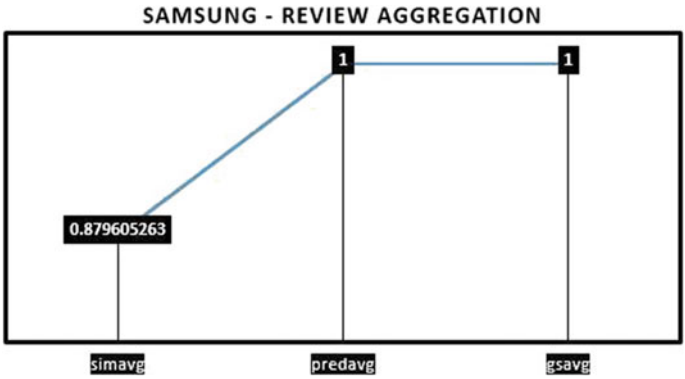


Fig. 5. Aggregation graph—SimAvg, PredAvg and GsAVg

SimAvg computes the aggregated polarity as a simple average of individual review polarity; that is $u(T_i(P))$ is always 1. PredAvg calculates the aggregated polarity as a weighted average of isolated review polarity; the weight allocated to each review T_i is its predicted usefulness score $u(T_i(P))$. GsAvg determines the aggregated polarity again as a weighted average of discrete review polarity; the weight assigned to each review T_i is its Gold Standard Usefulness Score $u(T_i(P))$. Review polarity is calculated with the help of the star rating given by the shopping e-commerce site to the reviews. Figure 5 represents the graph drawn on the values obtained from SimAvg, PredAvg and GsAVg.

5 Conclusion

Product Recommendation System using Support Vector Machines Regression. User is supposed to enter the product Name. Having the product name, reviews are fetched from the shopping site, preprocessing is done. Gold Standard Usefulness Score is calculated. Features such as LexSim and SSF are estimated. Usefulness Prediction is done based on two standards Support Vector Machine Regression. Three types of Review Aggregation, SimAvg, PredAvg, and GsAvg is done.

6 Future Work

In future, Correlation ranking will be calculated against the aggregations and sales rank. This ranking correlation will be done using Kendall's and Spearman's Ranking Correlation. Comparison between the ranking correlations can be done.

References

1. Garg A, Jayram TS, Vaithiyanathan S, Zhu H (2004) Generalized opinion pooling. In: Proceedings of the 8th international symposium on artificial intelligence and mathematics, vol 30, no 3, pp 155–167
2. Wiebe J, Wilson T, Bruce R, Bell M, Martin M (2004) Learning subjectivity language. Comput Linguist 30(3):277–308
3. Yu H, Hatzivassiloglou V (2003) Towards answering opinion question: separating facts from opinions and identifying the polarity of opinion sentences. In: Proceedings of the 2003 conference on empirical methods in natural language processing (ACL 03), pp 129–136
4. Turney PD (2001) Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th annual meeting association for computational linguistics (ACL 02), Association for computational linguistics, pp 417–424
5. Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the conference on empirical methods in natural language processing (EMNLP 02), Association for computational linguistics, pp 79–86
6. Popescu AM, Etzioni O (2005) Extracting product features and opinions from reviews. In: Conference on empirical methods in natural language processing (EMNLP 05), Association for computational linguistics, pp 339–346
7. Liu B, Hu M, Cheng J (2005) Observer: analyzing and comparing opinions on the web. In: Conference on empirical methods in natural association for computational linguistics, pp 342–351
8. Kim S-M, Hovy E (2006) Automatic identification of pro and con reasons in online reviews. In: Proceedings of the COLING/ACL on main conference on poster sessions, association for computational linguistics 2006, pp 483–490

A Master Map: An Alternative Approach to Explore Human's Eye Fixation for Generating Ground Truth Based on Various State-of-the-Art Techniques

Ravi Kant Kumar(✉), Jogendra Garain, Dakshina Ranjan Kisku, and Goutam Sanyal

Department of Computer Science and Engineering, National Institute of Technology, Durgapur, India
{vit.ravikant, jogs.cse, drkisku, nitgsanyal}@gmail.com

1 Introduction

Visual saliency is the perceptual eminence that makes an object, region, person, or pixel standalone from its surrounding and thus it drags our attention. Visual saliency guide our attention based on two standard mechanisms called bottom-up and top-down [1, 2].

In the practice of visiting a scene, selective visual attention [3] plays a crucial role to filter out unimportant information from that scene. It directs our brain to process only visually overlooked portions of a scene in the deeper level. Visual saliency [4, 5] gives intimation to any vision system for automatic selection of a region from a scene. Subsequently, our brain extracts the visual information from this region for the further processing. This mechanism allows extracting only "noticeable" information and ignores extraneous uninspiring information [6]. This proficiency of human vision system makes our organism to quickly gaze toward the precise target like acquaintances, strangers, stunning or horrible personalities, etc.

In order to provide the similar capability to the computer vision system, like humans, enormous researches have been established.

Based on bottom-up mechanism [7, 8], a biologically plausible basic framework has been developed by Koch and Ullman [9]. Inspiring from Koch and Ullman, Itti et al. [10] determine center-surround contrast using a Difference of Gaussians (DoG) approach.

Unlike, the biological-inspired model some researchers developed different aspects of saliency models which are purely computational [11–14]. On the basis of center-surround feature distances, Ma and Zhang [11] and Achanta et al. [14] had developed their saliency model. Hu et al. [12] determine salient location through heuristic measures. Some authors e.g. Hou and Zhang [13] computed saliency in the frequency domain.

In the third category of saliency model, authors have incorporated both biological-inspired model as well as the computational model altogether. These methods are Graph-Based Visual Saliency [15], Degree Centrality-Based Model [16].

Among various states of art models, we found Frequency-tuned Salient Region Detection [17], Minimum Barrier Salient Object Detection [18] and Context-Aware

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_50

Saliency [19] as the better one. Therefore, these models have been incorporated in our proposed work.

Even though several methods have been validated with the available ground truth, But in real time no any vision system has been developed which can deal with the generic vision task like we humans do. The main reason behind it is, the existing vision system has been developed by keeping the target to match with a particular ground truth which may be obtained through volunteer's voting or eye fixation experiment [20]. Many times this ground truth may be very specific.

Further, sections are arranged as: In Sect. 2, Requirements for a saliency is discussed. Various states of art methods for generating the saliency map have been discussed in Sect. 3. Section 4 described about our proposed saliency map for generating the ground truth. Experimental validation has been explained in Sect. 5. At last, Sect. 6 confers the concluding remark.

2 Requirement of Combined Saliency Map

We set the following requirements for a saliency detector:

- To develop a generalized saliency map.
- To get a single master map which covers the entire salient region obtained from various state of the art.
- Uniformly bring into light upon all prominent regions.
- Create precise boundaries of salient objects.
- Alternate approach for preparing ground truth data by using reverse engineering.
- Reduction in time and cost because of reusability.

3 Various State-of-the-Art Methods to Compute Saliency Map

To follow up the above requirements (mentioned in Sect. 2), some standard and well verified techniques has been considered for the fusion. These states of art techniques namely, Frequency-tuned Salient Region Detection [17], Minimum Barrier Salient Object Detection Methods [18] and Context-aware saliency [19] have been described in the following subsections.

3.1 Frequency-Tuned Salient Region Detection

In this technique, authors have filled the drawback of some state-of-the-art methods [5, 15, 16] for saliency detection. These methods depict salient regions with low resolution, poorly borders and also it is expensive to generate. Moreover, some other method [11, 13] highlights the object edges instead of highlighting the whole object, which fallout to consume entire spatial frequency of the input image. In this work, the author investigates the spatial frequencies and commences a frequency-tuned technique for computing the center-surround difference with intensity and color features and

proposes some benefits over existing techniques. It uniformly exposes salient location with clear borders, good resolution, and better computation. Authors compared the accuracy of the saliency maps with methods [10, 11, 13–15].

In this method, saliency map S for an image I of width W and height H pixels can thus be formulated as

$$S(x, y) = \left| I_{\mu} - I_{w_{hc}}(x, y) \right| \tag{1}$$

where I_{μ} is the arithmetic mean pixel value of the image and $I_{w_{hc}}$ is the Gaussian-blurred version of the original image to remove deep texture details and noises. The rule of difference is used because authors are finding the magnitude of the differences. This is quite efficient to compute. Also, operating this approach on the original image, without any downsampling a full resolution saliency map can be obtained. Rewriting the Eq. (1) for incorporating color and luminance, as:

$$\boldsymbol{S}(\boldsymbol{x}, \boldsymbol{y}) = \left\| \mathrm{I}_{\mu} - \mathrm{I}_{\mathrm{whc}}(x, y) \right\| \tag{2}$$

where I_{μ} is the mean feature vector, $\mathrm{I}_{\mathrm{whc}}(x, y)$ is the corresponding pixels values in Gaussian version (obtained by using 5 * 5 binomial mask) of the original image, and "‖ ‖" is the L_2 norm.

3.2 Minimum Barrier Salient Object Detection Methods

Here, authors have recommended a very efficient and prevailing, technique which is based on the Minimum Barrier Distance (MBD). In order to compute pixel-value fluctuation, MBD transform is found to be robust and effective to perform on original pixels without region generalization. In this work, an approximate MBD transform algorithm with 100X speedup has been implemented. After applying fast MBD transform algorithm, the proposed "salient object detection" process runs at 80 FPS, and performs significantly better than some state-of-the-art techniques [21, 22] on four large benchmark datasets with similar speed. Furthermore, this work is extended by a technique (which is based on color whitening) to control the cue. This comprehensive version improves the work and performs faster.

3.3 Context-Aware Saliency

This work proposes a novel kind of saliency called "context aware saliency" which identified the arousal region of the image. The saliency is computed based on four principles available in the psychosomatic prose. They are:

- Local low-level concern, which includes contrast and color of local regions.
- Global reflection, which incorporates repeatedly appearing features, also maintaining those features which are deviated from the norm.
- Visual association rules, which describe that visual appearance, may acquire one or numerous centers of gravity about which it is organized.
- Inclusion of high-level aspect such as human faces.

Further, including the above aspects, authors proposed a technique for saliency computation. There exists a variety of applications where the context of the prevailing objects is considered as crucial. Moreover, the contribution of context-aware saliency has been described in two applications namely retargeting and summarization.

4 Proposed Saliency Map for Generating Ground Truth

Composite Saliency map has been generated by using various state- of- art techniques mentioned in Sect. 3. These techniques have been implemented over different dataset having the large no of images. Results (Saliency map) obtained through these states of art techniques have also been validated with the proper ground truth. Accuracy has been measured and found to be satisfactory. Any particular method is not considered as a generic method. For a particular kind of images may be a particular approach is more suitable. Similarly, for another kind of images, any other method is more appropriate. Therefore, in this proposed work, we have incorporated the composite saliency map based on three effective state-of-the-art saliency model. The reason behind the selection of these models is to cover the necessary aspects like human psychological principle through [19], computing pixel-value fluctuation using [18], and frequency-tuned approach [17] for getting sharp edge and high resolution of salient region. As these models are considered as better than various leading state-of-the-art methods, therefore, these model altogether can enhance the performance significantly.

4.1 Saliency Map Generation Using Various Methods and Resizing

Saliency map using method [17–19] are shown in the Fig. 1. In the next step, we are combining these maps to generate a composite master map. In order to merge the corresponding pixels of these saliency maps, we have made them into the same size.

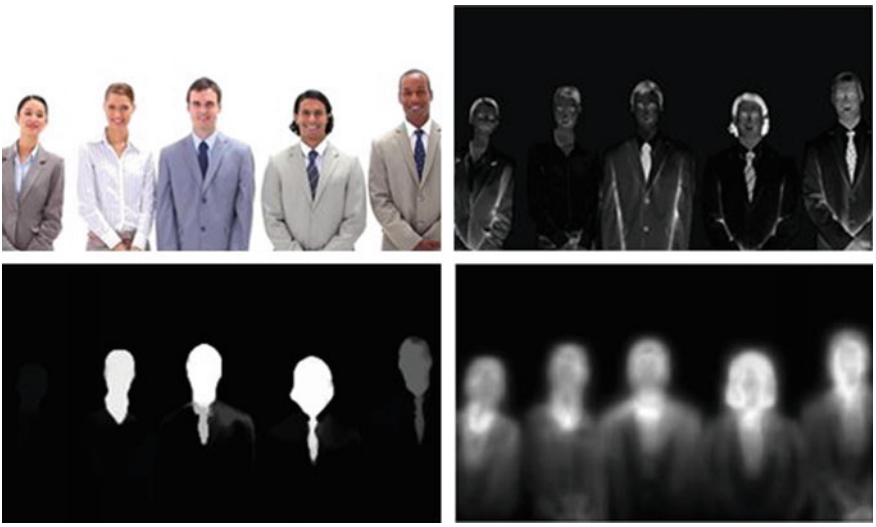


Fig. 1. Top-left: input image, top-right: saliency map using [17], bottom-left: saliency map using [18], bottom-right: saliency map using [19]

4.2 Proposed "Master Map" Through "Saliency Maps" Fusion

"Master map" of an image is obtained by fusing the corresponding saliency maps have been obtained by methods [17–19]. In this work, images are fused by using normal image addition process.

If "S_A", "S_{MB}", and "S_C" are saliency map generated by methods [17–19], respectively, then the proposed master saliency map "S_M" has been obtained as:

$$S_M = S_A + S_{MB} + S_C \tag{3}$$

The result obtained after combining the saliency map "S_A" and "S_{MB}" and then combining the saliency map "S_A" and "S_{MB}", and "S_C" have been shown in Fig. 2.

Fig. 2. Left: composite saliency map after fusing [17] and [18]. Right: master saliency map after fusing [17–19]

5 Experiment Validation and More Results

In this experiment, composite saliency map has been generated by fusing three leading state-of-the-art techniques [17–19]. The experiment has been performed on a large dataset. Result's validation through these techniques has already described in [17–19], with the proper ground truth. Accuracy in terms of matching with the ground truth has been measured and found to be satisfactory. As we have adopted these standard state of art methods and fused them together to generate the combined saliency map, therefore the resultant map grabs the properties of all its component maps. In other words, we can say, in the large scale, the resultant map has also been validated indirectly with the ground truth. The various results obtained from our proposed work have been shown below. In Fig. 3, we can clearly visualize that the proposed saliency map covers all the salient regions obtained from its component maps (Fig. 1). Therefore, the proposed map can be considered as the generalized map which occupies all properties of the state-of-the-art. Some more results obtained using our fusion techniques have been depicted in Fig. 3.

Fig. 3. Row wise—first column: input image, second column: saliency map using [17], third column: saliency map using [18], fourth column: saliency map using [19], fifth column: saliency map after fusing methods [17, 18], sixth column: proposed master saliency map after fusing methods [17–19]

6 Conclusion

In order to detect the salient regions in a scene, so far various saliency models have been developed. Many of these have been validated with the ground truth obtained from eye fixation experiment or based on volunteer's views. Therefore, these saliency models are considered as a state-of-the-art in the literature. In this paper, these techniques are utilized to generate composite saliency map and may be considered as the ground truth information. Preparing ground truth in such a way is inspired by reverse engineering and may be considered as an alternate approach (other than the volunteer's voting or eye fixation experiment) for preparing ground truth. In the complex images, where salient locations are very confusing and therefore preparing ground truth is also very difficult. In such images, results are varying with different state-of-the-art saliency model. To address this problem, this paper implements different combine strategies to achieve composite saliency map which can incorporate the properties of every individual map. Fusion of saliency maps in this way can be utilized to generate good ground truth information which can be alternate to base on volunteers or through eye fixation machine. As this approach incorporates the concepts of reusability, therefore it reduces the time and cost in the preparation of ground truth.

References

1. Pal R, Srivastava R, Singh SK, Shukla KK (2013) Computational models of visual attention: a survey. In: Recent advances in computer vision and image processing: methodologies and applications, pp 54–76
2. Itti L, Koch C (2001) Computational modelling of visual attention. Nat Rev Neurosci 2 (3):194–203

3. Koch C, Ullman S (1985) Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiol 4:219–227
4. Itti L, Koch C (2001) A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Res 40:1489–1506
5. Itti L, Koch C, Niebur E (1998) A model for saliency based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20:1254–1259
6. Gong Z (2015) Quantitative analysis of processing capacity in visual search: examining the impact of visual salience, involvement, and prior knowledge on selective attention to printed advertisements and memory formation. Doctoral dissertation. Texas Tech University
7. Toet A (2011) Computational versus psychophysical bottom-up image saliency: a comparative evaluation study. IEEE Trans Pattern Anal Mach Intell 33(11):2131–2146
8. Itti L, Rees G, Tsotsos J (2005) Models of bottom-up attention and saliency. In: Neurobiology of attention, p 582
9. Koch C, Ullman S (1987) Shifts in selective visual attention: towards the underlying neural circuitry. In: Matters of intelligence, pp 115–141. Springer, Netherlands
10. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20(11):1254–1259
11. Ma Y-F, Zhang H-J (2003) Contrast-based image attention analysis by using fuzzy growing. In: ACM international conference on multimedia
12. Hu Y, Xie X, Ma W-Y, Chia L-T, Rajan D (2004) Salient region detection using weighted feature maps based on the human visual attention model. In: Pacific Rim conference on multimedia
13. Hou X, Zhang L (2007) Saliency detection: a spectral residual approach. In: IEEE conference on computer vision and pattern recognition
14. Achanta R, Estrada F, Wils P, Susstrunk S (2008) Salient region detection and segmentation. In: International conference on computer vision systems
15. Harel J, Koch C, Perona P (2006) Graph-based visual saliency. In: Proceedings of annual conference on neural information processing systems (NIPS), pp 545–552
16. Pal R, Mukherjee A, Mitra P, Mukherjee J (2010) Modelling visual saliency using degree centrality. Comput Vision IET CV, 4(3):218–229
17. Achanta R, Hemami S, Estrada F, Susstrunk S (June 2009) Frequency-tuned salient region detection. In: Computer vision and pattern recognition, CVPR 2009, pp 1597–1604
18. Zhang J, Sclaroff S, Lin Z, Shen X, Price B, Mech R (2015) Minimum barrier salient object detection at 80 fps. In: Proceedings of the IEEE international conference on computer vision, pp 1404–1412
19. Goferman S, Zelnik-Manor L, Tal A (2012) Context-aware saliency detection. IEEE Trans Pattern Anal Mach Intell 34(10):1915–1926
20. Rayner K (2009) Eye movements and attention in reading, scene perception, and visual search. Quar J Exp Psychol 62(8):1457–1506
21. Zhu W, Liang S, Wei Y, Sun J (2014) Saliency optimization from robust background detection. In: CVPR
22. Jiang B, Zhang L, Lu H, Yang C, Yang M-H (2013) Saliency detection via absorbing markov chain. In: ICCV

Application of Fuzzy Clustering for Selection of Coating Materials for MEMS Sensor Array

Anurag Gupta(✉), T. Sonamani Singh, and R. D. S. Yadava

Sensors & Signal Processing Laboratory, Department of Physics, Institute of Science, Banaras Hindu University, Varanasi 221005, India
{anuragphyzics, soanamani.2065}@gmail.com, ardius@bhu.ac.in

1 Introduction

Electronic nose is a bioinspired engineering system that operates analogous to human nose for odor sensing. Like odorant sensing receptor neurons in human nose it consists of an array of chemical sensors imparted with broad range of chemical selectivities. The sensor array outputs are patterns of electrical signals encoded with odorants information. Application of data processing methods like scaling, normalization, feature extraction and pattern recognition analogous to biological functionalities of olfactory bulb and cortex part in human brain makes an Electronic Nose (E-nose) instrument [1].

A number of chemical sensor technologies are available, often with competing performance, which can be used for making an E-nose instrument. The sensors based on electro-acoustic and electro-mechanical principles such as surface or bulk acoustic wave (SAW or BAW) oscillators and microelectromechanical system (MEMS) cantilevers have been extensively investigated for E-noses for VOCs (volatile organic compounds) detection and identification. The chemical selective functionality in these sensors is imparted by parametric sensitization, most commonly by depositing a thin polymer film on the device surface. The odorants (chemical analytes) sorption into the polymer film generates a change in acoustic wave velocity, surface stress and/or elastic stiffness of the vibrating structure. An electrical read out circuitry converts these changes into electrical signal. A set of sensors coated with different polymers define the sensor array. The set of sensor array outputs upon exposure to a chemical sample define the array response pattern. The odor identity information is hidden in the response pattern. An analysis of response pattern by pattern recognition methods reveals the odor identity. The odor recognition capability comes from the proper selection of polymer coatings that can generate maximally discriminating information about the target chemical analytes. In other words, the polymers that provide divergent chemical interaction possibilities with target analytes will be the proper choice for making an E-nose. This is the most crucial step in polymeric E-nose design [2]. The conventional approach for the selection of sensing elements in an array is the empirical optimization, where various sensors are fabricated by using different polymers, their various combinations are evaluated as sensor array, and best among them are selected by the method of elimination. This involves long expensive experimentation, large man power

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_51

and material resource, hence makes the development of an E-nose instrument a costly affair.

Our group has been active over the last couple of years to develop some computational intelligence method that can explore the available polymer data base for making polymer selection (or at least short listing) prior to real sensors fabrication. This will reduce the development time and cost. The data base considered was the vapor-polymer partition coefficients. The partition coefficient (K) is defined as the ratio of analyte concentration in polymer phase (C_P) to that in vapor phase (C_V) in thermodynamic equilibrium in the limit of infinite dilution, $K = C_P/C_V$. The set of polymers that are maximally divergent with regard to partition coefficients of the target chemical analytes is expected to encode maximally divergent information the analytes.

Based on this idea, our group applied different clustering algorithms of computation intelligence to segregate polymers according to some measure of similarity among them in respect of their interaction with target chemical analytes. A polymer from each cluster is selected to represent maximally dissimilar (chemically divergent) subset for making the sensor array [3–8]. The clustering methods used are the principal component analysis (PCA), hierarchical clustering (HC), fuzzy c-means (FCM) clustering, fuzzy subtractive clustering (FSC), and particle swarm optimization (PSO). Each method employs a different measure of similarity. Therefore, it is expected that each one will lead to a different subset. The goodness of the selection is evaluated by analyzing the virtual (model based) sensor array responses for discrimination of different target analytes in presence of interferents specific to particular applications. The analyses in these works ranged from discrimination between living and dead human body odor [3], body odor for biometrics [4], explosives [5–7], chemical weapon agents, narcotics and drugs of abuse [7], food spoilage and freshness (milk and fish) [8–12], and oxidative stress [13]. The sensing platforms used were either SAW oscillators or MEMS cantilevers.

All the polymer selection methods resulted in 3–6-element sensor array. The analyses of synthetic array data generated by each method resulted in qualitative discrimination between various targets and associated interferents. However, the fuzzy clustering (both FCM and FSC) methods appeared to generate comparable and best results. It was difficult to decide the superiority of one over the other. Using the SAW sensing platform it was found that in respect of the discrimination between the target and interferents the FSC yields better results whereas in respect of the specific volatile marker detection the FCM is superior. It was conjectured that perhaps FCM and FSC reveal some complimentary information, and a method of combining or fusing them must be developed to obtain optimum results for both detection and identification. In the present study, we extend this effort using MEMS cantilever sensing platform and apply both FCM and FSC selection procedures by taking the spoilage/freshness detection of fish as a case study. The purpose of this study is (i) to ascertain the relative complementarity of the FCM and FSC selection procedures, and (ii) to examine whether their efficacies have to do something with the type of sensing mode.

2 Fish Spoilage Markers and Polymer Selection

Fish is a widely consumed food item. Ascertaining the stage of its degradation (spoilage) is important for human health safety. The degradation occurs due to microbes-linked biochemical processes in dead fish body which produce several volatile metabolites (biogenic amines). Most significant among the latter are trimethylamine (TMA) that is produced by bacterial reduction of TMA oxide, dimethylamine (DMA) generated from autolytic enzymes, ammonia due to deamination of amino acids, and nucleotides [14, 15]. The DMA and ammonia emissions become significant in the advanced stages of degradation, hence are unreliable for monitoring the level of spoilage threshold unfit for human consumption. In many research reports, the TMA has been established to be the most reliable volatile marker for monitoring the initial stages of degradation [16]. The TMA emits from the living fish also in low concentration, and its emission rate increases with time after fish death. Therefore, detection and concentration estimation of TMA in headspace of a fish sample is taken to be most reliable indicator of fish freshness or degradation level [17]. Typically, in fresh fish headspace the TMA concentration is in ppt–ppb range. As the degradation advances the concentration level goes up in ppb to ppm range. It may be remarked that some variation in TMA emission rate occurs due to types of fish species, geographical location, processing, and storage methods [18].

The problem of fish spoilage monitoring by SAW E-nose was investigated in an earlier study from our group [8]. Based on the available literature this paper identified 16 fish freshness/spoilage markers, 26 prospective polymers for sensor array coatings, and tabulated the partition coefficient data (Table 2, Table 3, and Table 5 respectively in [8]). The volatile markers and polymers are listed as follows. ***Fish freshness markers***: (1PTL) 1-Pentanol, (1HXL) 1-Hexanol, (1 OCL) 1-Octanol. ***Fish spoilage markers***: (TMA) Trimethylamine, (MM) Methyl mercaptan, (ET) Ethanol, (1PPL) 1-Propanol, (1BTL) 1-Butanol, (2BTN) 2-Butanone, (AC) Acetone, (TO) Toluene, (EB) Ethyl benzene, (MX) m-Xylene, (PX) p-Xylene, (OX) o-Xylene, (ST) Styrene. ***Prospective polymers***: (PIB) Polyisobutylene, (PECH) Polyepichlorohydrin, (OV-202) Polytrifluoropropylmethylsiloxane, (OV-25) 75% phenyl 25% methyl polysiloxane, (PEI) Polyethylenimine, (PDMS) Polydimethylsiloxane, (PMCPS) Polymethylcynopropylsiloxane, (PMPS) Polymethyl phenylsiloxane, (ZDOL) Fomblii-ZDOL, (PVPR) Polyvinyl propionate, (PVA) Polyvinyl acetate, (PMA) Polymethyl Acrylate, (FPOL) Fluoropolyol, (PVTD) Polyvinyltetradecanal, (PEM) Polyethylenemaleate, (SXCN) Polybiscyanopropylsiloxane, (SXFA) Hexafluoro-2-propanol substituted polysiloxane, (SXPYR) Alkylaminopyridyl-substituted polysiloxane, (P4V) Poly(4-vinylheduorocumyl alcohol, (SXPHB) phenyl methyl variant of polysiloxanes, (PMHS) Polymethylhydrosiloxane, (PLF) Linear functionalized polymer, (PBF) Branched functionalized polymer, (PMFTPMS) Trifluoropropylmethylsilox. The details of FCM and FSC algorithms are described in [8] and [9]. In [8] only FCM was used and in [9] both FCM and FSC were compared. Following the same approach, we obtained the following selection subsets in the present study.

Table 1. Polymers selected for making MEMS sensor array

Selection method	Selected polymers (mass density ρ_2 in 10^3 kg/m^3 [8])					
FCM	OV202 (1.25)	PEI (1.05)	P4V (1.44)	SXPYR (1.00)	SXFA (1.48)	PBF (1.30)
FSC	PMPS (1.12)	PEM (1.35)	P4V (1.44)	SXPHB (1.15)	PMHS (0.99)	PLF (1.43)

3 MEMS Sensor Model

MEMS cantilever coated on one surface with a thin polymer film responds to the vapor exposure via changes in the mass and volume of the polymer and surface stress at the polymer–cantilever interface. The volume change occurs due to the addition of free volume associated with the process of vapor solvation. Since the film is attached to the cantilever surface it can not expand in planes parallel to the surface. If the volume change (polymer swelling) is assumed to be isotropic this constraint has two consequences. (i) The polymer film thickness is increased because it is free to expand in the surface normal direction. (ii) Due the interfacial nonexpandability the film is elastically stretched in the planer directions. The latter produces a difference in surface stresses across the cantilever thickness causing downward bending of the cantilever beam. The gravimetric effect due to mass loading is usually small, and is ignored in calculating the bending effect. The mass loading and the volumetric swelling, however, change the cantilever mass and stiffness constant, hence the cantilever resonance frequency [19].

These effects are commonly used in two types of sensor operating modes: static and dynamic. These are schematically shown in Fig. 1. In the static mode, the cantilever bending due to differential surface stress is read by the laser beam deflection method. The vapor sorption induced deflection with reference to the undeflected condition defines the sensor signal, shown as Z_L. In the dynamic mode, the cantilever is made movable plate of a parallel plate capacitor against a rigid ground surface, and is driven by an external AC source in resonance condition. In this mode the cantilever is usually biased by a DC voltage so that it remains in the state of a static deflection state before the AC drive is applied. The amplitude of the AC drive is kept small so the entire vibratory sweep of the cantilever remains always on one side of the neutral position towards the ground surface, and it does not reach the pull-in condition [20]. This is done to avoid frequency doubling and instability. The resonance frequency is read by a laser Doppler vibrometer. The vapor sorption induced change in the resonance frequency defines the sensor output in dynamic mode, shown as Δf_{res} in Fig. 1.

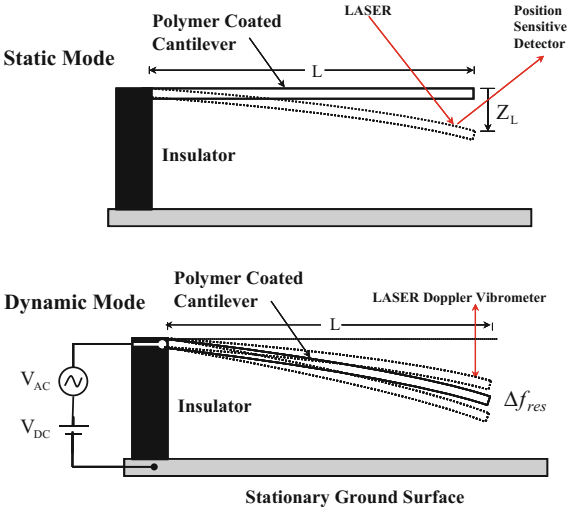


Fig. 1. Schematic of MEMS cantilever sensor operation in static and dynamic mode

The static mode bending is calculated by using the Stoney formula which relates the beam curvature with the differential surface stress across the beam thickness. Since the cantilever length is much larger than the width the interfacial nonexpandability constraint will result in mainly longitudinal strain. Integrating this formula twice and applying the fixed-free boundary conditions of the cantilever tip deflection (Z_L) is obtained as

$$Z_L = \left[3L^2 \frac{h_2(h_1+h_2)}{h_1^3} \frac{(1-v_1)}{E_1}\right]\sigma \tag{1}$$

with

$$\sigma = \frac{1}{3} E_2 V_a K C_V \tag{2}$$

where (L, h_1, E_1 and v_1) denote, respectively, the cantilever length, thickness, Young's modulus and Poisson's ratio, (h_2, E_2) denote, respectively, the polymer film thickness and Young's modulus, V_a denotes the specific volume of the sorbed vapor in ml/mol, K denotes the equilibrium partition coefficient and C_V denotes the chemical analyte concentration in the vapor phase in mol/ml; σ represents longitudinal stress at the polymer film cross section.

The resonance frequency of the coated MEMS cantilever under the conditions of soft polymer coating and equilibrium sorption is given as ([21], Eq. (12)).

$$f_{res} = \frac{\alpha^2}{2\pi L^2}\sqrt{\frac{\widetilde{E}_1 h_1^3}{12(\rho_1 h_1 + \rho_2 h_2)}} \times \left(1 + \frac{h_2(h_2^2 + 3(h_1 + h_2)^2}{2h_1^3}\frac{\widetilde{E}_2}{\widetilde{E}_1}\right) \tag{3}$$

where ρ_1 and ρ_2 are the mass densities of the cantilever and the polymer materials, respectively, $\widetilde{E}_1 = E_1/(1 - v_1)$, and α is constant that depends on the mode of vibration. For free vibrations in the fundamental mode (no axial loading) the value of mode constant $\alpha = \alpha_0 = 1.875$. However, the polymer swelling due to vapor sorption generates longitudinal film stress (σ) as discussed in the preceding given by (2). This acts as axial load on the cantilever beam. The axial load per unit beam length will be given as

$$\overline{\sigma} = \frac{\sigma(wh_2)}{L} \tag{4}$$

where w denotes the beam width. The effect of surface stress produced axial loading on α has been theoretically calculated in [22]. It is shown that α gets modified as

$$\alpha = \alpha_0\left(1 + \frac{2\overline{\sigma}L^3}{\pi^2 E_1 I_1}\right)^{1/4} \tag{5}$$

where $I_1 = \frac{1}{12}wh_1^3$ is the cross-sectional moment of inertia.

The changes in Z_L and f_{res} under exposure to vapor samples define the sensor outputs. From (1) and (3), it can be seen that these result from the changes in the polymer mass density and thickness. Denoting the latter by ρ_2' and h_2', it can be seen that

$$h_2' = h_2/(1 - V_a K C_V) \tag{6}$$

$$\rho_2' = (\rho_2 + M_a K C_V)(1 - V_a K C_V) \tag{7}$$

where M_a denotes the molar mass (g/mol) of the sorbed vapor. The calculation of Z_L and f_{res} by using (6) and (7) in (1) and (3) and subtracting from these their values prior to vapor exposure defines the output from these sensors. In the analyses to follow the normalized tip deflection Z_L/L and the fraction change in resonance frequency $\Delta f/f_{res}$ will be taken to be the signal from these sensors.

4 Data Generation

Assuming rectangular cross-section Si-MEMS cantilever to be the sensing platform, the two 6-element sensor arrays were defined by the two sets of 6 polymers selected by the FCM and the FSC methods, Table 1. The sensor array responses were calculated by

using (1) and (3) for both the static and the resonant modes. The geometrical and material parameters of the cantilever beam are $L = 200$ μm, $w = 40$ μm, $h_1 = 1$ μm, $\rho_1 = 2330$ kg/m^3, $E_1 = 169 \times 10^9$ N/m^2, $\nu_1 = 0.22$. The elastic parameters of all the selected polymers were not available. Therefore, $E_2 = 1 \times 10^8$ N/m^2 and $\nu_2 = 0.3$ (typical of rubbery polymers at $T = 298$ K) were assumed for all the polymers [23, 24]. For the dynamic response calculations the value of E_2 was assumed to be an order of magnitude higher than this to represent the high frequency limiting elastic behavior of the polymers [25]. The mass densities of the polymers (ρ_2) are listed in Table 1. The values of vapor-polymer partition coefficients were taken from the Table 5 in [8]. The molar specific volumes of the analytes V_a were estimated from their liquid phase densities ρ_L (g/ml) and molar mass M_a (g/mol) as $V_a = (M_a/\rho_L)$ ml/mol by data from [26]. The concentrations of the analytes and interferents in vapor phase were varied over the same range as specified in Table 2 of [8]. In synthetic data generation additive noise sources were included at the outputs of each sensor. For the static mode, the noise source is assumed to be uniformly distributed thermal fluctuations of the cantilever tip characterized by the root mean square deflection $\Delta Z = \sqrt{\frac{4}{3}k_B T/k}$ where k_B Boltzmann constant, T absolute temperature, and k represents the cantilever beam stiffness [27]. For the dynamic mode, Gaussian frequency fluctuation was assumed with mean zero and variance defined as $<\Delta f^2> = f_{res}/2Q$ with Q being the quality factor. A typical value of $Q = 500$ was assumed for all calculations. The fundamental frequency of the cantilever without coating was calculated to be $f_{res} = 38.939$ kHz, and the polymer coatings generated frequency shifts over [4.92, 7.17] kHz. The array response data were generated for 100 samples of each analyte vapor for ppb–ppm range of concentration as in [8]. Thus, each data matrix consisted of 1600 rows and 6 columns.

5 Results and Discussion

The raw data were preprocessed by mean centering, scaling normalization, and logarithmic scaling as done in previous studies [3, 7, 8]. The data matrix consisted of the vapor samples in rows and the polymer-coated sensors in columns were processed by principal components analysis (PCA) using MATLAB function '*princomp*'. The principal components with largest eigenvalues carry the most information. The principal component (PC) scores are the projections of data vectors onto the principal component eigenvectors. Figure 2 shows the PC score plots (PC1-PC2) for both the FSC and the FCM selected arrays in static and resonant modes as indicated.

We can make some specific observations from these plots either by comparing the results for the static and the resonant mode sensing for a given polymer selection or for the FSC and FCM selection methods for a given sensing mode. From the results in Fig. 2a, b for the FSC selection it can be seen that the spoilage markers (labeled 1–6) are clearly separated from the interferents and the freshness markers, and the separation is distinctly better for the dynamic mode sensing. However, it can be noted that the intraclass compaction within the spoilage group is much better with the static mode sensing. This has the implication that if in an application only the qualitative detection of spoilage is desired the resonant mode sensing will perform better. However, if a

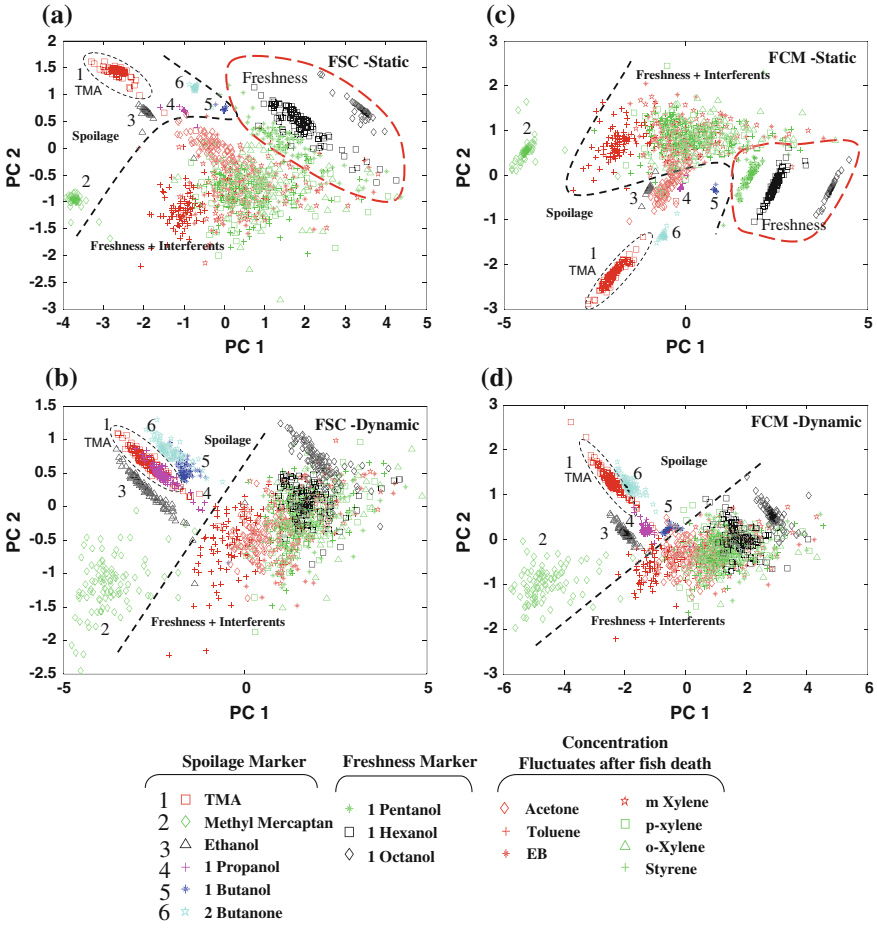


Fig. 2. PC1-PC2 score plots for the data generated by both the static mode and resonant sensors arrays: **a** and **b** comparing the results for FSC selection, and **c** and **d** for the FCM selection

quantitative estimation of the stage of spoilage is also desired then the static mode sensing will produce better results because the quantitative estimation is usually done by monitoring and calibrating a single marker analyte.

In several studies on fish spoilage detection trimethylamine (TMA) has been identified to be the most reliable marker. From Fig. 2a, c one can notice that TMA is well separated from the spoilage-interferents boundary as well as from the other intr-aclass markers for the static mode sensing for FSC and FCM selections. However, qualitative separation between spoilage and freshness/interferents is the best FSC-dynamic sensing. The FCM selection yields comparable results in sensing modes but the FSC selection appears to score better. Therefore, if TMA alone is considered for the detection of fish spoilage the FSC-static sensing is the obvious choice. However, the cost and complication associated with the static mode readout (laser reflection

vibrometer) may be much higher than the dynamic mode readout (frequency shift of electronic oscillators). Therefore, a trade-off between cost and performance is demanded then FCM-dynamic system may be the choice, compare the results in Fig. 2b, d. In terms of performance alone another point in favor of the static mode sensing is that the separation between three freshness markers (1-pentanol, 1-hexanol and 1-octanol) from the interferents, Fig. 2a, c is much better than that in dynamics mode.

6 Conclusion and Future Work

The analysis concludes that the FSC selection of polymers in combination with the static-mode MEMS cantilever sensing yields the best results for both qualitative detection as well as quantitative estimation of the stage of the fish spoilage. However, qualitatively the results for all the selection and sensing combinations (FSC-static/dynamic, FCM-static/dynamic) are comparable, and for a trade-off between cost and performance the FCM-dynamic system may be preferred. A closer scrutiny by focusing on the individual markers reveals that FSC and FCM selections throw up polymers with complimentary characteristics. Therefore, more exhaustive simulation studies based on different target detection cases and sensing platforms are needed. In our earlier studies mainly SAW sensing platforms were employed. Therefore, a comparative analysis between SAW and MEMS E-noses is imminent.

Acknowledgments. The author Anurag Gupta is thankful to UGC (Government of India) for providing JRF Fellowship and T. Sonamani Singh for providing BSR Fellowship.

References

1. Schiffman SS, Pearce TC (2003) Introduction to olfaction: perception, anatomy, physiology, and molecular biology. In: Pearce TC, Schiffman SS, Nagle HT, Gardner JW (eds) Handbook of machine olfaction, Chapter 1, pp 1–31. Wiley-VCH, Weinheim
2. Yadava RDS (2012) Modeling, simulation and information processing for development of a polymeric electronic nose system. In: Korotcenkov G (ed) Chemical sensors: simulation and modelling, 3: solid-state devices, Chapter 10, pp 411–502. Momentum Press, New York
3. Jha SK, Yadava RDS (2011) Designing optimal surface acoustic wave electronic nose for body odor discrimination. Sens Lett 9:1612–1622
4. Jha SK, Yadava RDS (2011) Statistical pattern analysis assisted selection of polymers for odor sensor array, 2011. In: IEEE international conference signal processing, communication, computing and networking technologies (ICSCCN 2011), pp 42–47
5. Jha SK, Yadava RDS (2012) Data mining approach to polymer selection for making SAW sensor array based electronic nose. Sens Transducers J 147(12):108–128
6. Jha SK, Yadava RDS (2011) Application of fuzzy inference system based on fuzzy subtractive clustering for, quantitative recognition in electronic nose systems. In: Proceedings 16th national seminar on physics and technology of sensors (NSPTS-16), 11–13 Feb 2011, pp 309–313. Lucknow, India

7. Verma P, Yadava RDS (2013) A data mining procedure for polymer selection for making surface acoustic wave sensor array. Sens Lett 11:1903–1918
8. Verma P, Yadava RDS (2015) Polymer selection for SAW sensor array based electronic noses by fuzzy c-means clustering of partition coefficients: model studies on detection of freshness and spoilage of milk and fish. Sens Actuators B Chem 209:751–769
9. Singh TS, Verma P, Yadava RDS (2017) Fuzzy subtractive clustering for polymer data mining for SAW sensor array based electronic nose. In: Proceedings 6th international conference on soft computing for problem solving (SocProS 2016), 23–24 Dec 2016, Patiala, India, AISC Series, Springer, 546, pp 245–253
10. Singh TS, Yadava RDS (2017) Application of PSO clustering for selection of chemical interface materials for sensor array electronic nose. In: Proceedings conference on soft computing: theories and applications (SoCTA 2016), 28–30 Dec 2016. AISC Series, 583. ed M Pant, K Ray, T K Sharma, S Rawat and A Bandyopadhyay (Singapore: Springer Singapore) pp 449–456
11. Verma P, Yadava RDS (2014) Application of fuzzy c-means clustering for polymer data mining for making SAW electronic nose. In: Satapathy SC, Udgata SK, Biswal BN (eds) Proceedings FICTA 2013, Advances in intelligent systems and computing, Springer AISC Series, 247, pp 1–8
12. Singh TS, Gupta A, Yadava RDS (2017) On development of electronic nose for fish spoilage detection. SST J Appl Phys 8(1):9–13
13. Gupta A, Singh TS, Yadava RDS (2017) Polymer-coated MEMS chemical sensor array for monitoring oxidative stress by breath analysis. In: Proceedings international conference on innovations in information, embedded and communication systems (ICIIECS 2017), 17–18 March 2017, Coimbatore, India, IEEE Xplore, pp 1–8
14. Haugen JE (2001) Electronic noses in food analysis. In: Rouseff RL, Cadwallader KR (eds) Advances in Experimental medicine and biology. Springer, US, Boston, MA, pp 43–57
15. Natale CD, Macagnano A, Davide F, D'Amico A, Paolesse R, Boschi T, Faccio M, Ferri G (1997) An electronic nose for food analysis. Sens Actuators B Chem 44:521–526
16. Schweizer-Berberich P-M, Vaihinger S, Göpel W (1994) Characterisation of food freshness with sensor arrays. Sens Actuators B Chem 18:282–290
17. Sadok S, Uglow RF, Haswell SJ (1996) Determination of trimethylamine in fish by flow injection analysis. Anal Chim Acta 321:69–74
18. Rouseff R, Cadwallader K (2001) Headspace techniques in foods, fragrances and flavors: an overview. In: Rouseff RL, Cadwallader KR (eds) Adv Exp Med Biol. Springer, Boston, MA, USA, pp 1–8
19. Wenzel MJ, Josse F, Heinrich SM, Yaz E, Datskos PG (2008) Sorption-induced static bending of microcantilevers coated with viscoelastic material. J Appl Phys 103:64913
20. Younis MI (2011) MEMS linear and nonlinear statics and dynamic. Springer, New York
21. Dufour I, Lochon F, Heinrich SM, Josse F, Dominique R (2007) Effect of coating viscoelasticity on quality factor and limit of detection of microcantilever chemical sensors. IEEE Sens J 7:230–236
22. McFarland AW, Poggi MA, Doyle MJ, Bottomley AL, Colton JS (2005) Influence of surface stress on the resonance behavior of microcantilevers. Appl Phys Lett 87:53505
23. Ferry JD (1948) Viscoelastic properties of polymer solutions. J Res Natl Bur Stand 41:53–62
24. Shen C-Y, Shen Y-T, Wu L (2002) Viscoelastic properties of polymer films on surface acoustic wave organophophorous vapor sensors. J Mater Sci 37:296–301

25. Yadava RDS, Kshetrimayum R, Khaneja M (2009) Multi frequency characterization of viscoelastic polymers and vapor sensing based on SAW oscillators. Ultrasonics 49:638–645
26. Mackay D, Shiu WY, Ma K-C, Lee SC (2006) Handbook of physical-chemical properties and environmental fate for organic chemicals, 2nd edn. CRC Press, I - IV
27. Butt H-J, Jaschke M (1995) Calculation of thermal noise in atomic force microscopy. Nanotechnology 6:1–7

Design of H Infinity (H_∞) Controller for Twin Rotor MIMO System (TRMS) Based on Linear Matrix Inequalities

Sumit Kumar Pandey(✉), Jayati Dey, and Subrata Banerjee

National Institute of Technology, Durgapur, India
{skpdmk, deybiswasjayati}@gmail.com,
bansub2004@yahoo.com

1 Introduction

In the recent days, many techniques have been studied frequently in order to control the unmanned air vehicle (UAV). The TRMS is a type of setup used for the experimental purpose because it resembles the helicopter model and it is difficult to test the designed controller directly to helicopter [1, 2]. In TRMS there is two inputs and two outputs and these inputs and outputs are coupled to each other.

Linear-state variable feedback controller is designed for nonlinear system with parametric uncertainty using linear matrix inequality [3]. This work is also applied to inverted pendulum system. Model predictive controller is implemented in nonlinear system with uncertainties using linear matrix inequality technique [4]. A linear matrix inequality (LMI)-based controller is designed for inverted pendulum system [5]. The balancing of the two-wheeled mobile manipulator has been done by using H infinity controller [6]. The developed controller is robust with respect to perturbations. The authors present a neural network based observer for the TRMS to estimate the unknown nonlinearities [7].

Linear matrix inequality is a competent computational method to solve the decisive control problem by algebraic approach. It is a part of a new generation computational tool which converts the control problem into a semi-definite programming problem. In general it is found during the literature review that LMI process is particularly used to solve the theoretical numerical problems. But since the past few years the demand of application of LMI in industrial engineering is increasing which motivate the authors to design the H infinity controller for TRMS on the basis of LMI in the present work. First, the feasibility condition with controller is checked and when the feasibility condition is satisfied the algorithm provides the controller gain matrix. The remainder of the paper is arranged as below. Next section reports the mathematical modeling of TRMS. Section 3 represents the decoupling technique of TRMS. In Sect. 4 design of H infinity controller is described. Section 5 covers the simulation results while Sect. 6 consists of conclusion section.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_52

2 Mathematical Modeling

In the helicopter there are two degree-of-freedom. One is the rotation of the helicopter body with respect to the horizontal axis and second one is rotation around the vertical axis. The TRMS is nonlinear in characteristics.

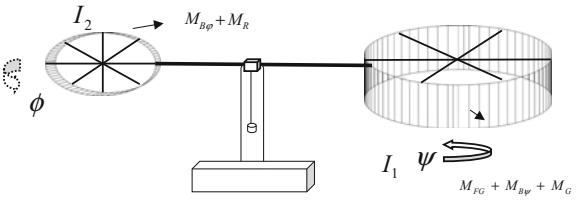


Fig. 1. TRMS phenomenological model

The nonlinear equations are obtained by an electromechanical diagram as shown in Fig. 1 [8]. Momentum along the pitch angle is derived as

$$I_1 \cdot \dot{\psi} = M_1 - M_{FG} - M_{B\psi} - M_G \tag{1}$$

where M_1 is as given as follows:

$$M_1 = a_1 \cdot \tau_1^2 + b_1 \cdot \tau_1 \tag{2}$$

The gravitational torque equation is derived as follows:

$$M_{FG} = M_g \sin\psi \tag{3}$$

The frictional torque is as follows:

$$M_{B\psi} = B_{1\psi} \cdot \dot{\psi} + B_{2\psi} \cdot sign(\dot{\psi}) \tag{4}$$

Due to the effect of the carioles force gyroscopic torque is produced in TRMS.

$$M_G = K_{gy} \cdot M_1 \cdot \dot{\phi} \cdot \cos\psi \tag{5}$$

The motor momentum is as follows:

$$\tau_1 = \frac{K_1}{T_{11}(s) + T_{10}} \cdot u_1 \tag{6}$$

The net torque of tail rotor is as below.

$$I_2 \cdot \ddot{\phi} = M_2 - M_{B\phi} - M_R \tag{7}$$

where,

$$M_2 = a_2 \cdot \tau_2^2 + b_2 \cdot \tau_2 \tag{8}$$

Frictional torque is shown below.

$$M_{B\psi} = B_{1\phi} \cdot \dot{\psi} + B_{2\phi} \cdot sign(\dot{\phi}) \tag{9}$$

The cross-reaction momentum M_R (Table 1)

$$M_R = \frac{K_c \cdot (T_0 s + 1)}{(T_p s + 1)} \cdot \tau_1 \tag{10}$$

$$\tau_2 = \frac{K_2}{T_{21}(s) + T_{20}} \cdot u_2 \tag{11}$$

Table 1. Specification of TRMS

System parameter specification	Value
I_1-vertical rotor moment of inertia	68×10^{-2} kg m^2
I_2-horizontal rotor moment of inertia	2×10^{-2} kg m^2
a_1-static characteristic	0.0135
a_2-static characteristic	0.0924
b_1-static characteristic	0.02
b_2-static characteristic	0.09
m_g-gravity momentum	0.32 N m
$B_{1\psi}$-friction momentum function	6×10^{-2} N m s/rad
$B_{2\psi}$-friction momentum function	1×10^{-2} N m s/rad
$B_{1\varphi}$-friction momentum function	1×10^{-1} N m s/rad
$B_{2\varphi}$-friction momentum function	10^{-2} N m s/rad
K_{gy}-gyroscopic momentum	0.05 s/rad
K_1-motor 1 gain	1.1
K_2-motor 2 gain	0.8
T_{11}-motor 1 denominator	1.1
T_{10}-motor 1 denominator	1
T_{21}-motor 2 denominator	1
T_{20}-motor 2 denominator	1
T_p-cross-reaction momentum	2
T_0-cross reaction momentum	3.5
k_c-cross-reaction momentum	−0.8

Now, TRMS transfer function is obtained by linearizing around origin as an operating point as below.

$$G(s) = \begin{bmatrix} \dfrac{1.246}{s^3 + 0.9215s^2 + 4.77s + 3.918} & 0 \\ \dfrac{1.482s + 0.4234}{s^4 + 6.33s^3 + 7.07s^2 + 2.08s} & \dfrac{3.6}{s^3 + 6s^2 + 5s} \end{bmatrix} \tag{12}$$

From the transfer function, it is clear that heavy influence is present across $u_1 - y_1, u_1 - y_2$ and $u_2 - y_2$ but no influence is detected across u_2 and $y_1 \cdot G_{11}, G_{22}$ are main rotor and tail rotor transfer functions, respectively.

3 Design of Decoupler

Here, a decoupler is designed for TRMS based on generalized decoupling technique to rectify the coupling effect associated with the plant. In the generalized decoupling technique the design of decoupler for any square plant *G(s)* is based on the formula as described below.

$$G_D(s) = G_I(s) * G_R(s) \tag{13}$$

where $G_D(s)$ = decoupling matrix of *G(s)*, $G_I(s)$ = inverse matrix of *G(s)* and $G_R(s)$ = diagonal matrix of $G(s)$

By considering the Eqs. (14) and (15), the decoupling matrix $G_D(s)$ is obtained as below for the TRMS.

$$G_I(s) = \begin{bmatrix} 0.7358s^3 + 0.7338s^2 + 3.522s + 3.148 & 0 \\ \dfrac{-0.3305s^6 - 2.313s^5 - 5.212s^4 - 12.55s^3 - 16.39s^2 - 7.07s + 4.056e^{-14}}{s^3 + 5.909s^2 + 4.545s + 1.11e^{-14}} & 0.2778s^3 + 1.667s^2 + 1.389s \end{bmatrix} \tag{14}$$

$$G_R(s) = \begin{bmatrix} \dfrac{1.246}{s^3 + 0.9215s^2 + 4.77s + 3.92} & 0 \\ 0 & \dfrac{3.6}{s^3 + 6s^2 + 5s} \end{bmatrix} \tag{15}$$

$$G_D(s) = \begin{bmatrix} 1 & 0 \\ \dfrac{-0.4492s^6 - 3.143s^5 - 7.083s^4 - 17.06s^3 - 22.28s^2 - 9.60s + 5.51e^{-14}}{s^6 + 6.906s^5 + 15.22s^4 + 37.09s^3 + 47.03s^2 + 19.44s + 4.749e^{-14}} & 1 \end{bmatrix} \tag{16}$$

4 Design of H Infinity Controller

Figure 2 depicts the H_∞ control method. Here P(s) is the plant and K(s) is the controller gain to be evaluated. The variables y and z are control variables and interest variables respectively. Once gamma $\gamma > 0$ is assigned compensator gain K(s) is

calculated on the basis of H_∞ to stabilizes the closed-loop system of Fig. 2. The state equations of the system P(s) are given as below.

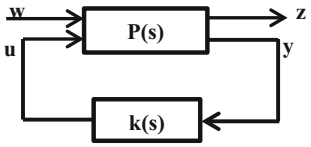


Fig. 2. Scheme of robust control

$$\begin{aligned} \dot{x} &= Ax + B_1 w + B_2 u \\ z &= C_1 x + D_{11} w + D_{12} u \\ y &= C_2 x + D_{21} w + D_{22} u \end{aligned} \tag{17}$$

The system to be controlled can be written in matrix scheme as

$$P(s) = \begin{bmatrix} A & B_1 & B_2 \\ C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{bmatrix} \tag{18}$$

LMI minimization problem is considered with linear objective defined by the following constraints.

$$\begin{bmatrix} N_{12} & 0 \\ 0 & I \end{bmatrix}^T \begin{bmatrix} AR + RA^T & RC_1^T & B_1 \\ C_1 R & -\gamma I & D_{11} \\ B_1^T & D_{11}^T & -\gamma I \end{bmatrix} \begin{bmatrix} N_{12} & 0 \\ 0 & I \end{bmatrix} < 0 \tag{19}$$

$$\begin{bmatrix} N_{21} & 0 \\ 0 & I \end{bmatrix}^T \begin{bmatrix} A^T s + sA & sB_1^T & C_1^T \\ B_1^T S & -\gamma I & D_{11}^T \\ C_1 & D_{11} & -\gamma I \end{bmatrix} \begin{bmatrix} N_{21} & 0 \\ 0 & I \end{bmatrix} < 0 \tag{20}$$

$$\begin{bmatrix} R & I \\ I & s \end{bmatrix} \geq 0 \tag{21}$$

where N_{12} and N_{21} are the bases of the null spaces, respectively, of (B_{T2}, D_{T12}) and of (C_2, D_{21}).

In case of main rotor, the transfer function s determined as below by the Eq. (14).

$$G_{11}(s) = \frac{1.246}{s^3 + 0.9215s^2 + 4.77s + 3.918} \tag{22}$$

Now converting the transfer function into state space form as described by Eq. (23), the state variables A, B, C, and D is obtained as below.

$$\dot{x} = Ax + Bu, y = Cx + Du \tag{23}$$

$$A = \begin{bmatrix} -0.9215 & -4.77 & -3.92 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, C = [0 \quad 0 \quad 1.246] \tag{24}$$

Now for obtaining the H infinity controller gain matrix the state equation is written in the form of Eq. (17) and system matrix is obtained by (18) in MATLAB is as below. The H infinity controller gain matrix is obtained and is converted to state space data form (a1, b1, c1, d1) as below.

$$a1 = \begin{bmatrix} -34.38 & 35.21 \\ 18.81 & -22.06 \end{bmatrix}, b1 = \begin{bmatrix} -193.46 \\ 113.44 \end{bmatrix}, c1 = [-20.39 \quad 35.71], d1 = [0.0034] \tag{25}$$

By adopting the same procedure the controller gain matrix of tail rotor of TRMS is determined.

5 Simulation Results

5.1 Time Response Analysis

Simulation results have been obtained from nominal model of TRMS with H infinity controller in MATLAB as displayed in Fig. 2. Figure 3a, b exhibits the main rotor system response and control voltage for a step input. Figure 4a, b exhibits the tail rotor system response and control voltage for a step input.

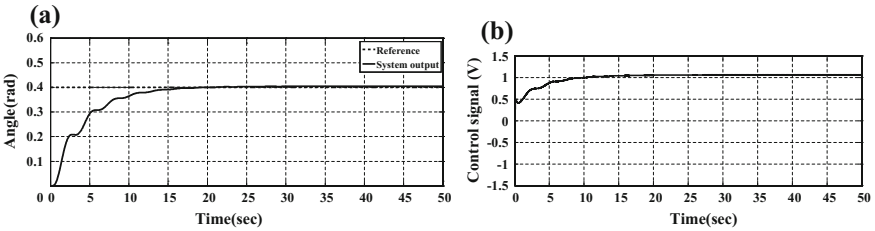


Fig. 3. **a** Step response for the Main rotor. **b** Main rotor control voltage

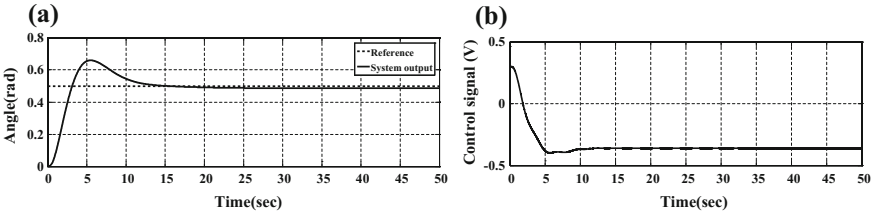


Fig. 4. **a** Step response for the Tail rotor. **b** Tail rotor control voltage

5.2 Robustness Analysis: Plots for Nominal Model with 50% Increment and Decrement in Moment of Inertia of Main and Tail Rotor

It is observed by simulation results that up to 50% increment and decrement in system parameters which is considered here as moment of inertia and the TRMS system remain stable. Figure 5a, b exhibits the main rotor system response and control voltage for step input with 50% decrement in moment of inertia. Figure 6a, b exhibits the tail rotor system response and control voltage for step input with 50% decrement in moment of inertia. Figure 7a, b exhibits the main rotor system response and control voltage for step input with 50% increment in moment of inertia. Figure 8a, b exhibits the tail rotor system response and control voltage for step input with 50% increment in moment of inertia.

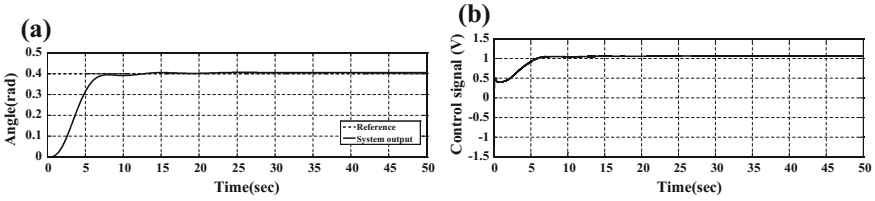


Fig. 5. **a** Step response for the Main rotor with 50% decrement in moment of inertia. **b** Main rotor control voltage with 50% decrement in moment of inertia

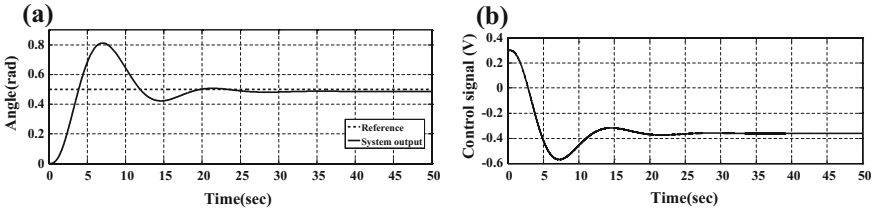


Fig. 6. **a** Step response for the Tail rotor with 50% decrement in moment of inertia. **b** Tail rotor control voltage with 50% decrement in moment of inertia

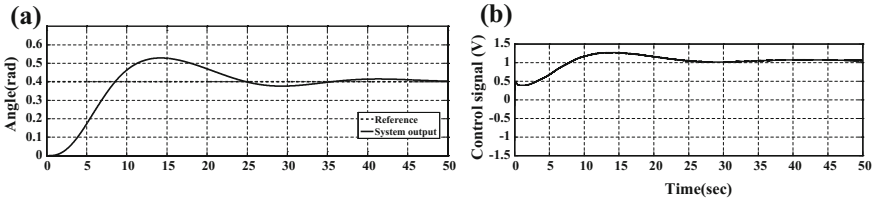


Fig. 7. **a** Step response for the Main rotor with 50% increment in moment of inertia. **b** Main rotor control voltage with 50% increment in moment of inertia

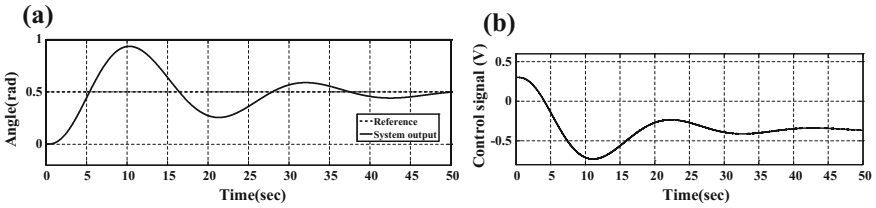


Fig. 8. **a** Step response for the Tail rotor with 50% increment in moment of inertia. **b** Tail rotor control voltage with 50% decrement in moment of inertia

6 Conclusion

In this work, H infinity controller is designed for TRMS system. A mathematical modeling of the system has been done. The decoupler is designed to eliminate the cross-coupling effect of TRMS. The designed controller is implemented in MATLAB/Simulink and implemented in simulation with nonlinear model of TRMS. The results show the satisfactory performance with respect to time domain analysis and parametric uncertainty.

Appendix

Minimization of objective function in MATLAB: Solver for linear objective minimization under LMI constraints

Iterations:	Best objective value so far
1.	2.670002
2.	0.019524
3.	0.019524
4.	9.005440e-03
5.	9.244373e-04
6.	9.244373e-04
7.	9.244373e-04
8.	4.591259e-04
9.	6.551457e-05
10.	6.551457e-05
11.	6.551457e-05
12.	1.942346e-05
13.	1.942346e-05
14.	1.942346e-05
15.	1.130349e-05
16.	1.130349e-05

(*continued*)

(*continued*)

Iterations:	Best objective value so far
17.	1.115295e-06
18.	1.306157e-07
19.	2.445126e-08
20.	2.445126e-08
21.	2.445126e-08
22.	1.645019e-08
23.	1.525763e-08
24.	1.525763e-08
25.	9.306554e-09
26.	8.736946e-09
27.	8.397046e-09
28.	8.397046e-09
29.	7.318684e-09
30.	7.178155e-09
31.	7.055896e-09
32.	6.979828e-09

References

1. TRMS 33–949S User Manual, Feedback instruments Ltd., East Sussex, UK
2. Ahmad SM, Chipperfield AJ, Tokhi MO (2003) Dynamic modeling and linear quadratic Gaussian control of a twin-rotor multi-input multi-output system. Proc I Mech E Part-I: J Syst Control Eng 217(I3):203–227
3. Sarkar C, Sengupta A (2016) LMI-based LSVF control of a class of nonlinear systems with parametric uncertainty: an application to an inverted pendulum system. Int J Control 89:2520–2532
4. Ojaghi P, Bigdeli N, Rahmani M (2016) An LMI approach to robust model predictive control of nonlinear systems with state-dependent uncertainties. J Process Control 47:1–10
5. Zeng Y, Xia C, Cao H (2009) A novel LMI based swing-up robust controller for a serial double inverted pendulum. In: Proceedings of the IEEE international conference on robotics and biomimetic, Guilin, China, 19–23 Dec 2009
6. Rao VS, George VI, Kamath S, Shreesha C (2015) Implementation of reliable H infinity observer-controller for TRMS with sensor and actuator failure. In: 10th Asian control conference (ASCC), Malaysia
7. Pratap B, Purwar S (2010) Neural network observer for twin rotor MIMO system: an LMI based approach. In: Proceedings of the international conference on modelling, identification and control, Okayama, Japan
8. Pradhan JK, Ghosh A (2013) Design and implementation of decoupled Compensation for a twin rotor multiple-input and multiple-output system. IET Control Theory Appl 7(2):282–289

Development of a Prototype Skin Color Monitor for Noninvasive Estimation of Blood Bilirubin

Sayantan Pal(✉) and Surajit Bagchi

Heritage Institute of Technology, Kolkata, India
sayantan.pall9@gmail.com, surajit.bagchi@heritage.edu

1 Introduction

Jaundice or icterus occurs due to the breakdown of red blood cells, named hemolysis and causes cite-dependent color changes that include sclera of eyes, palm, and underneath the tongue. A liver function test (LFT) is conducted to detect jaundice. It requires a blood sample, an invasive process, to get a numeric value of bilirubin concentration in the blood. However, noninvasive approach is always preferable relative to its invasive counterpart. Several researches have been done on noninvasive measurement bilirubin level in the blood.

Gosset, in 1960, has proposed an Icterometer which is like a transparent ruler with different marks of yellowish color of the skin. Each color corresponds to a bilirubin level [1]. Hamza [2] proposed a laser-based bilirubin meter to monitor neonatal jaundice [2]. J.W. Kornberg [3] introduced a transcutaneous bilirubin detection device based on the absorption of light by the skin at particular wavelength [3]. Buttitta and Steven [4] developed a device in which two lights were transmitted through the infant skin. The differential absorption of light was indicative of bilirubin level in the blood skin [4]. Jayasingh and Sankaranarayanan [5] introduced colorimetric detection method using RGB sensor to assess bilirubin level in blood [5]. Penhakar et al. [6] proposed a photometric method for blood bilirubin measurements [6]. Saini et al. [7] proposed a noninvasive technique for bilirubin detection that relied on the yellow discoloration of the skin. They used a smartphone-based application that was developed to predict the jaundice level [7]. The present work proposes an inexpensive, reliable, and user-friendly jaundice detecting opto-electronic device by measuring the blood bilirubin concentration noninvasively using.

2 Theoretical Backgrounds

The color of a surface what we see depends on proportionate mixing of three fundamental colors: red, green, and blue. Every color is formed by a certain proportionate combination of red, green, and blue colors (RGB). The RGB color model is an additive color model in which red, green, and blue are added together in various proportions to produce a broad array of colors. Therefore, every different color is having a different

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_53

RGB percentage. When all three colors are projected together upon an object's surface, it reflects a particular amount of RGB, which is allowed to fall on a light dependent resistor (LDR) [5, 6]. In dark condition, the resistance of the LDR is very high (10 MΩ). But as intensity of the falling light increases, the resistance of LDR drops dramatically. Keeping the same measuring conditions, if RGB source (TCLED) is subjected one after another upon skin surface, the reflected light will be received by the LDR and it will show different electrical output for different skin samples. As the yellow discoloration of the skin varies, the percentage of the color constituents in the reflected RGB change that causes a modulation in the voltage drop across the LDR. As different human being is having different skin tone, it is very difficult to fix the measurement reference In case of jaundice, yellow discoloration occurs only at some specific skin sites, where elastic tissues belong. The differential measurement of skin color is done with respect to the nearby cite on the skin surface where discoloration does not occur. To fix the measurement range of the device, we used the Von Luschan's chromatic scale [].

3 Materials and Methods

3.1 Components

Components used to develop the proposed device are: a TCLED (4 pin, common cathode), a LDR (200 k), 2 resistors (100 k, 1/4 W), 3 resistors (1 k, 1/4 W), 3 presets (10 k, 1/4 W), an Arduino Uno development board, and an LCD (16 × 2).

3.2 System Overview

Figure 1 shows the schematic diagram of the proposed device. TCLED (used as the RGB source), LDR (optical sensor), and a separator in between are fitted within a cylindrical chamber. To eliminate the surrounding optical noises, the **chamber** is wrapped with the black anchor tape. Figure 2 shows the circuit diagram.

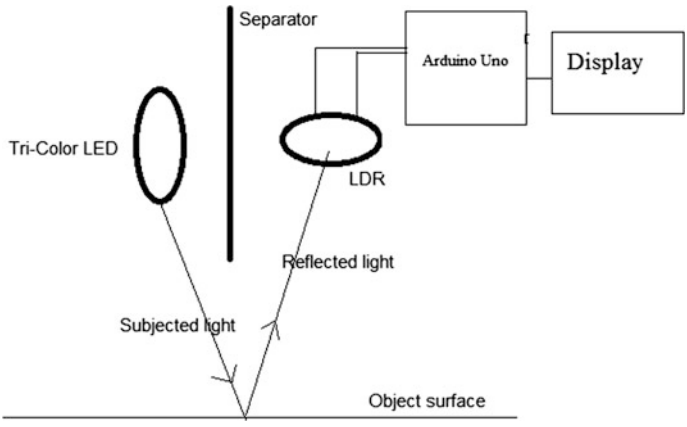


Fig. 1. System model

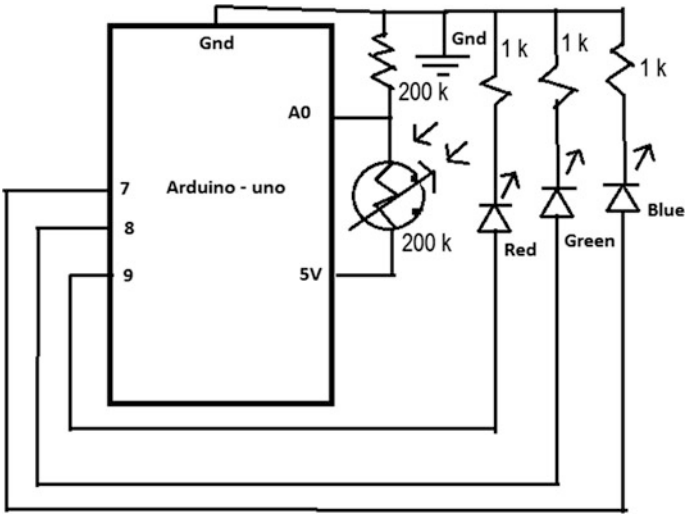


Fig. 2. Circuit diagram of the proposed device

The intensity of reflected light is calibrated in terms of the electrical output from the LDR. For different color samples, the individual electrical output for RGB is used to determine the output for combined RGB. Different color sample is having different combined output value which is fed into Arduino ADC. The 10-bit ADC (of the Arduino) maps the LDR output voltage (0–5 V) within a range of 0–1023. The smallest detectable change in voltage on the input pin is 5 V/1023 = 0.0049 V or 4.9 mV.

To determine the combined output value, Von Luschan's skin color chart has been scanned by the device and a multiple linear regression formula (Eq. 1) is derived.

$$\mathrm{Y_{com}} = 0.416843 * \mathrm{X_R} - 0.031851 * \mathrm{X_G} + 0.238606 * \mathrm{X_B} + 404.532014 \quad (1)$$

Y_{com} = Combined output value in term Arduino ADC, X_R = Output for red light over object's surface, X_G = Output for green light over object's surface, X_B = Output for blue light over object's surface.

The combined ADC output Y_{com} is almost similar to the actual ADC output for the combined light source, i.e., the device is performing linearly. The experimental set up of the proposed device is shown in Fig. 3.

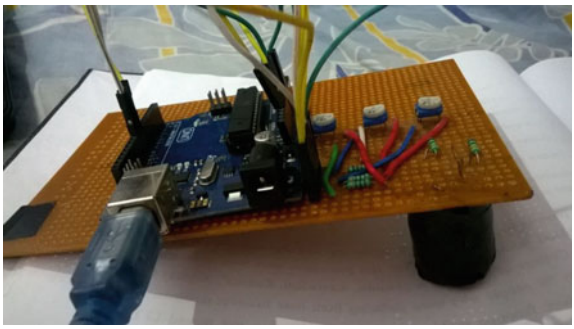


Fig. 3. The hardware setup

3.3 Program Algorithm

```
1     Turn on red LED
2     Delay for 10 seconds
3     Record LDR output
4     Turn on green LED
5     Delay for 10 seconds
6     Record LDR output
7     Turn on blue LED
8     Delay for 10 seconds
9     Record LDR output
10    Calculate the combined effect of three lights in-
      dividually using regression equation over the
      specific region (palm position of skin)
11    Turn off all lights and delay for 7 seconds and
      shift the measuring position
12    Repeat the steps from 1-9
13    Calculate the combined effect of three lights in-
      dividually using regression equation over the
      specific wrist position
14    Calculate the difference between the combined ef-
      fect of palm and wrist positions
15    Calculate the blood bilirubin amount using the
      difference value of the two measuring position
      and the liner equation: y= -19.33*x + 67.12
16    Display the output or the approximate blood bili-
      rubin amount
```

3.4 Calibrating the Device

The device is calibrated with black and white tiles for the adjustments of zero and span respectively. An ideal white colored tile reflects cent percent of red, cent percent of green, and cent percent of blue colors, whereas an ideal black color reflects zero percent of individual RGB colors. Presets are used to adjust the zero and span of the device, the variation of yellow color will reflect various proportions of RGB colors. The basic calibration stage has been done using Von Luschan's skin color chart.

3.5 Statistics of Subjects

Prior to check the performance of the proposed device, we took the permission of the ethical committee of Sagar Dutta College of Medicine and Hospital, Kamarhati, Kolkata, W.B., India. The study was conducted over the indoor patients, admitted to the department of medicine of the said hospital and were prescribed for the LFT tests. Before the tests, we described the working and objective of the project to the patients.

We collected noninvasive skin color-based data using the proposed device and blood bilirubin using LFT tests of 22 subjects.

4 Results

4.1 Selecting the Measuring Sites

The yellowish discoloration is easily detectable at the palm position of human skin. For safety reasons, we discarded the sensing from the eyes and tongue. To select the best sensing sites for differential measurements, we selected three positions on the palm and two on the wrist. We collected real-time skin color data from the selected positions and analyzed the data carefully. According to the analysis outcomes, the best sensing sites were fixed as shown in Fig. 4. The positions 1 and 2 refer to the measuring and reference sensing sites. We collected combined output data from position 1 and position 2.

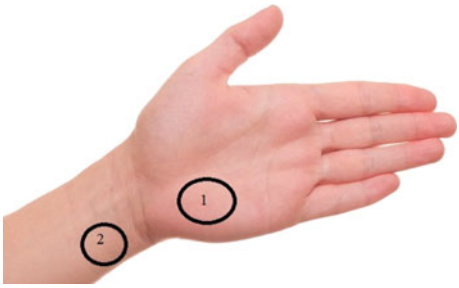


Fig. 4. Skin color sensing sites for measurements. Position 1: yellowish cite and position 2: normal skin cite

4.2 Performance Study

To select the best measuring position, we selected three positions from palm and two positions from wrist. We collected real-time skin color data from the selected positions following the proposed methodology. By trial and error method, the exact positions are fixed as shown in Fig. 4.

There is a measuring position on palm (position 1) and a reference position on wrist (position 2). We collected combined output data from position 1 and position 2 when the blood sample was taken for the test and recorded the blood bilirubin amount. By subtracting the outputs, we got the differential reflected light intensity of position 1 and position 2 in terms of voltage. The output is plotted in Fig. 5.

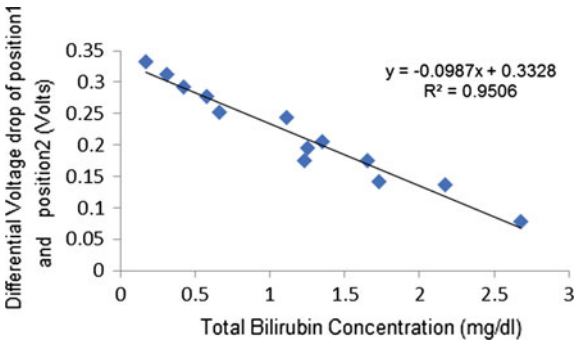


Fig. 5. Bilirubin concentration versus differential voltage drop (position 1 and position 2) graph

The response of the sensing system can be described by Eq. (2):

$$y = -0.0987 * x + 0.3328 \tag{2}$$

y Differential voltage drop between position 1 and position 2.
x Bilirubin concentration (mg/dl).

5 Discussion

Before we used the proposed device to access the color discoloration, the written consents were obtained from the healthy and jaundice abused patients. When Bilirubin rises above the threshold level (2 mg/dl), a color change occurs at some specific region. To measure the color change, we selected a nearby reference skin position. By plotting the differential voltage drop of measuring and reference position with respect to the real-time blood Bilirubin concentration, we found that as the Bilirubin concentration increases the differential voltage drop between measuring and reference position decreases almost linearly. The measurement is based on the RGB percentage of any color. The device is calibrated by following the RGB character of white and black color.

6 Conclusion and Future Scope

We have designed an inexpensive, easy-to-adopt noninvasive device to detect jaundice, which is based on the yellow discoloration of the bilirubin samples. The study was done taking 22 indoor patients who were prescribed for the LFT tests. The device obtained data and blood bilirubin concentration of the patients showed a good

correlation. In future, the device is to be tested with more clinical data. Moreover, an app for smartphone can be designed for early and easy detection of jaundice even at home ambience, it is required to gather.

References

1. Akman Ü, Arikan I, Bülgen HL, Kala S, Zek E (2001) Transcutaneous measurement of bilirubin by icterometer during phototherapy on a bilibed
2. Hamza M (1988) Laser transcutaneous bilirubin meter: a new device for bilirubin monitoring in neonatal jaundice. In: SPIE Proceedings
3. Kronberg JW (1993) Optical transcutaneous bilirubin. US Patent 5,259,382, 9 Nov 1993
4. Buttitta AD, Steven MS (1996) Librtyville, Illinois, Falk, Spring, Maryland, Non-invasive bilirubin monitor. European Patent Application. Ep 0 747 002 A1, 11 Dec 1996, pp 1–11
5. Jayasingh R, Sankaranarayanan K (2012) A novel colorimetric detection method for jaundice using RGB color sensor. Euro J Sci Res 91:563–568. [9] S. Kudavel
6. Penhaker M, Kasik V, Hrvolova B (2014) Advanced bilirubin measurement by a photometric method. Elektronika Ir Elektrotechnika, vol 19, pp 47–50
7. Saini N, Kumar A, Khera P (2016) Non-invasive bilirubin detection technique for jaundice detection using smart phones. IJCSIS 14(8). ISSN: 1947-5500
8. Assaad M, Yohannes I (2014) Design and characterization of automatic color sensor system. Int J Smart Sens Intell Syst 7(1). ISSN: 1178-5608

Noise Removing Filters and Its Implementation on FPGA

Ajay S. Mishra(✉) and Rutu Parekh

DA-IICT, Gandhinagar, India
ajaymishragecs@gmail.com, rutu_parekh@daiict.ac.in

1 Introduction

Image processing is very important field. It is used in various applications such as computer vision, medical imaging, object recognition, remote sensing, robotics, astronomy etc. [1]. Noise can be added into images during image acquisition or transmission stages. To remove such noises we need some real time filters. In this literature, we have implemented some real time filters on FPGA Spartan 3E. This requires a significant amount of computing resources [2]. Similarly these applications require a high speed. As FPGA has fast execution speed, large memory and reconfigure ability, it is generally used in real time applications [3]. Gaussian filter, Mean (Average) filter and Median filter have been implemented here. We have used both hardware and software implementation techniques to compare the results.

2 Filtering Algorithm

Algorithms required for these filters are explained in Sects. 2.1, 2.2, 2.3 and 2.4.

2.1 Convolution Operation

Convolution of the image I (x, y) and a mask g (x, y) of size w × l is given as below [2].

$$F(x,y) = \sum_{i=0}^{w-1} \sum_{j=0}^{l-1} G(i,j) I(x-i, y-j) \tag{1}$$

2.2 Gaussian Filter

The Gaussian filter is a 2D convolution operator which is used to smooth images and remove noise. It can be used to blur images. Gaussian filter is a filter whose impulse response is a Gaussian function. It can be considered as an ideal time domain filter just as the sinc function is the ideal frequency domain filter. Mathematically, Gaussian filter

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_54

modifies the input signal by convolution with a Gaussian function. Gaussian function in 2D can be given as below [4].

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{2}$$

where σ = Standard deviation.

We need to discretize the continuous Gaussian function to store it as discrete pixels. So, Gaussian functions are shown by different kernels. Kernels can have different sizes. Here we are using different kernel sizes which are 3 × 3, 5 × 5, 7 × 7, 9 × 9 and 11 × 11. Here 3 × 3 Gaussian kernel is as below:

$$\begin{bmatrix} 0.0113 & 0.0838 & 0.0113 \\ 0.0838 & 0.6193 & 0.0838 \\ 0.0113 & 0.0838 & 0.0113 \end{bmatrix} \tag{3}$$

These coefficients of the Gaussian kernel are floating point numbers. But hardware implementation of this kind of floating point Gaussian filter require large computational cost [1]. So we have converted this kernel into the fixed point numbers using the MATLAB command 'quantize'. After quantizing this kernel, we will get following mask which is containing only fixed point numbers. So now hardware implementation will be area efficient. The quantized kernel will be as below [1].

$$G(x,y) = \begin{bmatrix} 0 & 0.0625 & 0 \\ 0.0625 & 0.5625 & 0.0625 \\ 0 & 0.0625 & 0 \end{bmatrix} = \frac{1}{16} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 9 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{4}$$

2.3 Mean Filter

Mean (Average) filtering is a method of 'smoothing' images by reducing the amount of intensity variation between neighboring pixels. The average filter works by moving through the image pixel by pixel, replacing each value with the average value of neighboring pixels including it. The mean filter kernel is as below:

$$G(x,y) = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{5}$$

2.4 Median Filter

Median filter also does filtering like mean filter but it is better than Mean filter as it preserves the edges of the images. Median filter is very effective for salt and pepper noise. The median filter works by moving through the image pixel by pixel, replacing each value with the median value of neighboring pixels. The median value is calculated by arranging the pixels in increasing order and then selecting the center value. For this

3 × 3 sub-images are selected from left to right and then top to bottom and each and every central pixel value of those sub-images are replaced by their median value. Example is shown as below Fig. 1.

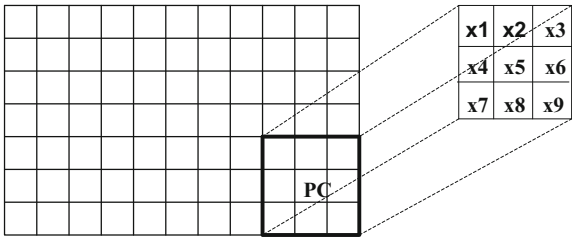


Fig. 1. Median filter working

3 Implementation

Filter implementation techniques are explained in following sections.

3.1 Implementation Steps for Gaussian and Average Filters on Xilinx ISE

Implementation of Gaussian and Mean filter is same only the difference is in between their kernel coefficients as shown in Eqs. 4 and 5. Filter will require to convolve the noisy image with the Gaussian or Mean mask. So we need to implement algorithm for convolution which is as shown in Fig. 2. Kernel is generated using MATLAB function 'fspecial' for Gaussian filter and then MATLAB quantization command 'quantize' is used to quantize the kernel as shown in Eq. 4 [1]. For convolution process each of the sub-image pixels are to be multiplied with the Gaussian kernel or Mean kernel (depending upon which filter we need to implement) values and then result of MAC (multiply and accumulate) is divided by16 for Gaussian filter or by 9 for mean filter. Here we have used standard multiplication '×' in verilog language.

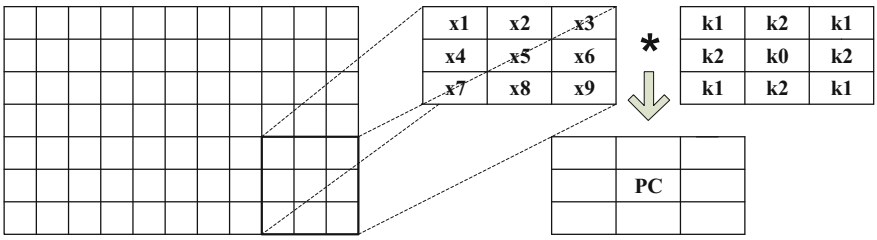


Fig. 2. Gaussian and mean filter working

A gray scale image is represented by a matrix of pixels with values ranging from 0 to 255. Each and every pixel is shown by 8 bits. In this design we are using a 128 × 128 image for filtering. For storing this image in Block RAM (BRAM) of FPGA, it requires the image to be converted into a vector of 16384 elements. The input to BRAM should be of the format '.coe' file. Using MATLAB different kind of noises can be added to image. Next step is to convert image to text containing 16384 elements using MATLAB. Then '.txt' file is converted to '.coe' file manually. Then '.coe' file is loaded into BRAM memory using Xilinx IP core generator.

Here an FSM has been implemented to fetch sub-image pixels from BRAM. This FSM will be in such a way so that mask (3 × 3, 5 × 5, 7 × 7 etc.) will be moved from left to right and then top to bottom. In this way the central pixel will be updated. Similarly this operation will be continued until all the pixels are updated. After simulation using Xilinx Isim, a text file will be generated containing the coefficients of filtered image.

In the above Fig. 2, PC = × 1 × k1 + x2 × k2 + x3 × k1 + x4 × k2 + x5 × k0 + x6 × k2 + x7 × k1 + x8 × k2 + x9 × k1. Here for Gaussian filter, k0 = 9/16, k1 = 0, k2 = 1/16 and for mean filter, k0 = k1 = k2 = 1/9. The text file can be converted to image by MATLAB coding. The resultant image will be the filtered image. Here each 8 bit pixels can be taken out from the UART.

3.2 Median Filter Implementation

To implement median filter, we need another algorithm which can find the median value of the sub-image [5]. Here the circuit to get median from any sub-image is shown in Fig. 3. Here Fig. 3a shows that nine pixels of a sub-image are given as input (x1, x2, x3, x4, x5, x6, x7, x8 and x9). In this circuit each node is working as a comparator which is shown in Fig. 3b. From this algorithm we will get the output pixel which will be median of these nine pixels.

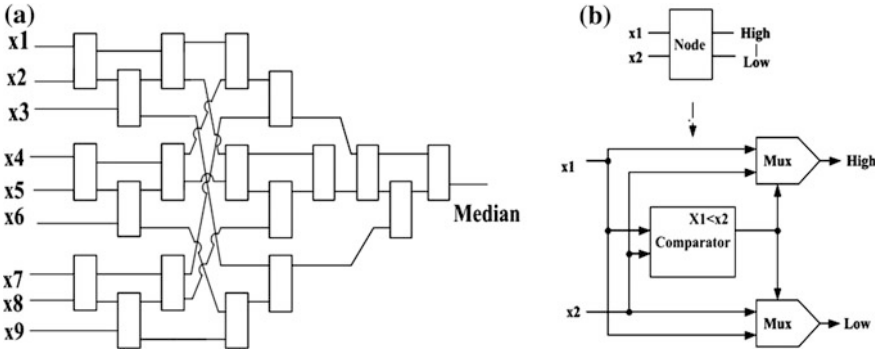


Fig. 3. **a** Median filter algorithm. **b** Basic node's working

Here we have implemented an FSM to fetch sub-image pixels from BRAM. This FSM will be in such a way so that a mask of size 3 × 3 will be moved from left to

right and then from top to bottom. In such way central pixel will be updated which will be the median pixel. Similarly this operation will be continued until all the pixels are updated. After simulation using Xilinx Isim, a text file will be generated containing the coefficients of filtered image. The text file can be converted to image by MATLAB coding. The resultant image will be the median filtered image. Here each 8 bit pixels can be taken out from the UART.

4 Results

Results for the implementation are explained in these sections. We have explained resource utilization and calculated MSE and PSNR values of images.

4.1 Resource Utilization

Here filters are implemented using Xilinx ISE 14.5, on XC3S500E of Spartan 3E FPGA. Following tables show the resource utilization by these filters. Here Median filter utilizes more resources in compared to the Gaussian and Mean filters. It concludes that with respect to resource utilization Gaussian and Mean filters are good. Gaussian and Mean both filters are implemented for 3 × 3, 5 × 5 and 7 × 7 kernel sizes and we are getting same resource utilization for these kernel sizes with our method.

Here we have implemented Gaussian filter for different values of σ (0.5, 1, 2, 4) as shown in Table 1. As σ increases, images are more smoothen. In our method we have used standard multiplication '×' in verilog language Table 2.

Table 1. FPGA resource utilization for Gaussian filter for different values of σ

Resource utilization	Gaussian σ = 0.5	Gaussian σ = 1, 2, 4	Mean	Median	Total available
Slices	60 (1%)	83 (1%)	60 (1%)	539 (11%)	4656
LUTs	116 (1%)	160 (1%)	116 (1%)	1000 (10%)	9312
IOBs	2 (0%)	2 (0%)	2 (0%)	170 (73%)	232
BRAMs	8 (40%)	15 (75%)	8 (40%)	8 (40%)	20

Table 2. Comparison of resource utilization of our method with Gaussian filter

Resource utilization	Our method	[6]	[2]
Slices	60	228	127
LUTs	116	2088	176
IOBs	2	30	–
DSP48Es (Multiplier)	0	6	9
Image size	128 × 128	137 × 109	8 × 8

4.2 Resultant Filtered Images

Figure 4 shows the resultant images (128 × 128) after filtering for different kernel sizes. To determine the quality of filtered images, we calculate the MSE and PSNR values. Equations 6 and 7 show the formulae to find MSE and PSNR values for image size w × 1 [7] (Fig. 5).

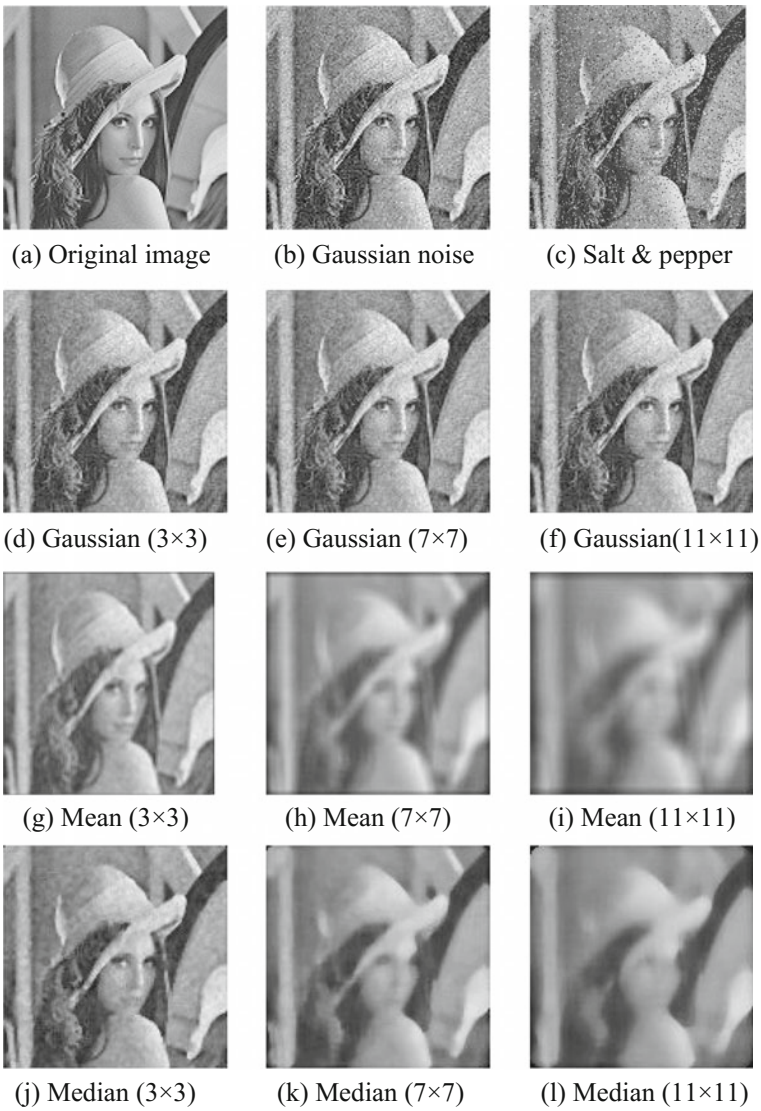


Fig. 4. MATLAB filtered images (Here images **a–c** are original and noisy images and **d–l** are filtered images with different filters and kernel sizes)

(a) Gaussian(σ = 0.5) (b) Gaussian(σ = 1) (c) Gaussian(σ = 2) (d) Gaussian(σ = 4)

(e) Mean filtered (f) Median filtered (g) Gaussian (salt& pepper) (h) Median (salt& pepper)

Fig. 5. Xilinx resultant images (Here images **a–h** are filtered images with different filters and σ values)

$$MSE = \frac{1}{w \times l} \sum_{i=0}^{w-1} \sum_{j=0}^{l-1} \|I_1(i,j) - I_2(i,j)\|^2 \tag{6}$$

$$PSNR = 10 \log_{10} \frac{256 \times 256}{MSE} \tag{7}$$

We have analysed different type of images with different noises. Median filter is good for the Salt and pepper noise. Comparing images having different complexity, we can conclude that as image complexity changes MSE and PSNR also changes. Tables 3, 4 and 5 shows the MSE and PSNR for these three filters. It can be seen that as σ increases, MSE increases and PSNR decreases (Table 4). Here difference between

Table 3. MSE and PSNR for all three filters with different size of kernels

Kernel size	Filter	MSE	PSNR (in dB)
3 × 3	Gaussian	46.59	31.4820006
7 × 7		58.38	30.5020808
11 × 11		79.42	29.1654914
3 × 3	Mean	27.53	33.7671274
7 × 7		48.85	31.2765085
11 × 11		68.76	29.7915171
3 × 3	Median	56.32	30.6583229
7 × 7		63.86	30.1125070
11 × 11		78.14	29.2359186

Xilinx and MATLAB results (Tables 4 and 5 compared) are due to the Quantization error as shown in Eq. 4. Here mean filter is giving least MSE (Table 5) but image quality is best for Median filter as it preserves sharp edges.

Table 4. MSE and PSNR for Gaussian filter for kernel size 3 × 3 with different σ values

σ	MSE	PSNR (in dB)
0.5	46.59	31.4820006
1	208.04	24.9832612
2	250.04	24.1846210
4	253.71	24.1213489

Table 5. MSE and PSNR comparison between three filters for Xilinx results

Filter type	σ	MSE	PSNR (in dB)
Gaussian	0.5	55.32	30.7356512
	1	224.73	24.6482540
	2	230.73	24.5319017
	4	235.40	24.4467816
Mean	–	19.17	35.3381811
Median	–	46.77	31.4655205

5 Conclusion

We have successfully implemented all three filters on FPGA XC3S500E Spartan 3E board and verified the results with MATLAB simulations. Here median filter is best among these filters in removing Salt and pepper noise but it is utilizing more resources in compared to other two filters. These filters have been analyzed with different kernel sizes, with different noises and also with different image complexity and compared the resource utilization and image quality too. Here Peak Signal to Noise Ratio (PSNR) from Xilinx is about +0.7463494 to −0.3254327 dB different than the MATLAB result because of quantization error.

References

1. Cabello F, León J, Iano Y, Arthur R (2015) Implementation of a fixed-point 2D gaussian filter for image processing based on FPGA. In: 2015 Signal processing: algorithms, architectures, arrangements, and applications (SPA), Poznan, pp 28–33
2. kabbai L, Sghaier A, Douik A, Machhout M (2016) FPGA implementation of filtered image using 2D gaussian filter. Int J Adv Comput Sci Appl 7(7)

3. Jain T, Bansod P, Kushwah CBS, Mewara M (2010) Reconfigurable hardware for median filtering for image processing applications. In: 2010 3rd International Conference on Emerging Trends in Engineering and Technology, Goa, pp 172–175
4. Rafael REW, Gooonzalez C (2009) Digital image processing using MATLAB, vol 24, no 11
5. Vega-Rodríguez MA, Sánchez-Pérez JM, Gómez-Pulido JA (2002) An FPGA-implementation for median filter meeting the real-time requirements of automated visual inspection. In: Proceedings of the 10th mediterranean conference on control and automation—MED2002 Lisbon, Portugal, 9–12 July 2002
6. Talbi F, Alim F, Seddiki S, Mezzah I, Hachemi B (2015) Separable convolution gaussian smoothing filters on a xilinx FPGA platform. In: Fifth international conference on the rinnovative computing technology (INTECH 2015), Galcia, pp 112–117
7. Kumar RN, Kumar UJ (2015) A spatial mean and median filter for noise removal in digital images. Int J Adv Res Electr Electron Instrum Eng 4(1)

A Novel Unsupervised Framework for Retinal Vasculature Segmentation

Samiksha Pachade(✉), Prasanna Porwal, and Manesh Kokare

Center of Excellence in Signal and Image Processing, Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, India
samikshapachade18@gmail.com
{porwalprasanna,mbkokare}@sggs.ac.in

1 Introduction

Retinal vascular structure has been extensively utilized in the analysis of various cardiovascular and ophthalmologic diseases. Change in retinal vasculature is an significant sign for diseases, such as diabetic retinopathy, arteriosclerosis, hypertension, and stroke [7]. The automatic detection and analysis of retinal vasculature structure is extremely important in order to avoid the visual impairment. Segmented retinal blood vessels and identification of branch points can be used for multimodal and temporal retinal image registration, optic-disc detection, fovea location, retinal image mosaic synthesis, and biometric identification [4].

Retinal blood vessel segmentation can be done both manually and automatically. Figure 1 shows manually detected vasculature of corresponding color fundus image. However, manual vessel segmentation is time-consuming, skill-demanding, and expensive. Automatic vessel segmentation has not attained wide acceptance due to several challenges: (1) irregular shaped pathologies such as exudates, hemorrhages, and drusens, (2) vessel diameter and length variability, (3) the central light reflex on large vessels, (4) low blood vessel contrast corresponding to the background, (5) vessel bifurcation and crossing.

The major contribution of this article is that we have proposed a novel unsupervised method for blood vessel segmentation by exploiting adaptive filter and mathematical morphology based operations. Initially, we employed a set of preprocessing steps to remove the noisy circular border and to enhance the vessel contrast. Further, we exploit large median filter for estimating the background. Finally vessels are segmented by iterative thresholding approach. This method was evaluated on extensive datasets and encouraging results are achieved.

The rest of the paper is organized as follows: Sect. 2 summarizes the earlier state-of-the-art approaches of vessel segmentation, Sect. 3 presents the proposed technique, Sect. 4 illustrates the performance of presented method and its comparison with other existing methods, Sect. 5 gives the conclusion.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_55

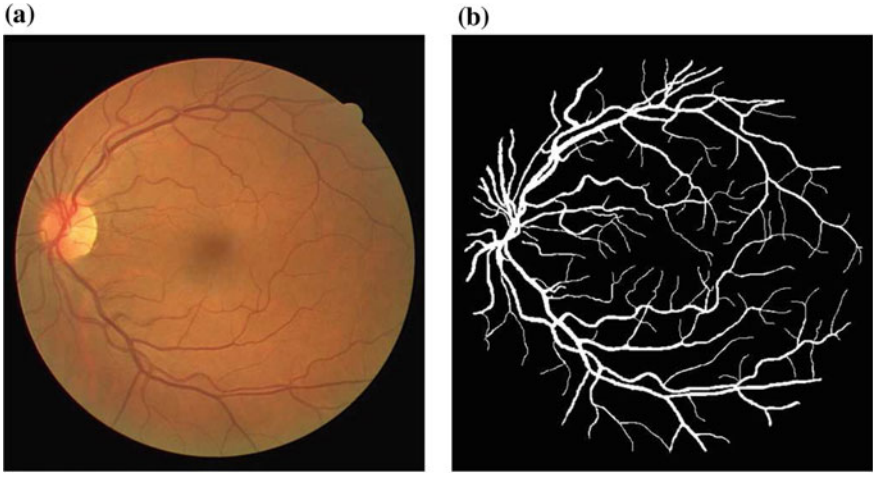


Fig. 1. a Color fundus image, **b** manually drawn ground truth image

2 Related Work

Several retinal vessel detection methods have been proposed earlier and are present in the literature. This substantial amount of work can be broadly divided based on two approaches: (a) *Supervised* and (b) *Unsupervised*.

Supervised segmentation method classifies each pixel of image as vessel and non-vessel using the ground truth data. Staal et al. [17] used the features of convex sets computed using ridge profile and feature selection is done in sequential forward manner and finally classification is done using K-Nearest Neighbor (KNN). Soares et al. [15] presented a pixel-based classification approach using Gaussian mixture model Bayesian classifier in combination with Gabor wavelet transform. In the algorithm of Ricci and Perfetti [14], line strength features were deployed using line operator and SVM was used as a classifier. Lupascu et al. [10] for each pixel generated a 41-dimensional feature vector and these features were used in an AdaBoost classifier. In Marín et al. [11], moment and intensity invariant features were extracted using mathematical morphology and neural network classifier is used for classification. Fraz et al. [4] employed classification based on a feature vector and ensemble of bagged decision tree. Annunziata et al. [3] presented a method for curvilinear segmentation using manually crafted features in combination with learned context filters.

Unsupervised segmentation techniques are based on matched filtering, model-based algorithm, multi-scale analysis, vessel tracking, and morphological processing. Hoover et al. [7] used matched filtering, it is based on 2-D Gaussian kernel rotated through 12 different orientations. A piecewise thresholding technique is used to obtain segmented vessels. Jiang and Mojon [8] proposed a multi-thresholding scheme in combination with a classification procedure to prune non-vessels and keep vessels like structures. Mendonca and Campilho [13] used

mathematical morphological coupling for centerline detection in combination with matched filter. Region growing process for vessel detection was exploited by Martinez-Perez et al. [12] based on information of edge and maximum principal curvature. Al-Diri et al. [1] used "Ribbon of Twins" for identification and width measurement of the retinal vessels. Lam et al. [9] proposed three multiconcavity measures to handle both bright and dark lesions in healthy and unhealthy retina. Yu et al. [18] used hessian-based method for retinal vessel detection. Annunziata et al. [2] used an inpainting filter which inpaints exudates and multi-scale Hessian approach for retinal vessel enhancement.

3 Proposed Method

3.1 Preprocessing

Preprocessing is done as a primary step to remove unwanted noise. Green plane g is chosen over red r and blue b planes for detection of vessels as it shows better contrast between the retinal vessels and the background. Green plane (see Fig. 2b) of the color fundus image I (see Fig. 2a) is defined as $g = (g(i))_{i \in \mathbb{R}^2}$. We use g to determine the field of view (FOV) as shown in Fig. 2c. FOV is obtained by mean estimation using the g as follows:

$$FOV = \begin{cases} 1, & g(i) \geq g(mean)/2 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

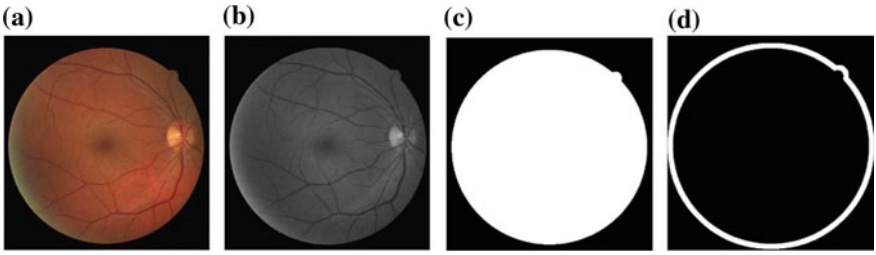


Fig. 2. a Color retinal fundus image, **b** green plane g, **c** FOV, **d** circular border

Finally, the circular border region is estimated to remove the bright border region. The circular border I_c is obtained by subtracting the eroded (ε) FOV from dilated (δ) FOV as given in Eq. 2. Figure 2d shows the circular border obtained.

$$I_c = [\varepsilon^{Es}(FOV)] - [\delta^{Es}(FOV)] \tag{2}$$

We call Es as structuring element E of size s.

3.2 Image Enhancement

Certain challenges in the identification of vessels are due to the uneven illumination which occurs while capturing the fundus image. Image enhancement increases the contrast between the vessel and background which creates ease in vessel detection. Image enhancement is done using an adaptive manifold filter [5]. In this filter, weighted average of the input color pixels is done and a new set of pixel colors is produced. The weights of the filter are given by kernel ϕ. Filtering the image with ϕ gives a new image y_i

$$y_i = \frac{\sum_{n_j \in s} \phi(\hat{n_i} - \hat{n_j}) g_i}{\sum_{n_j \in s} \phi(\hat{n_i} - \hat{n_j})} \tag{3}$$

Functional selection of ϕ is an axis-aligned Gaussian, as given below:

$$\phi \sum (\hat{n_i} - \hat{n_j}) = exp(-\frac{1}{2}(\hat{n_i} - \hat{n_j})^T \sum^{-1} (\hat{n_i} - \hat{n_j}), \tag{4}$$

where $\sum$ is a diagonal covariance matrix of size $d \times d$. The $\hat{n_i}$ and $\hat{n_j}$ are neighborhoods around the corresponding pixel. Figure 3b shows the image after enhancement.

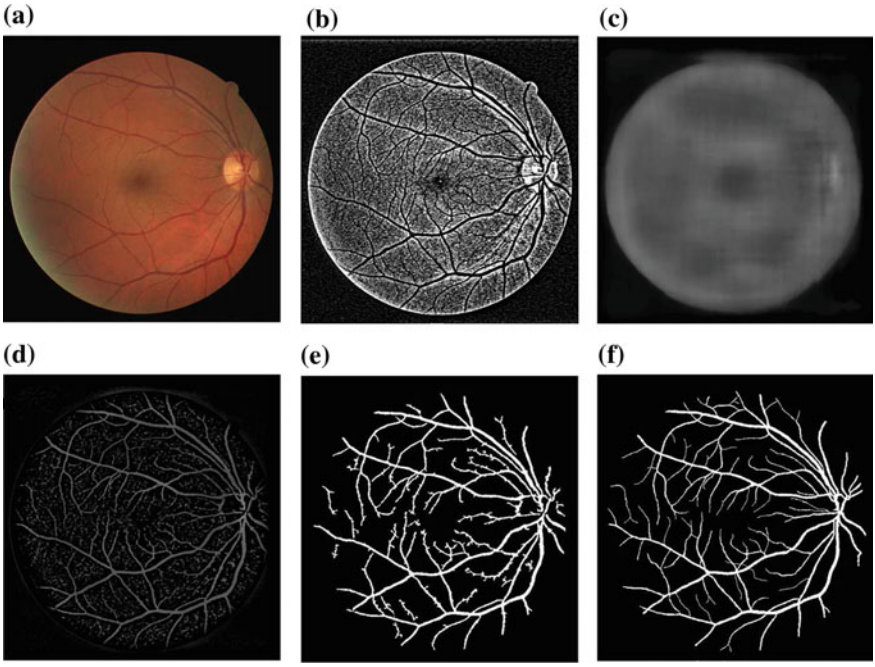


Fig. 3. **a** Color fundus image, **b** enhanced image, **c** background estimated, **d** retinal vessel with noise, **e** difference image, **f** ground truth image

3.3 Vessel Segmentation

Contrast limiting adaptive histogram equalization (CLAHE) is used to further enhance the contrast of the image obtained after previous step. Image y_i is splitted into blocks of size 8×8 and these small blocks are enhanced with 128 number of bins. Let the output image be I_{adpt}. Perform 2-D median filtering on image I_{adpt} and obtained image is I_{med}. Morphological opening I_{open} of I_{med} by disc structuring element E of size $s = 2$ is given mathematically by Eq. 5

$$I_{open}^{Es}(I_{med}) = \delta^{Es}[\varepsilon^{Es}(I_{med})] \tag{5}$$

where, δ and ε are morphological dilation and erosion operation respectively. Furthermore, a large 2-D median filter of size 69×69 is applied on I_{open}. Background is estimated and denoted by I_{back}. Difference image D is obtained by subtracting estimated background I_{back} from I_{med} given in Eq. 6. Figure 3d shows the difference image D.

$$D(x, y) = I_{med}(x, y) - I_{back}(x, y) \tag{6}$$

Thresholding of the difference image D is done by iterative approach. The following algorithm is utilized to estimate the threshold *level*:

Using the value of *level*, the image D is binarized using the threshold value as $level - 0.008$ for DRIVE and $level + 0.108$ for STARE database. Further from the binarized image, the bright border part is eliminated using circular border I_c and the connected components which are less than 100 pixels are removed.

Algorithm 1 Thresholding

```
function THRESHOLD(D)
    i ← iteration
    Let i = 1
    [C, N] = imhist(D)
    M ← mean intensity of D using histogram
    loop:
    MAT ← mean above threshold M
    MBT ← mean below threshold M
    i = i + 1
    T = (MAT + MBT)/2
    if T(i) ≠ T(i − 1) then
        goto loop
    end if
    level = (T − 1)/(N(end) − 1)
end function
```

Table 1. Comparative table of state-of-the-art methods with proposed method

Method		DRIVE			STARE		
		SE	SP	ACC	SE	SP	ACC
Supervised	Staal et al. [17]	–	–	0.9441	–	–	0.9516
	Soares et al. [15]	0.7332	0.9782	0.9461	0.7207	0.9747	0.9480
	Ricci and Perfetti [14]	–	–	0.9595	–	–	0.9646
	Lupascu et al. [10]	0.7200	–	0.9597	–	–	–
	Marín et al.[11]	0.7067	0.9801	0.9452	0.6944	0.9819	0.9526
	Fraz et al. [4]	0.7406	0.9807	0.9480	0.7548	0.9763	0.9534
Unsupervised	Hoover et al. [7]	–	–	–	0.6747	0.9565	0.9275
	Jiang and Mojon [8]	–	–	0.9212	–	–	0.9009
	Mendonca and Campilho [13]	0.7344	0.9764	0.9452	0.6996	0.9730	0.9440
	Martinez-Perez et al. [12]	0.7246	0.9655	0.9344	0.7506	0.9569	0.9410
	Al-Diri et al. [1]	0.7282	0.9551	–	0.7521	0.9681	–
	Lam et al. [9]	–	–	0.9472	–	–	0.9567
	Yu et al. [18]	0.7233	0.9745	0.9426	0.7112	0.9709	0.9463
	Annuizate et al. [2]	–	–	–	0.7128	0.9836	0.9562
	Proposed method	**0.7565**	**0.9793**	**0.9513**	**0.7706**	**0.9705**	**0.9553**

4 Experimental Evaluation

4.1 Performance Evaluation

The outcome of vessel segmentation after thresholding is binary in which the pixel is either classified as a vessel or non-vessel. The performance evaluation of vessel segmentation and quantification is done by using four events, true positive (TP) when pixel is a vessel and is classified as a vessel, otherwise they are false negative (FN), true negative (TN) when pixel belongs to background and is classified as non-vessel otherwise they are false positive (FP). *Sensitivity* and *Specificity* are two statistical measures used for evaluation, they are normalized measure of well-classified vessels and non-vessel respectively. The *accuracy* (ACC) is related to the segmentation confidence. These measures are defined in Eq. 7:

$$SE = \frac{TP}{TP + FN}, \quad SP = \frac{TN}{TN + FP}, \quad ACC = \frac{TP + TN}{TP + FN + TN + FP} \tag{7}$$

4.2 Vessel Segmentation Result and Analysis

The presented method is evaluated on retinal images of publicly available DRIVE [16] and STARE [6] database. The DRIVE database have two groups namely: training and testing. Both groups consist of 20 fundus images of size 565×584 pixels. The first observers gold standard hand-labeled images were used for the evaluation. In STARE database, fundus images are of size 605×700. Ground truths of 20 images for vessel segmentation are given and they are used for performance evaluation. Table 1 shows the comparative result of presented method with the other state-of-the-art methods on DRIVE and STARE database, respectively.

5 Conclusion

An improved unsupervised approach for the segmentation of retinal blood vessels using image enhancement and median filtering has been described. The proposed technique followed by simple thresholding gives results comparable to a state-of-the-art method on DRIVE (SE = 0.7565, SP = 0.9793, ACC = 0.9513) and STARE (SE = 0.7706, SP = 0.9705, ACC = 0.9553) database. Our method works better than the several methods listed and overcomes some of the listed challenges of vessel segmentation. The experimental result shows good performance on normal as well as pathological images.

References

1. Al-Diri B, Hunter A, Steel D (2009) An active contour model for segmenting and measuring retinal vessels. IEEE Trans Med Imaging 28(9):1488–1497

2. Annunziata R, Garzelli A, Ballerini L, Mecocci A, Trucco E (2016) Leveraging multiscale hessian-based enhancement with a novel exudate inpainting technique for retinal vessel segmentation. IEEE J Biomed Health Inform 20(4):1129–1138
3. Annunziata R, Kheirkhah A, Hamrah P, Trucco E (2015) Boosting hand-crafted features for curvilinear structure segmentation by learning context filters. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 596–603
4. Fraz MM, Remagnino P, Hoppe A, Uyyanonvara B, Rudnicka AR, Owen CG, Barman SA (2012) An ensemble classification-based approach applied to retinal blood vessel segmentation. IEEE Trans Biomed Eng 59(9):2538–2548
5. Gastal ES, Oliveira MM (2012) Adaptive manifolds for real-time high-dimensional filtering. ACM Trans Graph (TOG) 31(4):33
6. Hoover A (1975) Stare database. http://www.ces.clemson.edu/~ahoover/stare
7. Hoover A, Kouznetsova V, Goldbaum M (2000) Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans Med Imaging 19(3):203–210
8. Jiang X, Mojon D (2003) Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images. IEEE Trans Pattern Anal Mach Intell 25(1):131–137
9. Lam BS, Gao Y, Liew AWC (2010) General retinal vessel segmentation using regularization-based multiconcavity modeling. IEEE Trans Med Imaging 29(7):1369–1381
10. Lupascu CA, Tegolo D, Trucco E (2010) FABC: retinal vessel segmentation using adaboost. IEEE Trans Inform Technol Biomed 14(5):1267–1274
11. Marín D, Aquino A, Gegúndez-Arias ME, Bravo JM (2011) A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. IEEE Trans Med Imaging 30(1):146–158
12. Martinez-Perez ME, Hughes AD, Thom SA, Bharath AA, Parker KH (2007) Segmentation of blood vessels from red-free and fluorescein retinal images. Med Image Anal 11(1):47–61
13. Mendonca AM, Campilho A (2006) Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. IEEE Trans Med Imaging 25(9):1200–1213
14. Ricci E, Perfetti R (2007) Retinal blood vessel segmentation using line operators and support vector classification. IEEE Trans Med Imaging 26(10):1357–1365
15. Soares JV, Leandro JJ, Cesar RM, Jelinek HF, Cree MJ (2006) Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. IEEE Trans Med Imaging 25(9):1214–1222
16. Staal J, Abramoff M, Niemeijer M, Viergever M, van Ginneken B (2004) Ridge based vessel segmentation in color images of the retina. IEEE Trans Med Imaging 23(4):501–509
17. Staal J, Abràmoff MD, Niemeijer M, Viergever MA, Van Ginneken B (2004) Ridge-based vessel segmentation in color images of the retina. IEEE Trans Med Imaging 23(4):501–509
18. Yu H, Barriga S, Agurto C, Zamora G, Bauman W, Soliz P (2012) Fast vessel segmentation in retinal images using multiscale enhancement and second-order local entropy. In: SPIE medical imaging. International society for optics and photonics, vol 8315, p 83151B

Stochastic Resonance in Bagley-Torvik Equation

Aman K. Singh(✉), Vivek K. Verma, and R. D. S. Yadava

Sensors & Signal Processing Laboratory, Department of Physics,
Institute of Science, Banaras Hindu University,
Varanasi 221005, Uttar Pradesh, India
{aman.strgtr,vv0l06,ardius}@gmail.com

1 Introduction

Stochastic resonance (SR) is a nonlinear phenomenon in which the internal noise present in a nonlinear system interacts cooperatively with an external periodic stimulus to generate enhanced response [1]. In a variant, the noise can be fed externally also to a periodically stimulated nonlinear system [2]. Benzi et al. first suggested stochastic resonance in modelling the nearly periodic occurrence of earth's ice age due to climatic perturbations [3]. Since then it has found extensive applications in physics, chemistry and biology. The occurrence of SR has been experimentally demonstrated in several systems like the predator detection by crayfish, weak magnetic field detection by superconducting quantum interference device (SQUID) [4], ring laser [5], signal amplification and noise self-filtering [6], microelectromechanical sensor (MEMS) [7]. The major studies involving SR in noisy nonlinear electrical oscillators have been summarized in [8]. The existence of nonlinearity is considered to be essential for the occurrence of SR, even though some studies have reported the occurrence of SR in linear systems also in the presence of colored noise [9].

Recently SR has been investigated in fractional order systems [10]. The fractional order descriptions of certain nonlinear dynamical systems provide an effective method for modeling and analyzing their characteristics accurately [11, 12]. Bagley-Torvik equation with fractional order damping has been shown to represent the dynamics of a long thin plate in Newtonian fluid [11]. The solution of this equation has been analytically found in [11, 13]. The damping term in Bagley-Torvik equation is expressed as fractional order derivative. The fractional order systems have been found to model some other systems also having memory and viscoelastic damping [14, 15]. In the present work we consider random frequency fluctuations and analyze occurrence of SR in this system. The analyses presented here could be of direct relevance to the modeling of micro/nano cantilever devices [16].

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_56

2 Bagley-Torvik Equation with Frequency Noise

The Bagley-Torvik equation in dimensionless form is written as

$$\ddot{u} + \gamma \frac{d^{3/2}}{dt^{3/2}} u + \omega^2 u = a \cos \Omega t \tag{1}$$

This describes a harmonically driven damped harmonic oscillator system with fractional order damping. In this, γ denotes the damping coefficient, ω the fundamental frequency, a and Ω being the amplitude and frequency of the driving source. In order to investigate the possibility of stochastic resonance we consider a random fluctuation of the fundamental frequency ω and write the noisy Bagley-Torvik equation as

$$\ddot{u} + \gamma \frac{d^{3/2}}{dt^{3/2}} u + [\omega^2 + \xi(t)]u = a \cos \Omega t \tag{2}$$

The fractional order derivative is defined as [6].

$$D^{\alpha} u = \int_a^t \frac{(t-\tau)^{n-\alpha-1}}{\Gamma(n-\alpha)} u(\tau) d\tau \tag{3}$$

where n is an integer and α the order of derivative and $n > \alpha$. The system with $\alpha = 3/2$ defines the Bagley-Torvik equation. We consider an additive frequency noise $\xi(t)$ with the mean and autocorrelation function specified as

$$\langle \xi(t) \rangle = 0, \ \langle \xi(t)\xi(s) \rangle = \sigma \exp(-\lambda t) \tag{4}$$

where σ denotes the root variance and λ defines the correlation rate.

3 Solution

Equation (2) defines a stochastic differential equation. The solution is sought by using the methods of statistical differentiation and averaging as detailed in [10]. The method proceeds as follows. The first step is to write Eq. (2) in averaged form

$$\left(\frac{d^2}{dt^2} + \gamma \frac{d^{\alpha}}{dt^{\alpha}} + \omega^2 \right) \langle u \rangle + \langle \xi u \rangle = a \cos \Omega t \tag{5}$$

Next, multiply Eq. (2) by ξ and average to yield

$$\langle \xi \frac{d^2 u}{dt^2} \rangle + \gamma \langle \xi \frac{d^{\alpha}}{dt^{\alpha}} u \rangle + \omega^2 \langle \xi u \rangle + \langle \xi^2 u \rangle = 0 \tag{6}$$

For evaluation of the higher correlator $<\xi^2 u>$ the decoupling procedure as discussed in [9] is applied. Further simplification of Eq. (6) is made by using the Shapiro-Loginov formulae

$$\langle \xi \frac{d^2 u}{dt^2} \rangle = \left(\frac{d}{dt} + \lambda \right)^2 \langle \xi u \rangle \tag{7}$$

$$\langle \xi \frac{d^\alpha u}{dt^\alpha} \rangle = e^{-\lambda t} \frac{d^\alpha}{dt^\alpha} \langle \xi u \rangle e^{\lambda t} \tag{8}$$

Writing $\xi^2 = \sigma$ and making substitutions of Eqs. (7) and (8) in Eq. (6) we obtain

$$\left[\left(\frac{d}{dt} + \lambda \right)^2 + \omega^2 \right] \langle \xi u \rangle + \gamma e^{-\lambda t} \frac{d^\alpha}{dt^\alpha} (\langle \xi u \rangle e^{\lambda t}) + \sigma \langle u \rangle = 0 \tag{9}$$

The fractional order derivatives do not obey Leibnitz chain rule for differentiation [17]. Therefore, we apply Laplace transformation method for finding solutions of Eqs. (5) and (9). Substituting $X = \langle u \rangle$ and $Y = \langle \xi u \rangle$ in Eqs. (5) and (9) we obtain

$$\left(\frac{d^2}{dt^2} + \gamma \frac{d^\alpha}{dt^\alpha} + \omega^2 \right) X + Y = a \cos \Omega t \tag{10}$$

$$\left[\left(\frac{d}{dt} + \lambda \right)^2 + \omega^2 \right] Y + \gamma e^{-\lambda t} \frac{d^\alpha}{dt^\alpha} Y e^{\lambda t} + \sigma X = 0 \tag{11}$$

The Laplace transformation of a fractional order derivative is defined as [18]

$$L\left[\frac{d^\alpha f(t)}{dt^\alpha} \right] = s^\alpha L[f(t)] - \sum_{k=0}^{n-1} s^{\alpha-k-1} f^k(0) \tag{12}$$

where $f(t) = X(t)$ or $Y(t)$ and $f^k(0)$ is the kth order time derivative at the time of start and $s = \Sigma + j\Omega$ is the complex frequency. For long time behavior, the effect of initial condition can be ignored [10]. Under these conditions, denoting the Laplace transforms of Eqs. (10) and (11) as $L(X) = x, L(Y) = y$ we obtain,

$$c_1 x + y = a \frac{s}{s^2 + \Omega^2} \tag{13}$$

$$c_2 y + \sigma x = 0 \tag{14}$$

where $c_1 = s^2 + \gamma s^\alpha + \omega^2$ and $c_2 = (s+\lambda)^2 + \gamma (s+\lambda)^\alpha + \omega^2$. Solving Eqs. (13) and (14) we obtain

$$x = H(s)\, a \frac{s}{s^2 + \Omega^2} \tag{15}$$

with,

$$H(s) = \frac{1}{c_1 - \frac{\sigma}{c_2}} \tag{16}$$

Using $s = j\Omega$ where $j = \sqrt{-1}$ we rewrite Eq. (16) as

$$H(j\Omega) = \frac{c_2}{c_1 c_2 - \sigma} \tag{17}$$

Note that the coefficients c_1 and c_2 are functions of $j\Omega$. Therefore, these can be written in the complex notation as $c_1 = a_1 + ja_2$ and $c_2 = a_3 + ja_4$. The new variables a_1, a_2, a_3 and a_4 are related to the system parameters as follows

$$a_1 = \omega^2 - \Omega^2 + \gamma\Omega^\alpha \cos(\alpha\frac{\pi}{2}), a_2 = \gamma\Omega^\alpha \sin(\alpha\frac{\pi}{2}) \tag{18}$$

$$a_3 = \lambda^2 + \omega^2 - \Omega^2 + \gamma\, r^\alpha \cos(\alpha\theta), a_4 = 2\lambda\Omega + \gamma\, r^\alpha \sin(\alpha\theta) \tag{19}$$

with,

$$r = \sqrt{\lambda^2 + \Omega^2},\ \theta = \tan^{-1}\left(\frac{\Omega}{\lambda}\right) \tag{20}$$

Seeking the steady state response in harmonic form defined as $\langle u \rangle = A\ \sin(\Omega t + \phi)$ we can obtain for the output amplitude

$$A = |H(j\Omega)|a = a\sqrt{\frac{f_1^2 + f_2^2}{f_3^2 + f_4^2}} \tag{21}$$

and the output phase

$$\phi = \arg(H(j\Omega)) = \tan^{-1}\left(\frac{f_2 f_3 - f_1 f_4}{f_1 f_3 + f_2 f_4}\right) \tag{22}$$

where,

$$f_1 = a_3, \quad f_2 = a_4, \quad f_3 = a_1 a_3 - a_2 a_4 - \sigma, \quad f_4 = a_1 a_4 + a_2 a_3 \tag{23}$$

4 Results and Discussion

The amplitude responses of the system are shown in Fig. 1 through Fig. 3 for different values of the noise parameters (σ and λ) for a fixed value of the damping constant (γ). In Fig. 1 one can see that under no-noise condition ($\sigma = 0$) there is a well defined resonance peak in the vicinity of $\Omega = 1$. However, as the noise strength (σ) is increased, a new resonance peak starts appearing on the lower frequency side with simultaneous decline and shifting of the primary resonance peak. The amplitude of the lower frequency side peak grows with increasing noise strength. This is the stochastic resonance behavior. In going from Figs. 1 to 2 it can be noted that as the noise strength is increased further, the SR peak reaches a maximum, then starts diminishing. These behaviors are consistent with the SR studies in [7] and [9].

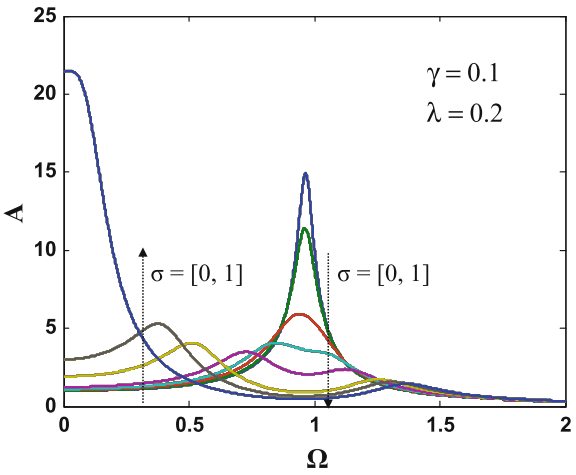


Fig. 1. Variation of amplitude (A) as a function of driving frequency (Ω) with noise strength (σ) as parameter. The range of σ variation is as indicated with the arrow pointing in increasing direction. Different curves correspond to $\sigma = 0, 0.01, 0.05, 0.1, 0.2, 0.5, 0.7, 1$. All the quantities are normalized by assuming $a = 1$ and $\omega = 1$

Figure 3 shows the effect of correlation rate for fixed values of noise strength and damping constant. The correlation rate is a measure of noise spectral bandwidth, higher the rate more is the bandwidth. It can be noted that at high correlation rates only the primary resonance peak occurs. As the correlation rate decreases (that is, the noise gets confined to narrower bandwidth) there emerge two new resonance peaks positioned symmetrically on either side of the primary peak position. This is a typical feature of signal mixing. To our knowledge, this type of behavior is not reported in the SR literature. In order to understand this behavior and to analyze the role of fractional order damping in it we need further analysis.

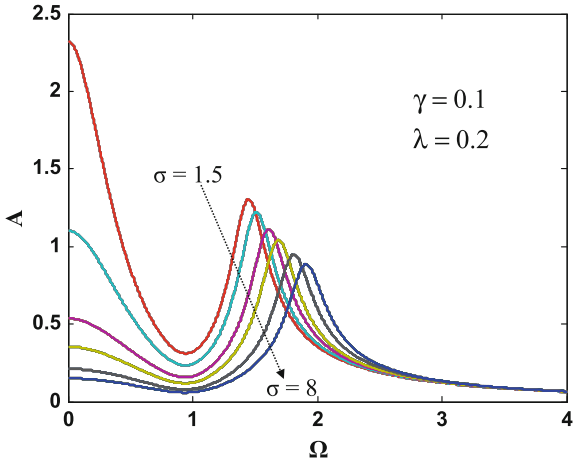


Fig. 2. Variation of amplitude (A) as a function of driving frequency (Ω) with noise strength (σ) as parameter. The range of σ variation is as indicated with the arrow pointing in increasing direction. Different curves correspond to σ = 1.5, 2, 3, 4, 6, 8. All the quantities are normalized by assuming $a = 1$ and ω = 1

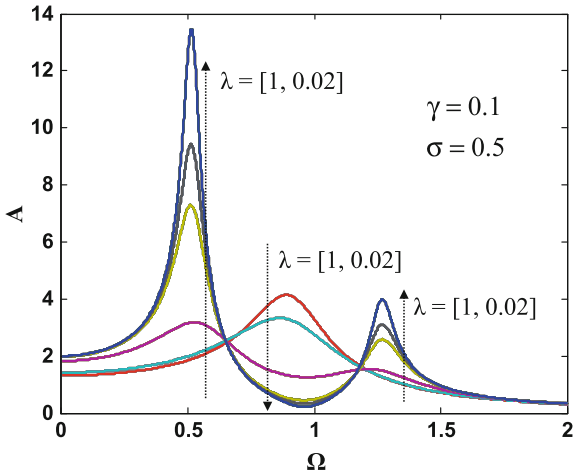


Fig. 3. Variation of amplitude (A) as a function of driving frequency (Ω) with noise correlation rate λ as parameter. The range of λ variation is as indicated with the arrow pointing in decreasing direction. Different curves correspond to $\lambda = 1, 0.8, 0.3, 0.08, 0.05, 0.02$. All the quantities are normalized by assuming $a = 1$ and ω = 1

5 Conclusion

We find that the frequency noise in a Bagley-Torvik system (which has the characteristics of a fractional order damping) gives rise to a new type of stochastic resonance, which looks like a traditional signal mixing (Fig. 3). This occurs for the noise processes with low correlation rates and moderate power level. For high correlation rate noises, the traditional stochastic resonance also occur (Figs. 1 and 2).

Acknowledgements. The authors AKS is thankful to U.G.C., India for providing research fellowships for pursuing this work.

References

1. Ando B, Graziani S (2000) Stochastic resonance: theory and applications. Kluwer Academic Publisher, Springer Sience+Bussiness Media
2. Chapeau-Blondeau F, Rousseau D (2002) Noise improvements in stochastic resonance: from signal amplification to optimal detection. Fluct Noise Lett 2:L221–L233
3. Benzit R, Sutera A, Vulpiani A (1981) The mechanism of stochastic resonance. J Phys A Math Gen 14:453–457
4. Wiesenfeld K, Moss F (1995) Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDs. Nature 373:33–36
5. McNamara B, Wiesenfeld K, Roy R (1988) Observation of stochastic resonance in a ring laser. Phys Rev Lett 60:2626–2629
6. Dylov DV, Fleischer JW (2010) Nonlinear self-filtering of noisy images via dynamical stochastic resonance. Nat Photonics 4:323–328
7. Verma VK, Yadava RDS (2016) Stochastic resonance in MEMS capacitive sensors. Sens Actuators B Chem 235:583–602
8. Gitterman M (2005) The noisy oscillators: the first hundreds years, from Einstein untill now. World Scientific
9. Gitterman M (2003) Harmonic oscillator with multiplicative noise: nonmonotonic dependence on the strength and the rate of dichotomous noise. Phys Rev E 67:1–4
10. Yu T, Zhang L, Luo M-K (2013) Stochastic resonance in the fractional Langevin equation driven by multiplicative noise and periodically modulated noise. Phys Scr 88:1–6
11. Podlubny I (1999) Mathematics in science and engineering. Fractional differential equations. Academic Press
12. Kilbas AA, Srivastava HM, Trujillo JJ (2006) Theory and applications of fractional differential equations. Elsevier BV
13. Rawashdeh EA (2006) Numerical solution of semidifferential equations by collocation method. 174:869–876
14. Du M, Wang Z, Hu H (2013) Measuring memory with the order of fractional derivative. Nat Sci Rep 3:1–3
15. Wahi P, Chatterjee A (2004) Averaging oscillations with small fractional damping and delayed terms. Nonlinear Dyn 38:3–22

16. Freundlich F (2013) Vibration of a simply supported beam with a viscoelastic material model support movement excitation. Shock Vib 20:1103–1112
17. Tarasov VE (2013) No violation of the Leibniz rule. No fractional derivative. Commun Nonlinear Sci Numer Simul 18:2945–2948
18. Liang S, Wu R, Chen L (2015) Laplace transform of fractional order differential equations. Electron J Differ Equ 2015:1–15

Classification of Brain MRIs Forming Superpixels

Nidhi Gupta(✉), Shiwangi Mishra, and Pritee Khanna

Computer Science and Engineering Discipline, PDPM IIITDM Jabalpur, Dumna Airport Road, Jabalpur 482005, MP, India
{nidhi.gupta, m.shiwangi, pkhanna}@iiitdmj.ac.in

1 Introduction

The abnormal growth of cells in the brain results in tumors. It is being observed that deaths caused by brain tumor are increased and it is a second leading cause for death nowadays. The article published by brain tumor association states that approximately 79,270 cases of brain tumor or related diseases will be diagnosed at the end of year 2017 [1, 2]. Brain tumor is more prominent in men as compared to women. It is a frequently found disease in children too [3]. Its early detection through image modalities has emerged as a necessary task in the medical science. Brain images can be captured through several modalities like magnetic resonance imaging (MRIs), X-rays, computed tomography (CT) scans, etc. Among these, brain MRI scans are the most frequently used among radiologists to detect tumorous images [4, 5]. Detection of brain tumor is a challenging task for the radiologists and clinicians, as it is not easy to detect various anomalies by simply visualizing medical images.

This work is aimed to classify tumorous and non-tumorous brain MRIs. The proposed classification is based on the generation of superpixels. Superpixels are more informative, and they adhere better to natural image boundaries. They also reduce susceptibility to noise (outliers) and redundancy [6, 7]. Run length of centralized patterns (RLCP) features is extracted and classified by naïve Bayes (NB) classifier. RLCP is proven robust against illumination changes and easy to implement. These features preserve most of the textural information and are also used to detect directional biasing of the textural patterns [8]. It is observed that the proposed approach using superpixel based classification produces significant results on two clinical datasets and outperforms existing methods. The work is organized in five sections. Section 2 highlights the existing works. The proposed methodology is discussed in Sect. 3. Results and discussion are presented in Sect. 4. At last, Sect. 5 concludes the work.

2 Previous Works

Based on the gravity of the situation, various researches are focused in this direction now. Segmentation is one of the important tasks for this type of categorical applications. Among various segmentation techniques, clustering is very common. Watershed,

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_57

k-means, fuzzy c-means (FCM), region growing, and Markov random field (MRF) are frequently used algorithms for segmentation task [9]. However, manual interruption and tuning of the parameters make them inconvenient for real-time processing. FCM with pixel intensities is used for clustering non-tumorous and tumorous MRIs [10] to achieve 88% similarity index. In a similar work, FCM is used with GLCM features to classify tumorous and non-tumorous MRIs and reported 98.67% accuracy [11]. In [12], watershed algorithm is used to extract the region of interest and 93.13% dice similarity index is obtained for segmentation results. In the similar way, [13] used region growing clustering for segmentation, which is followed by threshold to get segmented images. In the recent works, a few new algorithms for segmentation have been developed and it is observed that their performances are quite better as compared to conventional approaches [14]. In evolutionary methods, MRF model based segmentation techniques also perform better [15]. Voxels are extracted as features and 96% total lesion load is observed by MRF. In the recent works, [16] used superpixel-based segmentation on FLAIR images and classified images by extremely randomized tree to achieve 91% dice similarity coefficient. Similarly, texture-based and contour-based algorithms are used for automatic segmentation of tumors to achieve 96.80% accuracy on simulated FLAIR images [17]. In another work, wavelet transformation is used to extract texture features on BRATS dataset and 95% dice overlap is achieved on enhancing tumor regions [18].

To assists radiologists, various computer aided diagnosis (CAD) systems are developed to detect and classify tumorous and non-tumorous MRIs. In 2014, El-Dahshan et al. [19] proposed a technique by using discrete wavelet transformation (DWT) coefficients obtained from raw images and reduced feature length through principal component analysis (PCA) resulting in 98% accuracy using KNN. In 2015, Gupta and Khanna [20] developed a block based technique and extracted mean of each block. Threshold is used for classification of MRIs and 97.40% accuracy is observed. In 2016, selective block based approach is proposed and local binary patterns (LBP) are extracted from the selected blocks resulting in 99.67% accuracy [21]. FCM and watershed approaches are used to extract region of interest in [22] to observe 91% accuracy through NB classifier. Further, Gupta et al. [8] developed a technique using the fusion of different MR pulse sequences and extracted RLCP features from the segmented images. Images are classified by NB to achieve 97.83% on JMCD and 96.47% on BRATS dataset.

Although these works perform good but sometimes due to the selection of segmentation techniques, features, or classifiers, most of these hold a complex structure and computational costs. Conventional segmentation techniques, region growing approaches, and fuzzy c-means require manual tuning of parameters. This interruption delays the process. Features like Gabor suffer with correlation problem due to non-orthogonality [23], and GLCM yield poor directionality [24]. The complexity of classifiers and their biasness is another important issue. To overcome some of these issues, a simple yet efficient segmentation approach is used for the classification in this work and a CAD system is developed to assists radiologists for tumor detection on brain MRIs. The proposed approach uses the concept of superpixels. The performance

of RLCP is observed better than various conventional features. The proposed CAD system uses RLCP features extracted from the segmented images and classified by NB for detection of brain tumor.

3 Methodology

Basic steps of the proposed CAD system are depicted in Fig. 1 and discussed here. Superpixels capture image redundancy and provide a convenient primitive to compute image features. This greatly reduces the complexity of subsequent image processing tasks. This technique groups the pixels into perceptually meaningful regions, which can be further used to replace the rigid structure of the pixel grid. Instead of going through pixelwise extraction, superpixels are generated to efficiently utilize the property of each significant pixel [25]. The concept of superpixels is new in the image processing field and is successfully used for the segmentation task. Simple linear iterative clustering (SLIC) algorithm is used to create superpixels here. SLIC is an adaptation of k-means for superpixel generation. It produces compact superpixels in low computational cost of $O(n)$ having regular sized and excellent boundary adherence. SLIC is the fastest superpixel method and its advantage increases with the image size.

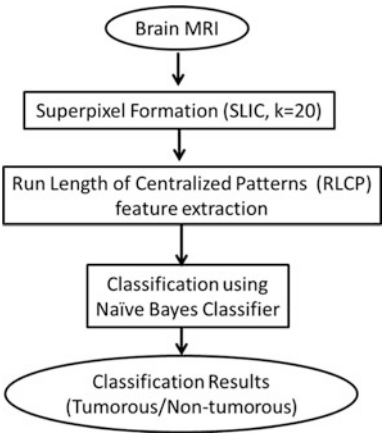


Fig. 1. Proposed methodology for brain MRI classification

3.1 Dataset Acquisition

The experiments are performed on two brain MRI datasets. A dataset is collected from NSCB Medical College, Jabalpur (referred as JMCD here) and the other is publically available Brain Tumor Segmentation Challenge (BRATS) dataset. JMCD contains images of 140 patients (80 non-tumorous and 60 tumorous) with 200–250 scanned MR

images of each patient having different kind of pulse sequences, i.e., T1-weighted, T2-weighted, T1-post contrast, and Fluid Attenuated Inversion Recovery (FLAIR). BRATS dataset contains four types of pulse sequences of 80 patients.

3.2 Segmentation

Superpixels are formed by the combination of significant pixels, which represent specific characteristics around the region. Superpixels are generated by simple linear iterative clustering with the adaptation of k-means clustering. The size and compactness are prescribed from the search space and weighted distance, respectively. Parameter k is initialized to get the desired number of superpixels. The grid interval is defined by $S = \sqrt{MN/k}$, where MN is number of pixels in the image. Each pixel is associated with the nearest clustering centers through the distance measure. A residual error E is measured by $L2$ norm to compute the locations of new cluster center and previous cluster center. This process repeats iteratively until the error converges, and assignments and updates are maintained [25]. Following this strategy, 20 significant superpixel grids are prepared for the analysis and their accuracy is observed. These segmented regions are further processed for feature extraction and classification. The formation of superpixels on both the datasets is shown in Fig. 2.

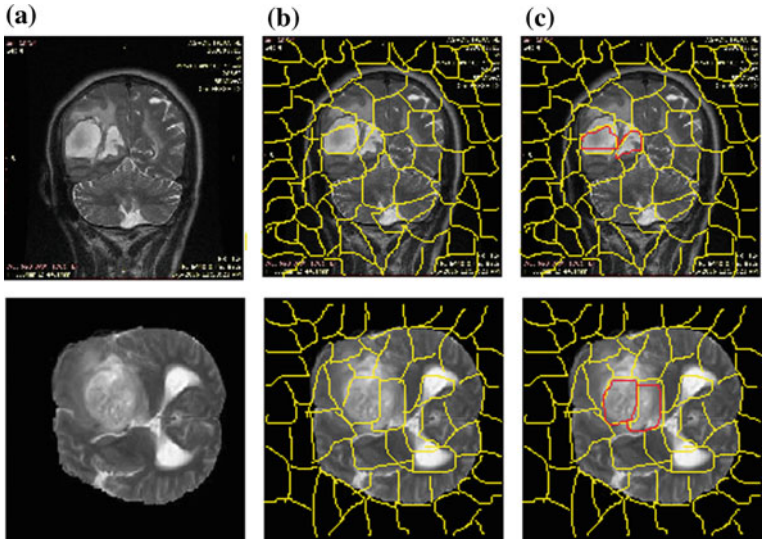


Fig. 2. Proposed approach (first row: JMCD and second row: BRATS) **a** original images **b** formation of superpixels **c** superpixels highlighting tumor

3.3 Feature Extraction and Classification

The combined strategy of Local Binary Pattern (LBP) and Gray-Level Run Length (GLRL) matrices gives a new feature named as Run Length of Centralized Patterns (RLCP) [8]. LBP is computed by matching the pixel value with its neighbor pixels.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p,\ s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases},$$

where g_c represents the central pixel and g_p denotes the neighbor on the circle of radius of R and P is the total number of the neighbors. $P = 8$ and $R = 1$ are selected here for experimentation. This parameter selection is significantly proven good for similar image processing tasks. The luminance of an image is signified by the joint distribution and their differences. LBP code is indexed and GLRL matrices are formed in each principle direction to count the run length occurrences for each gray level. Let *C(i,j)* denotes run-length matrix for indexed LBP image. *C(i,j)* defines the number of runs with pixels of gray level i and runlength j. For this run length matrix *C(i,j)*, let M_R is the number of gray levels, N_R is the maximum run length, n_r is the total number of run lengths, and n_p is the number of pixels in the indexed LBP image. RLCP features used in this work are summarized in Table 1. The observed textural information is quite important for the categorical task, specifically for medical diagnosis [8].

Table 1. RLCP features used in this work

Run percentage $RP = \frac{n_r}{n_p}$	
Run length nonuniformity $RLN = \frac{1}{n_r}\sum_{j=1}^{N_R}(\sum_{i=1}^{M_R} C(i,j))^2$	Gray-level nonuniformity $GLN = \frac{1}{n_r}\sum_{i=1}^{M_R}(\sum_{j=1}^{N_R} C(i,j))^2$
Short run emphasis $SRE = \frac{1}{n_r}\sum_{i=1}^{M_R}\sum_{j=1}^{N_R}\frac{C(i,j)}{j^2}$	Long run emphasis $LRE = \frac{1}{n_r}\sum_{i=1}^{M_R}\sum_{j=1}^{N_R} C(i,j) \times j^2$
Low gray-level run emphasis $LGRE = \frac{1}{n_r}\sum_{i=1}^{M_R}\sum_{j=1}^{N_R}\frac{C(i,j)}{i^2}$	High gray-level run emphasis $HGRE = \frac{1}{n_r}\sum_{i=1}^{M_R}\sum_{j=1}^{N_R} C(i,j) \times i^2$
Short run low gray-level emphasis $SRLGE = \frac{1}{n_r}\sum_{i=1}^{M_R}\sum_{j=1}^{N_R}\frac{C(i,j)}{i^2 \times j^2}$	Long run low gray-level emphasis $LRLGE = \frac{1}{n_r}\sum_{i=1}^{M_R}\sum_{j=1}^{N_R}\frac{C(i,j)\times j^2}{i^2}$
Short run high gray-level emphasis $SRHGE = \frac{1}{n_r}\sum_{i=1}^{M_R}\sum_{j=1}^{N_R}\frac{C(i,j)\times i^2}{j^2}$	Long run high gray-level emphasis $LRHGE = \frac{1}{n_r}\sum_{i=1}^{M_R}\sum_{j=1}^{N_R} C(i,j) \times i^2 \times j^2$

The features are concatenated to form a feature vector of size $20 \times 11 \times 4 (= 880)$. The processed feature vector is classified into tumorous and non-tumorous MRIs by NB classifier. This classifier beholds the mathematical properties and builds probabilistic model based on the available data. It predicts the missing label of the data from the established maximum probability [26]. The entire methodology is illustrated by the algorithm shown below.

Algorithm: Classification of brain MRI

Input: T2-weighted brain MR image
Output: Predicted class label

1. *Take input image I_in from dataset*
2. *Form superpixels*
   a. *Customize required parameters*
      [l_out] = slic (I_in, k, m, r, dis)
   b. *Select Optimum parametric value k*
3. *Feature extraction:*
   Feature vector = RLCP(I_out)
4. *Classification by Naïve Bayes classifier*
   Predicted label = NB(Feature vector)

4 Results and Discussion

Only T2-weighted MR images are considered for experimentation as tumors appear brighter than normal tissues in this sequence. These images are preferred by radiologists and clinicians also for the diagnosis of tumors. The experiments are performed using MATLAB® R2013a and simulated on Core i3 with 4 GB RAM.

4.1 Classification Rates and Performance Validation

Performance of the system is evaluated on various well-established parameters, i.e., accuracy, precision, sensitivity, and specificity [27]. The definition of each parameter and their observed values are given in Table 2.

Table 2. Classification rates (in %)

	Accuracy $\frac{TP+TN}{TP+TN+FP+FN}$	Precision $\frac{TP}{TP+FP}$	Sensitivity $\frac{TP}{TP+FN}$	Specificity $\frac{TN}{TN+FP}$
JMCD	99.84	100	98.6	100
BRATS	100	100	100	100

The proposed classification achieved 99.84% accuracy with 100% precision on JMCD dataset. On BRATS, the proposed approach outperforms and gives 100%

results. This shows that the proposed superpixel-based segmentation, RLCP feature extraction, and classification perform quite well on both the datasets.

The statistical testing is applied to validate the classified outcomes. Statistical student's *t*-test is applied and it is observed that the proposed CAD system achieves 99% confidence interval with p value 0.00568 (<0.01) on JMCD (99.84 ± 0.0049).

4.2 Comparison with the Other CAD Systems

Performance of the proposed system is compared with some existing CAD systems. The existing techniques are thoroughly implemented on a uniform platform and examined on JMCD and BRATS datasets. The observed values are summarized in Table 3.

Table 3. Comparison with existing techniques

Techniques	Accuracy (in %)	
	JMCD	BRATS
Dhashan et al. [19]	96.25	99.99
Gupta and Khanna [20]	97.30	98.90
Gupta et al. [21]	94.76	97.97
Subhashini et al. [22]	94.43	97.80
Nabizadeh et al. [17]	96.29	100
Gupta et al. [8]	98.14	97.73
Proposed system	99.84	100

It can be seen that the proposed approach outperforms other existing techniques. Though other systems also work well and provide accuracy more than 94% on JMCD. However, the proposed system achieves more than 99% accuracy. The accuracy achieved on BRATS dataset is quite high. Although [19] also performed better on BRATS dataset (99.99% accuracy), but does not do so on JMCD (96.25% accuracy). Wavelet transformation is used to extract features from raw images in [19] and classification is performed by KNN. The work by [22] used FCM with watershed algorithm to cluster the images and extracted area as a feature to achieve 94.43% accuracy on JMCD and 97.80% on BRATS. Similarly in another work, Gabor features are extracted and images are classified by SVM in [17]. Their system achieved 100% accuracy on BRATS, but again it did not perform well on JMCD and observed only 96.29% accuracy. The comparison by our recent works, i.e., [8, 20] and [21] also shows that the observed accuracies are better with the proposed approach. In [8], images are fused and adaptively threshold to segment the tumorous regions. RLCP features are extracted and classified by NB to observe 98.14% on JMCD. In a similar way, block-based segmentation is used in [20] and mean is extracted to achieve 98.90% accuracy on BRATS. Selective block based segmentation is used and LBP features are extracted in [21]. These are classified by SVM to achieve 94.76% on JMCD and 97.97% accuracy

on BRATS. The advantage of the proposed approach over these existing approaches is a simple strategy that gives comparable results on both the datasets. The system is well structured and performs effectively on both the datasets.

5 Conclusion

The proposed CAD system is efficient in classifying tumorous MRIs from non-tumorous ones. The classification of images is accomplished by applying the concept of superpixels before feature extraction. Features extracted from each superpixel are concatenated to form a feature vector and classified through naïve Bayes classifier. The accuracy achieved is 99.84% on JMCD and 100% on BRATS. The performance is also examined through statistical testing. The work shows that the classification approach using superpixels can be successfully used for diagnosis and prognosis of tumors. The concept may be checked for classification of other diseases too in future.

Acknowledgements. We thank senior radiologist Dr. Pushpraj Bhatele, NSCB Medical College Jabalpur for providing real dataset and ground truth for validating the outcomes of the proposed approach.

Conflict of Interest In compliance with the standard ethics of research, particular/personal information of patients has not been disclosed.

References

1. CBTRUS Fact Sheet. CBTRUS (2015)
2. American Cancer Society (2015) Cancer facts and figures. Georgia. https://doi.org/10.3322/caac.21254
3. Over 2,500 Indian kids suffer from brain tumour every year—The Hindu, pp 1–6 (2016). http://www.thehindu.com/sci-tech/health/over-2500-indian-kids-suffer-from-brain-tumour-every-year/article8721029.ece
4. Young RJ, Knopp EA (2006) Brain MRI: tumor evaluation. J Magn Reson Im 24(4):709–724. https://doi.org/10.1002/jmri.20704
5. Mehndiratta A, Giesel FL (2011) Brain tumour imaging. In: Abujamra AL (ed) Diagnostic techniques and surgical management of brain tumors, pp 1–17. https://doi.org/10.5772/23507
6. Wang H et al (2017) Improved image denoising algorithm based on superpixel clustering and sparse representation. Appl Sci 7(5):436
7. Neubert P, Peter P (2012) Superpixel benchmark and comparison. In: Proceedings Forum Bildverarbeitung
8. Gupta N, Bhatele P, Khanna P (2017) Identification of Gliomas from brain MRI through adaptive segmentation and run length of centralized patterns. J Comput Sci https://doi.org/10.1016/j.jocs.2017.02.009
9. Oke OA et al (2012) Fuzzy kc-means clustering algorithm for medical image segmentation

10. Boudraa AO, Dehak SMR, Zhu YM, Pachai C, Bao YG, Grimaud J (2000) Automated segmentation of multiple sclerosis lesions in multispectral MR imaging using fuzzy clustering. Comput Biol Med 30(1):23–40. https://doi.org/10.1134/s1054661808020235
11. Sharma M, Mukherjee S (2013) Fuzzy c-means and snake model for segmenting astrocytoma—a type of brain tumor. Int J Adv Eng Sci 3(3):30–35
12. Benson CC, Deepa V, Lajish VL, Rajamani K (2016) Brain tumor segmentation from MR brain images using improved fuzzy c-means clustering and watershed algorithm, pp 187–192
13. Deng W, Xiao W, Deng H, Liu J (2010) MRI brain tumor segmentation with region growing method based on the gradients and variances along and inside of the boundary curve. In: Proceedings—2010 3rd international conference on biomedical engineering and informatics, BMEI 2010 1:393–396. https://doi.org/10.1109/bmei.2010.5639536
14. Ren-jin L, Fu M, Fang J (2010) A new texture image segmentation algorithm. In: 2010 5th international conference on computer science and education (ICCSE). IEEE
15. Van LK, Maes F, Vandermeulen D, Colchester, Suetens P (2001) Automated segmentation of multiple sclerosis lesions by model outlier detection. IEEE Trans Med Im 20(8):677–688. https://doi.org/10.1109/42.938237
16. Soltaninejad M, Yang G, Lambrou T, Allinson N, Jones TL, Barrick TR, Ye X (2016) Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI. Int J Comput Assist Radiol Surgery, 1–21. https://doi.org/10.1007/s11548-016-1483-3
17. Nabizadeh N, Miroslav K (2017) Automatic tumor segmentation in single-spectral MRI using a texture-based and contour-based algorithm. Expert Syst Appl 77:1–10
18. Usman K, Kashif R (2017) Brain tumor classification from multi-modality MRI using wavelets and machine learning. Patt Anal Appl 1–11
19. El-Dahshan E-SA, Mohsen HM, Revett K, Salem A-BM (2014) Computer-aided diagnosis of human brain tumor through MRI: a survey and a new algorithm. Expert Syst Appl 41:5526–5545. https://doi.org/10.1016/j.eswa.2014.01.021
20. Gupta N, Khanna P (2015) A fast and efficient computer aided diagnostic system to detect tumor from brain magnetic resonance imaging. Int J Im Syst Technol 25(2):123–130. https://doi.org/10.1002/ima.22128
21. Gupta N, Seal A, Bhatele P, Khanna P (2016) Selective block based approach for neoplasm detection from T2-weighted brain MRIs. In: International conference on signal and image processing, pp 151–155
22. Subashini MM, Sahoo SK, Sunil V, Easwaran S (2016) A non-invasive methodology for the grade identification of astrocytoma using image processing and artificial intelligence techniques. Expert Syst Appl 43:186–196
23. Brown ML, Williams WJ, Hero AO (1994) Non-orthogonal Gabor representation of biological signals. In: Proceedings ICASSP '94, IEEE international conference acoustic speech signal process, IV/305-IV/308. https://doi.org/10.1109/icassp.1994.389742
24. Wei L, Hong-ying D (2016) Real-time road congestion detection based on image texture analysis. Procedia Eng 137:196–201. https://doi.org/10.1016/j.proeng.2016.01.250
25. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S (2012) SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans Pattern Anal Mach Intell 34 (11):2274–2281. https://doi.org/10.1109/TPAMI.2012.120
26. Murphy KP (2006) Naive Bayes classifiers generative classifiers. Bernoulli 4701:1–8. https://doi.org/10.1007/978-3-540-74958-5_35
27. Olson DL, Delen D (2008) Advanced data mining techniques. Springer, Heidelberg. https://doi.org/10.1007/978-3-540-76917-0

Damping Noise Induced Stochastic Resonance Improves Q-Factor of M/NEMS Resonators

Priyanka Singh(✉) and R. D. S. Yadava

Sensors & Signal Processing Laboratory, Department of Physics, Institute of Science, Banaras Hindu University, Varanasi 221005, India
{priyankabhul7,ardius}@gmail.com

1 Introduction

Stochastic resonance (SR) is a noise-induced resonance phenomenon in an oscillatory system [1]. The random fluctuations in oscillator parameters can produce cooperative effect to boost power level of a weak response. The phenomenon was first recognized by Benzi et al. while modeling the climatic fluctuations and the periodic occurrence of ice age on earth's surface [2]. Later, it found applications in several engineering disciplines like electronic circuits [3], laser [4] and superconducting quantum interference devices (SQUID) [1] and in biology like predator sensing by pray (crayfish response) [1]. Initially, it was thought that the existence of nonlinearity in the system dynamics is a basic prerequisite for the occurrence of SR. However, later in several studies, it was shown that it can occur in linear systems also with multiplicative noise [5], and with additive noise under periodic forcing [6]. The response of a periodically forced damped harmonic oscillator with random frequency and damping fluctuations has been analyzed in detail for its dependence on the noise strength and rate of fluctuation [7, 8]. We noted that the influence of SR on signal power (or amplitude) has been widely explored in these works, and its advantages are mentioned for relevant applications, a similar attention on the associated phase SR is not given.

In a recent study, noise-induced SR was analyzed for the case of vapor sorption-desorption associated fluctuations in surface mass density of the polymer-coated surface acoustic wave (SAW) chemical sensors [9]. The surface mass fluctuations are transformed into the SAW oscillator's frequency fluctuation. The SR was produced by superimposing a weak vapor loading modulation. It was found that the SR improves the signal-to-noise ratio of the SAW chemical sensors. In two other recent studies, the phenomena of SR was analyzed in micro/nanoelectromechanical (M/NEMS) cantilever sensors under the influence of frequency as well as damping fluctuations [10, 11]. In view of the applications for the chemical and biological sensing, the emphasis of their analyses had been on improving the sensor sensitivity by the amplitude SR. No work has come to our notice that explores utilizing SR in phase for sensing applications. In this paper, we consider M/NEMS cantilever resonator modeled as a driven damped harmonic oscillator under the influence of damping fluctuations and analyze its impact on the resonator frequency, amplitude, and quality factor (Q-factor). The findings here indicate that exploiting phase SR may provide better option for improving performances of chemical and biological sensors.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_58

2 Cantilever Dynamics and Damping Fluctuations

MEMS cantilever used as movable plate in a parallel plate capacitor is very common for using this device as resonator or varactor. The device is usually biased by fixed DC voltage and forced into vibrations by superimposing harmonic AC voltage. The AC drive is kept sufficiently small so as to keep the cantilever vibrations on one side of the neutral position and also to avoid pull-in instability. An electrostatically driven MEMS or NEMS cantilever device becomes a chemical or a biological sensor if on one surface of the cantilever is sensitized for the analyte binding. The sorption and desorption of analytes on the sensitized surface generate changes in the cantilever mass and stiffness. These in turn bring changes in the cantilever resonance frequency and the voltage dependence of the MEMS capacitance; both can be read as sensor signal.

The cantilever dynamics in fundamental mode is often simplified by an equivalent mass–spring–damper model whose motion under (DC + AC) actuation is described by the following differential equation [11]:

$$\ddot{u} + 2\gamma\dot{u} + \omega^2 u = (F_0/m_{eff})\cos(\Omega t) \tag{1}$$

where $u(t)$ represents displacement from the static equilibrium, overdot (.) denotes time derivative, and

$$2\gamma = \frac{c}{m_{eff}} \tag{2}$$

$$\omega^2 = \frac{1}{m_{eff}}\left(k_{eff} - \frac{\varepsilon A V_{DC}^2}{(d - x_0)^3}\right) \tag{3}$$

$$F_0 = \frac{\varepsilon A V_{DC} V_{AC}}{(d - x_0)^2} \tag{4}$$

with m_{eff} denoting the cantilever effective mass, k_{eff} the effective spring constant, c the damping constant, d the zero-drive plate separation, x_0 the static plate separation under DC bias, A the cantilever plate area, ε the dielectric permittivity of the capacitor filling medium. The (DC + AC) drive voltage is expressed as $V(t) = V_{DC} + V_{AC}\cos(\Omega t)$, where V_{DC} denotes the DC voltage, and V_{AC} and Ω denote, respectively, the AC drive voltage amplitude and angular frequency.

We consider random fluctuations in the damping that may arise due to several processes like the thermal, thermomechanical, and viscous forces in cantilever dynamics. We write the noisy equation of motion as

$$\ddot{u} + 2\gamma(1 + \xi(t))\dot{u} + \omega^2 u = (F_0/m_{eff})\cos(\Omega t) \tag{5}$$

where $\xi(t)$ represents the noisy part of the damping coefficient. We assume $\xi(t)$ to be a Gaussian random variable with zero mean and exponential autocorrelation is defined as

$$\langle \xi(t)\xi(t+\tau)\rangle = \sigma^2 \exp(-\lambda\tau) \tag{6}$$

where mean $\langle \xi(t)\rangle = 0$ and variance $\langle \xi^2(t)\rangle = \sigma^2$, and λ is the correlation rate.

The solution of (5) is obtained by the methods of statistical averaging and statistical differential calculus as detailed in [8]. By defining

$$\frac{du}{dt} = y \tag{7}$$

$$\frac{dy}{dt} = -2\gamma(1+\xi(t))y - \omega^2 u + (F_0/m_{eff})\cos(\Omega t) \tag{8}$$

and statistical averages over time, we can write

$$\frac{d\langle u\rangle}{dt} = \langle y\rangle \tag{9}$$

$$\frac{d\langle y\rangle}{dt} = -2\gamma\langle y\rangle - \langle \xi(t)y\rangle - \omega^2\langle u\rangle + (F_0/m_{eff})\cos(\Omega t) \tag{10}$$

Multiplying these equations by $\xi(t)$ and averaging over time yields

$$\langle \xi \frac{du}{dt}\rangle = \langle \xi y\rangle \tag{11}$$

$$\langle \xi \frac{dy}{dt}\rangle = -2\gamma\langle \xi y\rangle - 2\gamma\langle \xi^2(t)y\rangle - \omega^2\langle \xi u\rangle \tag{12}$$

Using the statistical differentiation formulae [12], we can obtain the equations for $\langle \xi u\rangle$ and $\langle \xi y\rangle$ as written below:

$$\frac{d\langle \xi u\rangle}{dt} = \langle \xi \frac{du}{dt}\rangle - \lambda\langle \xi u\rangle \tag{13}$$

$$\frac{d\langle \xi y\rangle}{dt} = \langle \xi \frac{dy}{dt}\rangle - \lambda\langle \xi y\rangle \tag{14}$$

Using (11)–(14), we obtain

$$\frac{d\langle \xi u\rangle}{dt} = \langle \xi y\rangle - \lambda\langle \xi u\rangle \tag{15}$$

$$\frac{d\langle \xi y\rangle}{dt} = -(2\gamma+\lambda)\langle \xi y\rangle - 2\gamma\sigma^2\langle y\rangle - \omega^2\langle \xi u\rangle \tag{16}$$

Now, by using (9), (10), (15), and (16), we obtain the following fourth-order differential equation for $\langle u\rangle$:

$$\begin{aligned} &\frac{d^4\langle u\rangle}{dt^4} + c_3\frac{d^3\langle u\rangle}{dt^3} + c_2\frac{d^2\langle u\rangle}{dt^2} + c_1\frac{d\langle u\rangle}{dt} + c_0\langle u\rangle \\ &= b_2(F_0/m_{eff})\cos\,\Omega t + b_1(F_0/m_{eff})\sin\,\Omega t \end{aligned}, \tag{17}$$

where

$$c_3 = 2(\lambda + 2\gamma) \tag{18}$$

$$c_2 = (2\omega^2 + (\lambda + 2\gamma)^2 + 2\lambda\gamma - 4\gamma^2\sigma^2) \tag{19}$$

$$c_1 = ((2\omega^2 + 2\gamma\lambda)(\lambda + 2\gamma) - 4\lambda\gamma^2\sigma^2) \tag{20}$$

$$c_0 = \omega^2(-\omega^2 + \lambda(\lambda + 2\gamma)) \tag{21}$$

$$b_1 = \Omega(\lambda + 2\gamma) \tag{22}$$

$$b_2 = (\omega^2 - \Omega^2 + \lambda(\lambda + 2\gamma)) \tag{23}$$

Equation (17) describes the average cantilever displacement. For small perturbations, one can assume that the solution remains close to steady motion under external drive. Therefore, one can assume the solution to be of the form

$$\langle u\rangle = a\,\cos(\Omega t - \phi) \tag{24}$$

where a and ϕ denote, respectively, the time-averaged amplitude and phase of the noisy oscillator. Making substitution of (24) into (17), using certain trigonometric identities, and equating the coefficient of $\sin(\Omega t)$ and $\cos(\Omega t)$, we obtain the following relations between a and ϕ:

$$f_3 a\,\cos\phi - f_4 a\,\sin\phi = f_2 \tag{25}$$

$$f_4 a\,\cos\phi + f_3 a\,\sin\phi = f_1 \tag{26}$$

with

$$f_1 = b_1(F_0/m_{eff}) \tag{27}$$

$$f_2 = b_2(F_0/m_{eff}) \tag{28}$$

$$f_3 = (\Omega^4 - c_2\Omega^2 - c_0) \tag{29}$$

$$f_4 = c_3\Omega^3 - c_1\Omega \tag{30}$$

Equations (25) and (26) can be solved for a and ϕ. The results are as follows:

$$a = \left(\frac{f_1^2 + f_2^2}{f_3^2 + f_4^2}\right)^{1/2} \tag{31}$$

$$\phi = \tan^{-1}\left(\frac{f_1 f_3 - f_2 f_4}{f_2 f_3 + f_1 f_4}\right) \tag{32}$$

These equations represent the amplitude and the phase of the cantilever vibrations under noisy damping environment. In the following, we examine these results with objectives for improving the performance of M/NEMS cantilever sensors by stochastic resonance.

3 Numerical Results and Discussion

Figure 1 shows the average amplitude (a) and phase (ϕ) responses of an electrostatically driven M/NEMS cantilever as a function of the AC drive frequency Ω. These responses are as defined through Eqs. (24), (31), and (32). Different quantities involved in these expressions are in the normalized forms as indicated in the figure. From Fig. 1a, one can see that as the noise strength (σ) is increased from 0 to 2.000, the resonance peak becomes sharper with increasing peak height. On further increase from 2.449 to 3.162, the height starts decreasing with peak broadening. This behavior in response to the addition of damping noise is the characteristic of amplitude stochastic resonance. From Fig. 1b, one can notice that the variation of phase near resonance peak becomes steeper with the increase in noise strength up to $\sigma = 2.000$. Beyond $\sigma > 2$, the steepness starts declining approaching toward low noise conditions. This has direct implications for the quality factor (Q) of the cantilever resonator.

One can recall that the one definition of the Q-factor relates it to the phase slope of the resonator at resonance frequency. This is written as

$$Q = \frac{1}{2}\omega \left|\frac{d\phi}{d\Omega}\right|_{\Omega=\omega} \tag{33}$$

This means that an increase in the magnitude of phase slope $|d\phi/d\Omega|$ must imply an enhancement in the Q-factor. The damping noise assisted enhancement in phase slope for noise strengths up to $\sigma = 2.000$ is contrary to the usual notion. Usually one expects the Q-factor to degrade with the addition of noise. The Q-factor improvement here is a direct consequence of the damping noise induced phase SR. At higher noise levels beyond $\sigma > 2.000$, the phase slope starts receding back and reaches the behavior of an overdamped resonator.

It is also noticeable that both the amplitude SR and the phase SR are correlated in the sense that the highest amplitude is associated with the steepest phase slope. Figure 2 illustrates this behavior more clearly. In this figure, the Q-factor as given by (33) has been plotted against the normalized drive frequency Ω/ω for three cases damping and noise strength σ used as parameter. The normalized damping coefficient

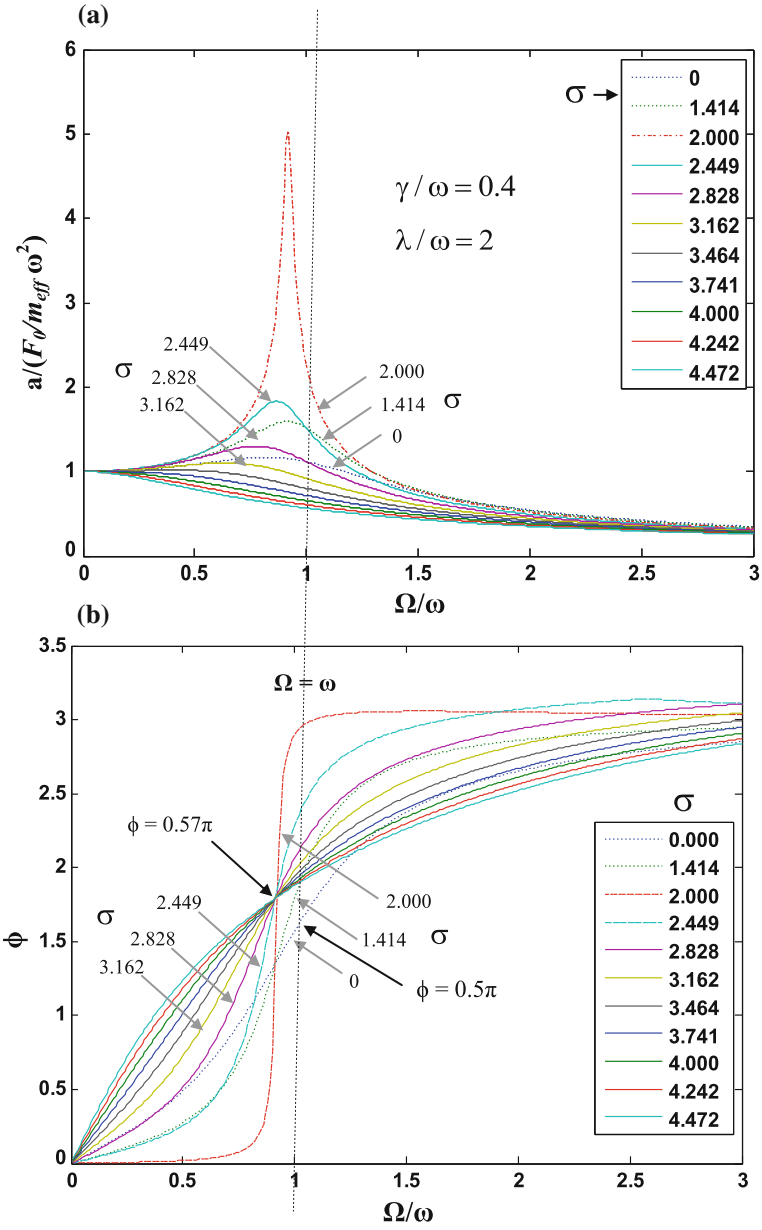


Fig. 1. Stochastic resonance in M/NEMS resonator with damping noise

values $\gamma/\omega = 0.04$, 0.4 and 1.0 can be taken to represent, respectively, the low, moderate, and high levels of damping. The normalized correlation rate $\lambda/\omega = 2$ has been fixed for all the cases. That means these cases represent a particular noise mechanism with varying noise powers. A distinct feature revealed by these plots is that

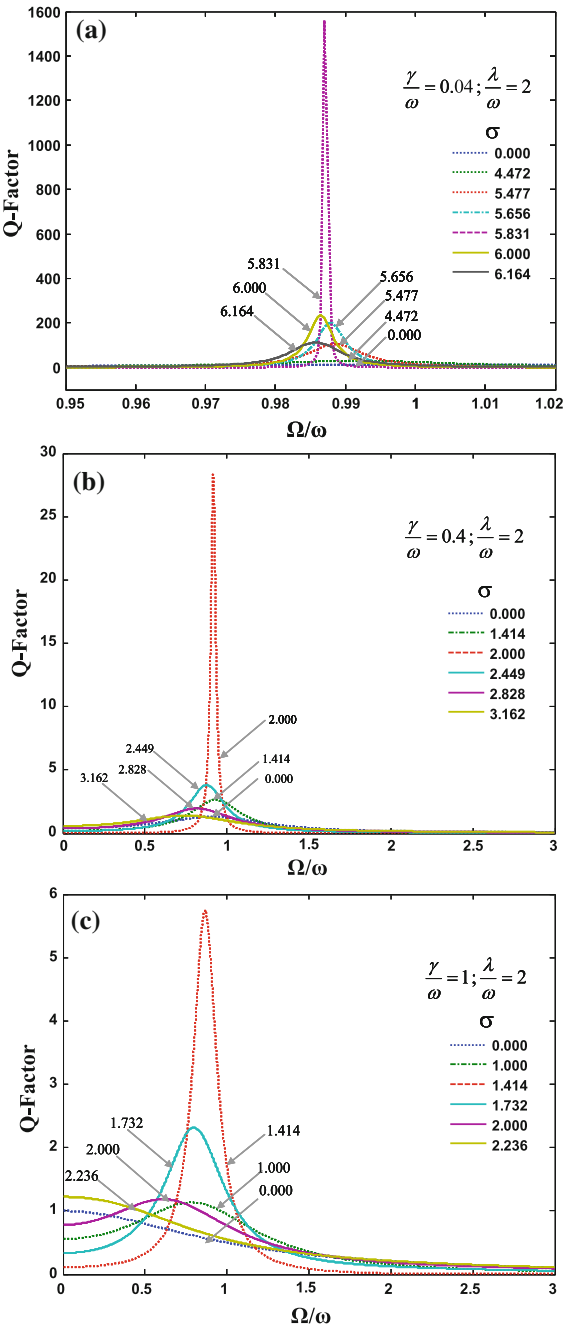


Fig. 2. Enhancement of Q-factor of M/NEMS cantilever resonators due to damping noise induced phase SR for three damping conditions: **a** low damping, **b** moderate damping, and **c** large damping

a high Q resonator (low γ) will need higher noise power for producing phase SR; see the range of σ values in each plot in comparison. Another noticeable point is that in the low damping resonator the Q-factor improves by orders of magnitude larger than the high damping resonator, see the trend from Fig. 2a–c.

In applications where high Q resonators are needed the design and development efforts are mostly centered on reducing all kinds of internal noises, and isolating the resonating component from external perturbations. This involves sophisticated precision oscillator technology. The analysis of damping noise induced phase SR presented here suggests that the quality factor can be enhanced by addition of the random damping noise and adjusting its power level. This may be highly useful in conditions where the resonator device has to face harsh operating environment such as fluid turbulence in chemical and biosensing applications. From Fig. 1b, one can notice that for taking advantage of the phase SR it may be prudent to operate the resonator sensors at some offset frequency from the fundamental. In MEMS-controlled feedback oscillators, the occurrence of phase SR implies that one can enhance the frequency sensitivity of the oscillator by tapping phase SR. This may be helpful in improving the performance of M/NEMS oscillator sensors. These aspects are however vague at present and need further investigation.

4 Conclusion

The paper concludes that the fluctuations in damping processes associated with the operation of M/NEMS resonators and electrical oscillators can be advantageously exploited for improving their quality factors. This feature of damping noise induced stochastic resonance in phase could be very helpful in enhancing the performance of M/NEMS chemical and biological sensors operating in harsh fluidic conditions.

Acknowledgements. The author Priyanka Singh would like to acknowledge UGC (Government of India, New Delhi) for the grant of BSR fellowship to pursue this work. The authors would like to thank Mr. Saumitra Mishra, Mr. Sonamani Singh, Mr. Anurag Gupta, and Mr. Aman Kumar Singh for their help and cooperation.

References

1. Wiesenfeld K, Moss F (1995) Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDs. Nature 373
2. Benzi R, Parisi G, Sutera A, Vulpiani A (1982) Stochastic resonance in climate change. Tellus 34:10–16
3. Mantegna RN, Spagnolo B (1994) Stochastic resonance in tunnel diode. Am Phys Soc 49 (3):1792–1796
4. McNamara B, Wiesnfeld K, Roy R (1988) Observation of stochastic resonance in ring laser. Phys Rev Lett 60(24):2627–2629
5. Gitterman M, Berdichevsky V (1996) Multiplicative stochastic resonance in linear systems: analytical solution. Europhys Lett 36(3):161–165

6. Xu W, Ning L (2007) Stochastic resonance in linear system drieven by multiplicative and additive noise. Phys A 382:415–422
7. Gitterman M (2003) Harmonic oscillator with multiplicative noise: nonmonotonic dependence on the strength and the rate of dichotomous noise. Phys Rev E 67:057103-1-4
8. Gitterman M (2004) Harmonic oscillator with fluctuating damping parameter. Phys Rev E 69: 041101-1-4
9. Yadava RDS, Verma VK (2014) A diffusion limited sorption-desorption noise model for polymer coated SAW chemical sensor. Sens Actuators B 195:590–602
10. Snyder P, Joshi A (2014) Modeling a nanocantilever-based biosensor using a stochastic perturbed harmonic oscillator. Int J Nanosci 13(2)
11. Verma VK, Yadava RDS (2016) Stochastic resonance in MEMS capacitive sensors. Sens Actuators B 235:583–602
12. Shapiro VE, Loginov VM (1978) Formula of differentiation and their use for solving stochastic equations. Phys A 91:563–574

Finding the Association of mRNA and miRNA Using Next Generation Sequencing Data of Kidney Renal Cell Carcinoma

Shib Sankar Bhowmick[1,2(✉)], Luis Rato[3], and Debotosh Bhattacharjee[2]

[1] Department of Electronics and Communication Engineering, Heritage Institute of Technology, Kolkata, Kolkata, India
shibsankar.bhowmick@heritageit.edu

[2] Department of Computer Science and Engineering, Jadavpur University, Kolkata 700032, West Bengal, India
debotosh@ieee.org

[3] Department of Informatics, University of Évora, 7004-516 Évora, Portugal
lmr@di.uevora.pt

1 Introduction

MicroRNAs (miRNAs) are small non-coding RNAs of approximately 19–22 nucleotides that act as posttranscriptional gene expression regulators and bind with complementary messenger RNAs (mRNAs) for stimulating their degradation or repressing translation. In this regards, innovation of next generation sequencing (NGS) [6] technology provides improved solution to the understanding of an entire human genome sequence than compared to the previously used technologies, i.e., Sanger sequencing or capillary sequencing. NGS technologies can also be applied to the RNA sequencing (RNA-seq) to directly sequence complementary DNAs (cDNAs). RNA sequencing permits the quantitative analysis of gene expression and transcript variant discovery. The primary aim of RNA-sequencing is to find out differentially expressed genes when applied to multiple samples at different biological conditions.

Numerous statistical approaches have been proposed over the years to analyze differentially expressed RNA-seq data, e.g., edgeR [15], DESeq [1], and sSeq [20]. In this connection, algorithms like [14,17] can be used to rank and select the differentially expressed miRNAs. Apart from this, hypothesis test [10], classifier based method [7], and information theory based measures [13] are used in gene ranking as well as in miRNA ranking. Parametric method like [18] is developed using the expression overlapping between different classes. Nonparametric ReDiscovery Curve (RDCurve) based method [12] studies the stability of various ranking methods.

Reviewing all these methods, here we have proposed an integrated analysis of the gene expression profiling of both mRNAs and miRNAs by using kidney renal

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_59

clear cell carcinoma data that includes tumor and control samples. The proposed method is a feature selection wrapper, consisting of random forests [5], particle swarm optimization (PSO) [8], and support vector machine (SVM) [4] classifier. Moreover, emphasis is given here to identify the differentially expressed mRNAs and miRNAs in order to find out mRNA-miRNA pairs. In particular, this kind of observations might be significant in the cancer diagnosis. Moreover, we validate some of our most interesting findings in different biological significance analysis.

The paper is organized as follows: Sect. 2 briefly describes the proposed method. Section 3 shows the empirical results. Finally, Sect. 4 concludes this paper with an additional note of future work.

2 Proposed Method

The proposed wrapper-based feature selection method starts with 1135 number of samples having 20, 531 mRNAs. The mRNA sample includes 534 tumors and 601 control patients whereas miRNA includes 254 tumor and 455 control sample. Since this experiment is an integrated analysis of mRNA and miRNA data. Hence, common tumor and control samples among mRNA and miRNA data are considered here that consist of 186 tumor and 71 control samples. Steps of the proposed method are described below.

The primary goal of feature selection is to avoid any kind of over fitting as well as to improve the model performance. To gain a deeper insight into the underlying processes of feature selection, random forest is used here. Several measures of variable importance can be identified by random forest. Hence, by iteratively fitting random forest at each iteration, smallest set of informative mRNAs is identified. In this regards, smallest *OOB* error rate is considered to be the underline criteria. Because of the iterative approach, the *OOB* error is biased down and mRNAs with smallest *OOB* error are selected as the reduced set of features. For this analysis, normalized transcript reads of mRNAs are considered.

After feature selection using random forest, in the next step of this work, t-test and fold change concepts are applied in order to segregate differentially expressed mRNAs from normal and malignant samples. For this purpose, cutoff like p-value < 0.05 and the fold change > 1 or < -1 are used as in [21]. In particular, change in expression profile is considered to be the underline criteria to find out differentially expressed mRNAs. Moreover, the differentially expressed mRNAs are mapped to their normalized mRNA expression dataset of tumor and control samples.

The particle swarm optimization (PSO) and support vector machine (SVM) provide a wrapper-based approach for feature selection. It can effectively identify significant mRNAs associated with kidney renal cell carcinoma. For that, initial particles of PSO are created using mRNA indices. Thereafter, the swarm is prepared from a number of such particles. Moreover, encoded indices of these particles are used to make a subset of data that are taken part during classification by SVM. Here, fivefold cross validation (FCV) is applied and accuracy is

considered to be the underline criteria for fitness evaluation. Block diagram of the proposed approach is presented in Fig. 1. Moreover, the optimal signature refers to be the best set of differentially expressed mRNA that provides highest classification accuracy. In this regards, most appeared mRNAs concerning all runs of the proposed methods are considered.

In this experiment, the expression datasets comprising the profiles of mRNAs and miRNAs are considered. For this analysis, expression values of 1047 miRNA in 186 tumor and 71 control sample are used. The selected mRNAs by PSO+SVM are considered to find their corresponding miRNA targets. For this validation, miRDB [19] database is used. Shortlisted target miRNAs are used to find their expression profiles from the normalized miRNA expression dataset of 186 tumor and 71 control samples. Any further analysis of miRNA is done using this reference expression vector in-order to find out differentially expressed miRNAs.

Likewise, the mRNAs, based on p-value and fold change in expression levels, up and downregulated miRNAs are identified. The miRNAs are known for suppressing the expression level of mRNAs. Hence, our method selected up and downregulated mRNAs and miRNAs that are used to make mRNA-miRNA pairs. For that, combination of upregulated mRNAs with downregulated miRNAs and vice versa is used.

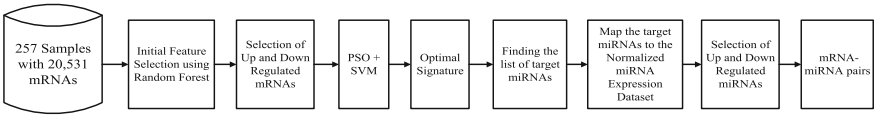


Fig. 1. A block diagram representation of the proposed workflow

3 Experimental Results

3.1 Datasets

The dataset comes from The Cancer Genome Atlas (TCGA).[1] It features mRNA and miRNA expression levels for kidney renal cell carcinoma (KIRC) of different tumor, and control samples. For KIRC, normalized expression mRNA and miRNA datasets consisting of 186 common tumors, and 71 common control samples are considered.

3.2 Results

Here, the problem is to identify mRNA-miRNA pairs for the investigated dataset. For this purpose, the expression profiles of 20,531 mRNAs over 186 kidney

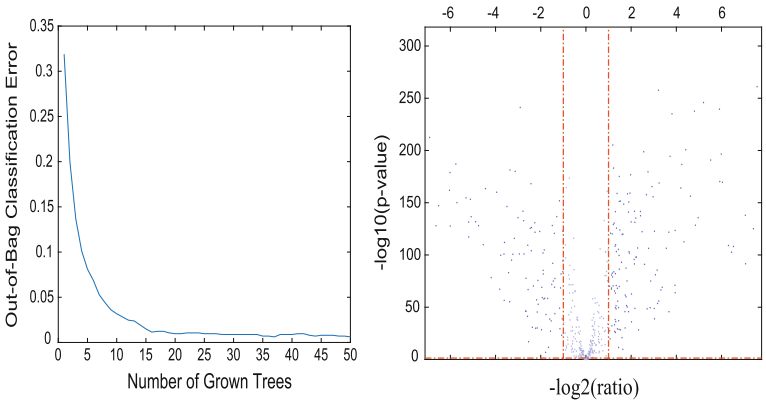
[1] https://tcga-data.nci.nih.gov/tcga/.

renal cell tumor samples and 71 control samples are considered. In order to find reduced set of features, random forest is used here. Because of its iterative approach, the *OOB* error is biased down and mRNAs with smallest *OOB* error are selected as the reduced set of features. This has been shown in Fig. 2a. By this process, 511 informative mRNAs are selected. In order to identify the up and downregulated mRNAs, significant changes in the expression profiles are measured. For this analysis, the volcano plot in Fig. 2b illustrates the differentially expressed mRNAs among the 511. By this process, 242 up and 191 downregulated mRNAs are identified. Now, these shortlisted mRNAs are process, using PSO+SVM. For this process, the parameters used in the proposed method are summarized in Table 1. Here, PSO+SVM executed 30 times (N_{exe}) on these differentially expressed mRNAs. According to the experimental results, PSO+SVM achieved 92.60% average classification accuracy for mRNA data. The average values of precision, sensitivity, specificity, F-measure and MCC are 92.57%, 92.58%, 92.62%, 92.58%, and 0.920%, respectively. For this process, 10 mRNAs are considered by PSO+SVM during each run. Moreover, as PSO+SVM was executed 30 times hence, in each run of the proposed method, a new set of differentially expressed mRNAs is identified. After the end of 30 executions, one list is prepared that includes all the appeared mRNAs. Thereafter, from that list most appeared mRNAs covering all the runs of the proposed method are identified. These mRNAs are considered to be optimal up and downregulated mRNAs associated with kidney renal cell cancer subtypes.

Thereafter, these selected mRNAs are used to find their corresponding miRNA targets and for this validation, miRDB [19] database is used. Moreover, these shortlisted target miRNAs find their expression profiles from the normalized miRNA expression dataset of 186 tumor and 71 control samples. A total of 450 miRNAs targets are identified from the miRGate database. Out of which, 417 miRNAs match their expression profile in the normalized miRNA expression dataset. Hence, the miRNAs for which expression profile information is known are only taken at this stage. Therefore, out of the 450 miRNAs in the list, 33 have been excluded because they did not have the expression profile information in our used normalized miRNA expression dataset. Thereafter, p-value and fold change analysis of the shortlisted 417 differentially expressed miRNAs, identified the up and downregulated one. By this process, 14 informative miRNAs are identified that include 12 downregulated and 2 upregulated miRNAs. Method selected downregulated miRNAs are *hsa-mir-200c, hsa-mir-181a-2, hsa-mir-196a-1, hsa-mir-183, hsa-mir-194-2, hsa-mir-196a-2, hsa-mir-10b, hsa-mir-138-1, hsa-mir-182, hsa-mir-192, hsa-'mir-135b, hsa-mir-199b* whereas *hsa-mir-141, hsa-mir-196b* are upregulated miRNAs.

Studies on miRNAs found that it can suppress the expression level of mRNAs. As miRNAs and mRNAs belonging to same sample (Tumor and Control) of KIRC dataset are used in this experiment. Hence, differentially expressed miRNAs and mRNAs of the same sample (Tumor and Control) can be paired. Moreover, upregulated miRNAs can be mapped with downregulated mRNAs and vice versa. In the process, many of the selected mRNAs have been excluded from the

list while miRNAs remain same. Final list of mRNA-miRNA pair includes 98 upregulated mRNA and 12 downregulated miRNA pair and 62 downregulated mRNA and 2 upregulated miRNA pair. Details of which are given in Table 2.



(a) *OOB* errors (b) Volcano plot

Fig. 2. a Variation of *OOB* errors with the number of grown trees in Random forest, **b** Volcano plot depicts 242 up and 191 downregulated mRNAs

Table 1. Parameters used in the experiment

Symbol	Value	Description	Symbol	Value	Description
N_{par}	50	Number of particles	φ_2	2	Social constant
N_{itr}	50	Number of iterations	ω	0.9	Inertia weight
L	10	Length of a particle	$\mathcal{C}$	0.01	SVM $\mathcal{C}$ constant
φ_1	2	Cognitive constant	N_{exe}	30	Number of executions

3.3 Biological Significance

Biological significance analysis of the mRNAs-miRNAs pairs has been carried out by KEGG pathway enrichment and gene ontology analysis. In this regards, miRSystem ver. 20160502 [11] and Enrichr [9] help in finding the associated pathways of the significant mRNA-miRNA pairs. For the gene enrichment analysis, the biological process (P), cellular component (C), and molecular function (F) hierarchies in gene ontology (GO) are considered. The most significant

Table 2. Method selected up and downregulated miRNA and mRNAs

Downregulated miRNA	Upregulated miRNA	Upregulated mRNA							Downregulated mRNA				
hsa-mir-200c	hsa-mir-141	ABCB7	BCL2L12	C9orf117	CHTF8	DUSP23	FGF2	HIST1H3E	B3GALT5	CHURC1	DNAH17	GEMIN7	ID4
hsa-mir-181a-2	hsa-mir-196b	ABCD4	BCL7C	CA11	COL4A6	DUT	FSTL5	HSPA12A	BSND	CIITA	DPYSL5	GIN1	IDH1
hsa-mir-196a-1		ACAD11	BICC1	CACNA2D2	CPD	EDEM1	FTL	IGSF3	CCNT2	CKAP2L	E2F3	GIPC1	
hsa-mir-183		ADCY2	BMPR1B	CALB2	CREG2	ENGASE	FZD3	IL17REL	CDC42	CLNK	ELL	GIT2	
hsa-mir-194-2		ADRA2C	BPGM	CCDC141	CRNKL1	EPB41L1	GABRE	IMMP2L	CDH10	CLUL1	ENOSF1	GLUD1	
hsa-mir-196a-2		AHCYL1	BRI3BP	CCDC152	CRTAM	EPPK1	GALNT6	IMPA1	CLNK	CNOT6L	FAM172A	GNAS	
hsa-mir-10b		ALCAM	BST1	CCDC85C	CRTAP	EXT1	GATM	INCENP	DNAH17	CNR1	FAM19A5	GNPTAB	
hsa-mir-138-1		ANKRD13A	BTBD2	CCDC88A	CSF2RB	FAIM2	GDF2	IQSEC1	DPYSL5	CPEB2	FBXO48	GPR37	
hsa-mir-182		ANKRD34B	BVES	CCND2	CYBRD1	FAM120C	GLT8D2		E2F3	CPEB3	FCGR3A	GRM5	
hsa-mir-192		ARGFX	C17orf107	CCNT1	CYLD	FAM126B	GPKOW		ELL	CWF19L1	FLRT2	GSTM2	
hsa-mir-135b		ARHGAP20	C19orf70	CDK10	CYP8B1	FAM189A1	GPR161		ENOSF1	CYP1A2	GAB2	HAGHL	
hsa-mir-199b		ARHGEF18	C4orf32	CGNL1	DIRAS1	FAM196A	GTPBP8		EPN1	DAZAP2	GAL3ST1	HBS1L	
		ARMC10	C4orf45	CHML	DNAJC14	FAM198B	H6PD		GIT2	DISC1	GBP2	HEY1	
		ATG2A	C5AR1	CHST5	DPP9	FBXL18	HACE1		HDDC3	DKK1	GDF11	HIVEP3	
		ATP2B2	C9orf114	CHST6	DRP2	FCHSD1	HEPH		CHST7	DLG5	GDF7	ID3	

Table 3. Most significant gene ontology terms concerning selected genes for biological process (P), cellular component (C), and molecular function (F) obtained through enrichment analysis via Enrichr [9]

	Term	P-value	Genes
GO Biological Process (P)	Cellular iron ion homeostasis (GO:0006879)	3.68e–05	HEPH, ABCB7, GDF2, CYBRD1, FTL
	N-acetylglucosamine metabolic process (GO:0006044)	5.16e–05	CHST6, CHST7, CHST5
	Sulfur compound metabolic process (GO:0006790)	9.36e–05	CHST6, CHST7, CHST5
	Positive regulation of cyclin-dependent protein serine (GO:0045737)	4.64e–04	CCNT2, CCND2, CCNT1
	Negative regulation of cytoplasmic translation (GO:2000766)	8.66e–04	CPEB3, CPEB2
GO Cellular Component (C)	Cyclin/CDK positive transcription elongation factor complex (GO:0008024)	1.60e–03	CCNT2, CCNT1
	Messenger ribonucleoprotein complex (GO:1990124)	1.60e–03	CPEB3, CPEB2
	Golgi membrane (GO:0000139)	1.66e–03	EXT1, CDC42, GALNT6, CHST6, CHST7, HACE1, GNPTAB, EDEM1, CHST5
	CCR4-NOT complex (GO:0030014)	6.59e–03	CNOT6L, CPEB3
	Neuron projection (GO:0043005)	1.36e–02	CALB2, CDC42, CPEB3, CPEB2

continued

Table 3. continued

	Term	P-value	Genes
GO Molecular Function (F)	N-acetylglucosamine 6-O-sulfotransferase activity (GO:0001517)	1.53e–05	CHST6, CHST7, CHST5
	7SK snRNA binding (GO:0097322)	8.66e–04	CCNT2, CCNT1
	mRNA 3'-UTR AU-rich region binding (GO:0035925)	2.55e–03	CPEB3, CPEB2
	Transforming growth factor beta receptor binding (GO:0005160)	2.87e–03	GDF11, GDF2, GDF7
	Translation repressor activity, nucleic acid binding (GO:0000900)	3.10e–03	CPEB3, CPEB2

GO terms, with lowest p-values (< 0.001) for the biological processes (P) associated with mRNA data are *cellular iron ion homeostasis (GO:0006879), N-acetylglucosamine metabolic process (GO:0006044), sulfur compound metabolic process (GO:0006790), positive regulation of cyclin-dependent protein serine (GO: 0045737), negative regulation of cytoplasmic translation (GO:2000766)*, etc. Significant top 5 gene ontology terms concerning selected genes for biological process (P), cellular component (C), and molecular function (F) are reported in Table 3. This table also includes a list of associated genes corresponding to each ontology term. Association of the obtained mRNAs with different biological pathways has been identified by KEGG pathway enrichment analysis. It has been observed that significant mRNAs are associated with many KEGG pathways. Among them, only 10 significant pathways with lowest p-values are presented in Table 4. For this analysis, 98 upregulated mRNA and 62 downregulated mRNA are considered. Overall, the results signify the importance of selected mRNA-miRNA pairs.

Table 4. Top 10 KEGG pathways associated with the selected up and downregulated mRNAs

Term	P-value	Genes
TGF-beta signaling pathway: hsa04350	0.0040	ID4, ID3, BMPR1B, GDF7
Signaling pathways regulating pluripotency of stem cells: hsa04550	0.0049	FZD3, ID4, ID3, BMPR1B, FGF2
Salivary secretion: hsa04970	0.0050	BST1, GNAS, ADCY2, ATP2B2
Pancreatic secretion: hsa04972	0.0065	BST1, GNAS, ADCY2, ATP2B2
Mineral absorption: hsa04978	0.0071	HEPH, CYBRD1, FTL
Retrograde endocannabinoid signaling: hsa04723	0.0077	GRM5, CNR1, ADCY2, GABRE
Dorsoventral axis formation: hsa04320	0.0182	CPEB3, CPEB2
Rap1 signaling pathway: hsa04015	0.0238	CDC42, CNR1, GNAS, ADCY2, FGF2
Phospholipase D signaling pathway: hsa04072	0.0253	GRM5, GNAS, ADCY2, GAB2
Adrenergic signaling in cardiomyocytes: hsa04261	0.0277	GNAS, CACNA2D2, ADCY2, ATP2B2

4 Conclusion

In this paper, the integrated analysis of mRNA and miRNA expression data identified a number of mRNAs and miRNAs that are differentially expressed in control and tumor samples. For this analysis, random forests is used as initial feature selector. Thereafter, with the help of particle swarm optimization and support vector machine classifier, the proposed method is optimized to have best sets of mRNAs. In addition to that, functional miRNA targets of selected mRNAs are also identified for the expression data of Kidney renal cell. Finally, differentially expressed mRNA-miRNA pairs are prepared. These selected mRNA-miRNA pairs are proven to be relevant according to gene ontology and pathway analysis tools. Hence, these mRNA-miRNA pairs are the potential biomarkers for the kidney renal cell cancer type. In conclusion, this analysis is proven to be helpful for the joint mRNA-miRNA biomarker identification and could also be used as miRNA marker [3,16] and gene selection [2].

References

1. Anders, S., Huber, W.: Differential expression analysis for sequence count data. Genome Biol. **11**(10) R106 (2010)
2. Bhowmick, S.S., Saha, I., Maulik, U., Bhattacharjee, D.: Biomarker identification using next generation sequencing data of RNA. In: 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 299–303 (2016)
3. Bhowmick, S.S., Saha, I., Maulik, U., Bhattacharjee, D.: Identification of miRNA signature using next-generation sequencing data of prostate cancer. In: Proceedings of the 3rd International Conference on Recent Advances in Information Technology, pp. 528–533 (2016)
4. Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Proceedings of the 5th Annual Workshop on Computational Learning Theory pp. 144–152 (1992)
5. Breiman L (2001) Random forests. Mach. Learn. 45(1):5–32
6. Grada A, Weinbrecht K (2013) Next-generation sequencing: methodology and application. J. Investig. Dermatol. 133(8):e11
7. Guyon I, Weston J, Barnhill S, Vapnik V (2002) Gene selection for cancer classification using support vector machines. Mach. Learn. 46(1–3):389–422
8. Kennedy J., Eberhart R.: Particle swarm optimization. In: Proceedings of IEEE International Conference on Neural Networks, vol. 4, pp. 1942–1948 (1995)
9. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al.: Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. p. gkw377 (2016)
10. Leidinger P, Keller A, Borries A, Reichrath J, Rass K, Jager SU, Lenhof HP, Meese E (2010) High-throughput miRNA profiling of human melanoma blood samples. BMC cancer 10(1):1
11. Lu, T.P., Lee, C.Y., Tsai, M.H., Chiu, Y.C., Hsiao, C.K., Lai, L.C., Chuang, E.Y.: miRSystem: an integrated system for characterizing enriched functions and pathways of microRNA targets. PloS one **7**(8), e42390 (2012)
12. Lu X, Gamst A, Xu R (2010) RDCurve: a nonparametric method to evaluate the stability of ranking procedures. IEEE/ACM Trans. Comput. Biol. Bioinform. 7(4):719–726
13. Peng H, Long F, Ding C (2005) Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. 27(8):1226–1238
14. Ray, S.S., Ganivada, A., Pal, S.K.: A granular self-organizing map for clustering and gene selection in microarray data (2015)
15. Robinson MD, McCarthy DJ, Smyth GK (2010) EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139–140
16. Saha, I., Bhowmick, S.S., Geraci, F., Pellegrini, M., Bhattacharjee, D., Maulik, U., Plewczynski, D.: Analysis of next-generation sequencing data of miRNA for the prediction of breast cancer. In: International Conference on Swarm, Evolutionary, and Memetic Computing, pp. 116–127 (2015)
17. Sehhati M, Mehridehnavi A, Rabbani H, Pourhossein M (2015) Stable gene signature selection for prediction of breast cancer recurrence using joint mutual information. IEEE/ACM Trans. Comput. Biol. Bioinfom. 12(6):1440–1448

18. Sharma A, Imoto S, Miyano S (2012) A top-r feature selection algorithm for microarray gene expression data. IEEE/ACM Trans. Comput. Biol. Bioinform. (TCBB) 9(3):754–764
19. Wong, N., Wang, X.: miRDB: an online resource for microRNA target prediction and functional annotations. Nucleic Acids Res. p. gku1104 (2014)
20. Yu D, Huber W, Vitek O (2013) Shrinkage estimation of dispersion in negative binomial models for RNA-seq experiments with small sample size. Bioinformatics 29(10):1275–1282
21. Zhuang, X., Li, Z., Lin, H., Gu, L., Lin, Q., Lu, Z., Tzeng, C.M.: Integrated miRNA and mRNA expression profiling to identify mRNA targets of dysregulated miRNAs in non-obstructive azoospermia. Sci. Rep. **5** (2015)

Investigations on Failure and Reliability Aspects of Service Oriented Computing Based on Different Deployment Techniques

Abhijit Bora(✉), Subhash Medhi, and Tulshi Bezboruah

Department of Electronics & Communication Technology, Gauhati University, Guwahati, India
{abhijit.bora0099, medhisubhash72}@gmail.com, zbt_gu@yahoo.co.in

1 Introduction

Development of enterprise system or invocation of remote method over network in service oriented architecture (SOA) paradigm provides a flexibility scope to the software practitioner, researchers as well as users. It supports atomic or aggregation of multiple services to perform a specific job or consolidated task that represents the overall process of business logic (BL) [1, 2]. Among the grooming techniques for execution of service oriented computing (SOC), web service (WS) is gaining comparatively more importance. Vidura et al. and Tang et al. had defined it as "software as service that supports interoperable execution of platform with de facto standard of interaction for distributed computational system" [3, 4]. Cunningham et al. stated that "Due to their decentralized, loosely coupled and highly interoperable architecture, multitier web applications have become the primary solution for business integration and collaboration problems such as e-commerce web sites" [5]. The widespread imposing of such system for critical service execution, the study of failure during massive requests is necessary for overall communication in between client and service tiers. Reliability on the other hand is the probability that the service will respond properly within the exposure period of its execution. While dealing with reliability, the focal point is on failure nature of the service. In our previous work, we had presented some aspects of performance that is usually associated with such a system [6–8]. In this paper, a novel reliability evaluation framework is introduced. It presents a thorough analysis considering the massive load of end users over the service execution.

2 Related Work

Reliability evaluation of WS has been carried out by many researchers for years. Their studies solely focus on model representation and implementation. However, the authors had presented performance results of such a system but there is a lack of study for the reliability of a multi service multi functional SOC for heavy load of consumers [8–10]. In the year 2004, Tian et al. had presented an approach for evaluating the reliability of

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_60

web-based software by using recorded web server log records [11]. The approach was for two existing websites. They established that by monitoring and analyzing the log records, defects can be extracted which in turn can be fixed to gain the quality of the web-based software. In 2007, Murlewski et al. had discussed an approach for testing the web-based system by using JUnit framework [12]. They also had stated that proper software testing and analyzing the performance metrics can enhance the quality aspects of such system. However, the study lacks reliability evaluation of such system. The theoretical analysis of the experimental arrangement is also omitted. In 2011, Kalita et al. had explained a methodology for implementation and comparative performance investigation of a web-based application using Microsoft .Net and Java technique separately [13]. They carried out a feasibility study of such a work for stability, scalability, and efficiency by load and stress testing. However, that study also lacks of reliability evaluation of such a system. In 2013, Janevski et al. had presented a framework for modeling the reliability of the web-based system for heavy-tailed workload from the perspective of user and service view [14]. The emphasis was given on handling loads in terms of consumer session that influences the reliability in serving incoming web request. In 2014, Oliveira et al. had presented a system for WS composition and evaluated the QoS metrics during run time [15]. The study had focused on fault tolerance technique in SOA based paradigm to enhance the availability and quality aspects of service. In 2017, Medhi et al. had presented a study on deployment of an e-ATM WS in .NET framework with different experimental constraints [16]. The emphasis was given on the study of some key metrics that imputes the quality of such implementation approach. However, the reliability evaluation of such deployment against massive consumers was omitted. The present work is novel from the previous work as emphasize is given on the execution of multi function multi-WS instead of single function single WS. The study is incorporated with multi-WS as respective business organizations utilize multiple WS that execute together for functional delivery to end users.

3 Key Objectives and Methodology

The objective of the proposed work is to study an insight about the failure and reliability aspects associated with the SOC-based system that can be expected when deployed in Apache Tomcat web server in one hand and Internet Information Service (IIS) on the other. The SOC-based system is developed considering the features of Java Application Programming Interface for Extensible Markup Language based WS (JAX-WS) and Microsoft Web service Communication Foundation (WCF) technology for necessary computational implementation. The JAX-WS and Microsoft WCF are deployed due to its popularity and adaptability nature over web. The reliability is estimated for different stresses of simultaneous users that can invoke the WS together for a specific time period. A prototype multiservice-based SOC is developed for the proposed study. It is composed of client, parent, and child role with simple object access protocol (SOAP) based WS. The execution of SOAP request and response in between WS communication is implemented in hierarchical manner. A business logic

(BL) method for addition of two numbers is implemented in the child WS. We have not developed and implemented any SQL programs, as in massive level of stress, execution of SQL in server side can generate extra overhead in the SOC based system, for which server-side resources utilization increases and additional overhead of system failure can generate [7, 8, 13]. A virtual user (VU) test script is created by using Mercury LoadRunner that can access the SOC-based system [17]. It passes two parameters and displays the result of SOC-based system. The SOC-based system deployed either with Apache Tomcat on one hand or IIS web server on the other hand is tested for different massive levels of VUs. During the test, we monitor the number of hyper text transfer protocol (HTTP) request made by the specific stress of VU towards SOC-based system. Out of the total request made, the number of passed and failed HTTP request processing in server side is also recorded. The experimental arrangement of the system is shown in Fig. 1.

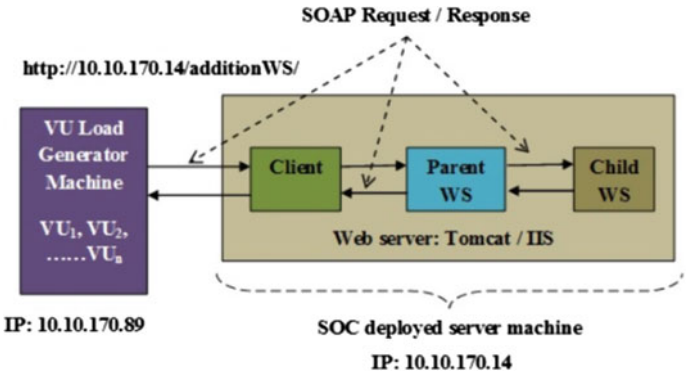


Fig. 1. Experimental arrangement for reliability evaluation of SOC

3.1 Algorithm and Time Complexity of BL Method

The following algorithm highlights the key steps for execution and development of Client WS:

Global WS name W_{Name}; web method W_{Method}; operation name O_{Name}; Parameter Y_1, Y_2, R; relation RL; integer Counter; Class W_{Class}
function ExecuteChildWS (X_1, X_2)

1. W_{Name}: = WS_Child
2. W_{Class}: = WS_Child
3. Y_1: = X_1, Y_2: = X_2, R: = Null
4. {W_{Name}, O_{Name}}: = MethodChildWS
5. R: = call **MethodChildWS** (Y_1, Y_2)
6. return R
7. Stop

function MethodChildWS (Y_1, Y_2)
1. R: = Y_1 + Y_2
2. Return R

From Step 1–4, it will initialize the necessary WS name, class file, supplying parameters along with the web method and operation name. The time complexity (Cx) of the algorithm is evaluated. The Cx is observed to be O(1), as it will be executed for once during the system development. Steps 5 and 6 invoke the function MethodChildWS and return the response to client. The function MethodChildWS performs necessary BL operation. For Step 5–6, the Cx is observed to be O(n), as their execution depends upon the number of call from client WS. Hence, the overall algorithm executes at Cx: = O(1) + O(n). Here it is observed that Cxn = O(n). As such, the overall functional execution is linear time complexity. The input of the algorithms is the number of supplied parameters. With increase in massive request, the number of supplying parameter increases. As such, overall functional time execution increases.

4 Testing and Analysis

The SOC with IIS and Tomcat web server are tested separately for the consecutive VU stress level of 50, 100, 500, 800, 1000, and 1200. The HTTP transaction pass and fail record is monitored for each service. Different responses for each stress level are given in Table 1. Here, the failure rate is observed at a stress level of 1200 VU for both the service. To better understand the failure record, the statistical analysis is carried out by considering the data sample for each service. The data sample is collected by considering a 30 repetitive test for 1200 consecutive VU. The data sample is arranged for both SOC with IIS and Tomcat web server, respectively. The correlation of observed failure rate and detected fault count of SOC in IIS and Tomcat web server is observed. It is shown in Fig. 2 and Fig. 3, respectively. It validates that the fault count and fault rate of the data sample are linear and strongly correlated. The correlation is about 90.4% and 93% for SOC in IIS and Tomcat web server, respectively. We observe the frequency ranges through histogram as shown in Fig. 4 and Fig. 5, respectively for

Table 1. HTTP transaction history for different VU levels (FR: failure rate)

VU	WCF with IIS				JAX-WS with Tomcat			
	Pass	Fail	Total	FR (%)	Pass	Fail	Total	FR (%)
50	1613	0	1613	0	1669	0	1669	0
100	2259	0	2259	0	6774	0	6774	0
500	11,604	0	11,604	0	22,400	0	22,400	0
800	34,569	0	34,569	0	67,120	0	67,120	0
1000	41,056	0	41,056	0	82,436	0	82,436	0
1200	79,080	15,293	94,373	16	193,662	91,806	259,025	32

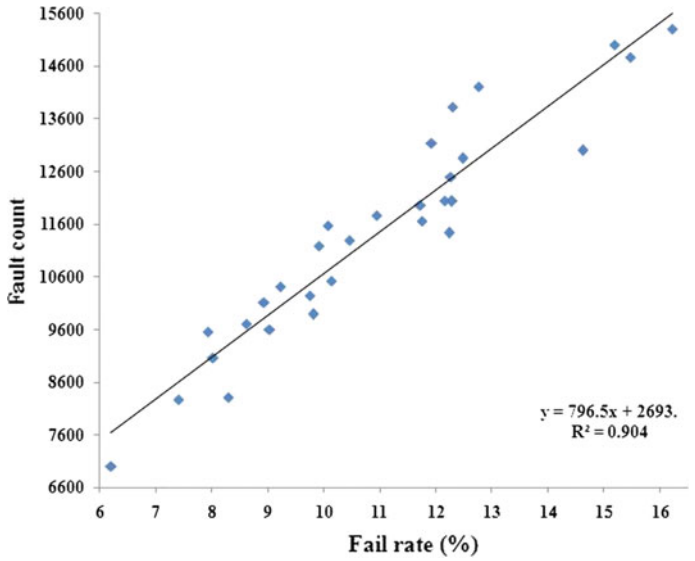


Fig. 2. The correlation of failure rate and fault count against IIS

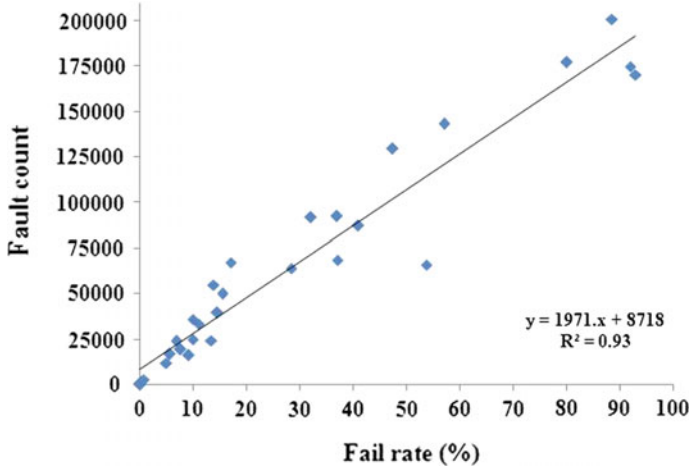


Fig. 3. The correlation of failure rate and fault count against Tomcat

each service. The highest fault count for SOC with IIS web server lies between 11,150 and 13221.5. However, it is in between 473 and 50,623 for Tomcat web server.

The histogram for SOC with IIS web server is left skewed. However, it is right skewed for SOC with Tomcat web server. As such, the fitted distribution of fault count is observed to be Weibull for both the services. To estimate the shape (SH) and scale parameter (SC) of the Weibully distributed data sample, we use EasyFit version 5.6 [18]. The SH and SC is observed to be 6.19 and 12,099 for SOC in IIS web server. In

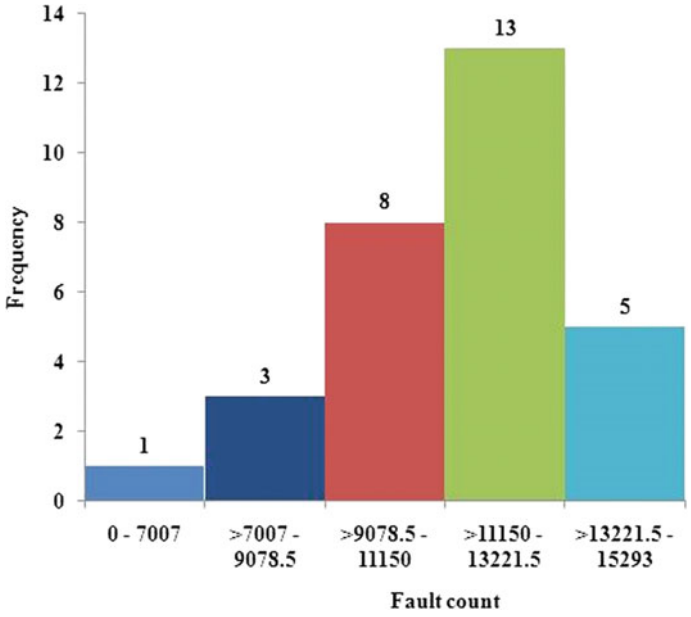


Fig. 4. Histogram of observed HTTP request failure against IIS

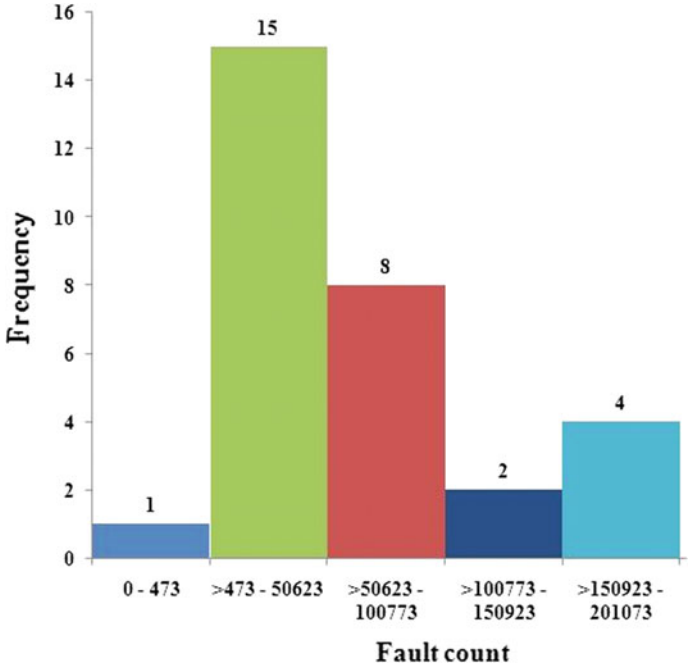


Fig. 5. Histogram of observed HTTP request failure against Tomcat

case of SOC in Tomcat web server, it is 0.69 and 63,879. The Weibull CDF of each fault count (Fc) is calculated by using Eq. 1 as defined elsewhere in [19].

$$CDF = 1 - \exp\{(-Fc/SC)SH\} \tag{1}$$

4.1 Reliability Aspects

The Weibull distribution reliability metric (R) is evaluated for each service by using Eq. 2 [20].

$$R = 1 - CDF \tag{2}$$

Equation 2 is evaluated as 1-(Mean of CDF). For IIS and Tomcat web server, the mean CDF of fault count is evaluated to be 0.5 and 0.52, respectively. As such, the reliability of SOC with IIS web server is observed to be 0.5. That means the SOC hosted with IIS web server will respond about 50% of its service period against a consecutive stress of 1200 VU in one day. However, in case of SOC with Tomcat web server, the reliability is evaluated to be 48%.

5 Results and Discussion

The interpretation of histogram shows that the population of Fc follows Weibull distribution. The correlation of failure rate and Fc for both the service is observed to be strong in the data sample. However, the efficiency of handling massive HTTP request for SOC with IIS web server is comparatively less than the service with Tomcat web server. For a stress level of 1200 VU, the failure rate is 16% and 32% respectively for SOC with IIS and Tomcat web server. However, failure rate is 0 below 1200 VU in both the cases. As such, strong reliability is observed below 1200 VU. However, at 1200 VU stress level, the reliability of SOC with IIS web server is observed to be better than its counterpart. It is observed that the SOC with Tomcat web server has processed 193,662 HTTP request out of 259,025. On the contrary, the SOC with IIS web server has processed 79,080 HTTP request out of 94,373. As such, even with lower reliability, the efficiency of SOC with Tomcat web server for handling massive request is better than its counterpart.

6 Conclusion

This paper presents a reliability evaluation framework with experimental arrangement and measurement for SOC with IIS and Tomcat web server. The reliability evaluation framework is observed to be effective and applicable. The SOC with Tomcat web server along with JAX-WS can handle more HTTP request in comparison to its counterpart SOC with IIS web server along with WCF technology. In both the cases,

the reliability of the service is observed to be strong up to a stress of 1000 VU. However, for a stress of 1200 VU, although the service with Tomcat web server is handling more HTTP request, the reliability of the service with IIS web server is observed to be better than its counterpart with Tomcat web server. It is observed that the overall efficiency and capacity of Tomcat web server along with JAX-WS for processing massive request are better than its counterpart. However, both services can be used for execution of BL through SOC. The overall assessment highlights some insights for deploying such services. However, such reliability trade-off should be compared with other methodology so as to identify the factors responsible for degrading the reliability.

References

1. Weerawarana S, Curbera F, Leymann F, Storey T, Ferguson D (2005) Web services platform architecture: SOAP, WSDL, WS-Policy, WS-addressing, WS-BPEL, WS-reliable messaging, and more. Prentice-Hall, Englewood Cliffs
2. Paolo B, Andrea DA (2011) A model-driven method for describing and predicting the reliability of composite services. Softw Syst Model 10(2):265–280
3. Vidur G, Abhay Z, Tari PB, Building Web services middleware with predictable execution times. World Wide Web, 15(5):685–744 (2012)
4. Tang D, Kumar D, Duvur S, Torbjornsen O (2004) Availability measurement and modeling for an application server. In Proceedings of international conference on dependable systems and networks, pp 669–678
5. Cunningham D, Anderson J, Medairy B (2004) Network-centric architecture to enable secure communications and discovery, In Proceedings of IEEE aerospace conference
6. Bezboruah T, Bora A (2015) Performance evaluation of hierarchical soap based web service in load balancing cluster-based and non-cluster-based web server. Int J Inf Retr Res 5(4): 20–31
7. Bora A, Bezboruah T (2015) Some aspects of QoS for interoperability of multi service multi functional service oriented computing, In: The Proceedings of IEEE international conference on research in computational intelligence and communication networks (ICRCICN), pp 363–368
8. Medhi S, Bora A, Bezboruah T (2017) Investigations on Evaluation of Some QoS Aspects of Service Oriented Computing System Based on Web Services. Sens Transducers J 209(2): 56–64
9. Saddik AE (2006) Performance measurement of Web Service based application. IEEE Trans Instrum Measur, 55(5):1599–1605
10. Gomez-Martinez E, Gonzalez-Cabero R, Merseguer J (2014) Performance assessment of an architecture with adaptative interfaces for people with special needs. J Empirical Softw Eng 19(6):1967–2018
11. Tian J, Rudraraju S, Li Z (2004) Evaluating Web software reliability based on workload and failure data extracted from server logs. IEEE Trans Softw Eng 30(11):754–769
12. Murlewski J, Wojciechowski J, Sakowicz B, Napieralski A (2007) Comprehensive approach to web applications testing and performance analysis. In: The proceedings of 9th International conference on CAD systems, pp 429–431
13. Kalita M, Bezboruah T (2011) Investigations on implementation of web applications with different techniques. IET Softw 6(6):474–478

14. Janevski N, Goseva-Popstojanova K (2013) Session reliability of web systems under heavy-tailed workloads: an approach based on design and analysis of experiments. IEEE Trans Softw Eng 39(8):1157–1178
15. Oliveira EM, Estrella JC, Kuehne BT, Filho DML, Adami LJ, Nunes LH, Nakamura LH, Libardi RM, Souza PSL, Reiff-Marganiec S (2014) Design and implementation of fault tolerance techniques to improve QoS in SOA. In: 10th International Conference on Network and Service Management (CNSM), pp 37–45
16. Medhi S, Bora A, Bezboruah T (2017) Investigations on some aspects of reliability of content based routing SOAP based windows communication foundation services. Int J Inf Retr Res 7(1):17–31
17. Mercury Load Runner tutorial. Mercury roadrunner quick start. https://qageek.files.wordpress.com/2007/05/loadrunner_tutorial.pdf. Accessed 22 Jan 2016
18. Easy fit distribution. http://www.mathwave.com/easyfit-distribution-fitting.html. 21 Sept 2016
19. Chandra M, Singpurwalla ND, Stephen MA (1981) Kolmogorov statistics for tests of fit for the extreme value and weibull distribution. J Am Stat Assoc 76(375):729–731
20. NIST/SEMATECH e-Handbook of statistical methods. www.itl.nist.gov/div898/handbook/apr/section1/apr162.htm. Accessed 29 Jan 2016

Data Fusion by Truncation in Wireless Sensor Network

Shrijana Pradhan(✉), Eshita Sinha, and Kalpana Sharma

Department of Computer Science and Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Manipal, Sikkim, India
Sriz_p@yahoo.com

1 Introduction

1.1 Wireless Sensor Network

Wireless sensor network (WSN) is a network of sensors [2], deployed in an area of interest for sensing various measurable parameters from the environment, upon analyzing which pivotal quality environmental decisions can be initiated. It is a rapidly emerging technology and is used in several applications, such as environment monitoring, military, medical systems, and many more [3]. The sensor nodes present in WSN are energy and resource constrained and have several other limitations, such as storage capacity, limited processing speed, and less communication bandwidth.

The sensor nodes organize themselves into different types of network infrastructure and collect information of their interest [4]. For efficient energy utilization, the sensed data in the network must be considerably reduced without compromising much with data integrity.

Thus, in WSN, there is a need of data fusion or data reduction technique which can reduce the amount of data sent to the base station, which can reduce the energy-related burden on the network and can increase the network life time.

1.2 Data Fusion in Wireless Sensor Network

Data fusion in WSN aims at eliminating the overhead of transmitting the redundant data, leading to reduction in transmission energy requirement which would consequently increase the life time of the sensor nodes [5]. Before sending the sensed data to the base station, the node reduces the volumetric burden of the data. This technique is very crucial in the WSN as it enables nodes to send fused data directly to the base station.

A WSN has nodes with variable energy level. The energy level of a node may be taken into consideration while assigning various responsibilities. Nodes with higher energy level should be selected as candidate for the role of cluster head and should perform data processing and onward communication whereas nodes with lower energy level should be engaged in sensing parameters from the proximity [6]. This would greatly increase the lifetime of the network through effective energy management. Further, an effective fusion based data aggregation would minimize energy

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_61

consumption by reducing the communication load between the cluster head and the BS. The overall performance of a hierarchical architecture can be improvised by implementing such strategies [7].

2 Literature Survey

Izad et al. [8] introduced a fuzzy-based data fusion approach for WSN with the aim of increasing the QoS while reducing the energy consumption of the sensor network. Further, only true value is distinguished and aggregated, thereby reducing the burden of processing the entire data at the sink. The proposed technique is also able to eliminate redundant data and consequently reduce energy consumption thus increasing the network lifetime.

Jayasri et al. [5] proposed the need for energy efficient data fusion. They have presented the overview of current challenges in the WSN, where it has been found that approximately 70% energy is consumed by data transmission; hence data transmission should be optimized in WSN for maximizing the lifetime of network. Data transmission can be optimized only with the help of effective protocol and effective ways of data fusion (aggregation). Various techniques of data fusion along with their advantages and disadvantages have been elaborately discussed.

Maulik et al. [3] discussed several data reduction methods to save energy in WSN. It is articulated that the data reduction helps in removing unnecessary data while transmission, which helps in saving energy and increasing the network life time. There are various data-driven approaches for the same, which includes data reduction by data compression, data prediction, and in-network processing. Other methods include adaptive filter based, tree based, cluster head based, data stream based, and hybrid data reduction.

Mashere et al. [9] presented a data reduction technique for WSN, which uses throttling techniques to reduce the data transmitted to the base station. This is done to minimize the energy consumption of the network. The scheme used is controlled duty cycle, capable of increasing the routing fairness, which in turn increases the network lifetime. The data reduction algorithms used are threshold level sampling and adaptive level sampling.

Zhang et al. [10] took into account the power consumption of WSN and reliability of the transmitted data and proposed a data fusion algorithm. The algorithm is based on the time-driven network data aggregation along with the scheduling and batch estimation of the sensor nodes. The proposed algorithm deals with the sensor nodes state, where some of them are put to sleep mode, while others are scheduled in cluster. The data fusion mechanism is achieved by the batch estimation.

3 Proposed Solution Strategy

The proposed data fusion technique aims at solving the problem at cluster head level, where cluster heads perform aggregation on the received raw data and send it to the base station. Performing aggregation function such as Min, Max, Sum, Average, Count, and Standard Deviation over cluster head causes significant energy wastage.

Considering the energy efficiency of the routing protocol, the overhead on the aggregators and the cluster heads will be reduced with the proposed technique. The sensed information from the sensor nodes will be transmitted to the aggregators initially by truncating the lower nibble of the data. The cluster head will plunge the higher nibble of last data into the lower nibble of the incoming data. Thus, this will reduce the size of the transmitted data in the network and in turn will also reduce the data traffic in the deployment area (Fig. 1).

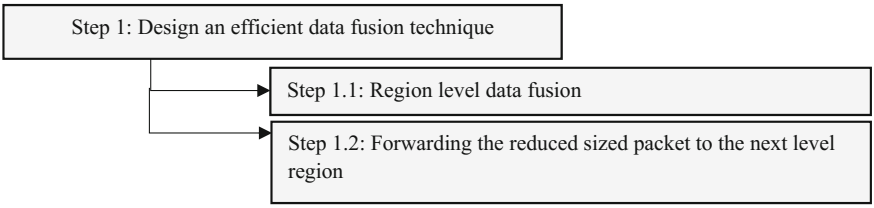


Fig. 1. Flow diagram

Proposed Algorithm:

i. Let total number of nodes be N, where the number of cluster heads elected in the entire network is 3.

ii. The remaining nodes are N−3 (Assuming all the sensor nodes need to send some data). Suppose, each cluster head contains equal number of sensor nodes as its members. Therefore, number of cluster members of each cluster head is (N−3)/3.

iii. Suppose, each cluster head contains equal number of sensor nodes as its members. Therefore, number of cluster members of each cluster head is (N−3)/3 (Fig. 2).

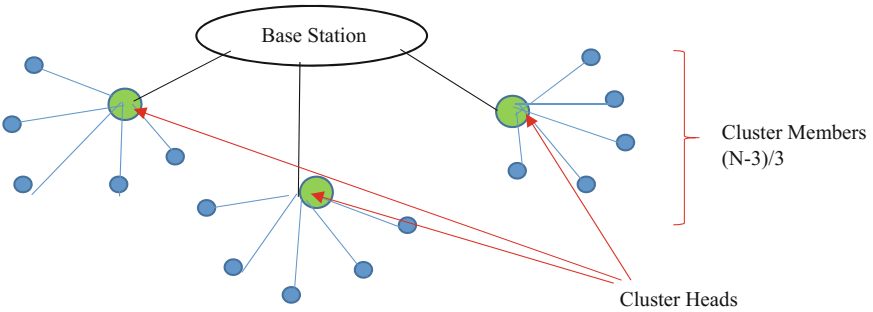


Fig. 2. Data transmission from cluster members to cluster heads

iv. All (N−3)/3 sensor nodes will send their data to their respective cluster head, where proposed data fusion will take place.

v. If (N−3)/3 is even, the number of data left with the cluster head after data fusion will be the half of the actual data, i.e. (N−3)/6.

vi. If (N−3)/3 is odd, then $\frac{(N-3)}{3} - 1$ data will become half. However, the (N−3)/3rd data will remain as it is and will be sent to the base station will its full byte data (Fig. 3).

vii. For example, the data to be transmitted from a sensor node to the aggregator is the number 50.

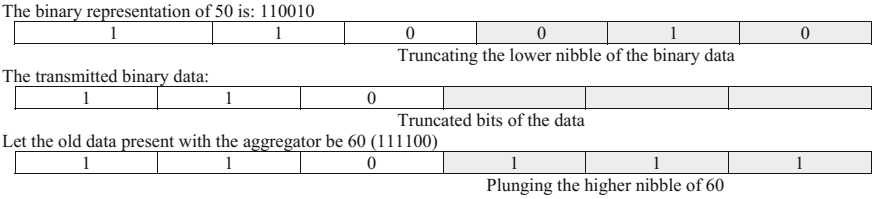


Fig. 3. Working of the data fusion technique

4 Inferences

The simulation of the proposed data fusion technique has been carried out by the Castalia (Version 3.2) framework of Omnet++ IDE (Version 4) (Table 1).

The simulation was done taking into consideration 50, 100, 150, and 200 nodes and various parameters were assessed (Table 2 and Fig. 4).

Table 1. Simulation parameter

Parameter	Value
Network area	100 m × 100 m
No. of nodes	50, 100, 150, 200
Initial energy of the nodes	0.5 J
ETX	50 nJ
ERX	50 nJ
Epsilon fs	10
Epsilon mp	0.0013

Table 2. Consumed energy

Number of nodes	Energy consumed (in Joule) LEACH without DFT	Energy consumed (in Joule) LEACH with DFT
50	55.7	55.5
100	57.4	56.5
150	53.9	53.1
20	58.2	54.8

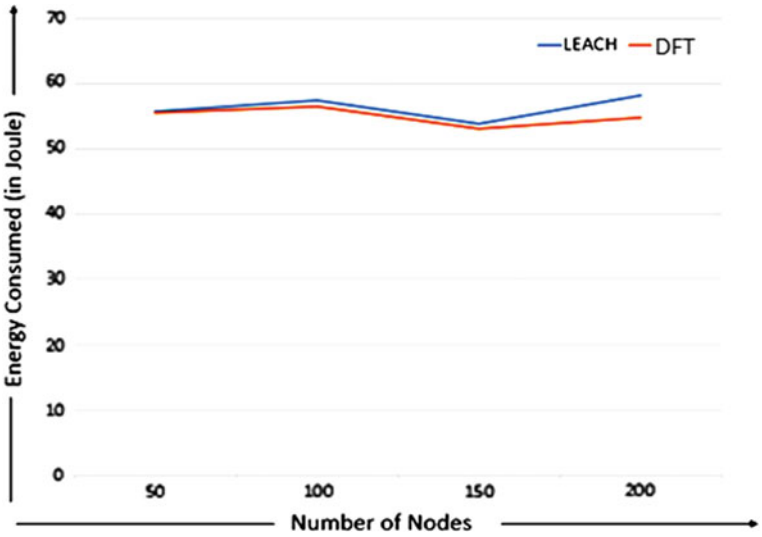


Fig. 4. Consumed energy assessment for data fusion

It can be inferred from the above that the energy consumption for the data transmission in the proposed data fusion technique is relatively lower than compared LEACH without the use of the proposed data fusion.

The second parameter assessed was network layer packet size (Tables 3 and 4, Figs. 5, and 6).

Table 3. Network layer packet size

Number of nodes	Network layer packet size (in bytes) LEACH without DFT	Network layer packet size (in bytes) LEACH with DFT
50	149	147
100	75	75
150	107	103
200	18	10

Table 4. Latency

Number of nodes	Latency (in ms) LEACH without DFT	Latency (in ms) LEACH with DFT
50	21	98
100	49	101
150	5	6.5
200	65	69.5

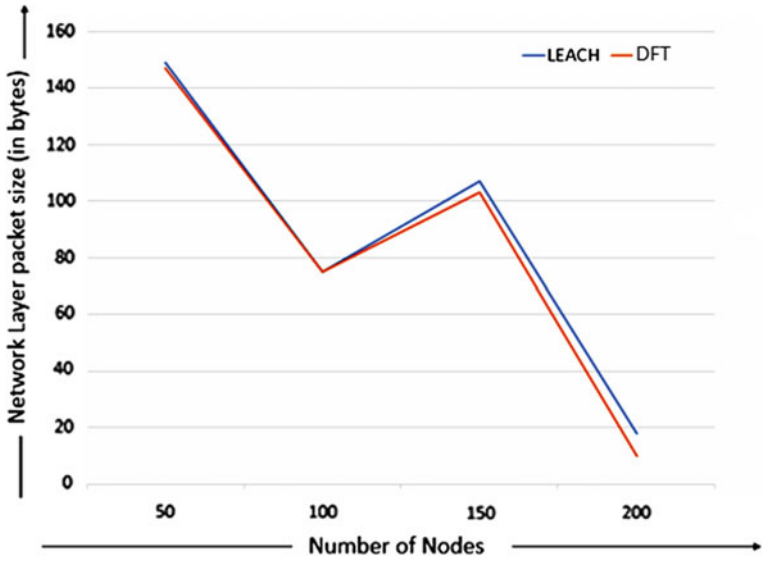


Fig. 5. Network layer packet size

It can be inferred from the above that there is a comparative increase in the application level latency due to the extra time taken to fuse the data.

Two Sample t-Test:

A t-Test with two samples is used for testing the difference between the samples when the variances of two normal distributions are not known. It is appropriate to run t-Test when sample size <30 and the samples are independent to each other (Table 5).

Consider a null hypothesis (H_o) = the proposed data fusion technique with LEACH is not better than LEACH without data fusion technique.

Since the performance of the proposed technique (LEACH with DFT) needs to be compared with respect to LEACH without DFT, the statistics values are not calculated. Thus, specified as NULL. Here, the t-statistics value is less than the t-critical value, the null hypothesis (H_o) is REJECTED. Hence, the performance of the proposed data fusion technique with LEACH is better than LEACH without the data fusion technique.

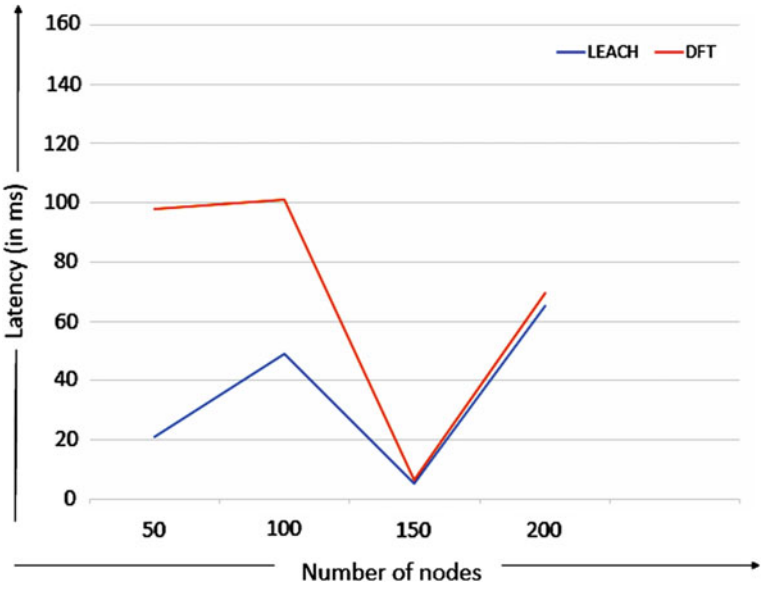


Fig. 6. Latency (in ms)

Table 5. t-Test for network layer packet size

	LEACH with DFT	LEACH without DFT
Mean	83.75	87.25
Variance	3295.58	3049.58
Observations	4	4
Hypothesized mean difference	0	NULL
Df	6	NULL
t Stat	−0.09	NULL
P(T $\leq$ t) one-tail	0.47	NULL
t-Critical one-tail	1.94	NULL

5 Conclusion

The ability of WSN depends on how effectively and efficiently the information is conveyed to the base station from the cluster heads with an eye on efficient processing, preserving data integrity as well as reducing transmission load. As energy requirements are mainly attributed by these prominent factors. Research initiatives may be directed toward devising more able techniques in these three aspects. This research initiative puts forward an effective data fusion technique to reduce the transmission load in the entire network. Comparative assessment concluded that the proposed technique stands tall. The ability of the proposed technique was also supported by the findings of t-Test. However, increase in the application level latency was observed because of data fusion.

6 Future Scope

The proposed work can be further extended to a better data fusion technique capable of reducing the application level latency to some extent.

References

1. Patil NS, Patil PR (2010) Data aggregation in wireless sensor network. In: IEEE international conference on computational intelligence and computing research
2. Shen C, Srisathapornphat C, Jaikaeo C (2001) Sensor information networking architecture and applications. IEEE Wirel Commun Mag 8:52–59
3. Thaker MB, Nagaraj U, Ganjewar PD (2014) Data reduction techniques in wireless sensor network: a survey 2(11)
4. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) Wireless sensor networks: a survey. Comput Netw 393–422
5. Jayasri BS, Rao GR (2014) Need for energy efficient data fusion in wireless sensor network 3(1)
6. Villalba LJG, Orozco ALS, Cabrera AT, Abbas CJB (2007) Routing protocols in wireless sensor networks. In: IEEE transactions on parallel and distributed systems, pp 919–931
7. Pal S, Bhattacharyya D, Tomar GS, Kim T (2010) Wireless sensor networks and its routing protocols: a comparative study. In: Proceedings of international conference on computational intelligence and communication networks (CICN), pp 314–319
8. Izadi D, Abawajy JH, Ghanavati S, Herawan T (2015) A data fusion method in wireless sensor network
9. Mashere MP, Barve SS, Ganjewar PD (2016) Data reduction in wireless sensor network for energy minimization. Int J Sci Eng Technol Res (IJSETR) 5(6). ISSN:2278–7798
10. Zhang K, Li C, Zhang W (2013) Wireless sensor data fusion algorithm based on the sensor scheduling and batch estimate. Int J Future Comput Commun 2(4)

Optimized Coordinated Economic Load Dispatch and Automatic Generation Control for an Interlinked Power System

Shimpy Ralhan[1(✉)], Neha Gupta[1], Shashwati Ray[2], and Mahesh Singh[1]

[1] Department of Electrical and Electronics Engineering, Shri Shankaracharya Technical Campus, SSGI, Bhilai, Chhattisgarh, India
{shimpys, nehagupta25061992, singhs004}@gmail.com
[2] Department of Electrical Engineering, Bhilai Institute of Technology, Durg, Chhattisgarh, India
shashwatiray@yahoo.com

1 Introduction

Power system is an electrical network to generate, transmit, and distribute the electrical power as one deployed in the various utilities to supply electricity. The system faces the problem of continuous supply of generation as input for matching the varying demands of load at minimum utility cost. This problem can be solved by maintaining the frequency of power system within a certain threshold value. The interlinked power system is divided into three units, viz., first is generation system, second is the transmission system, and the third one is the distribution system. The generators are referred to as power system control area. The smooth and economical running of interlinked power generation and distribution system depends hugely on the multiple layers of Automatic Generation Control (AGC) which makes sure that the generator output units follow as per the demand of electrical load. As the load changes in the system in continuous manner, generation is adjusted automatically according to the load to keep frequency to its nominal value. This is known as Automatic Generation Control (AGC). AGC works on the feedback system which is an error signal response [1–3]. These responses depend indirectly or directly on the various properties of the dynamics of governor and turbine incorporated in the power system. Here, we consider that the magnitude of transient frequency deviation should be minimum for the magnitude of load change. The individual generators present in the control area should share the total load for optimum economy of operation of system. If it is not required to maintain the frequency constant in a system, then the operator is not required to change the setting of the generator. But if constant frequency is required, the operator can adjust the velocity of the turbine by changing the characteristics of the governor when required. If a change in load is taken care by two generating stations running in parallel, then the complex nature of the system increases.

Economic load dispatch mechanism does a very important function of minimizing the overall operating expenses of total load demand in case of any power system by

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_62

regulating participation factor of each power generation unit [4]. Participation factors of power unit may be defined as the variation level of single power unit's output according to variation in complete power generation of each control system area of power system. The total generation is being controlled so as to meet the specified frequency error requirements but we have no control over the division of load among the generating units [5]. Hence, some of the generators might be commanded to generate more than their rating may permit. Thus, the main concern in our work is to solve the disparity problem of load produced and load demand, while maintaining the frequency up to a certain threshold limit. The speed governing mechanism employed in each generating unit maintains the constant speed of generators units by regulating the input supply of primary movers with respect to the control error frequency [6–8]. Also, the economic load dispatch acts as the cost minimizer for varying demand of loads. Economic load dispatch aligns the participation factor every minute to reduce the overall generation cost while load frequency control aligns the generation output every time.

This paper is organized as follows; in Sect. 2, problem of Load Frequency Control (LFC) with coordination of Economic Load Dispatch (ELD) is described. In Sect. 3, the mathematical modeling of interlinked power system is discussed. In Sect. 4, a detail of various algorithms of optimization methods investigated in our work is described. Lastly, in Sects. 5 and 6, comparison of the results obtained on two different test systems is discussed and the most effective method is recommended.

2 Problem Identification

There are two important parts in this work; one is to keep the frequency value low keeping the power value low and second is to keep the economics low. When the value of speed is decreased the error signal becomes smaller. One of the ways to regain the speed or frequency to its actual value is to add an integrator. This integrator will monitor the average error and overcome the offset. In an interlinked power system [9], the work of Automatic Generation Control (AGC) is to distribute the load among the systems, stations, and generations to achieve maximum economy of scale and maintain uniform frequency [10]. The task of coordination can be accomplished by designing the power units according to economic load dispatch. The logic of designing the power units as per economic load dispatch provides the information about the load sharing of total load demand for each power generating unit according to their economics. This idea provides the general view of each unit's participation share which may be called the participation factor for each unit [11].

3 Mathematical Modeling

In this work, we have adopted a sequential approach. First of all, a model of load frequency control is built using the MATLAB platform in SIMULINK. Then, different algorithms are applied in order to minimize the frequency response and to get the value

of optimum frequency. A model of economic load dispatch is built to get the value of participation factors. These values are then fed into the load frequency control system. Now, our system becomes a coordinated system comprising of load frequency control and economic load dispatch [12]. The various algorithms are applied to get the optimal value of frequency with economic load dispatch. In this work, model of the single control area and double control area are designed and simulated. The single-area system is designed with hydrothermal and gas power system and in two-area system, the first area is designed with hydrothermal and gas power system and second area is designed with hydro and thermal power system. The proposed control area used in this system comprises of many power producing units of various heavy input sources such as hydro, thermal based, nuclear based, gas based, etc. [13]. We know that the sum of participation factors of each unit will result in unity for each control area. While performing calculations involving economic load dispatch, if the total power generation is equal to the sum of the current unit generation, then this will be assigned as the base point generation (P_{ibase}). If any difference in load occurs from the base point generation, then that load will be distributed among each individual power generating units or generators using system participation factor. If the total deviation in load of the system is denoted by ΔP_D, then change in generation of each individual unit is represented as shown in (1), (2) and (3).

$$\Delta P_{Gi} = P_{fi} \times \Delta P_D \tag{1}$$

$$P_{ides} = \Delta P_{Gi} + P_{ibase} \tag{2}$$

$$\sum_{i=1}^{n} P_{fi} = 1, \tag{3}$$

where P_{Gi} is stepwise change from the base point generation of the individual unit
ΔP_{Gi} is incremental change from the base point generation other individual unit
P_{fi} participation factor for i^{th} unit and $i = 1, 2, 3, \ldots n$
P_{ides} expected output from i^{th} unit.

For calculating the participation factor of each generating unit, theories of ELD have been used.

The schematic diagram of proposed single-area interlinked power generation and distribution system which consists of thermal energy, hydro energy, and gas energy is shown in Fig. 1. Control error of frequency is calculated and its value is fed into the ELD to calculate the participation value of each unit. The change in electric load in power system is sensed and propagated back to the area control error, thus, it regulates the frequency of power system. Figure 2 represents the schematic diagram of two-area system which includes thermal, hydro and gas sources of power and their connections which include area control error, participation value from ELD, tie line, and their connection with power system.

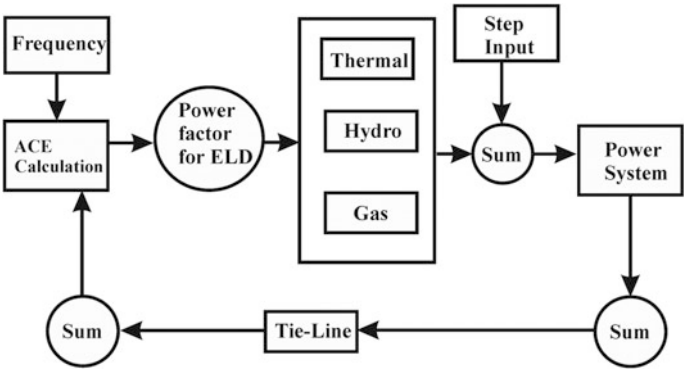


Fig. 1. Schematic diagram of single-area interlinked power system

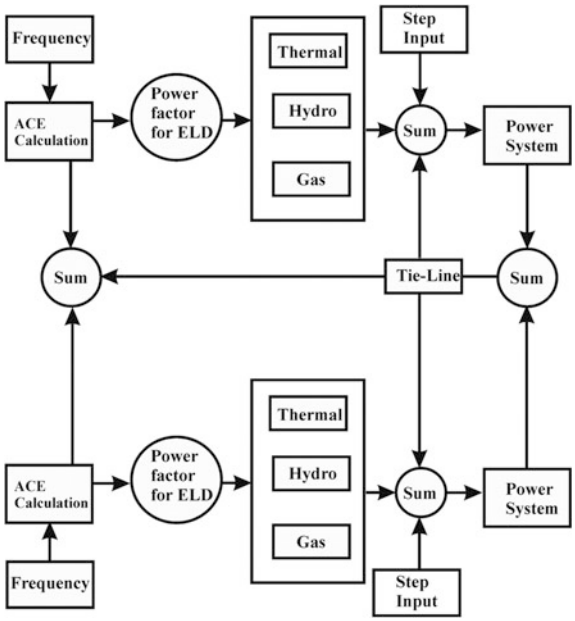


Fig. 2. Schematic diagram of two-area interlinked power system

4 Optimization Techniques

The primary goal to formulate the required objective function is to minimize the error response signal to improve the system stability, i.e., difference between the output and desired output. Error performance criteria are being applied for reducing the frequency error by implementing a PI controller. The integral error value is widely considered as a good quality measure for performance of system. The following are some most widely

used criteria based on the integral error value for a step set point or disturbance system response and integrated square of the error is given by (4).

$$IAE = \int_0^\infty |e(t)|dt,\ ISE = \int_0^\infty e(t)^2 dt,\ ITAE = \int_0^\infty t|e(t)|dt \tag{4}$$

The following optimization techniques are implemented in our work to minimize the error by above mentioned criterion. Best controller can be designed by selecting its parameters such that any of the above performance criteria is minimized.

4.1 Genetic Algorithm

Genetic algorithm (GA) method is based on the evolution theory of species and the ability of the fittest species to survive in adverse conditions. This algorithm guides the search of optimal solution through the solution space using natural selection of species and genetic operators such as crossover and mutation. This method is extensively used for improving the required optimal solution of the certain predefined problem space which is usually based on the searching solution via experimental operations in genetics and biological range of the system [14, 15]. The objective function for genetic algorithm is specified by recital parameters of the given problem space. The power system control strategy is similar to fulfill the group of criterion of some specified calculable parameters and the power generation and distribution system dynamics performance factors which specify the overall functional performance like peak time, settling time, steady-state error, etc.

4.2 Ant Colony Optimization

This technique was developed by Dorigo and his assistants in early 1990s. Ant colony optimization (ACO) is developed on the basis of the cooperation behavior of actual ants in their colonies. This behavior of ants in colonies enables them to find the shortest possible paths from their home to possible locations of food. This behavior of ant colonies is used to predict the behavior of artificial ant colonies which is used to get the most approximate solution of any discrete optimization problem, continuous optimization problem and any important problem of load balancing, frequency control, economic load dispatch, etc. This algorithm mimics the wandering and feeding behavior of ants. In this behavior, in order to develop the no direct communication among the ants, they secrete a chemical substance called pheromone trails, and this chemical enables them to discover the shortest possible path between their nests and their foods [16]. This behavior characteristic is used in ant colony optimization algorithm, which is used to find the approximate solution of any discrete optimization problem.

4.3 Firefly Algorithm

Firefly algorithm (FA) is biologically inspired technique which is developed on the biological conduct of fire fly. Firefly flashes light in the night with the help of

bioluminescence mechanism. This light has certain rhythm or pattern which is unique for a specific species of fire fly. This pattern of light can be used to formulate the objective function of any optimization problem [17]. This behavior provides the foundation ground of firefly algorithm.

4.4 Harmony Search Algorithm

This method is based on the music harmony. Music composers compose the harmony; they use various pitches stored in their memory which can be used to formulate an objective function using various inputs (pitch combinations) to optimize the problem (perfect harmony). Harmony search (HS) method inspired by harmony improvization, it has two distinct operators, one is harmony memory considering rate (HMCR) and another one is pitch adjusting rate (PAR) [2]. These two operators are used for generating and further mutating a solution, respectively.

5 Results and Discussions

The simulation of interlinked power system is done using MATLAB platform .The frequency response of single area system obtained by using all the four aforesaid optimization techniques is shown in Fig. 3.

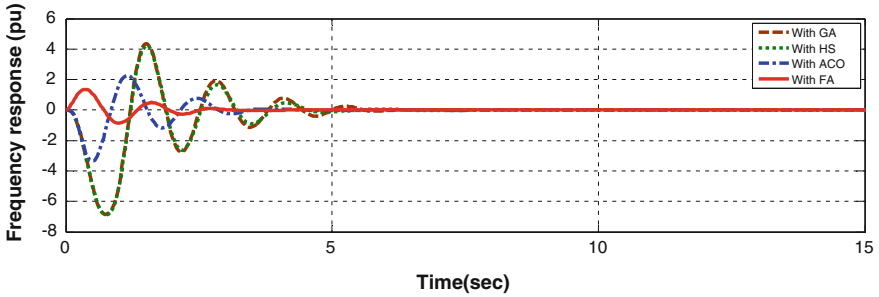


Fig. 3. Comparison of different optimization algorithm for single-area system

Four optimization techniques namely genetic algorithm, firefly algorithm, ant colony optimization, and harmony search algorithm are implemented in order to get the optimized frequency deviation. A comparison is done among the four techniques using different criteria such as integral square error, integral average error, and integral total average error criteria for a single area system which is depicted in Table 1.

For a two area system, the results obtained for the values of K_p and K_i by all the optimization techniques is shown in Table 2 and the comparision of the frequency response is shown in Fig. 4.

Table 1. Comparison of FA, ACO, GA and HS algorithm based on ISE, IAE and ITAE criteria for single-area system

Optimization algorithm	PI	PI error	Kp	Ki
GA	ISE	0.026	0.99	0.99
	IAE	8.23	1	1
	ITAE	12.5	1	1
HS	ISE	0.056	0.765	0.632
	IAE	8.23	0.99	0.99
	ITAE	9.12	1	1
ACO	ISE	0.012	0.876	0.546
	IAE	2.13	0.99	0.99
	ITAE	4.67	1	1
FA	ISE	0.001	0.003	0.01
	IAE	0.045	0.99	0.99
	ITAE	0.076	1	1

Table 2. Comparison of FA, ACO, GA and HS algorithm based on ISE, IAE and ITAE criteria for two-area system

Optimization algorithm	PI	PI error	Kp	Ki
GA	ISE	0.026	0.99	0.99
	IAE	8.23	1	1
	ITAE	12.5	1	1
HS	ISE	0.056	0.765	0.632
	IAE	8.23	0.99	0.99
	ITAE	9.12	1	1
ACO	ISE	0.012	0.876	0.546
	IAE	2.13	0.99	0.99
	ITAE	4.67	1	1
FA	ISE	0.001	0.003	0.01
	IAE	0.045	0.99	0.99
	ITAE	0.076	1	1

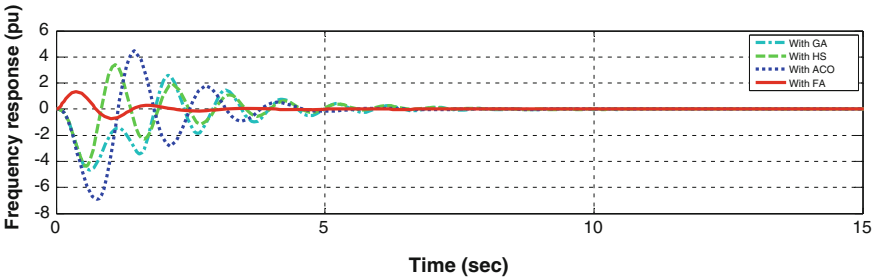


Fig. 4. Comparison of different optimization algorithms in two-area system

6 Conclusion

The results are obtained by simulation of interlinked power system based on both single- and two-area system under dynamic load conditions with varying frequency. Power system is coordinated with automatic generation control (AGC) and economic load dispatch (ELD) to optimize the load variation by minimizing frequency deviation. Four optimization techniques are used and compared based on different criteria. The technique showing least error is described as best method in this scenario. Here, the observations show that firefly algorithm gives the optimum result in both the cases as the values of integral square error, integral average error, and integral total average error obtained among all the methods are the least; so it can be suitably applied in order to minimize the frequency deviation. Further, the results of ACO, HS, and GA are also compared. This work can also be extended to three control area power system to optimize fluctuating load demands and to improve the quality of power supply in different circumstances.

References

1. Shankar R, Chatterjee K, Chatterjee TK (2014) Coordination of economic load dispatch and load frequency control for interconnected power system
2. Zhu JZ, Chang CS, Xu GY (1998) A new model and algorithm of secure and economic automatic generation control. Electr Power Syst Res 45(2):119–127
3. Mohammed Z, Talaq J (2011) Economic dispatch by biogeography based optimization method. In: 2011 international conference on signal, image processing and applications vol 21, pp 161–165
4. Han XS, Gooi HB, Kirschen DS (2001) Dynamic economic dispatch: feasible and optimal solutions. IEEE Trans Power Syst 16(1):22–28
5. Fan JY, Zhang L (1998) Real-time economic dispatch with line flow and emission constraints using quadratic programming. IEEE Trans Power Syst 13(2):320–325
6. Safari A, Shayeghi H (2011) Iteration particle swarm optimization procedure for economic load dispatch with generator constraints. Expert Syst Appl 38(5):6043–6048
7. Bhattacharya B, Mandal KK, Chakravorty N Cultural algorithm based constrained optimization for economic load dispatch of units considering different effects
8. Rao CS, Nagaraju SS, Raju PS, Reddy MR (2006) Optimized integral controller for economic load dispatch in a two area system based on hooke-jeeves algorithm
9. Hinojosa VH, Leyton C (2012) Short-term hydrothermal generation scheduling solved with a mixed-binary evolutionary particle swarm optimizer. Electr Power Syst Res 92:162–170
10. Barcelo WR, Rastgoufard P (1997) Dynamic economic dispatch using the extended security constrained economic dispatch algorithm. IEEE Trans Power Syst 12(2):961–967
11. Belleview WA (1979) Coordination of economic dispatch and load frequency control in electric power systems
12. Franze G, Tedesco F (2011) Constrained load/frequency control problems in networked multi-area power systems. J Franklin Inst 348(5):832–852
13. Saikia LC, Mishra S, Sinha N, Nanda J (2011) Automatic generation control of a multi area hydrothermal system using reinforced learning neural network controller. Int J Electr Power Energy Syst 33(4):1101–1108

14. Golpîra H, Bevrani H (2014) A framework for economic load frequency control design using modified multi-objective genetic algorithm. Electr Power Compon Syst 42(8):788–797
15. Blum C (2005) Ant colony optimization: Introduction and recent trends. Phys Life Rev 2(4): 353–373
16. Yang XS (2009) Firefly algorithms for multimodal optimization. In: International symposium on stochastic algorithms. Springer Berlin Heidelberg, pp 169–178, Oct 2009
17. Geem ZW, Kim JH, Loganathan GV (2001) A new heuristic optimization algorithm: harmony search. Simulation 76(2):60–68

Selection of Colour Correction Algorithms for Calibrating Optical Chronic Ulcer Images

Maitreya Maity[1(✉)], Dhiraj Dhane[2], Chittaranjan Bar[3], Chandan Chakraborty[1], and Jyotirmoy Chatterjee[1]

[1] School of Medical Science and Technology, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal, India
maitreya.maity@gmail.com
[2] Department of ECE, Indian Institute of Information Technology, Pune, India
dmd.ece@iiitp.ac.in
[3] Department of Dermatology, Midnapore Medical College, Midnapore, West Bengal, India
drcrbar@gmail.com

1 Introduction

A skin ulcer is a collective representation of a heterogeneous mixture of granulation, slough and necrotic tissues. The clinician does wound assessment using simple visual grading and spatial measurement. However, more accurate wound assessment could be achieved by accurate segmentation of target region (ROI) and also the precise classification of all tissues within it. Therefore, colour and textural features are the most important deciding parameters for dermatological evaluation [1]. Most of the studies in the literature for computer-aided wound diagnosis have evaluated the wound using photographs captured using the handheld digital camera. Moreover, the colours in the images vary significantly due to the influence of varying background and camera models. Therefore, it is difficult to obtain an identical image from different camera models in heterogeneous lighting conditions. Since colour plays a critical role in wound evaluation, the attention is needed in the study of colour correction for the acquired digital wound images.

Pereira et al. [2] and Molnar et al. [3] recommended the colour correction and normalization in wound image analysis. However, colour correction was neglected, since homogeneous illumination was practically unrealizable [3,4]. In few studies, pre-calibration methods were used for captured digital wound images. Wannous et al. [4] have proposed a standard 24-bit Macbeth colour checker for calibration of wound photographs. Plassmann and Jones [5] used a standard white patch which was placed besides the ulcer. The image was captured which included a white patch at the background and a wound region.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_63

The white balance was estimated using the placed white patch. Haeghen et al. [6] have introduced a 24-bit colour pattern to standardize wound image capture protocol. Pereira et al. [7] have also used colour patches and rulers to facilitate colour normalization and calibration. Perez et al. [8] attempted another colour calibration method, where a small white ruler was placed beside the target region for colour calibration.

It is evident from the literature that colour correction is a vital step for an accurate wound image evaluation. Most of the studies either use a white patch or 2–24-bit colour calibrator for colour alteration. The intent of this explorative study is to evaluate the performance of the existing colour correction models to select the best correction methods which performed closely with the manual colour checker. Twelve colour correction methods and their performance evaluation for best colour homogenization method selection using ten image quality metrics are described below.

2 Materials and Methods

2.1 Notations

The following notations are being used throughout this paper:

- $f(X)$ denotes original image where X is image matrix and p is pixel value.
- (P, Q) denotes the size of an image and c indicates a colour channel.
- e denotes the light source estimate and k is incident light constant.
- w denotes a weighted factor and K is an averaging kernel.
- R is the reference image obtained using 24-bit colour checker and O is the test image obtained from colour correction methods.

2.2 Chronic Wound Sample Image Database Development

The sample images were taken from Midnapur Medical College and Hospital, India (MMCH). Institutional ethical clearance was taken for this study. The images were acquired using multiple digital cameras in a hospital setting. The images were grabbed with and without flash to get variations in illuminations. Each raw image has a size of 4320×3240 pixels; however, the images were rendered conveniently to get maximal ulcer area with the minimum background. During Sample collection, total 68 images were captured from 11 patients with Diabetic ulcer, 24 images from 4 patients with Pressure ulcer, 34 images from 6 patients with Tropical ulcer, and 124 images from 21 patients with Surgical wounds.

2.3 Colour Constancy

In this study total, 12 existing colour correction algorithms were evaluated for correction of poorly illuminated original images. Each algorithm is described below.

2.3.1 Colour Corrected Reference Image Generation

Variability in camera model, configuration, lighting condition, etc., causes a huge colour variation among the captured images. Therefore, illumination correction or colour normalization is a necessary task. A 24-patch Macbeth colour checker pattern was used to correct the colour. The colour checker was placed besides the wound region and image was grabbed. Using a commercial calibration software provided by colour checker manufacturer, the captured images were corrected automatically. The generated corrected images were considered as a reference image for the further processing. In the Fig. 1, one sample captured wound image with Macbeth colour checker and corresponding corrected reference image is presented.

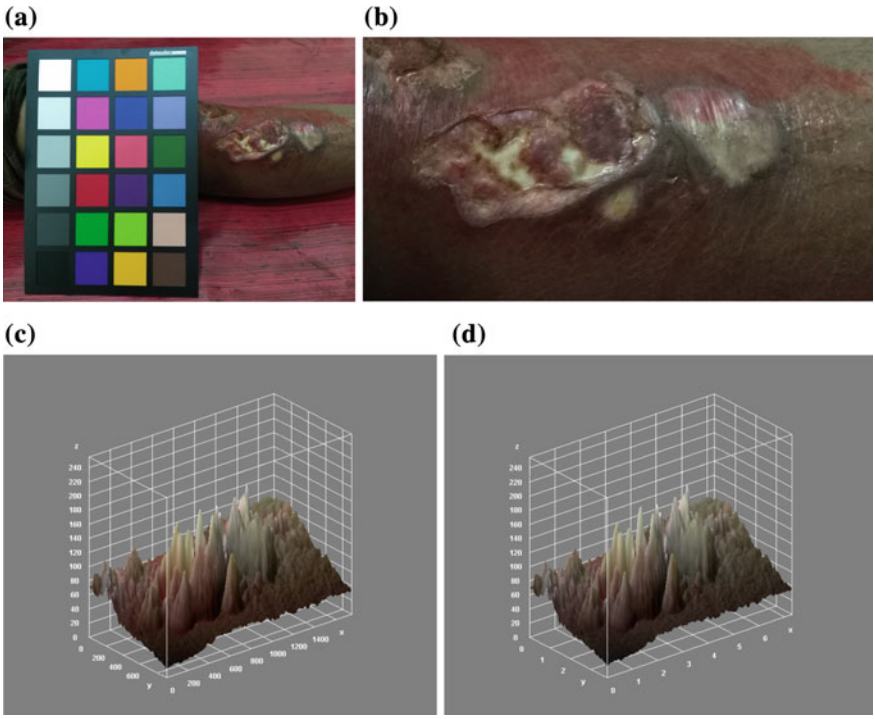


Fig. 1. a Original captured wound image, **b** corrected image (ROI), **c** 3D surface plot of original image, **d** 3D surface plot of corrected image

2.3.2 Grey World (A1)

Grey World is one of the popular and simple colour constancy methods. Here, it is argued that the average intensity of RGB channel in the scene should be equal [9]. Let an image have three colour channels like *R, G, B*. Then, a gain of red

and blue channel against the green channel can be computed from the average channel values.

$$Ch_{avg} = \frac{1}{PQ}\sum_{i=1}^{P}\sum_{j=1}^{Q} I_{ch}(i,j); Ch_{gain} : \hat{a} = \frac{G_{avg}}{R_{avg}}, \hat{b} = \frac{G_{avg}}{B_{avg}} \tag{1}$$

$$O = \left[\hat{a}I_r(i,j), I_g, \hat{b}I_b(i,j)\right] \tag{2}$$

2.3.3 Max RGB (A2)

It estimates the scene illumination by maximum camera response to the various colour channels [10]. The algorithm assumes that the reflectance of RGB channels are equal and the scene surface is uniformly reflective

$$max_x f(X) = ke \tag{3}$$

2.3.4 Grey Edge (A3)

The Grey Edge assumes that the edge difference in an image is achromatic. The basis of the algorithm is on the observation that the colour derivative distribution of images tends to show the largest variation in the direction of light source [11].

$$\left[\frac{\int f(X)dx)}{\int dx}\right] = ke \tag{4}$$

2.3.5 Shades of Grey (A4)

Shades of Grey was proposed by Finlayson and Trezzi [12]. It is a statistical corollary of the Grey World and MaxRGB algorithm based on the Minkowski norm.

$$\left[\frac{\int f(X)dx)}{\int dx}\right]^{\frac{1}{p}} = ke \tag{5}$$

2.3.6 Multi-Scale Retinex (A5)

Multi-Scale Retinex (MSR) algorithm combines many single local scaled retinex outputs to produce a single output image. The MSR enhances the image contrast in an image having wide brightness variations [13]. MSR can be explained as

$$O = \sum_{n=1}^{N} w_n \cdot \{\log\left[f_c(X)\right] - \log\left[f_c(X) * M_n(X)\right]\}, \tag{6}$$

where N is the scale M is surrounding kernel of a point.

2.3.7 Weighted Grey Edge (A6)

It is the modified version of Grey Edge algorithm in which distinct edges are considered to enhance the performance by calculating the weighted mean of the edges. The weights are calculated by photometric classification method [14].

$$\left(\int \left| w^k \cdot f_{c,x}(X) \right|^p dx \right) = k e_c \tag{7}$$

2.3.8 Colour Sparrow (A7)

Banic and Loncaric [15] used Random Spray Retinex (RSR) [16,17] to demonstrate a solution to global white balancing and illumination adjustment for the images. The colour sparrow estimates the global illumination using local RSR.

$$O = \frac{(f_c * K) \cdot p}{(R_c * K) \cdot p} \tag{8}$$

2.3.9 Light Random Spray Retinex (A8)

It is an improvised RSR algorithm by Banic and Loncaric (LRSR) [18]. In LRSR, the minimum number of sprays and pixels per sprays exhibits give and take relationship between quality of the image and computational cost.

$$Relative_{int} : C_{c,K}(p) = \frac{(f_c * K) \cdot p}{(R_c * K) \cdot p}; Final_{int} : \hat{C}_{c,K_1,K_2}(p) = (C_{c,K} * K_2) \cdot p \tag{9}$$

$$Corrected\, image : O = \frac{f}{\hat{C}} \tag{10}$$

2.3.10 Colour Rabbit (A9)

In [19], the authors incorporate the idea of maximum response for the distance distribution from the centre spray to the spray pixels. Here, pixel hunting is done for getting maximum channel response.

$$Relative\, intensity\, change : \; C_{c,K}(p) = \frac{(f_c * K) \cdot p}{(R_c * K) \cdot p} \tag{11}$$

$$Global\, illumination\, factor = \frac{1}{X} \sum_{1}^{X} [C_{R,K}, C_{G,K}, C_{B,K}]^T \tag{12}$$

2.3.11 Colour Badger (A10)

In this methods, a Tone Mapping Operator (TMO) is introduced which converts High Dynamic Range (HDR) image into Low Dynamic Range (LDR) image [20]. This effect can be reduced globally by minimizing the intensity gaps between the pixels.

$$Relative\, intensity\, change : \; C_{c,K}(p) = \frac{(f_c * K_1^{*r}) \cdot p}{(R_c * K_1^{*r}) \cdot p} \tag{13}$$

$$Final\, intensity\, change\, \hat{C}_{c,K_1,K_2,r_1,r_2}(p) = (Oc, K_1, r_1 * K_2^{*r_2}) \cdot p \tag{14}$$

2.3.12 Subsampled White Patch (A11)

Improvement over white patch method is suggested by Banic and Lonariac [21]. By considering the pixel or surface uniformity, only a few pixels are necessary to represent the whole area in the image.

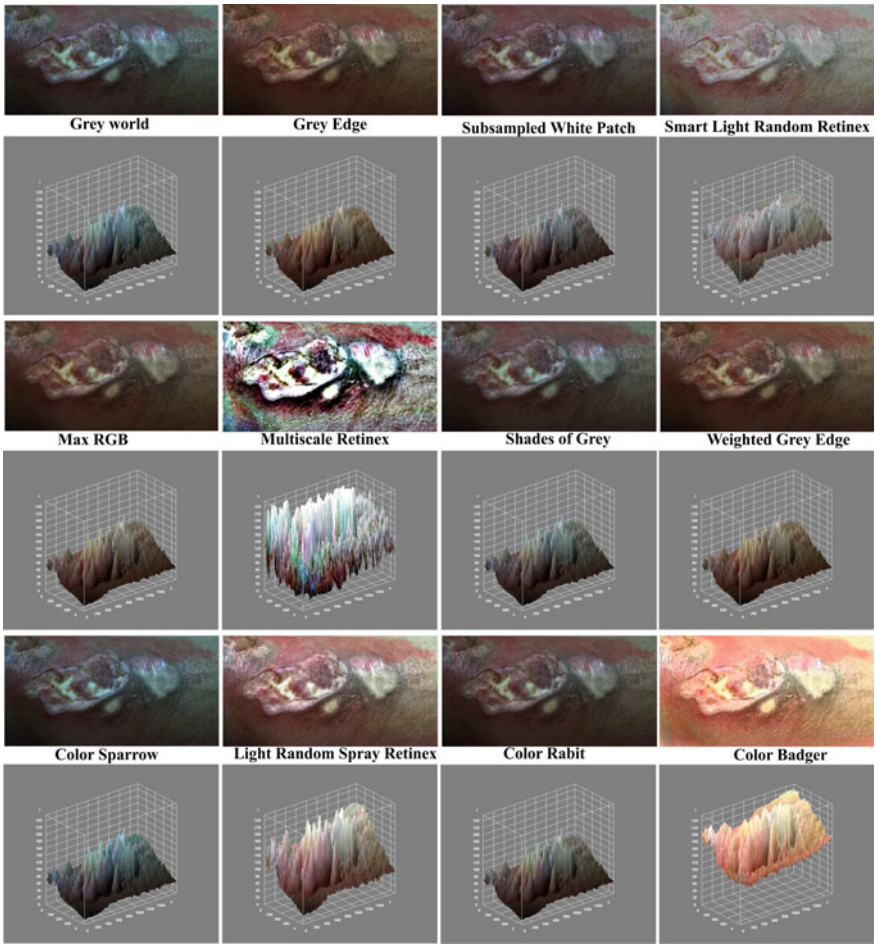


Fig. 2. Colour corrected images and corresponding 3D surface plots

2.3.13 Smart Light Random Memory Spray Retinex (A12)

Smart Light Random Memory Spray Retinex (A12): The authors in [22] proposed a method for brightness adjustment and colour correction by redefining LRSR method. The drawback of LRSR is the high computational cost due to per-pixel operations. This drawback is overcome in the memory sprays retinex method.

Table 1. Image quality performance measures of considered 12 colour correction methods

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
A1	1.1949	28.2984	42.2244	1.9274	0.9946	0.9957	0.9998	0.9979	0.9978	0.9659
A2	0.7309	33.054	46.68	1.1148	0.9971	0.9975	0.9999	0.9995	0.9981	0.9974
A3	0.8958	32.5972	46.5232	1.175	0.997	0.9975	0.9999	0.9995	0.9981	0.9987
A4	38.8173	22.28	11.6454	65.1517	0.247	0.3513	0.9696	0.3797	0.9542	0.4998
A5	0.7867	31.0512	44.9771	1.4039	0.996	0.997	0.9998	0.9991	0.998	0.9907
A6	**0.6402**	**3.974**	**49.9**	**0.9951**	**0.1972**	**0.2976**	**0.9999**	**0.9998**	**0.9992**	**0.9995**
A7	1.1875	28.3418	42.2678	1.9178	0.9947	0.9957	0.9998	0.971	0.9979	0.9659
A8	21.8609	13.4046	17.3306	33.8587	0.8013	0.9017	0.983	0.9204	0.9967	0.9241
A9	0.7463	31.1086	45.0346	1.3946	0.9962	0.9973	0.9998	0.9994	0.998	0.9970
A10	49.1397	12.8712	11.0548	69.7365	0.5239	0.7672	0.9478	0.8372	0.9669	0.8088
A11	0.7723	31.0995	45.0255	1.3961	0.9961	0.9971	0.9998	0.9993	0.998	0.9945
A12	23.9646	23.3291	17.2551	34.1543	0.7621	0.9044	0.9776	0.9339	0.9906	0.9362

3 Results and Discussion

To perform an experiment, overall 240 ulcer images from the developed database were considered. The calibration parameters were stored during calibration phase. Now, using these parameters, the reference image set (colour corrected using 24-bit colour checker) was generated. The 12 colour constancy algorithms were also applied on the same image set to generate colour corrected image set (colour corrected using twelve different algorithm). Finally, two sets of colour corrected image were compared with each other. The performance of these methods was evaluated using ten image quality measures. The considered image quality metrics are Mean Absolute Error (Q1), Signal-to-Noise Ratio (Q2), Peak Signal-to-Noise Ratio (Q3), Root Mean Square Error (Q4), Structural Similarity Index (Q5) [23], Multi-Scale Structural Similarity Index (Q6) [23,24], Luminance Comparison (Q7), Contrast Comparison (Q8), Structure Comparison (Q9), and Feature Similarity Index (Q10) [25]. The 12 colour correction algorithms based on corrected image outputs are shown in Fig. 2. The comparative study of colour correction methods is presented in Table 1. From the comparison results, Weighted Grey Edge illumination correction algorithms were able to correct images with minimum error in comparison with standard colour checker based corrected images. Therefore, Weighted Grey Edge technique could be considered as an alternative automated colour correction method for optical wound image colour correction.

4 Conclusion

A comparative study was made in the context of colour constancy methods where RGB images of skin ulcers were processed in order to choose the best method for colour correction. Although, colour checker is well-established and accurate method of colour correction, however, the 24-bit colour checker is quite costly. The experimental results show that, for chronic ulcer images, Weighted Grey Edge method is the best compared to remaining colour constancy methods. It is also evident from an experiment that algorithmic colour correction methods imitate the reference Macbeth colourmap. In photo-dermatology, this study is an appropriate low-cost alternative to commercially available resource-dependent methods.

Acknowledgements. The first author acknowledges CSIR for financial support (09/81(1223) /2014/EMRI dt. 12-08-2014). The second and third author would like to acknowledge ICMR, GoI, (Grant number: DHR/GIA/21/2014, dated 18th November, 2014).

Ethics Statement. The protocol was approved by the ethical board of Indian Institute of Technology (IIT) Kharagpur and MMCH with ref. no: IIT/SRIC/AR/2012 and ref. no: IEC/2012/1 respectively.

References

1. Veredas FJ, Mesa H, Morente L (2015) Efficient detection of wound-bed and peripheral skin with statistical colour models. Med Biol Eng Comput 53(4):345–359
2. Pereira SM, Frade MA, Rangayyan RM, Azevedo-Marques PM (2013) Classification of color images of dermatological ulcers. IEEE J Biomed Health Inform. 17(1):136–142
3. Molnar JA, Lew WK, Rapp DA, Gordon ES, Voignier D, Rushing S, Willner W (2009) Use of standardized, quantitative digital photography in a multicenter web-based study. Eplasty 9
4. Wannous H, Treuillet S, Lucas Y (2010) Robust tissue classification for reproducible wound assessment in telemedicine environments. J Electron Imaging 19(2):023002
5. Plassmann P, Jones T (1998) Mavis: a non-invasive instrument to measure area and volume of wounds. Med Eng Phys 20(5):332–338
6. Haeghen YV, Naeyaert JMAD, Lemahieu I, Philips W (2000) An imaging system with calibrated color image acquisition for use in dermatology. IEEE Trans Med Imaging 19(7):722–730
7. Pereira SM, Frade MAC, Rangayyan RM, de Azevedo Marques PM (2011) Classification of dermatological ulcers based on tissue composition and color texture features. In: Proceedings of the 4th international symposium on applied sciences in biomedical and communication technologies. ACM, p 68
8. Perez AA, Gonzaga A, Alves JM (2001) Segmentation and analysis of leg ulcers color images. In: Proceedings of international workshop on medical imaging and augmented reality. IEEE, pp 262–266
9. Buchsbaum G (1980) A spatial processor model for object colour perception. J Franklin Inst 310(1):1–26
10. Land EH, et al (1977) The retinex theory of color vision. Citeseer
11. Van De Weijer J, Gevers T, Gijsenij A (2007) Edge-based color constancy. IEEE Trans Image Process 16(9):2207–2214
12. Finlayson GD, Trezzi E (2004) Shades of gray and colour constancy. In: Color and imaging conference, society for imaging science and technology, vol 2004, pp 37–41
13. Rahman Z, Jobson DJ, Woodell GA (1996) Multi-scale retinex for color image enhancement. In: Proceedings of international conference on on image processing. IEEE, vol 3, pp 1003–1006
14. Gijsenij A, Gevers T, Van De Weijer J (2012) Improving color constancy by photometric edge weighting. IEEE Trans Pattern Anal Mach Intell 34(5):918–929
15. Banić N, Lončarić S (2013) Using the random sprays retinex algorithm for global illumination estimation. arXiv:13100307
16. Provenzi E, Fierro M, Rizzi A, De Carli L, Gadia D, Marini D (2007) Random spray retinex: a new retinex implementation to investigate the local properties of the model. IEEE Trans Image Process 16(1):162–171
17. Provenzi E, Marini D, De Carli L, Rizzi A (2005) Mathematical definition and analysis of the retinex algorithm. JOSA A 22(12):2613–2621
18. Banić N, Lončarić S (2013) Light random sprays retinex: exploiting the noisy illumination estimation. IEEE Signal Process Lett 20(12):1240–1243
19. Bani N, Lonari S (2014) Color rabbit: guiding the distance of local maximums in illumination estimation. In: 2014 19th international conference on digital signal processing (DSP). IEEE, pp 345–350

20. Banić N, Lončarić S (2014) Color badger: a novel retinex-based local tone mapping operator. In: international conference on image and signal processing. Springer, pp 400–408
21. Banić N, Lončarić S (2014) Improving the white patch method by subsampling. In: IEEE international conference on image processing (ICIP). IEEE, pp 605–609
22. Banić N, Lončarić S (2015) Smart light random memory sprays retinex: a fast retinex implementation for high-quality brightness adjustment and color correction. JOSA A 32(11):2136–2147
23. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
24. Rouse DM, Hemami SS (2008) Analyzing the role of visual structure in the recognition of natural image content with multi-scale SSIM. In: Electronic imaging, international society for optics and photonics, pp 615–680
25. Zhang L, Zhang L, Mou X, Zhang D (2011) Fsim: a feature similarity index for image quality assessment. IEEE Trans Image Process 20(8):2378–2386

Designing a Scalable Socio-Technical Method for Evaluating Large e-Governance Systems

Salim Saay[1(✉)] and Alex Norta[2]

[1] Computer Science Faculty, Kabul University, Kabul, Afghanistan
saay@tlu.ee
[2] Institute of Software Systems, Tallinn University of Technology, Akadeemia tee 15a, 12816 Tallinn, Estonia
alex.norta.phd@ieee.org

1 Introduction

This research is about designing a scalable socio-technical method for evaluating large e-governance systems. The design science research method [16] is mostly used for the architectural designing of socio-technical systems. Design science research is applicable in many different fields, including architecture, engineering, education, psychology, and fine arts.

Researchers use design science for rigorously developing artifacts, processes, or theories. Specific for design science research is that the starting point of investigations is a socio-technical case where people as members of organizations use technology to collaboratively solve challenging problems [4]. The evaluation of artifacts, theories, and processes is an important part of a design science research. The aim of a software-architecture evaluation is to determine quality attributes, identify potential risks, and understand the trade-off in the architecture design [11].

Several scenario-based evaluation methods already exist such as the Scenario-Based Architecture Analysis Method (SAAM) [14], Architecture Trade-off Analyses Method (ATAM) [6], Scenario-Based Architecture Re-engineering (SBAR) [3], Architecture-Level Modifiability Analysis (ALMA) [7], and Family-Architecture Assessment Method (FAAM) [2].

Software architecture re-engineering for quality attributes is the main objective of Scenario-Based Architecture Re-engineering (SBAR) which is using three steps for an evaluation. It supports multiple quality attributes and classifies the scenarios in development and operational, and qualitative analysis. ALMA is developed for evaluation of the maintenance cost, risk assessment, and comparison of the architecture, it is implemented in six steps and addresses the reusability of systems. FAAM also has six mail steps, it focuses on interoperability and extensibility of the systems [14]. Scenario-Based Architecture Analysis Method (SAAM)which is a predecessor of ATAM is used for analysis of risk and architec-

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_64

ture suitability, its implementation includes six steps [5]. The implementation of ATAM has seven steps [6], such as collection of scenarios, collection of requirements, description of the architecture, realization of the scenarios, analyses and specification of attributes, sensitivities identification, and trade-off identification. From the abovementioned scenario-based evaluation methods, only ATAM and SAAM consider industry involvement that it is very important for designing of a scalable socio-technical e-governance systems. However, a survey [11] shows that many open problems exist in established scenario-based evaluation methods. A full scale evaluation of architectures is very resource and time consuming, e.g., Patidar and Suman [11] claim that ATAM needs at least 200–400 man-hours for the evaluation of an architecture. They give very low consideration for quality attributes like availability, scalability, maintainability, and extensibility [1] of architectures. Existing methods require physical presence of expert groups, customer, and users. ATAM do not use tools for communication and negotiation between architecture developers, customer, and expert groups. Therefore, the aim of this research is to design an efficient and tools supported method for evaluation of scalable socio-technical e-governance systems. We design the SMIS architecture based on the Afghanistan Research and Education Network (AfgREN) infrastructure as a case in this research and implement an experimental evaluation on it. To organize the research, we answer the following questions. How to design a scalable socio-technical method for evaluating large e-governance systems? What systematic processes are necessary for the design and evaluation of a scalable socio-technical system? What are the main differences between the existing scenario-based architecture evaluation method and our developed architecture evaluation method? The paper is organized as follows: Section 2 explores the related work. In Sect. 3, we design an evaluation process by considering the design science research method. Section 4 mentions the experiments of the research and in Sect. 5, we conclude the results.

2 Related Work

In this research, we modify the formal Architecture Trade-off Analysis Method (ATAM). SMIS based on AfgREN is used as a case, and the proposed evaluation process address the evaluation of designed architectures. Design science research method is used for research of socio-technical systems that consider both social and technical components. It is a problem-solving paradigm that seeks to create novel ideas, practices, technical capabilities, and products through which the analysis, design, implementation, and use of information systems can be effectively and efficiently accomplished [15]. Design science research must have a clear contribution to the area and relevant environment. It relies on the application of rigorous methods in construction and evaluation.

Design science research must be implemented as a research process by considering laws, internal procedure, management, and technology to solve the problem in the relevant environment [8]. Figure 1 shows the structure of design science research method that has three main cycles, such as relevance cycle, design

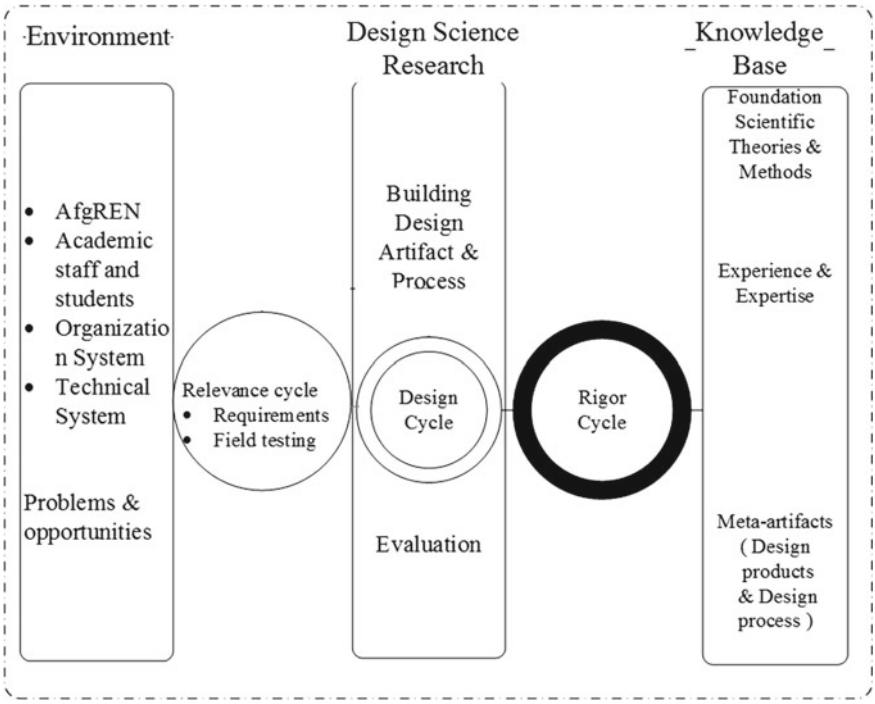


Fig. 1. Design science research

cycle, and rigor cycle. Study about application domain, design of problem solver artifact, and stepwise evaluation are the components of the research process. There are many different processes proposed for the application of design science research, e.g., Offermann et al. [10] propose a three-step design process which includes problem identification, solution and design, and evaluation. They used hypothesis, case study, expert survey, and a laboratory experiment for the artifact evaluation. Peffers et al. [12] propose a six-step process for the design and evaluation of the architectures. The problem identification and motivation, objectives of a solution, design and development, demonstration, evaluation and communication are their proposed steps. They are not using tools for the evaluation, which results in a time-consuming process. Design science research considers the users feedback, technical infrastructure, and relevant problems. So, it is the most suitable research method for designing a scalable socio-technical e-governance systems.

Evaluation is an important part of the design science research method to analyze the designed artifact or theory [17]. Evaluation supports the claims of the researcher and provides feedback for a further development. A researcher analyzes the designed artifact or theory to match with the requirements of the environment and to the expected quality attributes.

Venable et al. [17] propose a four-step framework for the evaluation in design science, explicating the goals, choosing a strategy or strategies for evaluation, determining the properties for evaluation, and designing the individual evaluation episode. In the first step, the evaluation of the designed architecture should be rigorous in order to reduce the development risk and to evaluate the efficiency of the software development. The second step focuses on the prioritization of software-risk development and is followed by the evaluation of the cost and complexity of the architecture. In the third step, the properties of the designed architecture, including usability, security, reliability, flexibility, portability, and interoperability need to be evaluated. Its fourth step, the evaluation, is based on environmental constraints and priorities of the resource.

Angeline Julia et al., proposed a new environmental based architecture evaluation method called Creative Innovative Pattern for Architecture Analysis (CIPA) that is implemented in seven steps, (1) Finding the stakeholders demands, and technical and managerial restrictions. (2) Collect the requirements, prioritize and analyze the architectural approach. (3) Categorize the quality attributes relevant to the architecture (4) apply CIPA (5) create architecture based on CIPA (6) Estimate the final result and get feedback from the client (7) Repeat the Steps 2–5 if needed [1]. Still the procedure of collecting requirements, analysis, and prioritizing the requirements is a complicated process in CIPA.

In this research, we use Requirements Bazaar [9] as a tool for collecting, analyzing and prioritizing the requirements. Requirements Bazaar is an online tool and has no limitation to time and place. The system developers add the title of the project, list the basic expected requirements, and send the link of the project to the target expert groups. The target groups add new requirements and can vote to already listed ones. Based on the Requirements Bazaar, the system developers not only collect the requirements but can also negotiate with large numbers of users and expert groups. In the following, we use Requirements Bazaar as a tool with customized ATAM for evaluation of a large scalable e-government system.

3 Proposed Evaluation Process

In this research, we modify ATAM and design an efficient and tools supported method for gathering quality attributes and evaluation of architectures. We were able to decrease seven steps of ATAM to three steps by using Requirement Bazaar with it. Figure 2 shows three main steps of the evaluation for design science research method, ATAM and Requirement Bazaar are used in parallel in this method. First, we implement the first ATAM workshop with policymakers, during the same workshop we start data collection and requirement gathering

with the Requirement Bazaar. In the second step, we implement the second ATAM workshop with the IT expert group, in the same workshop, we start the collection and identification of quality attributes. Third step is implementation of third ATAM workshop with the users, in the same workshop we prioritize the quality attributes based on the comments and votes of the users.

3.1 Requirements Gathering

Collecting proper and reliable data, supported by proper evidence, is always a time-consuming process for researchers. Without required data, a proper design and evaluation are not possible. In this step, we use the Requirements Bazaar parallel with the ATAM workshops. We write the goal and properties of the architecture, then add some functional and nonfunctional requirements, and then share the link of Requirements Bazaar to the target expert groups. They later add new requirements and comments, they vote on the requirements, and architecture designer can negotiate with the expert groups.

3.2 Identification of Quality Attributes

Evaluation of architecture is based on the quality attributes: Performance, modifiability, interoperability, security, scalability, data integrity, extensible, and so on. In our proposed evaluation method, the list of quality attributes will be based on IT experts and user preferences. In the second ATAM workshop, the IT experts list the quality attributes, during the same workshop, architecture designer also publishes them in Requirements Bazaar for communication and negotiation with the target groups of IT experts and users.

3.3 Evaluation of the Quality Attributes and Prioritization

In the third ATAM workshop, we discuss the quality attributes with the users and they can add quality attributes and requirements. At the same time, we use Requirements Bazaar where users can vote and comment on the quality attributes. Based on the votes and comments of the users and expert groups, architecture designer prioritizes the quality attributes.

Furthermore, Requirements Bazaar is used for communication and negotiation continuously between architecture developer, expert groups, and users during the design of architecture and during the evaluation process. Following the formal ATAM is compared with the proposed evaluation method.

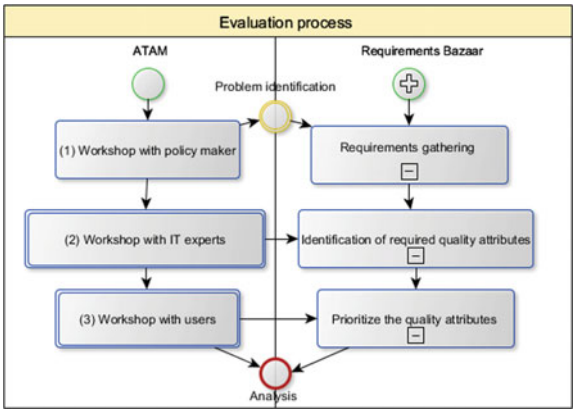


Fig. 2. Design science research evaluation process

4 Experiments

In order to analyze the proposed design and evaluation method, we implement it to the evaluation of SMIS architecture based on AfgREN. Figure 3 shows that three organizations are collaborating in this architecture, the SMIS infrastructure located in the Ministry of Higher Education (MOHE), Kabul University (KU), and Nangarhar University (NU). AfgREN infrastructure acts as the base for exchange of services. The broker component is brokering the services between SMIS systems.

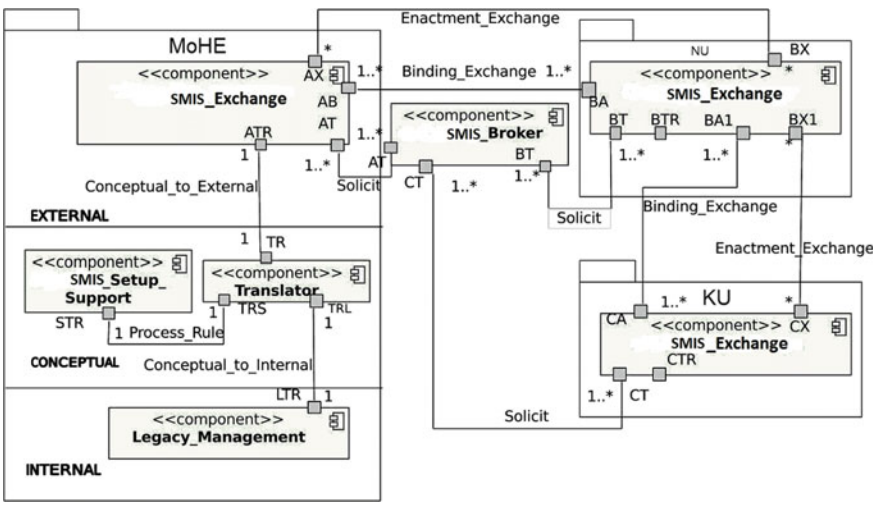


Fig. 3. AfgREN SMIS architecture

The SMIS architecture based on AfgREN has three main components that are collaborating peer to peer and are allocated in three layers including internal layer, conceptual layer, and the external layer. The internal layer is the base and all the operations of the systems are based on it. A Legacy_Management component operates as the base of collaboration between AfgREN members. It contents the rules, process, and organization structure. A Legacy_Management component has local process enactment engines that are wrapped as a web service, also there exist local rules engines and local workflow management system (WFMS). Local rule engine and local WFMS components coordinating through a local coordinator, furthermore they produce data to pass them to conceptual layer. The collaboration and data exchange authenticate based the on national law, organization structure, and cross-organization accepted procedures.

Figure 3, shows the collaboration between MoHE, NU, and KU. Data exchange between parties is based on a web service system, the system has control engine to work based on local rules and local workflow management system. A local coordinator is required to coordinate the local rules and local workflow management systems. After operation of local coordinator, exchangeable data would be isolated from conceptual layer.

Focusing on the fully depicted layers in AfgREN SMIS architecture, the conceptual layer comprises a SMIS_Setup_Support component, in which SMIS services are composed collaboratively between different organizations. On the conceptual layer of the AfgREN SMIS architecture, a pipes-and-filters pattern facilitates establishing communication channels between the external and internal layer via the conceptual layer. This pattern provides a structure for processing streams of heterogeneous data while filter components encapsulate each processing step. Hence, data passes through pipes between adjacent filters from the external layer to the internal layer and vice versa.

The translator component converts heterogeneous data formats between the external and internal layer. Translator component has two main parts that are translating data between internal, conceptual, and external layers. A conceptual external translator (CE_Translator) translates data between conceptual and internal layers to external layer and vice versa. The conceptual internal translator (CI_ Translator) translates date between external, conceptual to internal layer and vice versa.

The SMIS_exchange component located in the external layer, SMIS architectures exist in all collaboration parties, they synchronize each other during setup and approval time of an SMIS exchange. It is a border component for each collaboration party, it prevents the system from threats that may arise from outside. The external layer contains a component for distributed SMIS process prove, it includes a rules engine that supports the distributed SMIS process enactment engine. The SMIS_Exchange component comprises a façade pattern that is an unified interface, which offers to a collaborating counterpart access to a set of interfaces of a subsystem, namely the replicated components of the external layer. Hence, this supports the interoperability between SMIS

collaboration parties and enhances the security in a collaboration as it shields the legacy systems behind the façade of the external layer.

For SMIS collaboration, a Broker exists in the middle in the form of a SMIS_Broker component that facilitates the rapid matching of SMIS needs with requests of users. The SMIS_Broker component uses a publish/subscribe style in which publishers submit new SMIS services and all subscribers receive a notification automatically. The notifier forms the central component of a star topology where the publishers and subscribers are the leaves. The advantage of this style in a multiparty collaboration environment with large numbers of potential SMIS service consumers and providers has enhanced system performance because of reduced communication overhead and an enhancement of flexibility and integrability of additional national SMIS platforms. The user has access to students profile and the systems collaborate with each others.

We held six ATAM workshops [13], four of them were implemented without Requirements Bazaar and in two of them we used Requirements Bazaar. The first workshop implemented in EENeT in Tartu of Estonia, the author of designed architecture, technical manager of EENeT and two experts from e-learning center of Estonia participated in the workshop. The second ATAM workshop took place in Kabul University and at least 40 professors, Ph.D. candidates, masters and bachelor students and technical experts participated. Based on the designed architecture, we selected ATAM for evaluation. The third workshop was conducted in the e-learning board of Kabul Medical University, six member of e-learning board participated in the workshop. The ATAM process, architecture, scenario, and quality attributes were presented and the discussion for scenario ranking and prioritization was the result of this workshop. The fourth workshop was implemented in Avicenna private university where participated the 15 bachelor students in their final year who are working in industries, the e-Testing scenarios were explored and discussed in that workshop. We held the fifth workshop also in Avicenna private university. We gathered the relevant experts and presented the scenarios and mapped it to the architecture. Twenty experts participated in the workshop: the former manager of the AfgREN, professors of software engineering from four universities including Kabul University, Kabul Polytechnic University, Avicenna University and Kateb University, the IT advisor of the Ministry of Higher Education of Afghanistan and the IT advisor from the Ministry of Interior Affairs of Afghanistan, and other IT experts.

We used the Requirements Bazaar for adding comments and prioritizing the quality attributes of the designed architecture. Sixth workshop that was part of a seminar took place in the Rabia Balkhi private university, policymakers and around 50 computer science students, IT expert, computer science teachers from different universities participated in it. After the workshop a discussion panel assigned that architecture developer, the representative of the Ministry of Communication and IT of Afghanistan and a computer science professors participated in the panel. We discussed around one hour and collected the comments of other workshop participants. We found that the Requirements Bazaar make much easier the process of the evaluation of the architecture. It can provide a

virtual evaluator team, Based on the portal admin report AfgREN SMIS project became the second most popular project in the Requirements Bazaar.

The result of evaluation using Requirements Bazaar shows that it makes the process much faster, it can provide a large number of expert groups, it can provide a digital format evidence that researcher can be used for evaluation of large and scalable socio-technical systems. The proposed design and evaluation process are for a large scalable systems. It fulfills a large number of functional and non-functional requirements, reduces the time and cost for requirements realization, provides a large number of virtual expert groups, as well as, a negotiation opportunity, and architecture developers always can review the result and redesign the architecture.

Result of formal ATAM workshops and our proposed evaluation method that included Requirements Bazaar shows that our proposed method is more effective evaluation method than former ATAM. It covers environment-related quality attributes and considers the votes and comments of a large number of expert groups.

5 Conclusion

In this research, we designed an evaluation method that can be applied for evaluation of large and scalable e-governance systems. It is a modified version of ATAM that evaluates an architecture in three ATAM workshops and Requirements Bazaar is used in parallel with the ATAM workshops, while formal ATAM has seven steps and is applied without any extra tools. We designed SMIS architecture based on the AfgREN as a case in this research, the proposed evaluation method was applied for evaluation of SMIS based on AfgREN.

The result of evaluation of SMIS based on AfgREN shows that proposed evaluation method is resource and time effective method that supports tools and provides digitally documented evidence. With application of proposed evaluation method, designer of architecture can share the project with large number of expert groups. In this method, architecture designer can negotiate with the user and expert groups in different steps of the design and evaluation process. Using Requirements Bazaar together with the ATAM offers more opportunities for the researcher to collect more data with documented evidence, with no physical presence of the target expert groups and with accurate data. Using Requirement Bazaar instead of several workshops is the most suitable method for evaluation of architectures. However as the Requirements Bazaar is not matured enough and that is a new project, this method also needs to be evaluated in more projects.

References

1. Angeline Julia S, Rodrigues P (2016) Novel creative innovative patterns for architecture analysis (CIPA). Indian J Sci Technol 9(30) (2016)
2. Bengtsson P, Bosch J (1999) Architecture level prediction of software maintenance. In: Proceedings of the third European conference on software maintenance and reengineering, 1999. IEEE, pp 139–147

3. Bengtsson P, Bosch J (1998) Scenario-based software architecture reengineering. In: Proceedings of the fifth international conference on software reuse, 1998. IEEE, pp 308–317
4. Drechsler A (2015) A postmodern perspective on socio-technical design science research in information systems andreas, vol 9073, pp 152–167
5. Gampa S, Yazhini US, Manikandan N (2016) Methods for evaluating software architecture-a survey. Int J Pharmacy Technol 8(4):22395–22404
6. Kazman R, Klein M, Clements P (2000) ATAM : method for architecture evaluation. CMU SEI 4(August):83
7. Lassing N (2002) Architecture-level modifiability analysis. SIKS Dissertation (2002):1
8. March ST, Storey VC (2004) Design science in the information systems discipline: an introduction to the special. 32(1):725–730
9. Nussbaumer A, Kravcik M, Renzel D, Klamma R, Berthold M, Albert D (2014) A framework for facilitating self-regulation in responsive open learning environments. 1:1–41
10. Offermann P, Levina O, Schönherr M, Bub U (2009) Outline of a design science research process, p 7
11. Patidar A, Suman U (2015) A survey on software architecture evaluation methods. In: 2015 2nd international conference on computing for sustainable global development (INDIACom). IEEE, pp 967–972
12. Peffers K, Tuunanen T, Gengler CE, Rossi M, Hui W, Bragge J, Virtanen V (2006) The design science research process: a model for producing and presenting information systems research. J Manag Inform Syst 24(3):24
13. Saay S, Norta A (2017) An architecture for e-learning infrastructures on a national level : a case study of the afghanistan research and education network. 3:1–27
14. Shanmugapriya P, Suresh RM (2012) Software architecture evaluation methods-a survey. Int J Comput Appl 49(16)
15. Vaishnavi VK, Kuechler B (2010) Design research in information systems. In: MIS quarterly, vol 28. Springer, Berlin, pp 75–105
16. Vaishnavi VK, Kuechler W (2015) Design science research methods and patterns: innovating information and communication technology. CRC Press
17. Venable J, Pries-heje J, Baskerville R (2014) FEDS: a framework for evaluation in design science research. Eur J Inform Syst 25(Oct 2012):1–13

A Low-Noise Low-Cost EEG Amplifier for Neural Recording Applications

Pavan Varma Tirumani[1(✉)], Soukhin Das[2(✉)], Piyush Swami[1], and Tapan Gandhi[1]

[1] Department of Electrical Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India
paltirumani@gmail.com
[2] Department of Electrical Engineering, National Institute of Technology Durgapur, M.G. Avenue, Durgapur 713209, India
soukhind2@gmail.com

1 Introduction

There is a constant demand for low-cost biopotential amplifiers among the neuroscientists and doctors in order to observe the neural activities in brain. Cerebral signals in the form of EEG are one of the most important biopotential signals from the human body which is becoming a standard practice in basic neurological research and paving the path for neuroprosthetic applications. Recent advances in biomedical instrumentation have produced very small microelectrodes (less than 4 mm^2) containing hundreds of recording electrodes. The next generation of neural amplifier demands the development of systems capable of observing thousands of neurons in a single implanted unit [1–3]. Despite advancement in biomedical instrumentation and integration of electronics, the use of EEG has been largely limited due to the inconvenience, cost, and inadequate results in the EEG recording technologies. In the past, there has been a lot of work in electronics in the development of small-scale biopotential amplifiers [4–14], most of them do not produce optimum results which are often unacceptable due to high-noise levels in the recordings or consume large amount of power thus making it unsatisfactory to be implemented for large-scale multielectrode neural recordings. This paper describes the design, development, and testing of a biopotential amplifier used for EEG recordings. The circuit offers the best signal–noise tradeoff of any such amplifier reported. The amplifier is developed to record brain signals noninvasively in the range of 0.5–45 Hz [15, 16].

P. V. Tirumani and S. Das have equally contributed to this chapter.

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5_65

2 Methodology

2.1 System Design

The developed system is an eight-channel EEG amplifier. Signals (EEG/EOG) are acquired from the subject through noninvasive gold (Au) electrodes. Figure 1 shows the high-level schematic of the system. The acquired signals can be compared to a voltage source V_S ranging between ± 100 μV, consisting of the required EEG data and noise from surroundings and acquisition. The main challenge is to design the system which is able to amplify the low amplitude EEG signal and suppress noise and unwanted artifacts acquired. The initial part of the system is an ultra-high impedance instrumentation amplifier that cancels the differential component of the signals and provides high CMRR providing an amplitude gain of 152.5 to the signal. One major problem of using instrumentation amplifier is the continued presence of unpredictable DC drifts (in μVs) which tends to saturate the signal at some point in time. This small DC drift if not removed will affect the signal as the amplification rate of the later part of the circuit is very high. In order to counter the drifts, the signal is coupled to a first-order Sallen key high-pass filter having a cutoff frequency of 0.5 Hz. On the later stage of the system, the signal passes through a fifth-order low-pass Butterworth filter. Butterworth filter is used for its maximally flat frequency response in its passband. The cutoff frequency of the filter was set to 46.8 Hz with a gain of 22.2 dB to remove the unwanted high-frequency noise and amplify the signal. The passband mainly comprises of the information related to EEG as acquired from the subject. The output of the Butterworth filter is passed through two twin T notch filters of frequency 50 Hz each of second order to eliminate the power line noise from the signal (power line frequency is 50 Hz in India). The main aim of using a fourth-order notch is to increase the notch frequency depth. Finally, the signal is passed through a 0.5 Hz high-pass filter to remove any small DC drift associated with the signal. The final output signal is in the range of ± 1 V.

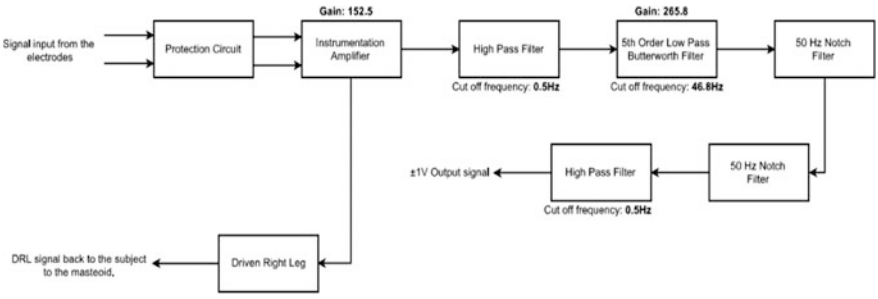


Fig. 1. Schematic of the system showing the gain and cutoff frequency at each stage

2.2 Circuit Description

During the design of the amplifier, three main requirements were considered (Fig. 2 and Table 1).

1. All the signal acquired from the subject was protected from the rest of the amplifier through the implementation of diodes.
2. Linear phase condition in the passband was achieved.
3. High common mode rejection ratio for filtering out most of the noise that is generated during acquisition.

One diode was connected between the amplifier channel and the subject in order to block any reverse current such as electric faults surges to protect the subject. The linear phase condition in the signal was maintained with the implementation of a fifth-order low pass Butterworth filter. Butterworth filter is a filter designed to have as flat as

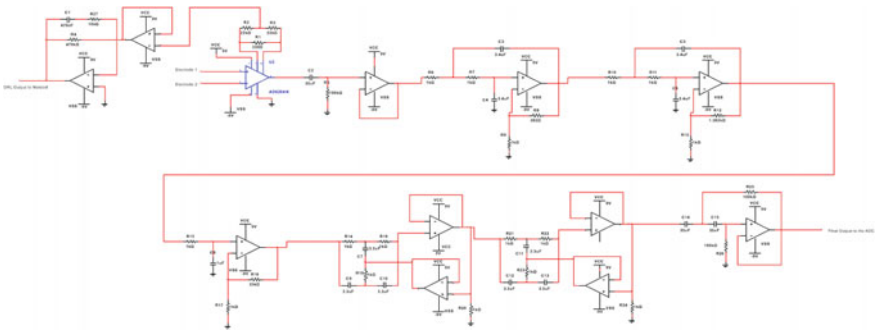


Fig. 2. Circuit diagram of one amplifier channel

Table 1. Experimental characteristics of EEG amplifier

Parameter	Measured
Supply voltage	±9 V
Supply current	6 mA
Gain	91 dB
Bandwidth	44.5 Hz
Low-frequency cutoff	0.5 Hz
High-frequency cutoff	45 Hz
CMRR	100 dB
Resolution	16 bit
Channels	8

frequency response as possible in the passband. Finally, the high common mode rejection ratio (CMRR) was achieved by using INA 114 as the instrumentation amplifier. INA 114 has a high CMRR in the range of 85–120 dB [17]. The signal acquired from the body is first passed through a diode to isolate the subject from the amplifier and to block any digital noise.

Initial Stage Amplifier

The signal is provided as inputs to the instrumentation amplifier. For this purpose, INA 114 by Burr-Brown [18] is used which is well known for its very high input impedance, low offset voltage (50 μV), and high common mode rejection (CMR).

The gain of the amplifier can be written as

$$Av = 1 + \frac{50 * 10^3}{Rg} \tag{1}$$

For our purpose, Rg was set to 330 Ω. The gain of the instrumentation amplifier was 43.6 dB. The output of the INA 114 sometimes offers some DC drifts in the range of 150 μV. To counter the drifts offered by the instrumentation amplifier, the output is passed through a first-order Sallen key high-pass filter set at 0.5 Hz.

Secondary Amplification Stage

This part of the system has been made with high input impedance operational amplifiers (op-amps). TL074 ICs have been used which consist of four high-precision op-amps. This IC has been chosen for its high-precision and very low DC drift [19].

The signal is then passed through a fifth-order Butterworth filter for low-pass filtering and removal of unwanted high-frequency bands in the signal. A Butterworth filter offers a linearly flat response in the passband [18]. The filter offers a gain of 48.49 dB amplifying the signal without any ripple in frequencies. The signal now contains the important EEG information devoid of unwanted high-frequency noise and other unwanted artifacts.

Since the amplifier has to be operated in normal conditions, the signal also has a significant amount of power line noise, i.e., noise at 50 Hz. So the amplified signal from the Butterworth filter has to be filtered to get rid of the 50 Hz frequency noise. The signal is filtered with two twin T 50 Hz notch active filters. The main aim of using two notches in the circuit is to make the stop band notch deeper and remove the power line frequency efficiently without the loss of important biopotential data. Finally, the signal is passed through a high-pass filter of cutoff frequency 0.5 Hz to remove any DC drift added by the op-amps. The final signal at the output varies in the range of ±1 V.

3 Experimental Results and Discussion

The circuit was tested in two stages, in the first stage, the output was recorded for various simulated signals and then in the second stage, the amplitudes of open/closed eyes test are recorded. For the first stage, pure sinusoidal signals (1, 10, 20, 30, 40, 50 and 100 Hz) of amplitude ±100 μV are given to the two input terminals of the circuit, the DRL terminal was not used for this part of the simulation. Now the output was measured at the final stage high-pass filter and amplitude variations are recorded. These tests are conducted to measure the linearity in the passband, the roll-off in the stop band, the linearity in phase response and to measure the simulated power supply noise (50 Hz sinusoidal signal).

In the second stage, the open/closed eyes test was conducted and relative amplitudes were measured. The test can be conducted in both unipolar and bipolar modes. In unipolar mode, the voltage difference at a node (say P_4) with respect to a reference node (F_{pz}) was recorded, whereas in bipolar mode, the voltage difference between a pair of points like O_1, O_2 was measured. For our testing, the EEG signals were recorded in a unipolar mode. The occipital lobe present at the back of the head is responsible for visual processing and a significant change in alpha signals (EEG signals of band 8–13 Hz) in different states of eyes (closed and open). When eyes are open, the alpha signal amplitude is low and in the closed state, the alpha signals are more prominent [20]. To measure this, a subject with healthy brain activity was chosen and the positive input electrode was put on the scalp at either O_1 or O_2 position (say O_1) of the subject and the negative input electrode was placed on the F_{pz} of the subject, the reference node. The DRL electrode should be placed on the subject's mastoid (the temporal bone of skull behind the ear) to drive the body with common mode signal [21]. Proper care was taken to ensure that the electrodes remained in electrical contact with the scalp using conductive gel. It was also ensured that electrodes do not slide or change position during the course of recording. Now after asking the subject to relax for 5 min, he/she was asked to keep eyes open and closed alternatively for periods of 10 s on the command of the experimenter. Figure 3 represents the recorded signal consisting of the alpha rhythm activity in open and closed eyes state.

The frequency response of these signals was taken to detect the peak in alpha range (8–13 Hz) in both open and closed states. In Fig. 4, the FFT of the signals records the frequency response of the system from 0.5 to 60 Hz in which maximum response has been recorded in the biopotential frequency range. In both eyes-open and -closed situation contrasting alpha rhythm activity has been observed. Also, the signal–noise tradeoff as inferred from Fig. 4 supports low-noise output of the designed system.

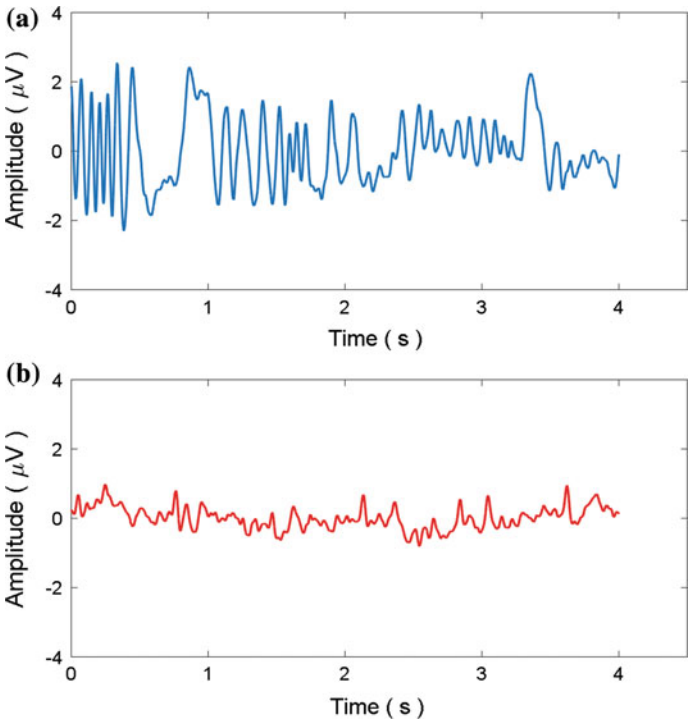


Fig. 3. EEG acquired during, **a** eyes-closed and, **b** eyes-open conditions

Student's t-test was conducted to statistically test the difference in the EEG recordings during the eyes-open versus eyes-closed conditions. The t-test outcome revealed highly significant p-value (<0.001). The findings were also compared with other previous works. The noise levels measured in our system have been found to be significantly less than that in other works. The FFT response in Fig. 4 and the t-test results in Fig. 5 clearly revealed the difference as compared with the results in Fig. 12 of [3], Fig. 8 of [12], etc.

This novel initiative was aimed at the development of a low-cost system. Each amplifier channel is manufactured in a daisy chain configuration so that extra amplifier channels can be added or removed to the system as and when needed so that there is no limitation in the number of channels. Further advantages and disadvantages of the system have been discussed in the following table (Table 2).

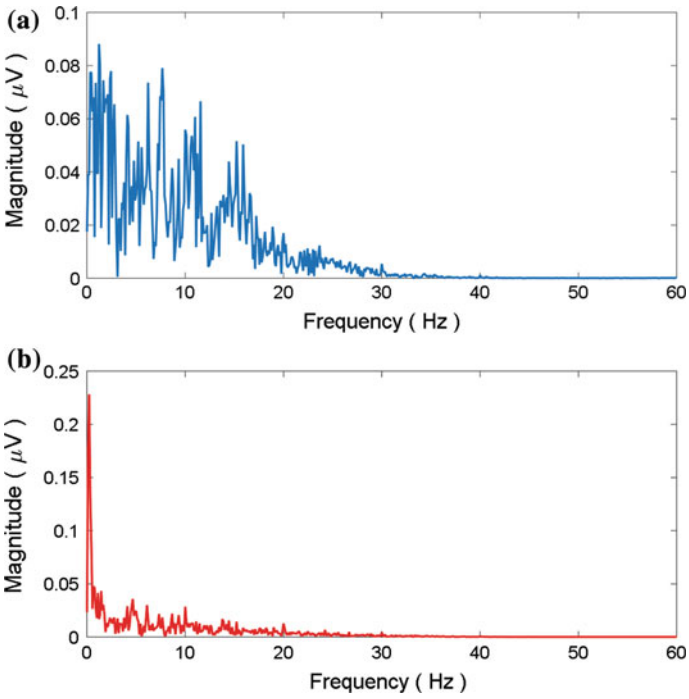


Fig. 4. FFT of EEG acquired during, **a** eyes-closed and, **b** eyes-open conditions

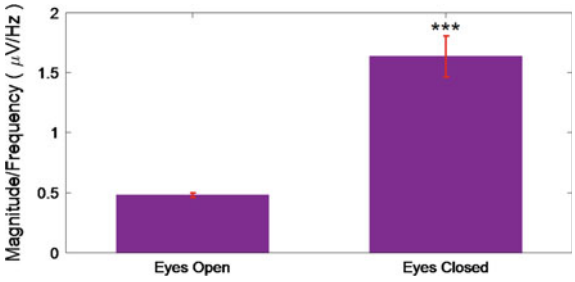


Fig. 5. t-test analysis of eyes-open versus eyes-closed states of EEG

Table 2. Advantages and disadvantages of the current system

Advantages
(1) The biopotential amplifier demonstrated in this manuscript provides a cost-effective solution for acquisition of scalp EEG
(2) The output data showcased highly significant ($p < 0.001$) change while capturing α-rhythms during eyes-open and eyes-closed conditions (shown in Fig. 5)
Disadvantages
(1) Present amplifier design is limited to eight channels. However, the number of channels of the existing design could be extended by daisy chaining additional amplifier channels

4 Conclusion

The final output signal is digitized through an A/D converter (ADS8588). It has eight channels for simultaneous sampling. The novel biopotential amplifier described in this paper was found to acquire the EEG signals in a real-time environment in daylight situation. This is considered as a major improvement over the existing biopotential amplifiers that have restricted neurological research and BCI applications [22, 23] from being widespread. The amplifier presented in this paper is designed to pass signals in 0.5–45 Hz bandwidth. In addition to attenuating high-frequency noise and power line noise, the fifth-order Butterworth filter implemented in this paper acts as an anti-aliasing filter which permits low-frequency signals such as EEG or ECG. When the amplifier is connected in a daisy chain configuration, the ADS8588 is set to lower sampling frequency. The system finds its application in brain–computer interfacing, detection of epileptic seizures, and real-time clinical tests.

In future, the circuit can be extended to 64 channels for use in neurological research and clinical tests. Also opto-couplers can be implemented in the system in place of diodes for better suppression of digital noise and isolation of the amplifier circuit form the subject as well as power line.

References

1. Hoogerwerf AC, Wise KD (1994) A three-dimensional microelectrode array for chronic neural recording. IEEE Trans Biomed Eng 41:1136–1146
2. Nordhausen CT, Maynard EM, Normann RA (1996) Single unit recording capabilities of a 100-microelectrode array. Brain Res 726:129–140
3. Harrison Reid R, Charles Cameron (2003) A low-power low-noise CMOS amplifier for neural recording applications. IEEE J Solid State Circuits 38(6):958–965
4. Degrauwe M, Vittoz E, Verbauwhede I (1985) A micropower CMOS instrumentation amplifier. IEEE J Solid State Circuits SC-20:805–807
5. van Peteghem PM, Verbauwhede I, Sansen WMC (1985) Micropower high-performance SC building block for integrated low-levelsignal processing. IEEE J Solid-State Circuits SC-20:837–844

6. Dorman MG, Prisbe MA, Meindl JD (1985) A monolithic signal processor for a neurophysiological telemetry system. IEEE J Solid State Circuits SC-20:1185–1193
7. Harrison RR (2007) A versatile integrated circuit for the acquisition of biopotentials. In: IEEE 2007 Custom integrated circuits conference, CICC'07. IEEE
8. Tseng Y et al (2012) A 0.09 μW low power front-end biopotential amplifier for biosignal recording. IEEE Trans Biomed Circuits Syst 6(5):508–516
9. Ji J, Wise KD (1992) An implantable CMOS circuit interface for multiplexed microelectrode recording arrays. IEEE J Solid-State Circuits 27:433–443
10. Pancrazio JJ, Bey PP Jr, Loloee A, Manne S, Chao HC, Howard LL, Gosney WM, Borkholder DA, Kovacs GTA, Manos P, Cuttino DS, Stenger DA (1998) Description and demonstration of a CMOS amplifier-based-system with measurement and stimulation capability for bioelectrical signal transduction. Biosensors Bioelectron. 13:971–979
11. Martins R, Selberherr S, Vaz FA (1998) A CMOS IC for portable EEG acquisition systems. IEEE Trans Instrum Meas 47:1191–1196
12. Gargiulo G et al (2010) An ultra-high input impedance ECG amplifier for long-term monitoring of athletes. Med Devices (Auckland, NZ) 3:1
13. Chandran AP, Najafi K, Wise KD (2002) A new DC baseline stabilization scheme for neural recording microprobes. In: Proceedings of IEEE BMES/EMBS conference, 1999, p 386
14. Yazicioglu RF et al (2007) A 60 μW 60 nV/√Hz readout front-end for portable biopotential acquisition systems. IEEE J Solid-State Circuits 42(5):1100–1110
15. Nagel JH (2000) Biopotential amplifiers.In: Bronzino JD (ed) Biomedical engineering hand book, 2nd edn. Springer, New York, pp 70–71
16. Neuman MR (1998) Biopotential amplifiers. In: Medical instrumentation: application and design, pp 292–296
17. www.ti.com/product/TL074
18. Van Vollenhoven E, Reuver H, Somer J (1965) Transient response of butterworth filters. IEEE Trans Circuit Theory 12(4):624–626
19. http://www.ti.com/lit/ds/sbos014/sbos014.pdf
20. Feige B et al (2005) Cortical and subcortical correlates of electroencephalographic alpha rhythm modulation. J Neurophysiol 93(5):2864–2872
21. Gudaitis AM (1995) Virtual right leg drive and augmented right leg drive circuits for common mode voltage reduction in ECG and EEG measurements. US Patent 5,392,784 28 Feb 1995
22. Deedwania S, Gandhi TK (2016) An ensemble approach for brain computer interface applications. In: 2016 IEEE Region 10 Humanitarian technology conference (R10-HTC). https://doi.org/10.1109/r10-htc.2016.7906812
23. Mohseni P, Najafi K (2002) A low power fully integrated bandpass operational amplifier for biomedical neural recording applications. In: Proceedings of IEEE EMBS/BMES conference, pp 2111–2112

Author Index

S. Bhattacharyya et al. (eds.), *Advanced Computational and Communication Paradigms*, Lecture Notes in Electrical Engineering 475,
https://doi.org/10.1007/978-981-10-8240-5