Lecture Notes in Electrical Engineering 765

Prabin K. Bora
Sukumar Nandi
Shakuntala Laskar *Editors*

Emerging Technologies for Smart Cities

Select Proceedings of EGTET 2020

Springer

Lecture Notes in Electrical Engineering

Volume 765

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

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

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

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

China

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

India, Japan, Rest of Asia

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

Southeast Asia, Australia, New Zealand

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

USA, Canada:

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

All other Countries:

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

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

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

Prabin K. Bora · Sukumar Nandi ·
Shakuntala Laskar
Editors

Emerging Technologies for Smart Cities

Select Proceedings of EGTET 2020

Springer

Editors
Prabin K. Bora
Department of Electrical and Electronics Engineering
Indian Institute of Technology Guwahati
Guwahati, India

Sukumar Nandi
Department of Computer Science and Engineering
Indian Institute of Technology Guwahati
Guwahati, India

Shakuntala Laskar
Department of Electrical and Electronics Engineering
Assam Don Bosco University
Guwahati, India

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-16-1549-8 ISBN 978-981-16-1550-4 (eBook)
https://doi.org/10.1007/978-981-16-1550-4

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

Preface

The International Conference on Emerging Global Trends in Engineering and Technology (EGTET) 2020 is the first International conference in the EGTET series, organized by Assam Don Bosco University as a continuation of its efforts to promote an environment of research and development in Northeast India. EGTET 2020 has provided a platform to scientists and researchers from all over India and abroad to present and share their contributions towards the achievement of sustainable technology.

EGTET 2020 attracted researchers not only from varied fields of Engineering and Technology but also from Physics and Chemistry. A total of 63 papers were received from various institutions all over India and neighbouring countries. A double-blind review process was then adopted for selecting the papers. After proper scrutiny of the research papers, around 30 papers were accepted out of which 23 papers were presented at the conference. Apart from the papers by the researchers, there were around 6 invited lectures presented at the conference by renowned scientists from India and abroad.

The proceeding of EGTET 2020 is a collection of 23 good-quality articles. These articles cover a plethora of topics such as IoT-based technology, Machine Learning, Sustainable Renewable Energy, Deep Learning, Biomedical Engineering, Reinforced Concrete Technology and Antenna Design.

We would like to convey our heartfelt gratitude to the Honorable Vice-Chancellor of Assam Don Bosco University, Fr. (Dr.) Stephen Mavely, for his constant support and motivation. We also would like to thank the Pro Vice-Chancellor of Assam Don Bosco University, Fr. Joseph Nellanatt, and Director—School of Technology, for their guidance and support in organizing this event. Our sincere gratitude to all keynote address presenters, invited speakers, session chairs, editors and other high officials from various organizations for their gracious presence on the campus on the occasion.

We thankfully acknowledge all the researchers for their contribution to this conference. We acknowledge all the organizations who have extended financial help towards this event.

We hope that the researchers from varied fields of Engineering and Technology will be benefited from the articles of this proceeding. Technologists from Industry would also find this volume to be a good source of reference.

Guwahati, India

Prabin K. Bora
Sukumar Nandi
Shakuntala Laskar

Contents

Integration of IoT and Blockchain Technology for Smart Cities 1
Jerry Casper Kharbhih, Kausthav Pratim Kalita, and Rup Kumar Deka

An IoT and Machine Learning-Based Crop Prediction System for Precision Agriculture 9
Saria Parween, Arunangshu Pal, Itu Snigdh, and Vinay Kumar

A Smart Feature Reduction Approach to Detect Botnet Attack in IoT 17
Rup Kumar Deka, Kausthav Pratim Kalita, Dhruba Kumar Bhattacharyya, and Debojit Boro

An Approach to Handle Heterogeneous Healthcare IoT Data Using Deep Convolutional Neural Network 25
Kishore Medhi, Md. Arifuzzaman Mondal, and Md. Iftekhar Hussain

Designing of *NimbleArm*—A Low-Cost, Semi-autonomous Interactive Robotic Arm 33
Abhishek Sarkar and Gypsy Nandi

Smart Refri: SMART REFRIGERATOR for Tracking Human Usage and Prompting Based on Behavioral Consumption 45
Aditya Sharma, Ashmita Sarkar, Aazaz Ibrahim, and Rupam Kumar Sharma

Cascade-Based Pedestrian Detector Using Edge and Pattern Features 55
Amlan Jyoti Das, Navajit Saikia, and Simantika Choudhury

Nanomaterial-Based Microstrip Patch Antenna Array for X Band Operation 65
Parismita A. Kashyap, Smriti Rekha Das, and Sunandan Baruah

Study of TiO_2 and ZnO Nanostructures as Wide Band Gap Semiconductor for the Application of Nanotechnology-Based Solar Cells 77
Karen Das, Akib Khan, Arnab Sarkar, Pronoy Baishya, Priyanka Kakoty, and Sunandan Baruah

AGC of Hybrid Solar-Hydro-Thermal System with GWO-based Conventional Secondary Controllers 85
Asadur Rahman, Lalit Chandra Saikia, and Yatin Sharma

Study on Substrate Dependency of Graphene-Based Patch Antennas for Gigahertz and Terahertz Applications 97
Dilruba Khanam, Kaustubh Bhattacharyya, Kandarpa Kumar Sarma, and Sunandan Baruah

Novel Approach to Reduce Rate of False Fire Alarms Through Multivariate Characterization of Fire 109
Pranjal Kumar, Subham Naskar, and Arpit Raj

A Graph-Based Band Selection Method for Hyperspectral Images Using Correlation Matrix 119
Jintu Kumar Das, Christopher D. Tholou, Alok Anand Minz, and Sonia Sarmah

A Study on Using Deep Learning for Segmentation of Medical Image 127
Lal Omega Boro and Gypsy Nandi

Attention Recognition System in Online Learning Platform Using EEG Signals 139
Swadha Gupta and Parteek Kumar

Comparative Study of Jute Fiber and PET Fiber-Reinforced Concrete 153
Rony Kutum, Piyush Singh, and Anirban Saha

Reused-Based Replacement Policy for Last-Level Cache with Minimum Hardware Cost 161
Purnendu Das and Bishwa Ranjan Roy

Revealing Structural and Optoelectronic Properties for Bi-Doped $CuGaS_2$ Chalcopyrite: A Density Functional Investigation 171
Karina Khan, Aditi Gaur, Amit Soni, U. Ahuja, and J. Sahariya

Designing Lightweight S-Box Using Simplified Finite Field Inversion Mapping 179
Mebanjop Kharjana, Fabiola Hazel Pohrmen, and Goutam Saha

Review of Performance of a Single Basin Passive Solar Still 187
Ranbir Kalita, Parimal Bakul Barua, and Deva Kanta Rabha

Flat Plate Solar Thermal Collectors—A Review 197
Dhrupad Sarma, Parimal Bakul Barua, Deva Kanta Rabha, Nidhi Verma, Soumyajyoti Purkayastha, and Sudipta Das

About the Editors

Prof. Prabin K. Bora is currently a Professor in the Department of Electronics and Electrical Engineering, Indian Institute of Technology, Guwahati. He obtained his B.E. (Electrical Engineering) from Gauhati University (Guwahati), M.E. (Electrical Engineering) and Ph.D. from the Indian Institute of Science, Bangalore. He has nearly four decades of teaching experience. Prior to joining Indian Institute of Technology, Guwahati, he had worked in premier institutions of North East India like Jorhat Engineering College and Gauhati University. His major areas of research interests include Signal Processing; Image Processing, Probability and Random Processes. He has published 25 papers in reputed international journals. Prof. Bora is also actively involved in National Level Extension Activities like developing NPTEL-I Web-based course on Probability and Random Processes. He has completed many sponsored research projects and currently has few ongoing projects.

Prof. Sukumar Nandi is currently a Professor in the Department of Computer Science and Engineering and Head Centre for Linguistic Science and Technology, Indian Institute of Technology, Guwahati. He obtained his B.Tech (Electrical Engineering) from Calcutta University, M.Tech (Computer Science) from Calcutta University and Ph.D. from the Indian Institute of Technology, Kharagpur. He has more than two decades of teaching experience. His major areas of research interests include Networks (specifically QoS, wireless networks), Computer and Network Security, Distributed Computing, Data Mining, VLSI Design and Test, Computational Linguistic. He has published 86 papers in reputed international journals and co-authored one book. Prof. Nandi is a Fellow of Indian National Academy of Engineering, Senior Member ACM, Senior Member IEEE, Fellow of the Institution of Engineers (India) and Fellow of the Institution of Electronics and Telecommunication Engineers (India). Currently, he is serving as a member of School Board of SOCIS, IGNOU and Member Senate, CIT Kokrajhar.

Prof. Shakuntala Laskar is currently the Head and Professor in the Department of Electrical and Electronics Engineering, School of Technology, Assam Don Bosco University, Guwahati. She obtained her B.E. (Electrical Engineering) from Gauhati University (Guwahati), M.Tech (Electrical Engineering) and Ph.D. from the

Indian Institute of Technology, Kharagpur. She has nearly four decades of teaching experience. During her teaching career, she had worked in reputed institutes like Assam Engineering College, Guwahati and Samundra Institute of Maritime Studies, Lonavala. Her major areas of research interests include Electro-optics, MEMS and Sensors. She has published quite a number of papers in reputed national and international journals. Prof. Laskar is also the Editor-in-Chief of ADBU Journal of Engineering and Technology (AJET).

Integration of IoT and Blockchain Technology for Smart Cities

Jerry Casper Kharbhih, Kausthav Pratim Kalita, and Rup Kumar Deka

Abstract In this modern era of the digital revolution, urban development is growing rapidly and many cities around the world are continuing to develop in many areas be its health sector, government sector, education sector and so on. But there is no denial that information and technology (IT) is one of the main contributors to this development. With the growing demand for smart devices such as smartphones along with various internet services, there is no doubt that communication has been made extremely easy for everyone. IoT is one of the emerging technology that can contribute to a smart city. It is seen that many IoT-based systems are currently used in various sectors from healthcare to parking systems and is proved to be efficient. But IoT suffers from security issues and is vulnerable to attacks. Blockchain on the other hand is a technology that is secured by design. It uses cryptography techniques to secure the data and is meant to be a trusted technology in a trust-less environment. In this paper, we discuss the existing systems that integrate these two technologies and mention the challenges faced when such systems are deployed in a working environment.

Keywords Smart cities · Internet of things · Block · Blockchain · Distributed · System

1 Introduction

If we see in recent years, urban growth has increased due to the increase in population. More and more people are moving into cities mainly because the opportunities

J. C. Kharbhih (✉) · R. K. Deka
Assam Don Bosco University, Azara, Assam, India
e-mail: jerrykharbhih@gmail.com

R. K. Deka
e-mail: rup.deka@gmail.com

K. P. Kalita
Tezpur University Napaam, Tezpur, Assam, India
e-mail: koztov.project@gmail.com

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_1

in rural areas are quite less comparatively. So to earn a living, people from rural areas are moving into cities to get more opportunities so as to make a living. With the increase in population, many challenges also arise. Modern technology can help in coping with these challenges by optimizing the resources of the city, reducing the cost, provide better facilities that are reliable in various sectors such as health, transport, finance, etc. In the present world, we observe that there is significant growth in the field of IoT and wireless communication. With these technologies, we can easily connect various sensor devices and store the information in the cloud. We can also transmit data more easily between the various devices in a more optimized way with a minimum affordable cost. If these smart systems are integrated with the infrastructure of a city, then that city can be called a smart city. But one of the issues that arise when it comes to wireless networks and IoT is security. Many threats and attacks have been reported in trust-less or not fully reliable. Many approaches have been proposed to secure an ecosystem from such threats and one of the ways where such a system can be secured is by integrating blockchain technology into the smart system. Blockchain technology is secured by design as it uses cryptography techniques such as asymmetric cryptography, hashing and also digital signature to preserve the integrity of stored content. Its properties such as immutability, auditability, transparency, persistence and decentralization make the ecosystem better and more trusted.

1.1 Blockchain

Blockchain represents a chain of information where a single block stores transactions or information about the transfer of digital assets, thus, creating a shared ledger. Blockchain technology is decentralized in nature and there is no central server or central authority. Hence, the peers in the network need to keep the existing of the ledger by storing copies of it in their own machines. To make sure that the correct copy of the blockchain is considered, a consensus has to be reached between the various participants in the peer-to-peer network so that only the legitimate copy of the chain is considered. To solve this, various consensus algorithms have been proposed by different researchers. Among them, proof of stake and proof of work has been extensively used by various blockchain platforms. The blockchain basic element is a block. A block is composed of two distinct components which are the header and the body. The block header keeps the hash of the previous block which is the fundamental principle that allows restoring the immutability feature in this technology. It also enhances the security of the ecosystem. The last field in the block header stores application-specific details. This field is different for different blockchain technologies. Some of the information are block signature, nonce value, other data. The data in the block body depends on the services where blockchain is used. For example, if the service is a cryptocurrency service, then the data will be transaction records, if the service is related to smart contracts, the data will contain data related to the contracts and so on [1, 2].

1.2 Internet of Things

Internet of Things (IoT) serves as a network of energy-efficient devices that communicate among themselves using the internet. IoT technology uses various sensors to collect data from different types of sources ranging from environmental conditions to human health-related measurements. The data once accumulated is stored in a server (preferably could storage) and from the stored data, information is extracted which enables the system to take appropriate actions. Many IoT based system has been developed over the years and has helped a lot in various areas such as healthcare, agriculture, road safety, home security and so on. Research states that by 2020, almost all smart devices will use IoT at its core architecture. A smart city is one that integrates IoT into its infrastructure meaning, almost all of the sectors in the city will use IoT such as smart parking meters, smart healthcare, smart homes, smart environment and so on. But one of the prime challenges of IoT is security. IoT tends to suffer from various types of network attacks. Many methods have been proposed for securing the network and our study reveals that numerous research work suggests the utilization of blockchain to construct a secured platform to run IoT-based applications. With these two technologies, smart applications can be built for smart cities that are user-friendly, trusted and secure as well [3, 4].

2 Existing Systems

In smart cities, IoT applications are the main part of their infrastructure [3]. Blockchain on the other hand is a technology that is secured by design. In this section of the paper, we will see how these two technologies can be integrated and how they can contribute to the infrastructure of a smart city. A smart city security framework using IoT and blockchain is made of four layers namely the physical layer which consists of sensors and actuators, communication layer which consist mainly of network devices and wireless technologies, database layer which is represented by a shared ledger (private or public) and an interface layer containing various smart applications [3]. Each and every layer in the framework is secured using blockchain technology for example in the communication layer, blockchain is used for providing security during data transmission by various wireless communication.

[5] proposes an IoT and blockchain integrated system called “The CitySense System” whereby data from the city like temperature, humidity, traffic and are collected with the help of sensors and also a CitySense mobile app is used to collect data from users who are willing to co-operate and contribute to the system. The mobile app will also allow the transfer of information from the various sensors. Other functionalities of the app include a feedback and reporting system. The data collected is stored in a blockchain and will be analyzed. Based on the data collected, decisions are made for the benefit of the city. A user using the CitySense mobile application will also get the benefit of knowing the status of the environment, quality

Table 1 Existing systems integrating the internet of things and blockchain technology

Smart City issues	Proposed system	Blockchain used	Consensus algorithm	References
Vehicular network	Block-VN	Public	Proof of Work	[10]
Review Of The City's Environment	CitySense	Public/private	Proof of Stake	[5]
Energy Grid	Blockchain based Smart Energy Grid	Public	Proof of Identity	[11]
Data Management	Blockchain based hybrid network architecture for a smart city	Public	Proof of Work	[6]
Traffic and Vehicle Management	Speedy Chain	Private		[12]
Real Estate	Smart Contract For Real Estate	Private	Proof of Work	[13]
Law Enforcement	Custody Tracking System	Public	Proof of Concept	[14]
Efficient electrical energy transaction between prosumers	Blockchain-based peer-to-peer (P2P) energy transaction platform	Public/private	Proof of Work/Proof of Stake/Proof of Concept	[15]

of a particular area in the city and also the problems that are currently going on in the city. This system follows the same security framework in [3] and uses blockchain to secure all four layers.

According to the requirements for a smart city to be well established, data must be transmittable not only within a closed environment but also across cities. In [6], the authors propose a hybrid architecture for smart cities where technologies like blockchain and software-defined networks are merged to enable a scalable platform. The consensus mechanism used in their architecture is proof of work. The network is composed of two categories, namely, the core network and the edge network. Other proposed systems that integrates IoT and blockchain technology systems that can be can be summarized in Table 1.

3 Challenges in Blockchain Technology and IoT Integration

Integration of IoT and Blockchain technology will enable the development of smart systems that are user friendly, trusted, secure and reliable But as with any other technologies, IoT and Blockchain integration have some challenges of which some are discussed below. city, reducing the cost, provide better facilities that are reliable

in various sectors such as health, transport, finance, etc. In the present world, we observe that there is a significant growth in the field of IoT and wireless.

3.1 Storage Capacity

IoT devices when implemented in real-time can generate data in GBs per day. Blockchain may not be able to handle the rate at which the data is generated as it takes some time for a transaction to be verified by the nodes in the network. The consensus mechanism used also contributes to this factor. So hence this will be one of the challenges faced when integrating IoT and blockchain [7].

3.2 Security

One of the main issues of IoT is security. Although using blockchain can solve this issue, but one thing to keep in mind is that the data in the blocks are the only reliable data. There is a possibility that the data from IoT devices are not reliable or may be corrupted, hence using blockchain will end up with a chain of blocks containing non-reliable data or corrupted data which is not desirable. Hence this is another challenge faced when integrating IoT and blockchain [7].

3.3 Processing Power

Systems using IoT consist of various devices connected to each other in a network. These devices have different processing time and power. This means that the various devices will run the same encryption algorithm at their own time and speed. This might result in a case where the speed and time taken by any of the IoT devices for running the encryption algorithm will not meet the required speed and time [8].

3.4 Legal Issues

As Blockchain aims to develop systems that are decentralized and not under the control of any authority or third party, the manufacturers and the service providers will have certain issues. This is one of the reasons many are not adopting blockchain in their business and applications [7, 8]. Apart from the challenges mentioned, IoT and blockchain technology have challenges on their own as well. Examples of issues in IoT are mainly related to security such as jamming attacks, replay attacks, wormhole

attack, sybil attacks, etc. On the other hand, example issues related to blockchain are 51% attack, nothing at stake problem, double spend attack and forks [4, 2].

4 Conclusion

Seeing the advancement in the field of IoT, it will not be long one will see the growth of smart cities in the coming years. With smart devices and smart systems coming up and also with the coming up of 5G wireless technology, it is evident that cities will start adopting smart systems as they are fast and more efficient than traditional systems. It is also seen that a lot of research is done on blockchain technology and there are also some current applications of it such as cryptocurrencies(bitcoin, nxt coin, ppcoin), smart contracts, hyperledger [2, 9]. With these two technologies combined, smart systems can be made that are not only efficient but also secure, reliable and trusted. In this paper, we have discussed about the Internet of Things, blockchain technology, the existing system that integrates both the technologies and finally the challenges faced when integrating both the technologies.

References

1. Zheng Z, Xie S, Dai HN, Chen X, Wang H (2018) Blockchain challenges and opportunities: A survey. Int J Web Grid Serv 14(4):352–375
2. Lin IC, Liao TC (2017) A survey of blockchain security issues and challenges. IJ Netw Secur 19(5):653–659
3. Biswas K, Muthukkumarasamy V (2016, December) Securing smart cities using blockchain technology. In: 2016 IEEE 18th international conference on high performance computing and communications; IEEE 14th international conference on smart city; IEEE 2nd international conference on data science and systems (HPCC/SmartCity/DSS). IEEE, pp 1392–1393
4. Khan MA, Salah K (2018) IoT security: review, blockchain solutions, and open challenges. Fut Gen Comput Syst 82:395–411
5. Ibba S, Pinna A, Seu M, Pani FE (2017, May) CitySense: blockchainoriented smart cities. In: Proceedings of the XP2017 scientific workshops. ACM, p 12
6. Sharma PK, Park JH (2018) Blockchain based hybrid network architecture for the smart city. Fut Gen Comput Syst 86:650–655
7. Reyna A, Mart´ın C, Chen J, Soler E, D´ıaz M (2018) On blockchain and its integration with IoT. Challenges and opportunities. Fut Gen Comput Syst 88:173–190
8. Atlam HF, Alenezi A, Alassafi MO, Wills G (2018) Blockchain with internet of things: benefits, challenges, and future directions. Int J Intell Syst Appl 10(6):40–48
9. BitFuryGroup (2015) Proof of Stake versus Proof of Work
10. Sharma PK, Moon SY, Park JH (2017) Block-VN: A distributed blockchain based vehicular network architecture in smart City. JIPS 13(1):184–195
11. Pieroni A, Scarpato N, Di Nunzio L, Fallucchi F, Raso M (2018) Smarter city: smart energy grid based on blockchain technology. Int J Adv Sci Eng Inf Technol 8(1):298–306
12. Michelin RA, Dorri A, Steger M, Lunardi RC, Kanhere SS, Jurdak R, Zorzo AF (2018, November) SpeedyChain: A framework for decoupling data from blockchain for smart cities. In: Proceedings of the 15th EAI international conference on mobile and ubiquitous systems: computing, networking and services. ACM, pp 145–154

13. Karamitsos I, Papadaki M, Al Barghuthi NB (2018) Design of the Blockchain smart contract: a use case for real estate. J Inf Secur 9(03):177
14. Swan M (2018, June) Blockchain enlightenment and smart city cryptopolis. In: Proceedings of the 1st workshop on cryptocurrencies and blockchains for distributed systems. ACM, pp 48–53
15. Park L, Lee S, Chang H (2018) A sustainable home energy prosumer-chain methodology with energy tags over the blockchain. Sustainability 10(3):658

An IoT and Machine Learning-Based Crop Prediction System for Precision Agriculture

Saria Parween, Arunangshu Pal, Itu Snigdh, and Vinay Kumar

Abstract With the advancement of intelligent devices we stand among a plethora of technologies, tools, state-of-the-art techniques, and proof of concepts for a number of applications that essentially use a huge volume of data. Our precision agriculture system aims towards low input, high accuracy with the help of machine learning and the Internet of things towards sustainable agriculture. This article presents results that show that the prediction of fertilizer with different classifiers can be calculated accurately with corresponding heatmaps. We show that Naive Bayes is more accurate as it depends on probabilistic features. Hence, this classifier can be used for better crop prediction.

Keywords SVM · Naive bayes · Machine learning · IoT · Precision agriculture

1 Introduction

IoT is defined as a system that is built for bigger things rather than smartphones and wireless devices, connected by communication infrastructure with a range of software and work according to the sensed environment without human intervention. Such applications find wide use in Climate Science, Neuroscience, Environmental Science, Precision Agriculture, Epidemiology/Health care, Traffic Dynamics, Crime data, etc. These applications need sensor networks to collect data for the necessary analysis and prediction. IoT (Internet of Things) in smart agriculture refers to the use of cameras,

S. Parween · A. Pal (✉)
Department of CSE, Birla Institute of Technology, Mesra, India
e-mail: palarunangshu@gmail.com

S. Parween
e-mail: saria.parween@gmail.com

I. Snigdh (✉) · V. Kumar (✉)
Department of CSE, National Institute of Technology Jamshedpur, Jamshedpur, India
e-mail: itusnigdh@bitmesra.ac.in

V. Kumar
e-mail: VKumar.cse@nitjsr.ac.in

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_2

sensors, and other devices wherein every action has been done corresponds to related data. India's start-up ecology is the 3rd largest technical ecology in the world with 60-65% start-ups where 70% of them are not more than 7 yrs old. These startups focus on customers as well as agricultural and industrial areas [1] since the Industrial sector contributes 31% of the Indian GDP while the agriculture sector contributing around 16% of the Indian GDP.

Smart agriculture is a need of our modern world to grow more food with effective use of the land. Based on this idea many countries have started the implementation of smart agriculture. Researchers from different parts of the world have proposed and published their ideas to solve the traditional farming challenges. Since there is a need to evaluate and monitor industrial agricultural enterprises, as it is currently a high-tech sector of the economy, precision agriculture finds its important place in the Internet of things applications. Smart agriculture has been realized with diverse technologies, some of which are computationally constrained like wireless sensor networks approach or computationally intensive like machine learning models. Internet of Things is the technology that employs both sensors as well as for analytics for implementing precision agriculture.

Current technologies provide common practices done at the farm level with the help of geo-referencing or meteorological data. Some of them include electromagnetic soil mapping, soil sample collection, and crop yield data collection. Others focus on soil types, soil characteristics, drainage level, Arial imagery and crop and soil index mapping [2, 3].

With the use of intelligent systems and without the requirement of regular human intervention, monitoring the agricultural lands and products would create an efficient system. Smart agriculture applications adapt to handle monitoring and decision-making remotely through intelligent devices like sensors and actuators. The current techniques for precision architecture are enumerated as follows:

1. Use of GPS

Access to position locating satellite enables the farmers to closely monitor their crops against weeds, intruders and weather notifications enabling convenience. However, these require expensive and bulky equipment to be installed in the fields wherein the monitoring of devices becomes an overhead.

2. VRT(Variable Rate Technology)

This enables the variable application of input. It controls the number of inputs like fertilizers, pesticides that farmers apply to a specific location. It allows the soil to recover its lost nutrients, maintains soil fertility, prevents soil erosion. The cultivation of different crops in a sequenced season is done in this type.

Moisture, temperature, pH and soil water are important aspects of plant growth. As the demand for food increases day by day, we need to increase food production with the efficient use of limited resources like freshwater. In this view, IoT can be a game-changer because the IoT devices, especially the sensors, can constantly monitor the different environmental conditions as well as the soil condition by continuously

collecting data about them. With the help of these data, IoT can assist an automatic system to control irrigation, pest control, etc. in a better direction to grow more and quality food [4, 5].

2 Related Work

Wireless sensor-based smart agriculture use microcontrollers to monitor the soil moisture with the help of ground- implanted sensors. The system sends the data from sensor to the database server through the internet and also automatically notifies the users if the moisture level reached a certain level. In [6], the author reviewed different use of WSN technology used in the aeroponic method of cultivation. The aeroponic system is a new and modern plant cultivation technique of agriculture, where plants are cultivated under a completely controlled condition in a closed chamber by giving a small amount of mist of nutrient solution in place of the soil. In a periodical basis, the nutrient mist is ejected through atomization nozzles. They discussed that using WSN we can easily detect and diagnosis the fault at an early stage without depending upon the laboratory test. It is also mentioned that by using WSN remotely a farmer can able to control the aeroponic system. Similarly, IoT-based systems have been used to overcome the constraints of bandwidth and power of sensor nodes, by designing a state of the art for LoRaWAN [7], where sensor nodes are enabled with the feature of low power consumption and long-range communication. They also proposed to learn and adopt an algorithm to extend the operational lifetime of the sensor nodes as well as the WSN. In this work, each sensor node has the capability to decide which data must be sent to the fusion center, for that author proposed a two-step decision algorithm.

Computational intelligence [8–10] is a powerful tool to develop an automated decision-making system for precision agriculture. Many researchers have been shown the advantages of neural networks in many applications like simultaneously handle quality and quantitative information in agro studies, which can be effective for handling nonlinear and linear responses. It has been accepted by researchers that Artificial Neural Networks (ANN) based model can be trained based on different physical soil information to monitor and predict soil health. Some researchers predict the crop product with the help of combining satellite image information, electrical conductivity value of soil particles, soil fertility, and elevation. However, their proposed system was unable to forecast the spatial variation in the yield because the soil fertility and soil conductivity properties are closely correlated to yield. Machine learning is one of the most suitable methods for solving multi-valued statistical problems and it can also work in unsupervised mode with the help of clustering data. Non-linear-based algorithms have also been proposed by researchers for yield prediction, but the results obtained are not very promising. Also, laboratory methods are very time-consuming and expensive.

Ongoing progressions in Machine learning have furnished the science and engineering community with an adaptable and quick forecast system, demonstrating a

huge potential effect. Machine learning in agriculture is a set of well-defined models that collects specific data and apply specific algorithms and steps for the achievement of best-expected results. Machine learning provides more accurate and faster results in analyzing cropping patterns, accurate detection and classification of cropping patterns, and helps to understand the dynamics of the environment. Different algorithms of ML are used to classify cropping patterns like Artificial Neural Networks (ANN), Random forest, Support Vector Machines (SVM), Naive Bayes, and other Bayesian algorithms. In ANN, data are passed through many different layers, all consist of nodes from where the data is passed in a multistep process for pattern recognition. Every node of each layer gets trained on different feature sets depending on the output of the pre-layer. In random forest algorithm, which is a supervised classification algorithm, is used for quantitative analysis in cropping patterns. Contextual information of crops is taken and from this approach, it proceeds in pattern recognition of crops.

3 Proposed Work

Ideally, farmers are already aware of the weather patterns, soil temperature, humidity, expected growth, and other factors. However, their decisions are purely based on human experience and expertise which may sometimes be wrong. It has also been observed that farmers rotate crops to improve diversity and monitor irrigation rates so that salts do not accumulate. Other precision agriculture practices are applying nutrients, water, seeds, and other agricultural inputs to grow more crops in a wide range of soil environments. For this, a large number of field data is taken to analyze the performance of crops in various temperatures, rainfall, sunlight hours. These characteristics are correlated and finally, a probable model is built which decides what crops will most likely contribute with following climate and soil factors. For our prediction algorithm, we employ Naive Bayes and SVM algorithm approaches for feature extraction and classification purposes. These machine learning techniques are implemented and tested finally, and a comparative analysis is recorded for optimal cropping. For the collection of the requisite data, we have used GIS datasets from GIOVANNI websites.

3.1 Naive Bayes [11]

Application areas of naive Bayes in agriculture are crop classification, site-specific crop management, cattle management, early crop yield estimation, disease detection in leaves, weed detection and vegetation classification. These are probabilistic classifiers. A dynamic self-reliance assumption among different features is made. This approach is to ascertain the particular crop of data transmitted by the sensor. Classifiers can be trained in an effective manner in supervised learning which very efficient

in complex real-world situations. A small amount of trained data is only required for the estimation of parameters which is important for classification purposes.

3.2 *Naive Bayes [11]*

SVM is a risk minimization algorithm based on learning. It is an effective method for agricultural data classification. SVMs can reduce redundant information and sensors only collects data that further need to be classified on extracting the feature from the collected data. It is a binary classifier that constructs a linear separating hyper plane for the classification of instances of data. They have been used for clustering, regression, and classification. SVMs are used for feature extraction, predicting yield and also for estimating crop quality. We use the Kernel trick for transforming the data and then extracting the desired features. Our algorithm outputs an optimal hyper plane that categorizes new features. The original input data are mapped into dimensional feature space to get a separated hyper plane, following which classification is done with the help of N-dimensional hyper plane constructed. Experiments on real agricultural datasets have been performed as depicted in Fig. 1 where post-extraction to features of crops, naive Bayes algorithm predicts the probable crop.

Here naive Bayes [11] classifier is used, this method is for representing knowledge and beliefs using probabilities. The main objective of using a naive Bayes is to calculate the conditional probability of an object with a feature vector X1,X2, …, Xn and belongs to particular class Ci according to the relation P(Ci/X1,X2…Xn) = {P(X1, X2, … Xn/Ci). P(Ci)}/P (X1, X2, … Xn);

$$\text{For } 1 \Leftarrow i \Leftarrow k \tag{1}$$

Probabilistic model Algorithm

1. Assume conditional probability (Pi) of a crop with n feature vectors
2. Assume N feature vectors where N = {X1, X2, X3, … Xn} and n ∈ N

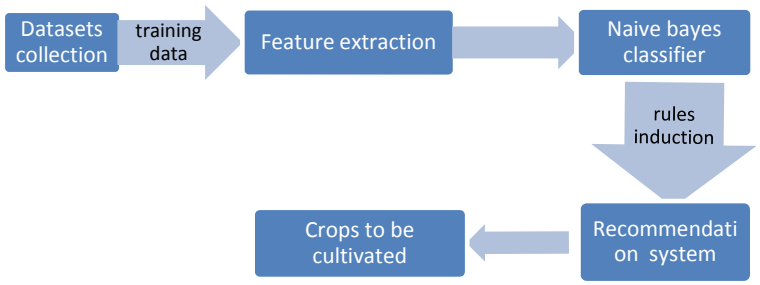


Fig. 1 Block diagram for crop prediction

3. Assume 'i' classes that each crop belongs from a set of classes {C1, C2, …, Cn}
4. The Naive Bayes classifier predicts Crop 'X' to belong to 'C' with a class efficiency given by

$$P(C/X) = P(X1/C) * P(X2/C) * \ldots * P(Xn/C) * P(C)$$

SVM (support vector machine is a classifier that is constructed for linear separating hyperplane or a decision boundary to classify instances of data. The nature of the hyperplane is that it should be equidistant from both the side negative and positive sides of the decision boundary.

Step 1: Hyperplane can be expressed as

$$W0 + WT * Xpositive = 1 \tag{2a}$$

or

$$W0 + WT * Xnegative = -1 \tag{2b}$$

Step 2: Subtraction of eq. 2a and 2b gives

$$WT\ (Xpositive - Xnegative) = 2 \tag{3}$$

Step 3: On Normalizing this by vector W

$$W = \sum mj = 1Wj_2 \tag{4}$$

Step 4: Compute result:

$$\{WT(Xpositive - Xnegative)\}/\|W\| = 2/\|W\| \tag{5}$$

The optimization function requires maximizing $(2/\|W\|)$ with the constraint of the given samples being extracted properly.

We assume 'X' features wherein X = {soil type, rainfall, temperature, pH, sunlight, class} are vectors. The 'Class' decides which type of crop is best to be grown with the given vectors. There are 'i' classes given as (C = C1, C2,…, Ci). The Classes that we have considered for our analysis are rice, wheat, cotton, oilseeds, sugarcane. 'C' is considered as the evidence of belonging to a class here, hence the probability of evidence of a selected crop is 'P(C)'. The vectors are the deciding factors on which class will depend, and their probabilities are assumed as P(X), that is, prior probability of evidence. Therefore, P(C/X) is called posterior i.e. occurrence of the particular crop when the given vector is true. P(X/C) is called likelihood, that is, probability of observing all the vectors with given parameters. Where

P(C) = probability of evidence
P(X) = prior probability of evidence
P(C/X) = posterior
P(X/C) = likelihood
Our algorithm Posterior as

Posterior = {(Likelihood) * (Probability evidence)}/Prior Probability of evidence.

We have assumed the parameters of temperature in degree Celsius, rainfall in cm, sunlight in hours as provided by the datasets. The data is first to split into train and test phases and then by applying Naive Bayes as a classifier for prediction. The prediction of the test data through applying the Naive Bayes classifier in our model gives an accuracy of 96.00% and applying Support Vector Machine (rbf kernel) gives the prediction with 76% accuracy. In this experiment, a confusion matrix has been used for getting the result. Figure 2 depicts the visualization of the performance of algorithms adopted. The X-axis represents the true label while the Y-axis is the predicted label. This shows performance measurement for machine learning classification. The number of correct and incorrect values of fertilizers to be applied to the soil for the cultivation of crops as specified in the dataset tells about the predicted result and incorrect values tell about errors faced during prediction.

From Fig. 2 we can see that the naïve Bayes heat map has a value 7 with dark color which shows this has a more accurate result than SVM as SVM has value 5 which is dark in color and value 0 show that these points are far away from prediction point. Thus, we can infer that Naive Bayes is better in classifying the dataset than SVM. As Naive Bayes gets more accuracy and this classifier depends on probabilistic features so this classifier can be used for better crop prediction, and farmers can use this in the agricultural field for their crop prediction system to get better productivity.

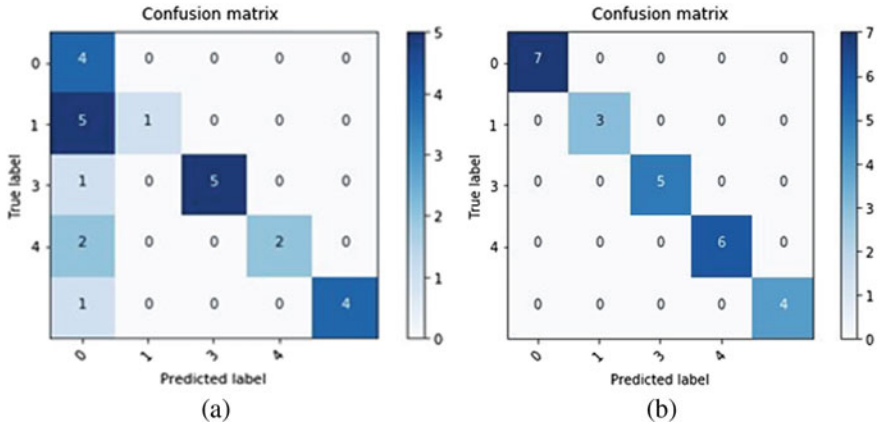


Fig. 2 **a** heatmap of SVM, **b** heatmap of Naive Bayes

4 Conclusion

Precision agriculture or site-specific agriculture adopts many modern farming practices to increase the productivity of the available land. With smart agriculture, the misapplication of products is greatly reduced with an improvement in crop and farm efficiency. With our proposed algorithm, the prediction of crops that would effectively grow in the given set of soil characteristics and minerals has been presented. Results show that the prediction of fertilizer with different classifiers calculating accuracy and heatmap. These results can be used to estimate the amount of fertilizer or minerals that can be provided to the soil for the desired crop which is the future extension of our work.

References

1. Patil KA, Kale NR (2016, December) A model for smart agriculture using IoT. In: 2016 International conference on global trends in signal processing, Information Computing and Communication (ICGTSPICC). IEEE, pp 543–545
2. Giri A, Dutta S, Neogy S (2016, October) Enabling agricultural automation to optimize utilization of water, fertilizer and insecticides by implementing Internet of Things (IoT). In: 2016 international conference on information technology (InCITe)-The next generation IT summit on the theme-internet of things: connect your worlds. IEEE, pp 125–131
3. Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D (2018) Machine learning in agriculture: A review. Sensors 18(8):2674
4. Pudumalar S, Ramanujam E, Rajashree RH, Kavya C, Kiruthika T, Nisha J (2017, January) Crop recommendation system for precision agriculture. In: 2016 Eighth international conference on advanced computing (ICoAC). IEEE, pp 32–36
5. Chettri R, Pradhan S, Chettri L (2015) Internet of things: comparative study on classification algorithms (k-nn, naive bayes and case based reasoning). Int J Comput Appl 130(12):7–9
6. Gondchawar N, Kawitkar RS (2016) IoT based smart agriculture. Int J Adv Res Comput Commun Eng 5(6):838–842
7. Kerns SC, Lee JL (2017, September) Automated aeroponics system using IoT for smart farming. In: 8th International Scientific Forum, ISF 2017. UNCP, USA
8. Pierce FJ, Nowak P (1999) Aspects of precision agriculture. In: Advances in agronomy, vol 67. Academic Press, pp 1–85
9. Murase H, Ushada M (2006) Machine vision applications for micro-precision agriculture
10. Van Alphen BJ, Stoorvogel JJ (2000) A functional approach to soil characterization in support of precision agriculture. Soil Sci Soc Am J 64(5):1706–1713
11. Ng AY, Jordan MI (2002) On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In: Advances in neural information processing systems, pp 841–848

A Smart Feature Reduction Approach to Detect Botnet Attack in IoT

Rup Kumar Deka, Kausthav Pratim Kalita, Dhruba Kumar Bhattacharyya, and Debojit Boro

Abstract To classify sensor data correctly and quickly has a very sound impact on areas such as performance monitoring, user behavior analysis, and user accounting and intrusion detection in IoT (Internet of things). This work is an approach to reorganize the features in a dataset of 114 features depending on the relevancy and non-redundancy of an attribute or feature. Thus having a minimal set of 30 features can create a linear SVM classifier model to detect TCP attack in IoT botnet attack traffic.

Keywords Feature selection · IoT · Botnet · TCP · SVM

1 Introduction

The network traffic generated or collected by sensors of IoT devices can be voluminous and in few minutes it can accumulate and that is why it becomes difficult to monitor and analyze the attack traffic to detect the abnormality. Researchers face a lot of challenges while gathering appropriate information from large samples of data having lots of attributes/features for classification [1], [2], [3]. We aim for the reduction of attributes from data samples collected by IoT sensors based on relevancy and non-redundancy of an attribute or feature without degrading the classification

R. K. Deka (✉) · K. P. Kalita
Department of Computer Science and Engineering, Assam Don Bosco University, Azara Campus, Azara 781017, India
e-mail: rup.deka@gmail.com

K. P. Kalita
e-mail: koztov.project@gmail.com

D. K. Bhattacharyya · D. Boro
Deaprtment of Computer Science and Engineering, School of Engineering, Tezpur University, Napaam, Tezpur 784028, India
e-mail: dkb@tezu.ernet.in

D. Boro
e-mail: deb0001@tezu.ernet.in

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_3

accuracy. Mirai and BASHLITE were used in the DDoS attack on 20 September 2016 on the Krebs on Security site[1] which reached 620 Gbit/s [4] and Ars Technica[2] said a 1 Tbit/s attack on French web host [5]. It is a very big concern and important aspect to deal with such amount of attack.

2 Background Study

An optimal subset of features can be obtained by ranking or evaluating the features to achieve high classification accuracy [6]. Correlation and dispersion measures can provide the relevancy and non-redundancy of a feature with respect to a class. Bravi et al. [7] calculate scores of features and associated with them to assess their relevance. According to Morn-Fernndez et al. [8] feature selection methods can be centralized or distributed. Aljawarneh et al. [9] used the voting algorithm with *Information Gain* to select the important features to increase the classification accuracy. Moustafa et al. [10] applied *AdaBoost* ensemble learning using three machine learning techniques, that is, *decision tree, Naive Bayes* (NB), and *artificial neural network*, to evaluate the effect of features and detect malicious events.

2.1 Motivation

To select a feature, which is going to be included in the classification process, must be evaluated and it is necessary to see how many features can be included to be an effective classification. Having a reduced set of features can provide necessary arrangements to detect abnormality in IoT in less time.

3 Problem Statement and Framework

3.1 Problem Statement

To reorganize features of a dataset in decreasing order of relevancy and non-redundancy, so that a minimized set of features provide the same accuracy as the whole.

[1] https://arstechnica.com/, accessed on 21 November 2019.

[2] https://arstechnica.com/, accessed on 21 November 2019.

3.2 Framework

In Fig. 1., the framework to reduce the number of features in an iterative way. Three primary steps are included.

1. Reorganizing the attributes of the dataset as per relevancy and non-redundancy by calculating three measures:

   - **Correlation measure** can provide dependency of an attribute on all other attributes. An attribute can be selected for the classification having a higher value of correlation coefficient for a given class. We have used the Pearson Correlation coefficient (p) to measure the dependency of the attributes, x and y.

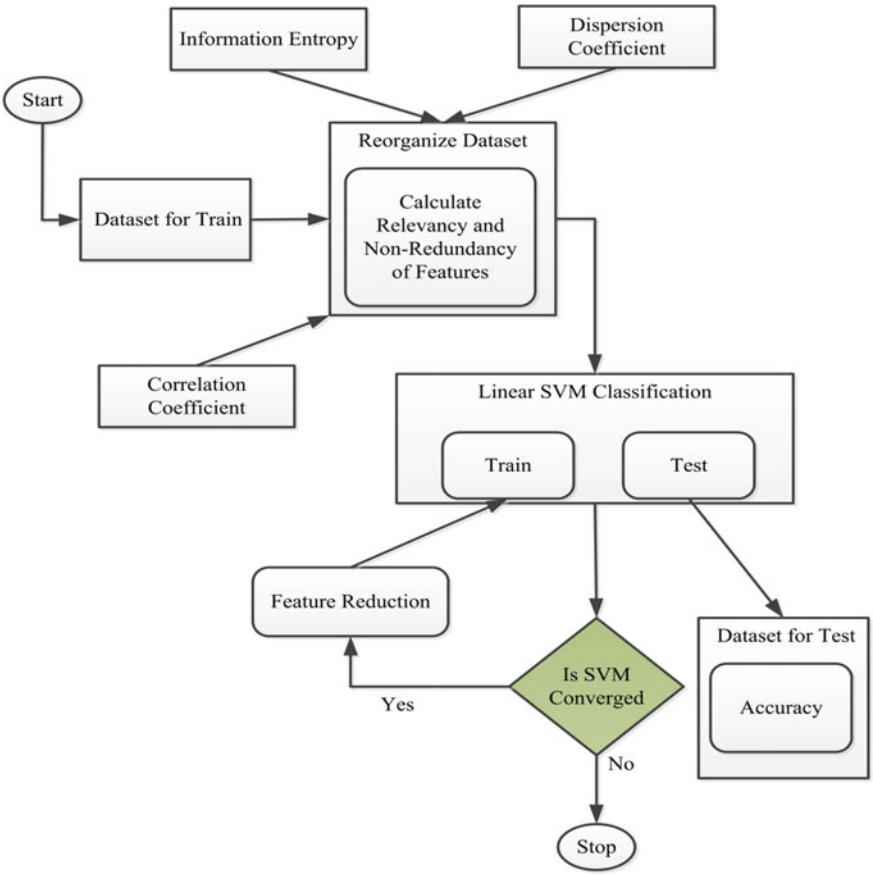


Fig. 1 Proposed framework for feature reduction

$$p = \frac{(n * \sum(x * y) - (\sum x) * (\sum y))}{\sqrt{(n * \sum(x^2) - (\sum x)^2) * (n * \sum(y^2) - (\sum y)^2)}} \quad (1)$$

- **Dispersion value** of data samples for a given attribute allows us to quantify whether a set of observed samples of an attribute or feature are more dispersed or less dispersed (we can say, clustered) for a given class. If the value of dispersion is higher than the spread will be bigger. This idea can be inherited to provide preference to those attributes, which are more appropriate in terms of relevancy and non-redundancy. The dispersion value (q) of a feature is given by:

$$q = \frac{\text{variance}}{\text{mean}} = \frac{\sigma^2}{\mu} \quad (2)$$

- If an attribute x produces a low-probability value, the event carries more information than other attributes with a high-probability value. **Information entropy** value (r) of a given feature is given by:

$$r = -\sum \text{prob}(x)\log(\text{prob}(x)) \quad (3)$$

2. Train and test on the dataset using Linear SVM. Accuracy values are stored.
3. After reducing the number of features in every step, if SVM doesn't converge we stop the procedure.

4 Experiment and Results

4.1 Dataset Description

We have collected real traffic data, gathered from 9 commercial IoT devices authentically infected by Mirai and BASHLITE [11, 12] from the UCI repository[3]. There are 10 different types of attacks carried by these two botnets. This dataset can be used for multi-class classification. In our experiment, we selected normal traffic and TCP type of attack only. For the experiment, we have used 60% of the traffic to be trained and the rest for the test out of 149,931 samples.

[3] https://archive.ics.uci.edu/ml/index.php, accessed on 21 November 2019.

4.2 Experimental Set-up

Experiments are performed on a workstation with a 2.10 GHz processor, 4 GB RAM and a 64bit-Windows 10 operating system. We use MATLAB R2015a 64 bit edition for our experiments.

4.3 Support Vector Machine(SVM) Classifier

Basically, training and testing stages are there in SVM Classifier. For learning purpose, training data samples are selected or updated and thus the learning algorithm constructs the hyper plane using those data samples in the training stage.

The model created by the training phase is applied on data samples for classification during the testing phase.

4.4 Results and Discussion

Table 1 and Fig. 2 show the experimental results. It can be seen that, by decreasing the total number of features from 114 to 30, the accuracy does not deteriorate. Thus, we can say that by having those 30 features only we can achieve the same level of accuracy. Also, it claims that those 30 features are more relevant and non-redundant with respect to their classes for the classification.

Table 1 Accuracy vs. Decreasing number of features

Sl. No.	Number of features included in classification	Linear SVM Classification Accuracy (fivefold Verification)
1	114	99.4
2	110	99.2
3	100	99.3
4	90	99.5
5	80	99.4
6	70	99.2
7	60	99.1
8	50	99.2
9	40	99.5
10	30	99.4
11	20	SVM not converging

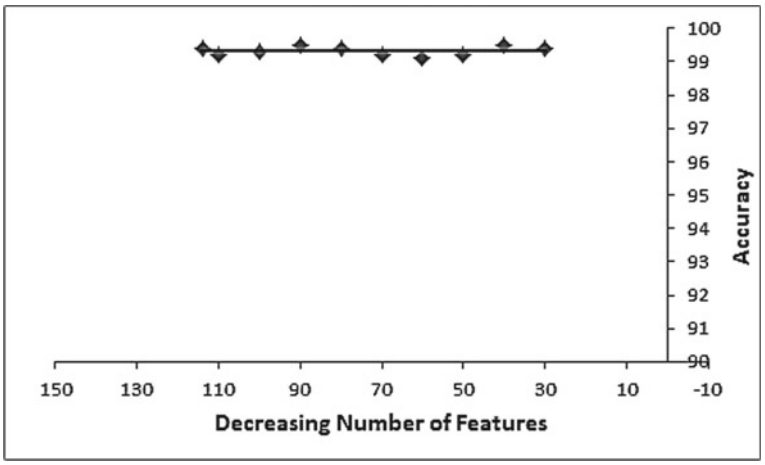


Fig. 2 Accuracy versus Decreasing number of features

4.5 Comparisons

In the field of data mining we can see a lot of works has been done till now to reduce the features or to select an appropriate feature or to extract necessary features for the classification process. But, to detect botnet attacks in IoT, there are few works in recent years. We have compared our work with [13], [14] and [15] to point out the significant differences, as shown in Table 2.

Table 2 Comparison with Few Existing Approaches

Author (s)	Year	Approach	Dataset (s)	Results
Aksu et al. [13]	2018	Fisher Score Algorithm	CICIDS2017	Achieved 99% accuracy on 2,972 samples having 80 features reduced to 30
Kumar and Bhama [14]	2019	Adaptive Filter using Neural Network	Self-generated Attack Traffic	Achieved 99.69%Accuracy on 50,000 test request
Habib et al. [15]	2020	Particle Swarm Optimization	UCI repository [16]	The minimum Classification error is 0.11%
Our Work	2020	Feature Ranking using correlation dispersion and information entropy	UCI repository	Achieved 99.5% accuracy on 149,391 samples having 114 features reduced to 30

5 Conclusion

At last, we conclude that by minimizing the number of features for classification to detect TCP attack from the IoT Botnet dataset having 114 features can provide the same result. A feature or attribute is relevant or not and redundant or not can be measured by using the correlation coefficient, dispersion coefficient and information entropy, as shown in our approach. Thus, a reduced set of features with high relevancy and non- redundancy can be effective in the classification of normal and attack traffic.

The dataset contains other types of attacks also. In future, we have to see how this procedure behaves with other types of attack.

References

1. Mwangi B, Tian TS, Soares JC (2014) A review of feature reduction techniques in neuroimaging. Neuroinformatics 12(2):229–244
2. Goswami S, Das AK, Guha P, Tarafdar A, Chakraborty S, Chakrabarti A, Chakraborty B (2019) An approach of feature selection using graph-theoretic heuristic and hill climbing. Pattern Anal Appl 22(2):615–631
3. Manoj RJ, Praveena MA, Vijayakumar K (2019) An ACOANN based feature selection algorithm for big data. Cluster Comput 22(2):3953–3960
4. Bonderud D (2016) Leaked Mirai malware boosts IoT insecurity threat level. www.securityintelligence.com
5. The Economist: The internet of stings (8/10/2016). Accessed 21 Nov 2019
6. Deka RK, Bhattacharyya DK, Kalita JK (2019) Active learning to detect DDoS attack using ranked features. Comput Commun 145:203–222
7. Bravi L, Piccialli V, Sciandrone M (2016) An optimization-based method for feature ranking in nonlinear regression problems. IEEE Trans Neural Netw Learn Syst 28(4):1005–1010
8. Morn-Fernndez L, Boln-Canedo V, Alonso-Betanzos A (2017) Centralized vs. distributed feature selection methods based on data complexity measures. Knowl Based Syst 117:27–45
9. Aljawarneh S, Aldwairi M, Yassein MB (2018) Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model. J Comput Sci 25:152–160
10. Moustafa N, Turnbull B, Choo KKR (2018) An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of internet of things. IEEE Internet of Things J 6(3):4815–4830
11. Mirsky Y, Doitshman T, Elovici Y, Shabtai A (2018) Kitsune: an ensemble of auto encoders for online network intrusion detection. arXiv preprint arXiv:1802.09089
12. Meidan Y, Bohadana M, Mathov Y, Mirsky Y, Shabtai A, Breitenbacher D, Elovici Y (2018) N-baiotnetwork-based detection of IoT botnet attacks using deep auto encoders. IEEE Pervasive Comput 17(3):12–22
13. Aksu D, Stebay S, Aydin MA, Atmaca T (2018) Intrusion detection with comparative analysis of supervised learning techniques and fisher score feature selection algorithm. In: International symposium on computer and information sciences. Springer, pp 141–149
14. Kumar CO, Bhama PRS (2019) Detecting and confronting flash attacks from IoT botnets. J Supercomput 75(12):8312–8338
15. Habib M, Aljarah I, Faris H, Mirjalili S (2020) Multi-objective particle swarm optimization for Botnet detection in Internet of Things. In: Evolutionary machine learning techniques. Springer, Singapore, pp 203–229
16. Dua D, Efi KT (2017) UCI machine learning repository

An Approach to Handle Heterogeneous Healthcare IoT Data Using Deep Convolutional Neural Network

Kishore Medhi, Md. Arifuzzaman Mondal, and Md. Iftekhar Hussain

Abstract Internet of Things (IoT) is going to be the next big technological revolution of mankind by connecting everything on the earth via the Internet. Mobile healthcare or remote healthcare is an important application of IoT, which provides a new platform to people for getting benefit regarding healthcare-related problems. In these applications, various sensing devices connect with the patient's body and generate an enormous amount of heterogeneous data over time. Due to the variety of data, extraction of knowledge from these data is not straight forward just like the conventional data mining process. In this research, we have proposed a Deep Convolution Neural Network (DCNN)-based classification method for performing data mining over heterogeneous data by taking an unstructured sensor dataset from the arrhythmia database of physionet. In the proposed method, the CNN feature extraction layer converts the ECG signals into numeric form by calculating its features without any human intervention. Similarly, categorical data are converted based on their respective categories. Finally, all the converted data together were added to the CNN classification part and arrhythmia disease is predicted with an accuracy of more than 98%. The simulation result shows that our proposed CNN-based architecture outperforms other handcrafted feature extraction techniques in terms of accuracy, sensitivity, and specificity.

Keywords ECG · Arrhythmia · CNN · Data mining · IoT

K. Medhi (✉) · Md. Arifuzzaman Mondal · Md. Iftekhar Hussain
North-Eastern Hill University, Shillong 793 022, India
e-mail: kishoremedhi2015@nehu.ac.in

Md. Arifuzzaman Mondal
e-mail: arifuzzaman161@nehu.ac.in

Md. Iftekhar Hussain
e-mail: ihussain@nehu.ac.in

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_4

1 Introduction

The main motivation of the IoT concept is to join all the physical objects of the earth with the Internet to make it smart [1]. IoT devices capture data from various types of sources, which makes the data heterogeneous. The design and development of smart healthcare is an important and active research area of IoT. Among all the healthcare issues, arrhythmia is one of the most important, due to which a huge number of people died every year. Therefore, a proper and efficient remote heart monitoring system is an essential application that needs to be developed. In this application, various sensors are attached to the human body to collect different types of information. During the development of this project, different researchers face different problems, among them handling of heterogeneous data is an important issue. It is not easy to capture the hidden information from these heterogeneous arrhythmia data using the conventional KDD system available today. The various authors try to identify heart diseases differently. Detail comparison of various techniques used to perform data mining over arrhythmia data has illustrated in Table 1. Park et al. [10] detected heart-beat from the ECG signal with 94% accuracy using the RR interval as a feature and SVM classification algorithm. Similarly, Dallali et al. used FCM and NN classification over-extracted RR interval data from the ECG signal to classify arrhythmia disease. They have successfully classified arrhythmia with 90% accuracy [6]. R peak and RR

Table 1 Comparison of various heart disease classification methods using ECG signal

Method	Features	Data Mining Algorithm	Result
(SVM-RR) (2008) [10]	RR interval	SVM	94% accuracy
(FCM-RR) (2011) [6]	RR interval	FCM, ANN	90% accuracy
(ANN-R, RR) (2011) [3]	R peak, RR interval	ANN	97.7% accuracy
(ANN-QRS) (2007) [9]	QRS complex	MLP	98% accuracy
(MLPNN-R) (2014) [7]	RR interval and R peak	MLPNN	97% accuracy
(ANN-R) (2015) [12]	Wavelet transform and R peak	ANN	94% accuracy
(SVM, KNN-QRS) (2016) [8]	QRS complex	SVM, KNN	98% accuracy
(SVM-RR) (2017) [4]	RR interval	SVM	98% accuracy
(SVM-RR) (2018) [13]	RR interval	SVM	94% accuracy

intervals were used by Ayub et al. [3] to classify ECG signals for detecting abnormalities. During the experiment, they have detected the ECG signal correctly with 97% accuracy. Ghongade et al. [9] analyses various feature extraction technique of ECG signal. During the research, they have used NN as a classifier and they found that DFT, DWT, and DCT techniques are more suitable for real-time diagnostic systems. In [7], temporal feature, wavelet transform & s-transform feature extraction techniques are compared with MLP using five different ECG beat. Similarly, Thomas et al. [12], Elhaj et al. [8], Chen et al. [4] & Venkatesan et al. [13] Identify ECG signal using RR interval and neural networks and they successfully identify with more than 94% accuracy. In all the above research experiments, researchers have tried to identify heart diseases or arrhythmia of a person from the ECG signal characteristics only. But, in real life along with ECG features, age, and gender of a person are also considered to identify heart diseases accurately [5, 11], which are not considered in the above-mentioned methods. Therefore, the main innovative elements of this research were, the use of ECG signal, age, and gender together as an input to predict heart diseases, which is not done by any existing methods. Moreover, from the literature review, it is clear that all the existing methods used handcrafted feature extraction techniques and then applied the data mining algorithms. But our proposed CNN method, extracts the features and performs classification simultaneously without any human intervention. The rest of the paper is organized as follows. Section 2 provides the details about the materials and the proposed method used to carry out our experiment. Finally, the result analysis and conclusion of the experiment is explained in Sect. 3.

2 Materials & Methods to Handle Heterogeneous Data

2.1 Materials

In this research experiment, to carry out our execution we have taken a heterogeneous arrhythmia dataset from the MIT-BHI arrhythmia database of PhysioNet [2].

It consists of 565 records where each record contains three different parameters of a patient such as age in numeric form, gender in categorical form and respective ECG signal of the patient. The recorded data were from 367 male and 198 female heart patients. In the above dataset, all the patients were in the age range from 1 to 89 and containing six different heart disease classes including normal sinus rhythm (NSR), pulmonary ventricular contractions (PVC), sinus tachycardia (ST), sinus bradycardia (SB), atrial fibrillation (AF), and atrial flutter (AFL). All the 10-second (3600 samples) fragments of the ECG signal were captured at a sampling frequency of 360 Hz [2].

2.2 *Proposed Method*

An overview of the developed multilayer CNN architecture is shown in Fig. 1. The architecture is divided into three main sections namely, input, feature extraction, and classification section. All the different sections are briefly discussed below.

1. **Input Section**: As shown in Fig. 1, the age, 1-D raw ECG signal, and gender of a patient are given as an input to the multi-layer CNN architecture to identify the presence of an arrhythmia.
2. **Feature Extraction**: The developed feature extraction architecture consists of 3 convolutional layers and each layer follows the ReLU activation function and

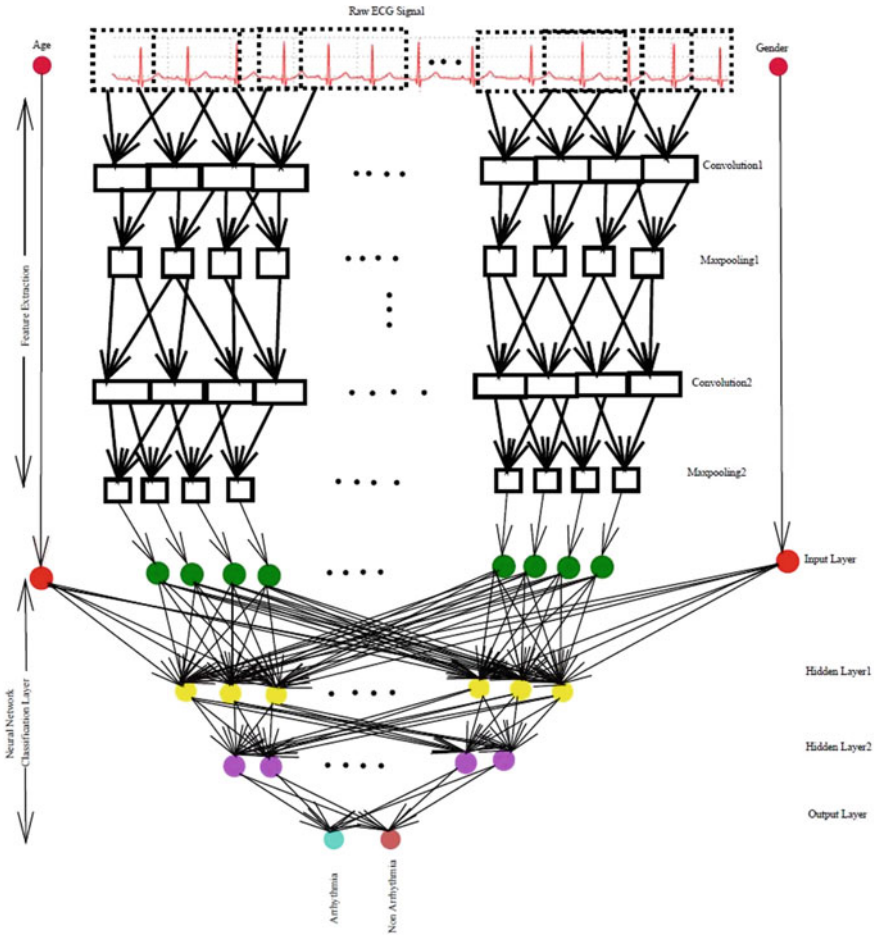


Fig. 1 Developed MLCNN Architecture

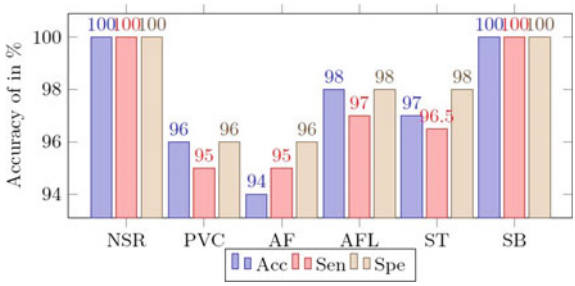


Fig. 2 Classification accuracy of various heart diseases using the proposed method

max-pooling layer. In the first layer, the 1-D convolution is performed with 128 filters with filter size 1×9. The convolution output is normalized and reduced by using the ReLU activation function and pooling layer of size 1×4, respectively. These processes are repeated in the successive two layers with 100 and 64 convolution filters. Finally, to solve the heterogeneity problem we have passed the other two inputs such as age and gender along with the extracted feature of the ECG signal to the classification section. To reduce the overfitting of the model we have set the dropout value is 0.2 and collected only the important five different features from the signal.

3. **Classification**: The features that were extracted by the feature extraction layers act as input to the classification layer. As shown in Fig. 1, to perform classification we have used NN (Neural Network). In the developed NN, the input layer consists of 7 neurons, where five features come from the ECG signal, and

two from the age and gender of the patients. There are two hidden layers and each contains 50 and 20 numbers of neurons, respectively. Based on our experiment requirement, we have set two neurons in the output layer. Where one neuron represents disease present and another neuron represents no disease.

3 Result Analysis and Conclusion

In this research, all the experiments were done in the Matlab environment. Here, we have tried to solve the heterogeneity issue by using multi-layer CNN, without converting all data in the beginning. To compare our results, we have implemented all the state-of-the-art methods using our dataset. As shown in Fig. 3, the best results of different state-of-the-art methods were compared with our proposed method by taking accuracy as the performance evaluation parameter. From the result, it is clear that our proposed CNN-based method predicts arrhythmia diseases more accurately compared to other existing methods with an accuracy of 98.4%. After predicting the arrhythmia, we have also identified the type of heart disease that the patient was suffering, and all the experimented results were shown in Fig. 2. From Fig. 2, it is clear that our proposed CNN based method can classify Normal Sinus Rhythm

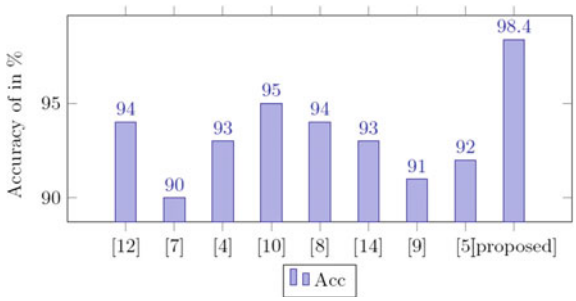


Fig. 3 Comparison of the proposed method with state-of-the-art methods (ANN, MLP, SVM, FCM, KNN)

(NSR) and Sinus Bradycardia (SB) ECG signal with 100% accuracy, similarly, it can classify premature ventricular contractions (PVC), atrial fibrillation (AF), atrial flutter (AFL) and sinus tachycardia (ST) ECG signal with 96%, 94%, 98%, and 97% accuracy respectively.

Therefore we can conclude that the heterogeneity issue can be solved easily by using the CNN technique without any prior conversion into a common format. In this research, we have used CNN technique to solve the heterogeneous data issue taking healthcare as an IoT application, but in the future, we will try to use different ANN and deep learning modules, such as RNN, VGG Net, etc., to improve our detection accuracy.

References

1. Ahmed E, Yaqoob I, Hashem IAT, Khan I, Ahmed AIA, Imran M, Vasilakos AV (2017) The role of big data analytics in internet of things. Comput Netw 129(2):459–471
2. Albert C-C, Yang AS (2019) MIT-BIH Arrhythmia Database. https://physionet.org/note. Accessed 25 Nov 2019
3. Ayub S, Saini J (2011) Ecg classi_cation and abnormality detection using cascade forward neural network. Int J Eng Sci Technol 3(3)
4. Chen S, Hua W, Li Z, Li J, Gao X (2017) Heartbeat classification using projected and dynamic features of ecg signal. Biomed Signal Process Control 31:165–173
5. Curry SJ (2019) Screening for coronary heart disease with electrocardiography. https://www.uspreventiveservicestaskforce.org/Home/GetFileByID/1883. Accessed 14 June 2019
6. Dallali A, Kachouri A, Samet M (2011) Fuzzy c-means clustering neural network wt and hrv for classi_cation of cardiac arrhythmia. ARPN J Eng Appl Sci 6(10):2011
7. Das MK, Ari S (2014) Ecg beats classi_cation using mixture of features. International scholarly research notices 2014
8. Elhaj FA, Salim N, Harris AR, Swee TT, Ahmed T (2016) Arrhythmia recognition and classi_cation using combined linear and nonlinear features of ecg signals. Computer methods and programs in biomedicine 127:52–63
9. Ghongade R, Ghatol A (2007) Performance analysis of feature extraction schemes for arti_cial neural network based ecg classification. In: International conference on computational intelligence and multimedia applications (ICCIMA 2007). vol 2. IEEE, pp 486–490
10. Park K, Cho B, Lee D, Song S, Lee J, Chee Y, Kim I, Kim S (2008) Hierarchical support vector machine based heartbeat classi_cation using higher order statistics and hermite basis function. In: 2008 Computers in cardiology. IEEE, pp 229–232

11. Paul S (2019) Cardiac (heart) screening. https://www.radiologyinfo.org/en/info.cfm?pg=screening-cardiac. Accessed: 14 June 2019
12. Thomas M, Das MK, Ari S (2015) Automatic ECG arrhythmia classification using dual tree complex wavelet based features. AEU-Int J Elctron Commun 69(4):715–721
13. Venkatesan C, Karthigaikumar P, Paul A, Satheeskumaran S, Kumar R (2018) Ecg signal preprocessing and svm classi_er-based abnormality detection in remote healthcare applications. IEEE Access 6:9767–9773

Designing of *NimbleArm*—A Low-Cost, Semi-autonomous Interactive Robotic Arm

Abhishek Sarkar and Gypsy Nandi

Abstract Over the millennium that has passed since the dawn of robotics, man has always dreamt of creating robots that could support him, ease his problem and make a better world. In these recent years, every iterative attempt has been made to create better and robust robots which are not only smart but can also make "dangerous and tedious tasks easier". This paper presents one such robotic system, the *NimbleArm* that can be implemented for picking and placing familiar as well as unfamiliar objects in industrial warehouses, factories and assembly lines. The system comprises six degrees of freedom robotic arm allowing it for wide flexibility and an autonomous platform thus allowing it to locomote from one place to another. Additionally, it is equipped with computer vision as well as an Artificial Intelligence-based framework.

Keywords Robotic arm · Degree of freedom · Inertial measurement unit (IMU) · Artificial intelligence

1 Introduction

Robotics is a world of "Dreams and Dreamers." The term robot has its origin from the Czech word *robota*, which when translated means "forced labor." This describes the majority of robots fairly well. Most of the twenty-first century robots in the world are designed to be able to mimic human beings in the uncanniest ways, for which today we have individual robots that can match and even exceed human capabilities at specific tasks, like heavy, repetitive manufacturing work or defusing explosives. Just as the human body is composed of numerous subsystems which are governed by an inherent structure, so are the robots the product of a tenuous design that piece together to form a complete structure capable to perform its task that is desired

A. Sarkar (✉)
Department of Computer Science Engineering, Assam Don Bosco University, Guwahati, Assam, India
e-mail: rigelinx@gmail.com

G. Nandi
Department of Computer Applications, Assam Don Bosco University, Guwahati, Assam, India
e-mail: gypsy.nandi@gmail.com

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_5

nothing more and nothing less. In the twenty-first century, the most popular robot that we often come across is the robotic arm. They have been used in the industrial sector for about more than a decade now. They are versatile and well suited for the heavy and repetitive tasks that are carried out in industries and factories. A robotic arm is an example of an articulated body.

The functionalities of the articulated body depend on its degrees of freedom (DOF), which can be defined as a specific mode in which a machine can perform for example a chalk duster has only one degree of freedom since it can be used in only one way to clean the chalk board. *NimbleArm* similarly has 6 DOF which means that it can perform a wider variety of tasks since it can access intricate places easily. The axis for each DOF can be stated as: Axis 1—Yaw, Axis 2—Pitch 1, Axis 3—Pitch2, Axis 4—Roll1, Axis 5—Pitch3 and Axis 6—Roll2. As shown in Fig. 1., let us consider a generic 4 DOF robotic arm. The arm is made up of four joints and three links. Each link depicts the structure of a human arm i.e. Link 0—Shoulder, Link 1—Elbow and Link 2—Wrist. At each joint, there is a motor connected to it such that it allows rotating parallel or in perpendicular to its axis. These motors are controlled by a computer thus acting harmoniously with one another in order to perform the task that it is assigned. The movement all together however relies on a set of functions that deals with kinematics.

The platform section of the *NimbleArm* consists of a 4 WD drive system that comprises powerful Johnson Motors with 60 RPM each and stalls torque of 20 kg. The system also consists of a *Mastcam* which allows it to have a better vision. The first prototype is limited to fix the positioning of the camera. Additionally, the system also contains inertial measurement units (IMUs) for each of the pitch axis of the arm. These IMUs feed the arm with necessary data regarding each of the assigned axes, allowing it to function autonomously. In the case of manual operation, the

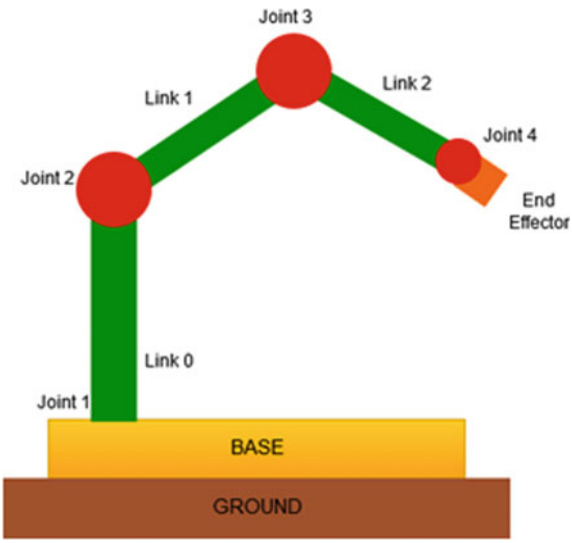


Fig. 1 DOF Robotic Arm

operator can easily control the system within a range of 800 meters. While being autonomous, the system can connect it to a nearby radio port and transmit as well as receive data through the radio network. These networks can be upgraded using radio trans-receivers with each port having a range of 1.5 km to 5 km or even more depending on the antenna power output. The overall design of the system is a crucial factor since using smart mechanical design along with object detection and Artificial Intelligence (AI) comes with a wide variety of choices as well as challenges to meet up, but we came up with the challenges and finally managed to develop a robust system.

The rest of the paper is designed as follows. Section 2 discusses the related work that has been accomplished in the domain of robotics, computer vision and machine learning. The related work has been summarized in chronological order in order to have a better understanding of the developments that have been carried out in the above-mentioned fields. In Sect. 3, we discuss the technical specification of the system of *NimbleArm* where we emphasized each of the units of the whole *NimbleArm*. Section 4 concludes the paper and also discusses the scope of further developing and modifying the robotic arm in order to have more functionalities which a keen objective on well-structured economic feasibility for large-scale production.

2 Related Work

Some of the related works that involved both the domains of robotics and computer vision are discussed in this section. As mentioned by authors in [1], cloud infrastructure is provided which acts as a learning and storing base for the system discussed allowing a vast array of data to be fed into a system under minimal efforts when iteratively accomplished. As mentioned by authors in [2], the system proposed by them is divided into two parts—the recognition and the reconstruction section. The rest of the objectives include pre-processing followed by feature extraction along with classification and localization. For training the database the authors used images of different objects collected from different angles with a total of 153 images for the test database having an all-black background. For feature extraction and object detection, the acquired images are resized and converted to grayscale which is further exposed to sobel filter and the missing or overflowing pixels are removed through dilation and erosion. In total, about 11 features are extracted for each object using MATLAB. The extracted features were classified using ANN and MLP which are a feedforward type of neural network and it examines the object co-ordinates from the given datasets to determine its location and generate the forward kinematic equations.

With the onset of AI's new capabilities for computers like active learning and crowdsourcing, the trend is upcoming, reliable and promising and has paved the way to efficiently build training sets for object recognition as mentioned by authors in [3], where techniques are tested in artificially controlled settings. Here, an approach is presented using the active learning method, in which the object detectors of the system autonomously refine its models by actively crowdsourcing from the Web. It

demonstrates an approach in which several experiments of unprecedented scale and autonomy are conducted. This reveals that the effort successfully improves the state-of-the-art for the most challenging objects when is tested in the PASCAL benchmark. In addition, their detector competes well with popular nonlinear classifiers. The goal is to enable online active crowd-sourced object detector training. Given the name of a class of interest, the system produces a detector to localize novel instances using automatically obtained images and annotations. To make this feasible, the authors first proposed a part-based linear SVM detector, and then showed how to identify its uncertain examples efficiently using a hashing scheme. In [4], we come across a system named *You Only Look Once* (YOLO) which a single convolutional network that can simultaneously predict multiple bounding boxes and class probabilities for those boxes. YOLO is trained on full images which directly optimizes its detection performance. It is extremely fast running at 45 fps at the base network. Such systems can utilize localization algorithms along with proper feature extraction algorithms for better performance. As in [5], the implemented system can be given a search region where the technique returns a bounded box in which the object of interest is present. When such a system is implemented for a robotic arm, the capabilities and the quality of the arm can be vastly increased to deploy multiple tasks in the industries such as a PCB milling machine or an automated welding machine.

Since the past few decades with the birth of AI and machine learning several efforts have been made to utilize these areas in various fields and one such prominent field is robotics. Now the systems discussed using computer vision can further be more enhanced if they are provided with deep learning and implementing neural networks as they provide a large database from which better classification can be done allowing more optimal performance. In [6], we come across such a system that is trained with a large database that helps the systems to cope up with a wider variety of solutions. One such system is designed to build a robotic arm that can be used as assistive platforms to help the elderly people with their day to day activities but in order to do that, we need to have a proper understanding of head-pose which has been clearly discussed in [7] and deep tracking using RNN and CNN that provides the system to predict several future possibilities from a single action. As discussed in [8, 9] some unique features like text recognition along with text rectification have been developed which allows the robotic arm to make out meaningful data from text without the hassle of the position from which it captures. Efforts have been made in [10] to develop a better face recognition algorithm using *Facial Recognition Technology* (FERET) that is implemented in order to make robotic arms deployable in new sectors. Table 1 depicts a comparison of *NimbleArm* along with similarly developed commercial systems.

Computer vision-based robots are not only restricted to industrial sectors, it can also be used in various other sectors such as agriculture as discussed in [11] where HSV space along with extreme learning machine algorithms have been implemented. In [12] efforts have been made in order to implement two robotic arms that work together synchronously using neural networks]. In [13] the authors proposed a virtual robotic arm and a Lab view instrument [3D—Picture Control] to control the arm by translating hand gesture movements using a leap motion sensor to capture the

Table 1 Market Study on similar existing systems

System Name	Seller	Price (INR)	Websites (Accessed on)
DOBOT M1	Amazon	83,930	6th August 2019
DOBOT MAGICIAN	Amazon	1,04,930	6th August 2019
MAKER ARM	Kickstarter	1,53,930	7th August 2019
NIRYO ONE	Niryo	1,50,500	7th August 2019

gestures. The leap motion sensor operates in two leap functions—Open Connection and Rea one swipe Direction. The subsequent stage of operations is related to forward and inverse kinematics in order to relate the movement of the arm with respect to gesture movements of the hand. Such systems are beneficial for people with disabilities and also serve as virtual training platform before working on an actual arm. A market study was done in order to get an insight into the value for similar robotic systems.

3 System Overview of *NimbleArm*

Building *NimbleArm* was a very adventurous task. The prototype designs were made and tested, and then redesigned, again and again, to remove imperfections, until we got the exact result we needed. Figure 2 shows the block diagram of the system. Here we can see that the whole system runs on an Arduino Mega 2560 and a Raspberry Pi Model 3B+. Both are connected to each other via UART communication.

The system uses a relay-based motor driver to run 4 60 RPM motors and also houses a 6 axis arm (*NimbleArm*). We will further discuss each of the sub-systems in the following sections. The end-effector currently attached to the *NimbleArm* is a two-finger gripper. However, suction grippers are also compatible with the arm. We used aluminum as well as PVC (Polyvinyl Chloride) pipe for designing the arm along with wood. The motors that are currently used are standard MG995 Servo motors, both 180 and 360 motors.

3.1 Inertial Measurement Unit (IMU)

IMUs are Inertial Measurement sensors based on MEMs, which collect the rotational and acceleration data of an object along an axis. The IMUs which we currently deployed for *NimbleArm* are MPU-6050. It has 3 axis gyroscope and 3 axis accelerometer. It has a 16-bit ADC integrated inside. In total there are 3 IMUs

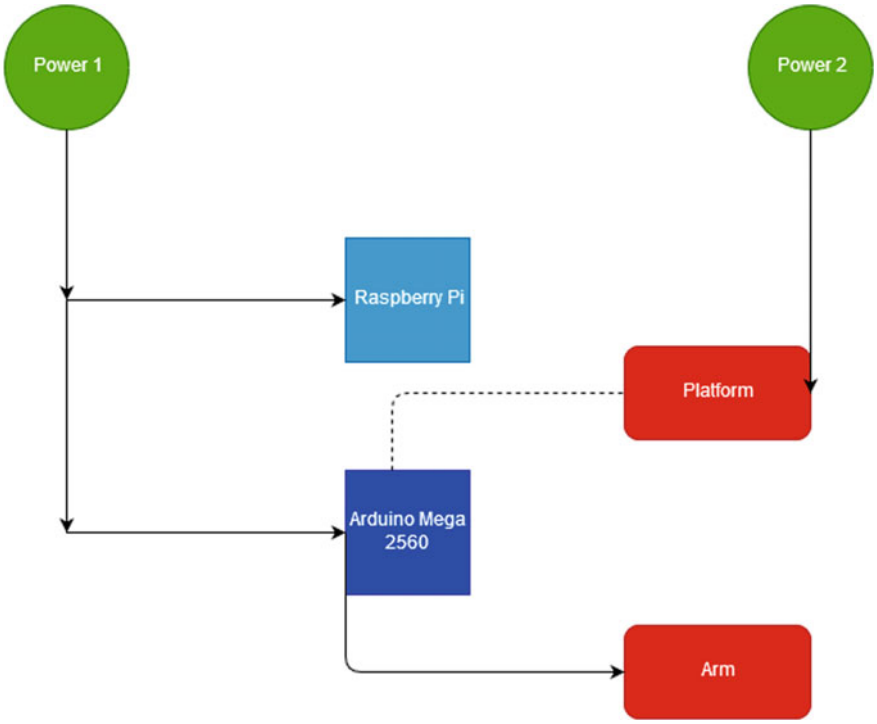


Fig. 2 NimbleArm Block Diagram

on the arm along 3 axes. We placed the IMUs along axis 2, axis 3 and axis 5, that is, along the pitch axes because any movement in the rest of the axis will alert the respective pitch axes, the last axis is software-driven directly for better clarity in end-effector control. The IMUs are I2C devices so we used an I2C multiplexer/de-multiplexer 4051 to de-multiplex the data from each of the IMUs. The IMUs can transmit data from address 0×68 since we had multiple IMUs we separated the SCL and SDA lines such that when we enter the first address of both the de-multiplexers we get the data from the first IMU and so on. The latency between each of the reading is 5 mille-seconds; it can be altered based on the user since it is software-driven. While reading the data from each of the IMUs, we kept 3 LEDs which will give a visual feedback regarding the current IMU from which the data is being read. The data is transferred back to the Arduino Mega 2560, which further transmits it to the Raspberry Pi via UART communication. The Pi records the data and uses it while it is in autonomous mode, which is important in order to cross validate the data that it is acquiring through computer vision.

3.2 Transmitter

NimbleArm can be operated in manual mode also. In this mode, the whole system including the platform can be controlled via a transmitter. The remote we designed for the system is a 12 channel [8 Analog and 4 digital] 2.4 Ghz radio transmitter with full-duplex capacity. The radio is a 3.3 V NRF2401L PA+SMA. The transmitter uses a 2S li-Ion battery with a standby time of up to 2 h. There is also an Led which acts as an indicator, once the transmitter connects with the receiver.

3.3 Power Systems

The whole system uses one 12 V lead-acid battery, two 3S li-ion batteries and two 2S li-ion cells. The power distribution is further divided into 5 V, 12 V and 3.3 V for feeding all the components.

3.4 Image Processing (Experimental Setup)

The system uses a 5MP Omni Vision OV5647 which is popularly known as the Pi Camera V1, it is a CMOS-based camera with low power consumption and capacity to directly feed digital data to the Raspberry Pi 3B+. For the Computer Vision, we use Tensorflow Lite along with OpenCV. OpenCV is not needed to run Tensorflow Lite however it is necessary in order to grab the images and generate the detected results. We used SSDLite–Mobilenet-V2 model to train the system. The model is based on the MSCOCO dataset that is provided by Google. It can detect up to 80 common objects. The detection model implements real-time detection by making predictions in a grid. The grids are responsible for detecting the object in the image and NMS (Non-Maximum Suppression) in post-processing. The SSDLite-Mobile net V2 is a CNN-based model which implements a novel knows as bottleneck residual block (BRB). The Camera takes individual snapshots every 5 mille-sec and performs a series of operations over them.

- **Image Filtering in Spatial Domain**—The image is then filtered for modifying and enhancing it. In order to do it several less satisfying features are excluded over more important features. The image is enhanced using smoothing, sharpening and enhancement techniques.
- **Convolution**—It involves using a convolution technique in which each pixel in the feature region is a weighted sum of the neighborhood input pixels.
- **Binarization**—It involves converting the convoluted images into the binary image to exclusively distinguish the recognizable objects and their distance
- **Thresholding**—In this step, the image provided by the camera is converted to three images for the RGB. These images are then converted to other color space

HSV and processed. A cut-off threshold can be set to reject any non-equivalent image.

- **Localization**—In this stage, the processed data from HSV is used and to detect the location and size of the pertained objects that are already present in its database. Once it detects it extracts features from the data and compares them to the database.

3.5 Control Flow of the System

Figure 3 shows the overall control diagram of the *NimbleArm.* The system operates in two modes which can be toggled depending on the switch signal the system boots up. If the switch is toggle OFF, then it will boot into a manual mode which will be controlled using a transmitter.

If the switch is toggled ON, the system boots up in autonomous mode in this mode the system implements the image processing algorithms discussed above along with a neural network and the data that is generated in order to detect and identify objects and grab them or drop them. The grabbing and dropping part is still under experimentation because of the limited computing power of the Pi, it can be improved using an accelerator like CORAL which is and ASIC based chip to improve the performance.

3.6 Test Cases

The robustness of the arm was tested based on few criteria that could provide feedback on the specification of the system. Table 2 shows the test cases for measuring robustness.

4 Conclusion

In this paper, we discussed about how the journey of developing *NimbleArm* using open-source hardware and about the calibration work to fully match the desired functions of the whole system The overall design of the arm can be innovated more with the addition of stepper motors in some specific axis. There is scope for adding more functionality like GPS and building a more robust neural network. The overall effective working area can be increased from (1 km to 5 km) by using LORA instead or 2.4 Ghz radiofrequency. The camera can be deployed with pan/tilt motion in order to get a wider angle of view. The deployment of Image Recognition for autonomous movement of the arm was experimental the results generated gave us insight on the improvements that can be made which includes adding accelerator hardware, improving the graph segmentation and image morphology. The data set running for

Fig. 3 Control flow diagram

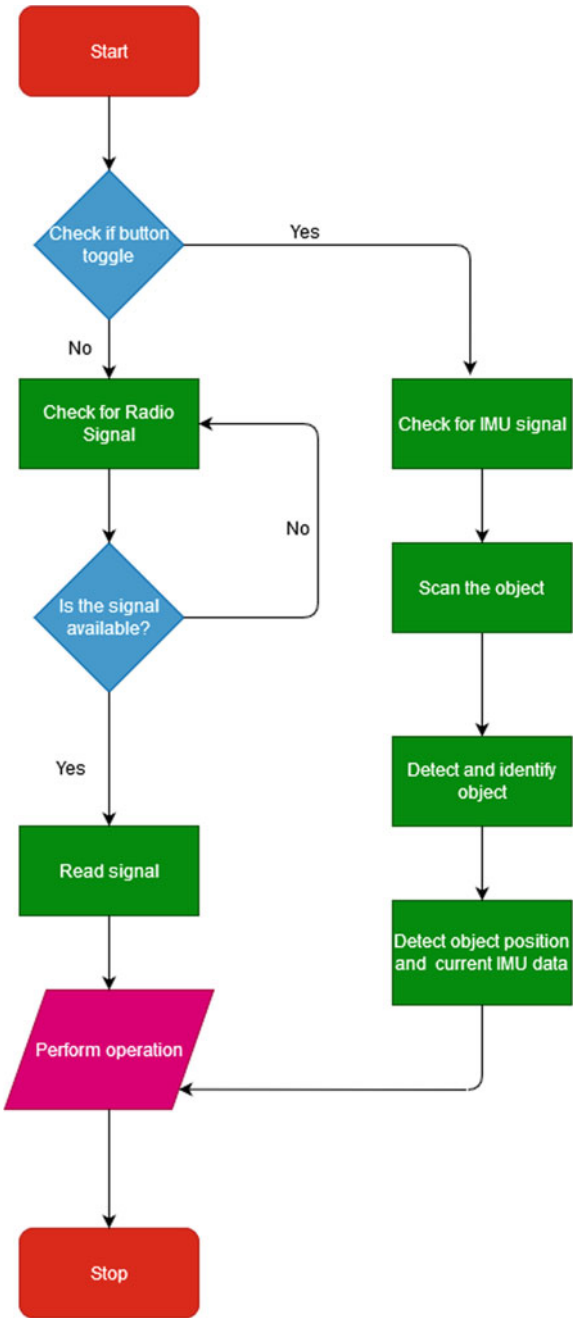


Table 2 Test Cases

Test Case#	Test Case Description	Test Data	Expected Result	Actual Result	Remarks
1	Maximum operational area of the system	The LED on the receiver would be ON if the system is within 800 mtrs from transmitter	The LED should glow as soon as the transmitter is ON	The LED started to faint after 779 meters from Receiver and went OFF after 812 meters	Pass
2	Maximum Reach of the arm	The maximum reach of the arm is 300 mm	The arm should be able to cover all the points within 300 mm radius	The arm was able to cover 97% points falling within a radius of 300 mm	Pass
3	Maximum payload capacity of the system	The maximum payload capacity of the system is 15 kg	The system should be able to locomote with payload less than 15 kg	The system was able to move without the motors stalling up to 14.5 kg	Pass
4	Power Consumption of the system	The overall power consumption of the system is 144 Watts	The system should draw near about power equivalent to 144 Watts	The actual power draw is 137 Watts	Pass

object recognition is based on the MSCOCO dataset from Google, it can be improved by training on a custom dataset. Currently, the detection is running at a low FPS around 16 which is not enough, the arm is able to move to a grabbing position but with some latency. All in all, it's a favorable result since it was an experimental setup. However, this system can be developed indigenously since the whole system can be developed under 50,000 INR (Indian currency). Thus this robotic design provides a one-step ahead in the contribution of robotic design for automation of factories and industries which is one of the requirements for the smart city concept.

References

1. Vijayanarasimhan S, Grauman K (2014) Large-scale live active learning: Training object detectors with crawled data and crowds. Int J Comput Vision 108(1-2):97–114
2. Iscimen B et al (2015) Smart robot arm motion using computer vision. Elektronika ir Elektrotechnika 21(6):3–7
3. Beksi WJ, Spruth J, Papanikolopoulos N (2015) Core: a cloud-based object recognition engine for robotics. In: Intelligent robots and systems (IROS)
4. Redmon J et al (2016) You only look once: unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition

5. Gidaris S, Komodakis N (2016) Locnet: Improving localization accuracy for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition
6. Ondruska P (2016) Neural robotics-a new perspective AAAI robotics fellowship
7. Rahmatizadeh R et al (2016) Trajectory adaptation of robot arms for head-pose dependent assistive tasks. In: FLAIRS conference
8. Saxena D Saritha SK, Prasad KNSSV (2017) Survey paper on feature extraction methods in text categorization. Int J Comput Appl 166(11):1–7
9. Banerjee D, Kevin Y, Aggarwal G (2018) Image rectification software test automation using a robotic arm. IEEE Access 6:34075–34085
10. Banerjee D, Kevin Y (2018) Robotic arm-based face recognition software test automation. IEEE Access 6:37858–37868
11. Hu Z, Liu B, Zhao Y (2018) Agricultural robot for intelligent detection of pests and diseases
12. Tripathi S et al (2018) Robotic arm controlled by voice
13. Rusanu OA et al (2019) Virtual robot arm controlled by hand gestures via Leap Motion Sensor. In: IOP Conference Series: Materials Science and Engineering. Publishing

Smart Refri: SMART REFRIGERATOR for Tracking Human Usage and Prompting Based on Behavioral Consumption

Aditya Sharma, Ashmita Sarkar, Aazaz Ibrahim, and Rupam Kumar Sharma

Abstract This paper focuses on a technical solution to a specific modern problem which is food wastage; which if not handled efficiently, may further lead to a plethora of socio-economic issues. This paper deals with the designing of a Smart Refrigerator, using the Internet of Things (IoT), that will help in the successful monitoring of food items kept inside the refrigerator. It defines the working of the Smart Refrigerator, which is equipped to determine by itself when a food item needs to be replenished and also the kind of food items stored inside the refrigerator. The system is built with Raspberry Pi, in addition to the camera(s), which are hooked up to the Raspberry Pi board, for capturing the images of the various food items. The proposed system design will also be smart enough to interact with the user through an Android application to notify the current status and quantity of food items stored inside the refrigerator. With this smart device, the user will be easily prompted about the food items that are less in quantity inside the refrigerator, which can help the user to refill the necessary food items as soon as possible.

Keywords Smart refrigerator · Internet of things (IoT) · Raspberry pi · Camera(s) · Android

A. Sharma (✉) · A. Sarkar · A. Ibrahim
Department of Computer Science and Engineering, School of Technology, Assam Don Bosco University, Guwahati, Assam, India
e-mail: aditya.sharma.adbu@gmail.com

A. Sarkar
e-mail: ashmita.sarkar.as@gmail.com

A. Ibrahim
e-mail: mails4aazaz@gmail.com

R. K. Sharma
Department of Computer Applications, School of Technology, Assam Don Bosco University, Guwahati, Assam, India
e-mail: Rupam.Sharma@dbuniversity.ac.in

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_6

1 Introduction

Health is a primary matter of concern. Food that we eat at home is mostly kept in the fridge. When we shop for groceries, we do not know exactly what is there in the refrigerator or the exact quantity of a particular food item. This problem exists for all people who use refrigerators. If we do not know exactly what is there in the fridge, we will not be able to buy the proper quantity or the essential items of groceries or other edibles. The refrigerators that we use in our daily life are not smart enough to understand our needs and convey the same to us. Their only role is to keep the food fresh and healthy. However, with the upcoming advanced generation, with digital technology, every device is getting smarter day by day. A device like a refrigerator can also be a smart device that makes the user's life easier by prompting about the items the user likes more and informing him for necessary procurement if they are missing.

Smart Refri: Smart Refrigerator for tracking human usage and prompting based on behavioral consumption.

The objectives of Smart Refri are as follows:

- To recognize the food items kept in the fridge.
- To observe the pattern followed by the user while taking out and storing food items in the refrigerator.
- To learn the behavioral patterns exhibited by the user while using the system.
- To prompt the user, if items discovered from his consumption behavior are found either in less quantity or missing.

2 Existing Systems

2.1 Smart Refrigerator –a Next-Generation Refrigerator Connected to IoT

The design of the proposed system of Smart Refrigerator [1] is based on the core concept of product identification utilizing the RFID technology. The researchers of this paper have considered a use case that in the nearby future, all or most products bought from the store will have a tracking RFID tag along with information stored in a global level database maintained by all or most manufacturers, that will eventually serve two purposes: Firstly, the manufacturers will be able to easily track their products from the assembly line, transportation route and finally on the store shelf, which will provide them with more information about assembly and transportation costs along with information about product visibility in the store itself. Secondly, the consumer will be able to easily access the above information, in time. The set of well-defined functionalities on which the concept of the proposed system is built are-

A. Identifying new products
B. Identifying removed products
C. Inventory and shopping list
D. Triggered alerts
E. Information output

However, the proposed system does not provide a solution for the identification of products stored in the refrigerator that is not contained in RFID-equipped packaging like fruits and vegetables.

2.2 IoT Based Smart Refrigerator System

The proposed design is to implement a smart refrigerator system [2], which is easy to use and economical for the user. The system is capable of notifying its owner about the activities going on inside the fridge via the android app that is developed to be used as a GUI for the user. Through the application, the user will be able to see the condition of the food items kept inside the refrigerator. An alert notification is sent to the user's mobile when the weight of the items is below the set threshold value. This helps the user to replenish any food item in the proper time. The proposed system is designed by following the steps mentioned below:

- The initial value (quantity) of a particular food item kept in the fridge is determined. This step is followed for all food items kept in the smart refrigerator.
- The food items are picked from the inventory tray of the fridge.
- The status of food items is compared (i.e. present value with threshold value).
- If the load value is lesser than the threshold value then a notification is sent to the user on the mobile phone.

The proposed smart refrigerator is designed for efficiently managing items stored in it and therefore enables a healthy lifestyle. However, it does not act on analyzing the behavioral patterns of the user and thus helps to minimize food waste.

2.3 Smart Refrigerator Based on Internet of Things (IoT)—An Approach to Efficient Food Management

The proposed concept of the smart refrigerator [3] is designed to include a variety of functionalities that help in both improved food management and also to keep track of the balanced diet for the consumer. The primary issue is to allow consumers to allow the distribution of the food to the neighbors that need the same grocery items before the expiration date has been reached. The application of this project will eventually allow improved food management and hence reduce food wastage. The consumer can input a pre-determined minimum quantity for a food item and when the amount falls

below that threshold value, this will trigger an automated connection to the central database to find whether the same item is available for selling by another consumer. The contact to the consumer can be made through a text message and the quantity at the price available for sale can be specified. Since the distribution is done within the same neighborhood, the cost of transport will be reduced and simultaneously it will be a much efficient way for food management instead of letting it expire on the fridge shelf without consumption. In case, the potential buyer is not interested in the purchase, an online option to purchase from a fresh grocery store or a farmer's market can be made available. The user can also have an option to discard the purchase altogether.

3 Proposed System

With the tremendous improvement in technology, all devices are connected to the internet, which forms the internet of things. The sensors are used to collect the data and send them to a host where it is intended to be processed over the internet. Improvement in technologies, make our day-to-day life simpler. The technologies implemented using IOT in electrical appliances at homes make it smarter. One of those technologies is a smart refrigerator which is used to store food items. Refrigerators are used to prevent the spoilage of food and keep it fresh. It reduces illness and makes our lifestyle healthier in the modern world.

The proposed IOT based device uses Raspberry Pi as the central server to make the fridge smarter. It will be able to detect the different items kept inside the refrigerator and also measure the quantity of the items present inside it. The device will learn the pattern in which the items from the refrigerator are being used in day-to-day life. A Mobile application will be developed to monitor the quantity of items in the refrigerator and notify the user about the items inside the fridge, also about the quantity and if there is any shortage.

When the refrigerator door is opened, the camera clicks a picture (Say img1). Figure 1 will become the input to the TensorFlow Object detection Classifier which will detect the objects (food items) in it. After the door is closed, the camera will click another picture (Say img2). Again the TensorFlow Object detection Classifier will detect the objects (food items) in it. After the objects are detected from both the images then a comparison will be made among the objects in both the images. The objects which will not be found in the Fig. 2 will be considered to be consumed. The objects which are consumed will be inserted into the database. Then Self Organizing Maps(SOM) will read from the database, learn the weights (quantity) of the objects in it and thus it will create a consumption behavior. The consumption behavior will then be added to the Behavioral Database. The alert system created in the Android app will read from the Behavioral Database about the objects that are less than the threshold value and thus it will alert the user about it.

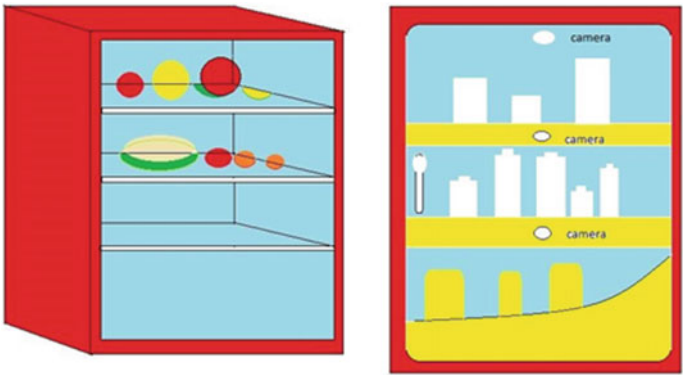


Fig. 1 The proposed model

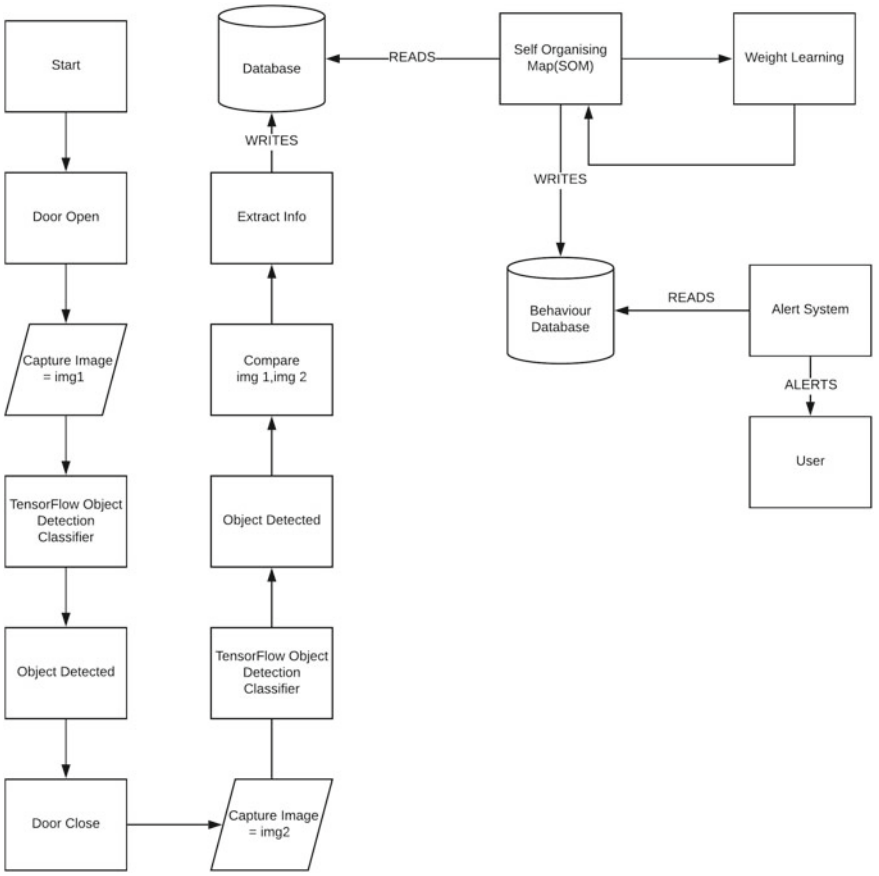


Fig. 2 Block Diagram for the proposed system

TensorFlow Object Detection Classifier. TensorFlow is an open-source deep learning framework created by Google Brain. For training the classifier, we have used TensorFlow's Object Detection API [4]. An image classifier is trained so that it can detect objects in an image and then classify the set of data into different classes or categories.

Self Organizing Maps (SOM). The Self-Organizing Map is one of the most popular neural network models and it belongs to the category of competitive learning networks. The SOM is based on unsupervised learning, which means that little needs to be known about the characteristics of the input data and that no human intervention is needed during the learning.

Behavior Database. The information produced as a result of actions, typically commercial behavior using a range of devices connected to the Internet, such as a PC, tablet, or smartphone is referred as a Behaviour Database. It is therefore not static. Behavioral data is valuable because it provides information above and beyond what static data can provide. The collection of all behavioral data of users is stored in the Behavioral Database.

Alert System. The Alert System is an Android application accessible to the user via a smartphone. A notification sent by the application to the user's device alerts him/her regarding the current circumstances in real-time.

4 Results and Discussion

The goal of the proposed system is to mitigate the problem of food wastage and provide an efficient solution to handle the same. It will allow the enhancement of IoT technologies in new ways to realize the vision of sustainable IoT applications that in return will enable things and objects to become more reliable, resilient, more autonomous and smarter [5].

The proposed system will result in a smart refrigerator system that will be able to help the user to look through the items that are present inside the refrigerator via an android application, which will also prompt the user if any of the items are required to be replenished. The system also learns the pattern in which the items are being consumed by the user.

This system comprises an object detection image classifier that detects the items kept inside the fridge, an android application that will connect the user with the refrigerator to look for items that are kept inside it, and it also includes a behavior learning system that will learn the user's behavior towards their daily usage of the items kept inside the fridge. An alert system is also included inside the android application which will prompt the user about the items that are to be refilled and also about their consumption behavior of the items (Fig. 3).

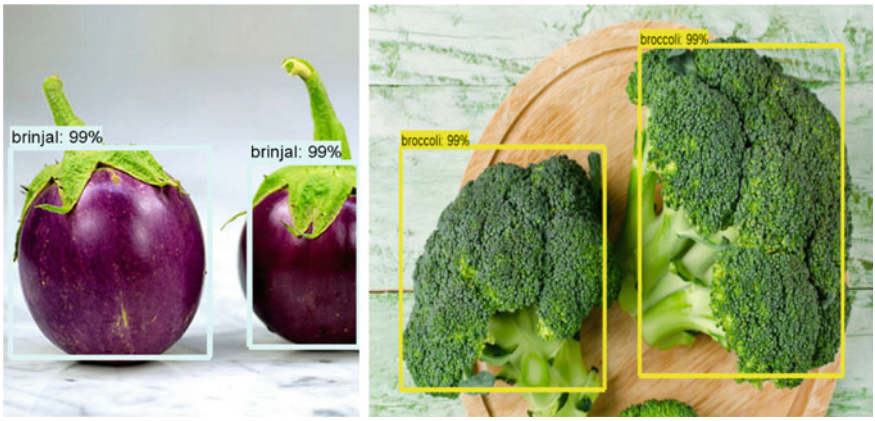


Fig. 3 Output for object detection

The object detection algorithm used to achieve the above outputs is Faster R-CNN. The steps followed by a Faster R-CNN algorithm [6] to detect objects in an image are summarized below:

- An input image is taken and passed to the ConvNet which returns feature maps for the image
- Region Proposal Network (RPN) is applied on these feature maps and object proposals are obtained
- ROI pooling layer is then applied to bring down all the proposals to the same size
- Finally, these proposals are passed to a fully connected layer in order to classify any predict the bounding boxes for the image

ConvNet. A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm that can take in an input image to assign importance (learnable weights and biases) to various aspects/objects in the image and then be able to differentiate one image from the other. Compared to other classification algorithms, the pre-processing required in a ConvNet is much lower. With enough training, ConvNets have the ability to learn the filters/characteristics while in primitive methods these filters are hand-engineered. Without losing features that are critical for getting a good prediction, the role of the ConvNet is to reduce the images into a form that is easier to process.

Feature Maps. The convolutional layer is the main building block of CNN which utilizes the concept of convolution. Convolution is a mathematical operation to merge two sets of information and in our case, the convolution is applied on the input data using a convolution filter to produce a feature map. An image is represented as a 3D matrix with dimensions of height, width and depth in reality, where depth corresponds to color channels (RGB). A convolution filter has a specific height and width, like 3

$\times$ 3 or 5 $\times$ 5 and by design it covers the entire depth of its input so it also needs to be 3D.

We perform multiple convolutions on a single input, using a different filter for each that results in a distinct feature map. Stacking all these feature maps together becomes the final output of the convolution layer.

Region Proposal Network (RPN). The purpose of the Region Proposal Network is to propose multiple objects that are identifiable within a particular image. Using the features that were computed by the CNN, it finds up to a predefined number of regions (bounding boxes), which may contain objects.

ROI pooling layer. Region of Interest (ROI) pooling is used for utilizing a single feature map for all the proposals that are generated by RPN in a single pass. It solves the problem of fixed image size requirements for object detection networks. The RoI Pooling layer is a type of max-pooling, where the pool size is dependent on the size of the input. It ensures that the output is always of the same size. This layer is used because the fully-connected layer always expects the same input size, but input regions to the fully-connected layer may have different sizes.

Fully Connected Layer. Fully connected layers are an essential component of Convolutional Neural Networks (CNNs). It has been proven to be very successful in recognizing and classifying images for computer vision. The CNN process begins with convolution followed by pooling, breaking down the image into features, and then analyzing them independently. The result of this process is fed into a fully connected neural network structure that drives the final classification decision.

The final decision of the object detection with a proper bounding box and labeling is done with the help of a functional matrix. It is a single dimension array, containing the scores with respect to the classes of objects (Fig. 4).

However, the above mentioned approach could not always deliver accurate results. The Faster R-CNN algorithm failed to detect multiple food items in a single image while providing accurate results in detecting single food items within an image. Therefore, a new approach was adapted to obtain better and accurate results for object detection. The object detection algorithm used in this approach is called YOLOv3. It

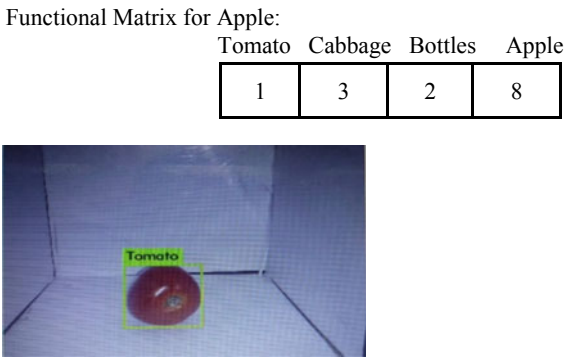


Fig. 4 Real time object detection

is the third object detection algorithm in YOLO (You Only Look Once) family with improved accuracy and is more capable of detecting small objects.

YOLOv3. The YOLOv3 algorithm works by applying a single neural network to the full image. This network divides the image into regions and then predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by the predicted probabilities meaning that each object detected in an image is bounded within a box with its corresponding probability in percentage. The YOLOv3 model has several advantages over classifier-based systems. Its predictions are informed by the global context in the image as it looks at the whole image at test time. Unlike systems like R-CNN which require thousands for a single image, YOLOv3 also makes predictions with a single network evaluation. This makes it extremely fast, faster than R-CNN (1000x) and Fast R-CNN (100x) [7].

5 Conclusion

A Smart Refrigerator system has been proposed for detecting the items kept inside the refrigerator, which learns the behavioral consumption of a single user and notifies the user if there is any shortage of food items. Also through this smart refrigerator, people can save their time in searching for the items inside the refrigerator. We are confident that such type of smart refrigerator will be an important component in the future smart home. The concept of a smart refrigerator is far more reaching than notifying the user about the contents of the refrigerator. A smart refrigerator is cost-effective and user-friendly. Thus, this system saves manual effort and time in predicting future needs. In the future, this system may be expanded to detect the freshness of the food items and also suggest the dishes that can be made from the available food items in the fridge considering the freshness of that item. This system may also be made to display the content inside the fridge from a remote place.

References

1. Floarea AD, Sgarciu V (2016) Smart refrigerator: A next generation refrigerator connected to the IoT. In: 8th International conference on electronics, computers and artificial intelligence (ECAI). https://doi.org/10.1109/ecai.2016.7861170
2. Singh D, Jain P (2016) IoT based smart refrıgerator system. Int J Adv Res Electron Commun Eng (IJARECE) 5(7)
3. Hossain S, Abdelgawad A (2018) Smart refrigerator based on ınternet of things (ıot) an approach to efficient food management. In: ICSDE'18 Proceedings of the 2nd international conference on smart digital environment, pp 15–18, Rabat, Morocco, 18–20 Oct 2018
4. Huang J, Rathod V, Sun C, Zhu M, Korattikara A, Fathi A, Fischer I, Wojna Z, Song Y, Guadarrama S, Murphy K (2017) Speed/accuracy trade-offs for modern convolutional object detectors, CVPR 2017
5. Kyriazis D, Varvarigou T (2013) Smart, autonomous and reliable Internet of Things, 2013. In: International workshop on communications and sensor networks (ComSense-2013)

6. Ren S, He K, Girshick R, Sun J (2016) Faster R-CNN: Towards real-time object detection with region proposal networks. arXiv:1506.01497v3 [cs.CV] 6 Jan 2016
7. Redmon J, Farhadi A (2018) Yolov3: An incremental improvement. arXiv

Cascade-Based Pedestrian Detector Using Edge and Pattern Features

Amlan Jyoti Das, Navajit Saikia, and Simantika Choudhury

Abstract Pedestrian detection plays a significant role in computer vision applications and has been a popular research topic. In this paper, a new pedestrian detector is proposed with a combination of the edge-based histogram of oriented gradient features and pattern-based dense local difference binary features. An important requirement in a pedestrian detector is computational speed. Cascade-based classifiers provide a good trade-off between accuracy and speed. This work uses a cascade of boosted classifiers to enhance the detection speed. To further boost the speed, the histogram of oriented gradients is computed using integral images. The proposed system is evaluated in terms of computational speed as well as precision versus recall and miss-rate versus FPPW/FPPI. The performance is also compared with similar existing pedestrian detectors.

Keywords Pedestrian detection · Cascade classifier · Edge-based feature · Pattern-based feature

1 Introduction

A vital area in computer vision is pedestrian detection which has wide applications in robotics, surveillance, intelligent vehicles, etc. Challenges involved in the detection of humans in different outdoor environments are variations in poses, occlusion, illumination, etc. and also real-time data processing. Pedestrian detectors employ either a handcrafted feature-based or data-driven approach. The handcrafted feature-based approaches require comparatively smaller training datasets and hardware with

A. J. Das (✉) · S. Choudhury
Gauhati University, Guwahati 781014, Assam, India
e-mail: amlandas78@gmail.com

S. Choudhury
e-mail: csimantika@gmail.com

N. Saikia
Assam Engineering College, Guwahati 781014, Assam, India
e-mail: navajit.ete@aec.ac.in

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_7

lower computational capacity than the data-driven approaches. However, it requires designing robust and discriminative features, and also selecting an efficient classifier. Popular handcrafted features include Histogram of Oriented Gradients (HOG) [2], Haar-like features [15], Local Binary Pattern (LBP) [10], Channel features [5], etc. People also combine different features to attain better detection performance [8, 13, 14]. As gradient information provides important cues for human shape, most of the combined feature sets use it as the primary component. Some widely used classifiers include variants of Support Vector Machine (SVM) [2], decision tree and boosted classifiers [5]. Two primary aspects of the pedestrian detector are detection performance and speed [5, 17]. This paper introduces a pedestrian detector that uses a cascade of boosted classifiers with a combination of HOG and dense Local Difference Binary (dense LDB) [3] features. Here we employ a cascade of boosted classifiers in order to enhance the detection speed. Integral images are also considered for faster feature computation.

The rest of the paper is organized as follows: Sect. 2 discusses about some related methods in pedestrian detection. A brief overview of the related concepts is presented in Sect. 3. Section 4 describes the proposed methodology and Sect. 5 presents the experimental results. Section 6 concludes the paper.

2 Related Works

A considerable amount of research works in the area of pedestrian detection is available in the literature. Most of these works are evaluated on some popular datasets like Caltech [4], ETH [9], Daimler [6], INRIA [2], etc. HOG, introduced in [2], is one of the most commonly used gradient-based feature descriptors because of its ability to describe pedestrians efficiently. Here, the classifier considered is Linear SVM. Since pedestrian detection is a complex problem, a single type of feature can barely solve all the challenges [7]. Researchers combine several features to enhance performance. Walk et al. [13] employ a combination of HOG, Histogram of Optical Flow (HOF) and color self-similarity. LBP features are combined with HOG to handle partial occlusions in [14]. In [3], HOG features are coupled with dense LDB descriptors to demonstrate the advantages of using pattern information along with gradient-based features. Viola et al. [12] use the frame differencing method and Haar features with a cascade of boosted classifiers to achieve a better speed of detection. The feature set in [8] uses color, texture and edge information. In [5], Integral Channel Features (ICF) are used by combining the gradient magnitude, gradient histograms and color channels. Many literatures using data-driven approaches have been reported and some of them are mentioned in [1, 16]. Unlike handcrafted feature-based techniques, data-driven techniques use large datasets for training like Caltech [4] and therefore are out of scope of this paper.

The basic objective of this work is to enhance the detection speed of the detector in [3] by using a similar set of features. In [3], HOG and dense LDB features are used with a linear SVM classifier. The HOG features represent gradient information

that closely characterizes local shape. Dense LDB features capture local pattern information and intrinsically eliminates noise. The detection speed of the detector in [3] is around 2.5 s with frames of size 640X480 which is beyond the scope of real-time implementation. This work proposes to employ a cascade of boosted classifiers in combination with HOG and dense LDB features to achieve a good trade-off between detection performance and speed for real-time applications. To achieve a faster detection speed, we propose to use the following:

- HOG features are to be computed using the integral image for faster computation as in [17]. Since the computation of dense LDB features also uses integral images, the overall feature computation will become faster.
- To efficiently capture different parts of a human, variable block sizes of different scales and aspect ratios are to be considered as in [17].
- Faster detection is proposed to be achieved by using a cascade of boosted classifiers in place of linear SVM in [3] as the early rejection stages in a cascade of boosted classifiers reduce the number of features to be evaluated per window.

The Section below describes related theories in brief.

3 Theoretical Background

3.1 HOG Features

The fundamental concept behind the HOG descriptor, introduced by Dalal and Triggs [2], is that it can be represented by the distribution of edge directions or intensity gradients describing the shape and local object appearances within the detection window. These features are evaluated by splitting the detection window into "cells" by considering 9 gradient directions of equal space and then by accumulating their weights, a local histogram of gradient directions is generated. In order to enhance the features' performance in presence of shadowing and variation in illumination, a larger spatial region known as "block" normalizes the contrast of the histograms. These normalized local histograms thus produced derive the HOG descriptor.

It is observed in [17] that when blocks of fixed sizes are used in HOG it fails to capture useful patterns of variable scales and aspect ratios. Hence, it is proposed to consider blocks of variable sizes, location and aspect ratios in order to provide a dense representation of the detection window. Moreover, the authors here calculate HOG features in a faster way with the use of integral images.

3.2 Dense LDB Features

These features are introduced in [3] for pedestrian detection. Dense LDB is a binary descriptor that is obtained by computing the difference of regional gradients and the average intensity of local cell pairs inside a block. This descriptor has inherent capability of eliminating noise. The dense LDB descriptor here uses variable-sized blocks as used in the computation of integral HOG features unlike the one used in [3] which uses fixed block size.

3.3 Cascade of Boosted Classifiers

A cascade classifier is a collection of stages, where each stage consists of weak classifiers [11]. These stages are trained by using a boosting technique. The weighted average of the choices of weak classifiers is taken which provides training of a strong classifier. The stages of a cascade are designed in such a way that they can reject negative patches as fast as possible. It is assumed that the majority of the detection windows do not possess the object of interest and are rejected at the initial stages [11].

4 Proposed Methodology

Figure 1 shows the block diagram of the proposed technique for pedestrian detection. The functions of the different blocks are discussed in the following.

Computation of integral HOG features: Here, HOG features are computed by using the integral image and variable block sizes are used to represent the detection window in a detailed manner. Here, we use integral HOG to keep a trade-off between the performance and speed of the detection system. We considered block sizes from 8×8 to 32×64 in a window of size 64×128 with aspect ratios of 1:1, 1:2 and 2:1. A total number of 36 features per block are HOG features. Based on observations, the step size of 4 is selected for blocks. The total of 1534 blocks per detection window is derived by using variable block size.

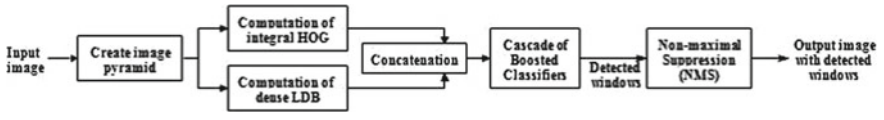


Fig. 1 System block diagram

Computation of dense LDB features: Here, the cell, block and window sizes considered are similar to the computation of HOG features using the same integral image. The input given to dense LDB is a gamma and color normalized image and a total of 18 dense LDB features per block are obtained, referring to our previous work [3]. These are then combined with the 'blocks' representative bits in the detection window to attain the LDB features. A total number of 1890 binary features per window is achieved.

Training of the Cascade Classifier: As mentioned above, the cascade of boosted classifiers are used for selection and learning of the integral HOG and dense LDB features using Adaboost. The cascade classifier here contains 21 stages having a total of 446 weak classifiers out of which the first 3 stages rejects almost 95% of the probable negative windows. Depth-2 decision tree is used as a weak classifier. The minimum detection rate and maximum false alarm rate are considered as 0.9975 and 0.3, respectively. The training steps are presented in Algorithm 1. The INRIA dataset is used for training which consists of 614 positive and 1218 negative images. By using 1208 annotated pedestrian windows in 614 positive images and their mirror images, positive windows for training are generated.

Algorithm 1 Cascade classifier training

Input: F_{target}: target overall false positive rate
f_{max}: maximum acceptable false positive rate per cascade level
d_{min}: minimum acceptable detection per cascade level
P: set of positive samples
N: set of negative samples
Initialization: $i = 0,\ D_i = 1.0,\ F_i = 1.0$
Loop_1 start: while $F_i > F_{target}$
$i++$
$f_i = 1.0$
Loop_2 start: while $f_i > f_{max}$
1) use P and N to train depth-2 decision tree classifier (weak) using Adaboost
2) add the best weak classifier into the strong classifier, update the weight in Adaboost manner
3) evaluate P and N by current cascaded (strong) classifier to determine F_i and D_i
4) decrease threshold for the current cascaded (strong) classifier until d_{min}holds
5) compute f_i in this threshold
Loop_2 end
$F_{i+1} = F_i \times f_i$
$D_{i+1} = D_i \times d_{min}$
$N = \phi$
if $F_i > F_{target}$ then evaluate the current cascaded (strong) classifier on the set of non-human images and add any misclassified samples into the set N.

Loop_1 end

Output: A i-levels cascade classifier whose each level has a boosted classifier of depth-2 decision trees.
Final training accuracy: F_i and D_i

5 Experimental Results and Discussions

The proposed pedestrian detector is evaluated by using the INRIA dataset with 288 positive test images in it. The performance is measured in terms of miss- rate versus FPPI (False Positive Per Image), miss-rate versus FPPW (False Positive Per Window), precision versus recall and computational speed. To determine the presence of a pedestrian in a test window, a final score at the output of each cascade stage is compared with a threshold computed for that stage. For eliminating nearby detections, a non-maximal suppression is applied at the final stage.

To examine miss-rate versus FPPW, 288 positive test images of the dataset are used, from which 588 pedestrian windows are considered. The mirror images of the 588 pedestrian windows are also used as positive test windows. From the negative test images, all the windows are considered as negative test windows. At 10^{-4} FPPW, the classifier attains a miss-rate of 91.6%.

For examining miss-rate versus FPPI, the classifier uses an image pyramid to detect pedestrians in multiple scales. The PASCAL criteria are used to handle partial detection. Performance is tested for scale factors ranging from 1.03 to 1.1 at a step size of 0.01. Miss-rate versus FPPI is plotted at different scale factors, is shown in Fig. 2. Due to clarity of presentation only, the three best performing detectors for different scale factors are shown. From the figure, at a scale factor of 1.05, the best detection performance is achieved. Table 1 shows the miss-rate of the detector at different FPPI values for different scale factors as observed from Fig. 2. The miss rate versus FPPI plot for 1.04 and 1.06 shows that it achieves a close miss-rate to 1.05 at lower values of FPPI. At 1 FPPI, the detector achieves 82% recall or 18% miss-rate.

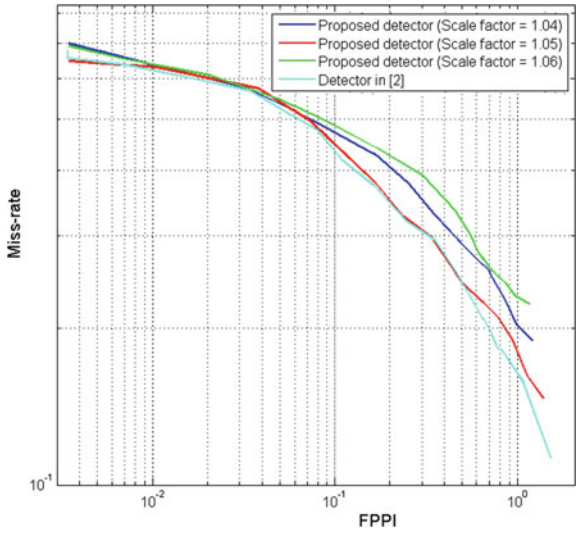


Fig. 2 Miss-rate versus FPPI plots

Table 1 Miss-rate of the detector using different scale factors at different FPPI values

Scale factor/FPPI	1.04	1.05	1.06
Minimum FPPI	70.0%	**64.7%**	69.1%
0.01	64.5%	**62.9%**	64.2%
0.1	46.5%	**43.0%**	48.9%
1	20.2%	**18.0%**	22.8%

Figure 3 shows the precision versus recall plot for the detector. It is seen from the figure that the cascade-based detector has similar and comparable recalls at all precision values. The performance is also similar to the work presented in [3] in terms of recall as shown in Fig. 2 (difference of around 1% at 1 FPPI). However, to be noted that the proposed detector processes a frame of size 640 × 480 at about 0.2 s as compared to 2.5 s in [3]. To maintain a good trade-off between detection accuracy and computational speed, it can be said that the proposed detector has achieved a better overall performance compared to [3] since there is a speedup of around 13 times. The reason behind this is due to the usage of cascade classifier which rejects negative test windows at an early stage and also the usage of the integral image while computing the features.

Table 2 provides a comparison with the existing detectors considered in [3]. As discussed in Sect. 2, we are not considering the data-driven approaches as they do not use INRIA dataset due to its small size. From Table 2, it may be observed that the miss-rate of the proposed detector is lower than the detectors presented in [2, 14] and slightly higher (by 1%) to the detector in [3] at 1 FPPI. At 0.1 FPPI, the performance

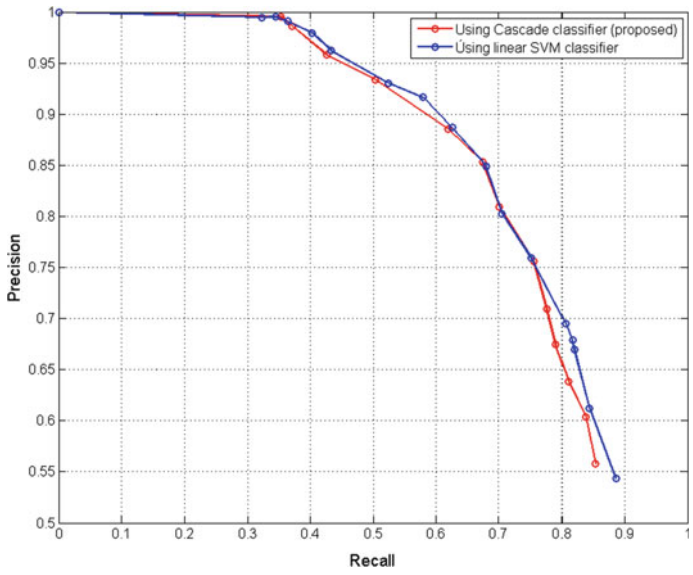


Fig. 3 Precision versus Recall plot

Table 2 Comparison of performance with similar existing detectors

Detector	Classifier	Training Dataset	% Miss-rate at			Sec/ frame
			1 FPPI	0.1 FPPI	Min FPPI	
HOG [2]	Linear SVM	INRIA	23	46	88	2.3
HOG + LBP [14]	Linear SVM	INRIA	19	39	70	9.1
HOG + Dense LDB [3]	Linear SVM	INRIA	17	41	68	2.5
Proposed method	Cascade of Boosted	INRIA	18	43	64.7	**0.2**

of this detector is marginally lower than the detectors in [14] and [3] and better than [2]. However, the proposed detector performs better than all the detectors presented in the table at minimum FPPI. The notable improvement of the proposed detector is in terms of computational speed. It can process at around 5 fps and outperforms the other detectors in Table 2. This makes the detector suitable for real-time applications.

6 Conclusion

This paper has introduced a new system for pedestrian detection based on the combination of integral HOG and dense LDB features to enhance the performance of detection. Cascade of boosted classifiers has been used for classification in order to improve computational speed. Variable block sizes have been considered to efficiently represent useful patterns. The proposed system has been trained and tested with the INRIA person dataset and the performance has been measured in terms of precision versus recall, miss-rate versus FPPI, miss-rate versus FPPW and computational speed. The experimental results show a detection rate of 91.6% at 10^{-4} FPPW and 82% at 1 FPPI. Although the detection rate at 1 FPPI of this system is reduced by 1% as compared to the detector in [3], the computational speed has increased by about 13 times and hence the proposed system may be used in real-time. This speed may be further increased by using GPGPU.

Acknowledgements This work is supported by the Department of Science and Technology, Govt. of India, under the INSPIRE fellowship program.

References

1. Benenson R, Omran M, Hosang J, Schiele B (2015) Ten years of pedestrian detection, what have we learned? In: Agapito L, Bronstein MM, Rother C (eds) Computer vision—ECCV 2014 workshops. Springer International Publishing, Cham, pp 613–627

2. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol 1, pp 886–893
3. Das AJ, Saikia N (2016) Pedestrian detection using dense LDB descriptor combined with hog. In: 2016 international conference on information technology (InCITe)—The next generation IT Summit on the Theme—Internet of Things: Connect your Worlds, pp 299–304
4. Dollar P, Wojek C, Schiele B, Perona P (2009) Pedestrian detection: a benchmark. In: 2009 ieee conference on computer vision and pattern recognition, pp 304–311
5. Dollar P, Tu Z, Perona P, Belongie SJ (2009) Integral channel features. In: BMVC
6. Enzweiler M, Gavrila DM (2009) Monocular pedestrian detection: Survey and experiments. IEEE Trans Pattern Anal Mach Intell 31(12):2179–2195
7. Mori G, Belongie S, Malik J (2005) Efficient shape matching using shape contexts. IEEE Trans Pattern Anal Mach Intell 27(11):1832–1837
8. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987
9. Ouyang W, Wang X (2012) A discriminative deep model for pedestrian detection with occlusion handling. In: 2012 IEEE conference on computer vision and pattern recognition, pp 3258–3265
10. Tuzel O, Porikli F, Meer P (2008) Pedestrian detection via classification on riemannian manifolds. IEEE Trans Pattern Anal Mach Intell 30(10):1713–1727
11. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, vol 1, pp I–I
12. Viola P, Jones MJ, Snow D (2003) Detecting pedestrians using patterns of motion and appearance. In: Proceedings Ninth IEEE international conference on computer vision, vol 2, pp 734–741
13. Walk S, Majer N, Schindler K, Schiele B (2010) New features and insights for pedestrian detection. In: 2010 IEEE computer society conference on computer vision and pattern recognition, pp 1030–1037
14. Wang X, Han TX, Yan S (2009) An hog-lbp human detector with partial occlusion handling. In: 2009 IEEE 12th international conference on computer vision, pp 32–39
15. Wojek C, Schiele B (2008) A performance evaluation of single and multi-feature people detection. In: Rigoll G (ed) Pattern recognition. Springer Berlin Heidelberg, pp 82–91
16. Zhang L, Lin L, Liang X, He K (2016) Is faster r-cnn doing well for pedestrian detection? In: Leibe B, Matas J, Sebe N, Welling M (eds) Computer vision—ECCV 2016. Springer International Publishing, Cham, pp 443–457
17. Zhu Q, Yeh MC, Cheng KT, Avidan S (2006) Fast human detection using a cascade of histograms of oriented gradients. In: 2006 IEEE computer society conference on computer vision and pattern recognition, vol 2, pp 1491–1498

Nanomaterial-Based Microstrip Patch Antenna Array for X Band Operation

Parismita A. Kashyap, Smriti Rekha Das, and Sunandan Baruah

Abstract Antenna designers around the globe have extensively researched upon the patch antenna and its array structures for achieving better performance for various operations. One of the novel techniques is to use nanomaterial composites over the substrate of a microstrip patch antenna to enhance its performance. Nanostructures can provide unique characteristics and have the potential to improve the performance of the patch antenna. This paper focuses on the study and design of a nanomaterial-based 1 × 2 rectangular microstrip patch antenna array for X band operation. The proposed antenna array is designed over an FR4 substrate where the individual radiating patches are realized using gold deposited Zinc Oxide (ZnO) nanorods. The comparison of performance parameters of the proposed antenna array with the conventional copper (Cu)-based patch antenna array is made. The results obtained show that there is an improvement in the return loss and bandwidth of the proposed array structure.

Keywords Microstrip patch antenna · Nanorods · Return loss · Bandwidth

1 Introduction

Microstrip antennas are the ideal choice for wireless devices because of its characteristics like low weight, low cost, and ease of fabrication [1]. However, microstrip antennas also have disadvantages like less bandwidth and low gain [2]. Intensive

P. A. Kashyap (✉)
Assam Don Bosco University, Azara, Guwahati 781017, Assam, India
e-mail: parismitak3@gmail.com

S. R. Das
Gauhati University, Guwahati 781014, Assam, India

S. Baruah
Assam Down Town University, Guwahati 781068, Assam, India

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_8

research has been done in recent years to develop different performance enhancement techniques of the microstrip patch antenna. These techniques include utilization of thick substrates with low dielectric constant, slotted patch, metamaterial-based antenna, antenna array, and nano-material-based antenna, etc. [3]. Microstrip antennas are very versatile and are also used, to synthesize a required pattern that cannot be attained with a single element. In addition, microstrip patch antennas are used to scan the beam of an antenna system which increases the directivity and performs various other functions which would be difficult with only one element [4]. Hence, array structures are used. But, as the patch antenna suffers from a number of drawbacks, it is of utmost importance to improve these parameters while using it in an array.

Designing of patch antennas using nanomaterials is a new trend because using nanomaterials the size of the antenna as well the return loss can be reduced [5]. The use of conductive carbon nanotubes (CNT) in radiating elements has led to the development of low-cost and lightweight patch antennas. However, the biggest challenges involved in working with nanomaterial like CNTs is the ability to make reliable electrical contacts with other conductive parts of the antenna such as feed line, matching transformer, stub, etc. This unreliability contact makes it difficult to determine the RF properties of the designed antenna [5].

In this paper, the performance of nanomaterial-based patch antenna array is investigated and compared with Cu-based patch antenna array for operation in the X band. A 1 $\times$ 2 antenna array with Cu-based patch using corporate feeding is first simulated in High Frequency Structure Simulator (HFSS) software in the X band and then the same structure is fabricated and the results are obtained. The proposed nanomaterial-based antenna array consisted of ZnO nanorods formed over FR4 substrate by hydrothermal process and gold depositions over the nanorods. These metallic nanorods serve as the radiating element or patch of the antenna. The results of the proposed antenna array and the Cu-based antenna array are compared to investigate the feasibility and effectiveness of nanomaterial-based antenna arrays for operation in the X band.

The rest of the paper is organized as follows: Sect. 2 discusses about the structure of the proposed antenna array, Sect. 3 includes the fabrication technique and methodology. Simulation and fabrication results are discussed in Sect. 4. A small acknowledgement to the organizations is provided without whose helping hands this work would not have been possible. Section 5 concludes the work.

2 Proposed Antenna Design

The geometry of the proposed 1 $\times$ 2 antenna array is shown in Fig. 1. In this work, a two-element antenna array is designed with copper as well as gold as metal nanoparticles deposited over semiconductor ZnO nanorods which are grown on patch with two different concentration, i.e., 5 mM and 10 mM. Initially, simulation of two-element antenna array is done. Table 1 shows all the physical dimensions of the two-element

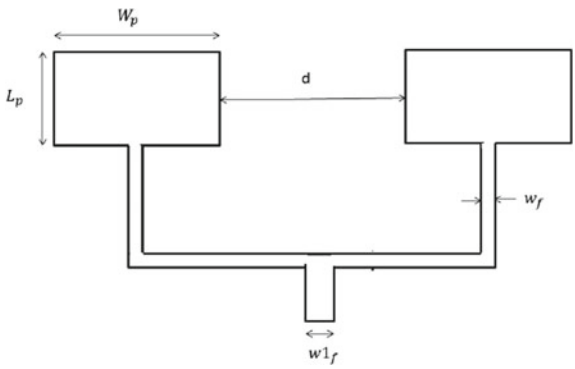


Fig. 1 Geometry of the 1 × 2 antenna array

Table 1 Physical Dimension of the 1 × 2 Antenna Array

Parameters	Values
Operating frequency,f_r	10 GHz
Length of the individual patch,L_p	6.4 mm
Width of the individual patch,W_p	9 mm
Length of the ground plane,L_g	26 mm
Width of the ground plane, W_g	40 mm
Height of the FR4 substrate, h	1.5 mm
Inter-element distance, d	15 mm
Width of the 50 Ω impedance line, w_{1f}	2.86 mm
Width of the 100 Ω impedance line, w_f	0.664 mm

array. The most critical aspect in designing antenna array is the spacing between the individual elements of the array, which can be between 0.2λ and λ. The proposed array is designed using corporate feeding technique.

3 Fabrication of the Nanomaterial-Based Antenna

The various steps during the fabrication of the nanomaterial-based antenna are discussed in the subsections below:

3.1 Synthesis of ZnO Nanomaterial

20 ml of 4 mM zinc acetate dihydrate ($Zn(CH_3COO)_2.2H_2O$) was stirred rigorously at 50 °C for 45 min and then it is diluted with 20 ml fresh ethanol. At room temperature, 20 ml of 4 mM NaOH is added drop-wise under mild stirring which is continued for

about 15 min. The mixture is then put in a water bath at 50° to 60 °C for 2-3 h [6]. The solution is then cooled to room temperature. The chemical reaction taking place is:

$$Zn[(CH_3COO)]_2 \cdot 2H_2O + 2NaOH \rightarrow ZnO + 2NaCH_3COO + H_2O\,[7] \quad (1)$$

3.2 Direct Seeding in FR4 Epoxy Substrate

At first, the copper layer on the substrate is etched out using Iron Chloride. After etching the substrate is rinsed with deionized water. After that masking is done on whole part of the substrate except the size of the patch. Then 2 ml of ZnO nanoparticle is added with 2 ml of deionized water and then it was dropped on the substrates which is marked as patch for 5 times at 100 °C.

3.3 Synthesis of ZnO Nanorods

Nanorods are grown in a sealed chemical bath containing equimolar solution of zinc nitrate hexahydrate ($Zn(NO_3)_2.6H_2O$) and hexamethylenetetramine ($C_6H_{12}N_4$) [6].

Solvent: Deionized water.

Stock solution of 100 ml of 5 mM ($Zn(NO_3)2.6H_2O$) and 100 ml of 5 mM hexamethylenetetramine ($C_6H_{12}N_4$) and 100 ml of 10 mM ($Zn(NO_3)2.6H_2O$) and 100 ml of 10 mM hexamethylenetetramine ($C_6H_{12}N4$) are prepared by taking two conical flasks or beaker. 100 ml of distilled water is taken in both of them and 0.297 g of ($Zn(NO_3)_2.6H_2O$) and 0.140186 g of ($C_6H_{12}N_4$) is put in the conical flask, respectively. The seeded substrates are arranged in such a way that the seeded side is kept upside down inside the petri dish with a slight elevation. Pour 30 ml each of solution of ($Zn(NO_3)2.6H_2O$) and ($C_6H_{12}N_4$) in a petri dish. The substrate should be completely inside the solution. The hot air oven is set at 90° C and the petri dish is kept inside it. The solutions are changed after 5 h. The growth is continued up to 15 to 20 h. The substrate is heated at 100 °C temperature for annealing to remove organic deposits. After hydrothermal growth of ZnO nanorods on FR-4 substrate, a metallic Au layer is deposited on the ZnO nanorods grown on the substrate.

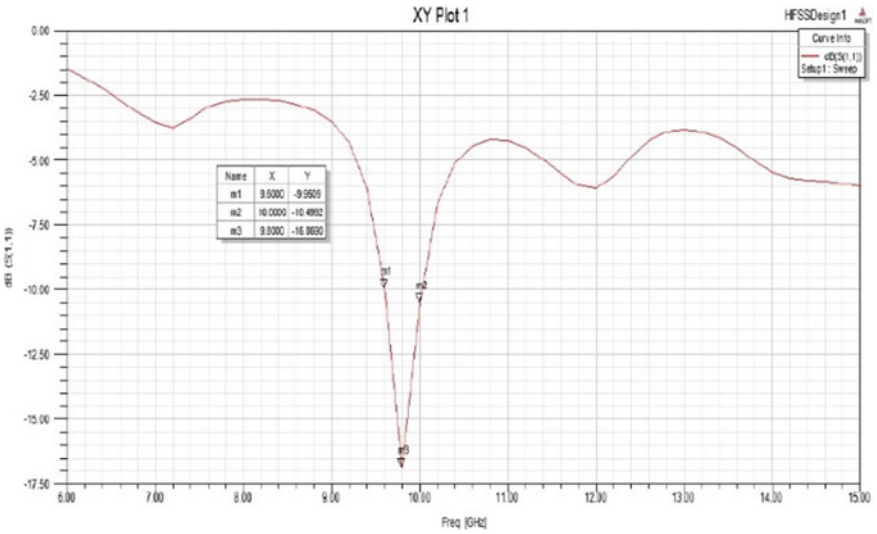


Fig. 2 Return loss of 1 × 2 antenna array

4 Results and Discussions

4.1 Simulation of Cu-Based Antenna Array

As mentioned, a 1 × 2 Cu-based antenna array is first simulated in HFSS software. The simulated return loss is shown in Fig. 2.

It can be seen that the return loss is about −16.86 dB at 9.8 GHz which is quite good for a microstrip patch array antenna. The bandwidth obtained is 0.4 GHz. Figure 3 shows the radiation pattern of the array. It can be seen that the pattern is directive with some about of losses in the form of sidelobes.

Figures 4 and 5 shows the gain and directivity of the simulated antenna array respectively. The gain is found to be 7.952 dB and the directivity is 9.337 dB, which is quite obvious as the gain and directivity are enhanced in an array structure compared to single antennas.

4.2 Measured Results of Cu-Based Antenna Array

The fabricated Cu-based 1 × 2 antenna array with corporate feed is shown in Fig. 6. Return loss of the fabricated Cu-based 1 × 2 antenna array is shown in Fig. 7. The return loss is found to be −15.79 dB at 11.18 GHz which is found to be quite similar with the simulation result. The shift in frequency is due to change in electrical length of the patch during the fabrication process. The bandwidth is found to be 0.47 GHz.

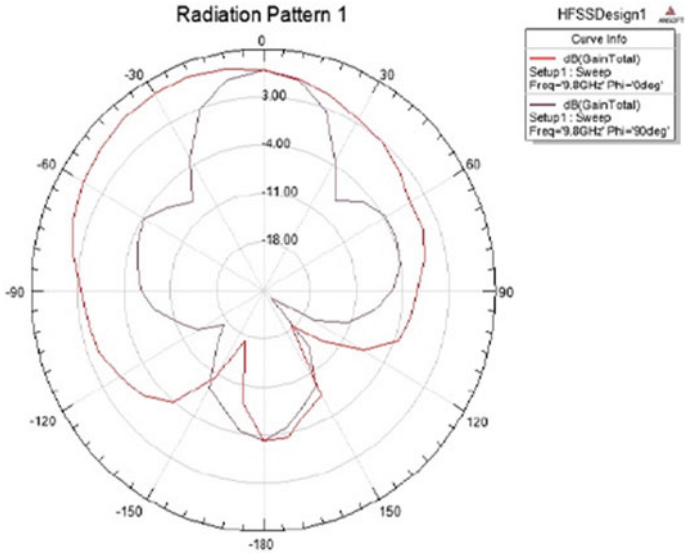


Fig. 3 Radiation pattern of the 1 × 2 antenna array

Fig. 4 Gain of 1 × 2 antenna array

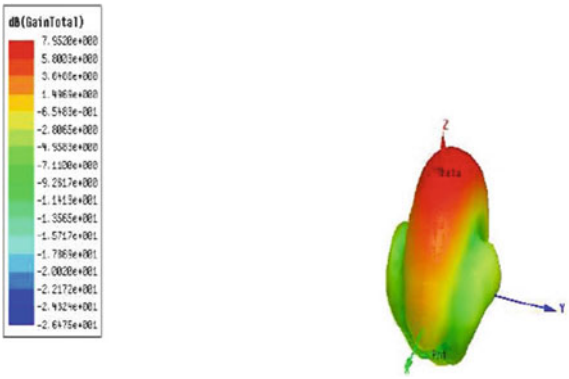


Fig. 5 Directivity of 1 × 2 antenna array

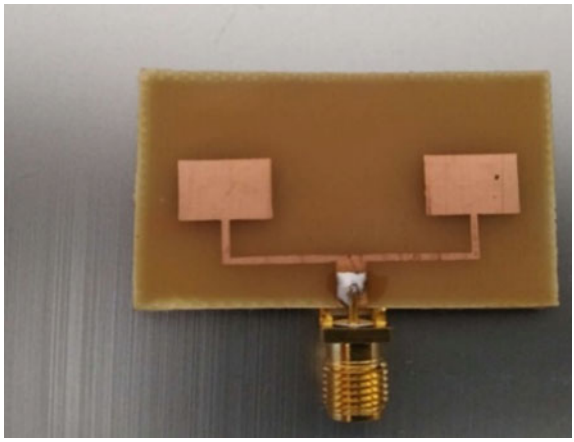


Fig. 6 Fabricated Cu-based 1 × 2 antenna array

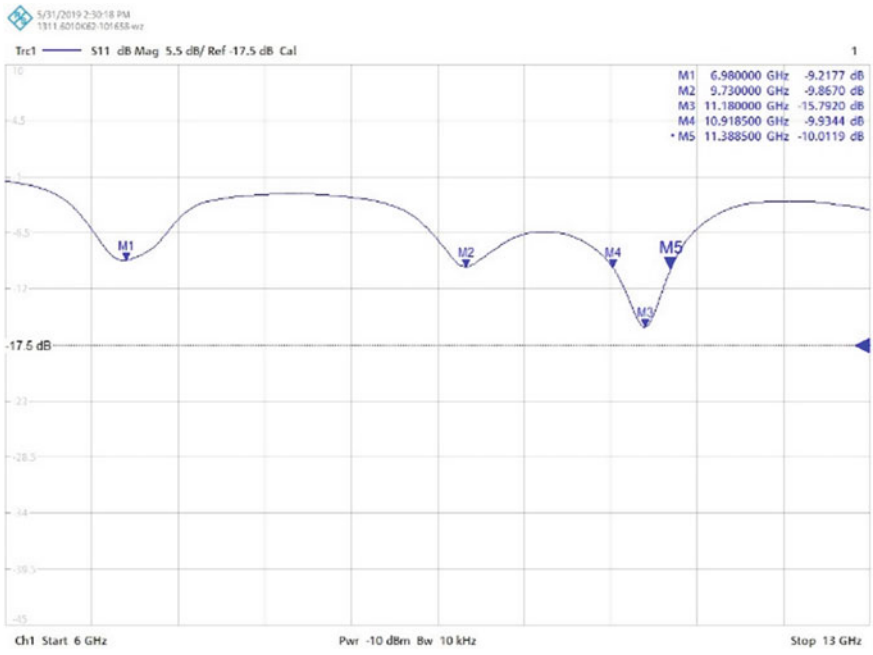


Fig. 7 Return Loss of Cu-based 1 × 2 antenna array

Figure 8 shows the radiation pattern of the array which is also similar to the one found in simulation.

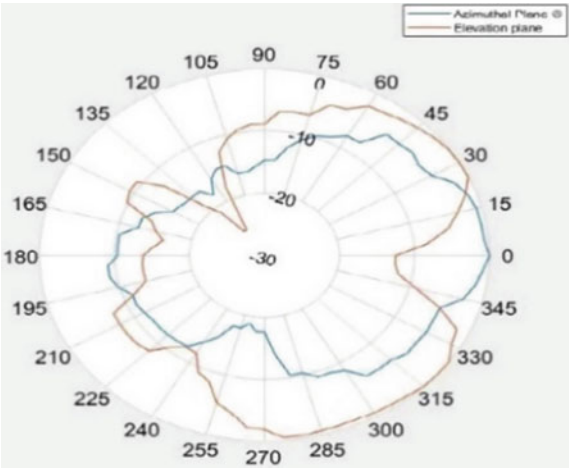


Fig. 8 Radiation pattern of Cu-based 1 × 2 antenna array

4.3 Measured Results of Proposed Nanomaterial-Based Antenna Array

The fabricated 1 × 2 antenna array with 5 mM and 10 mM concentration Au deposited over ZnO nanorods grown on patch is shown in Figs. 9 and 10, respectively.

The return loss of the fabricated arrays with 5 mM and 10 mM concentrations respectively, are shown in Figs. 11 and 12. It can be seen that the array with 5 mM concentration has a return loss of −17.02 dB at 12.12 GHz and a bandwidth of 1.16 GHz. The array with 10 mM concentration has a return loss of −21.10 dB at 9.05 GHz and a bandwidth of 0.805 GHz. In both cases, it is observed that the return loss and especially the bandwidth of the nanomaterial-based arrays are improved compared to the conventional Cu-based patch antenna array. The radiation pattern of both 5 mM and 10 mM concentrations arrays are also shown in Figs. 13 and

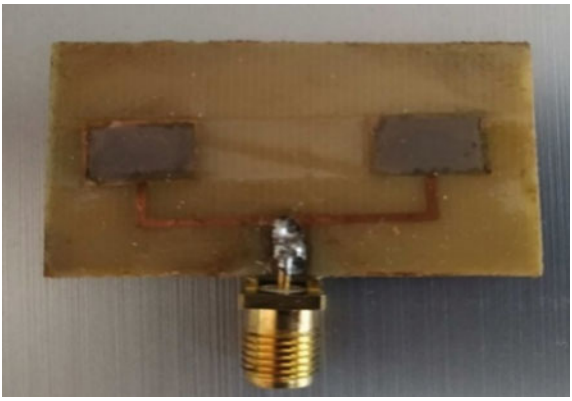


Fig. 9 Fabricated 1 × 2 antenna array with 5 mM Au deposited ZnO nanorods on patch

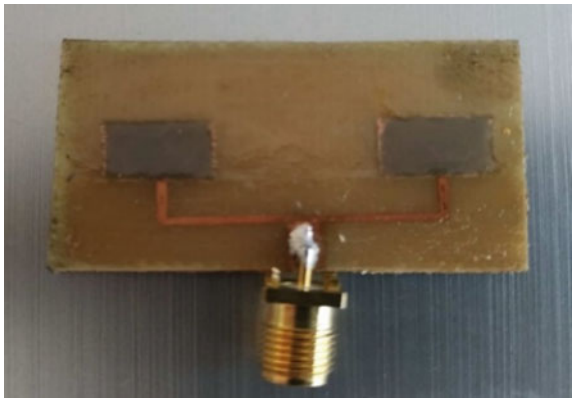


Fig. 10 Fabricated 1 × 2 antenna array with 10 mM Au deposited ZnO nanorods on patch

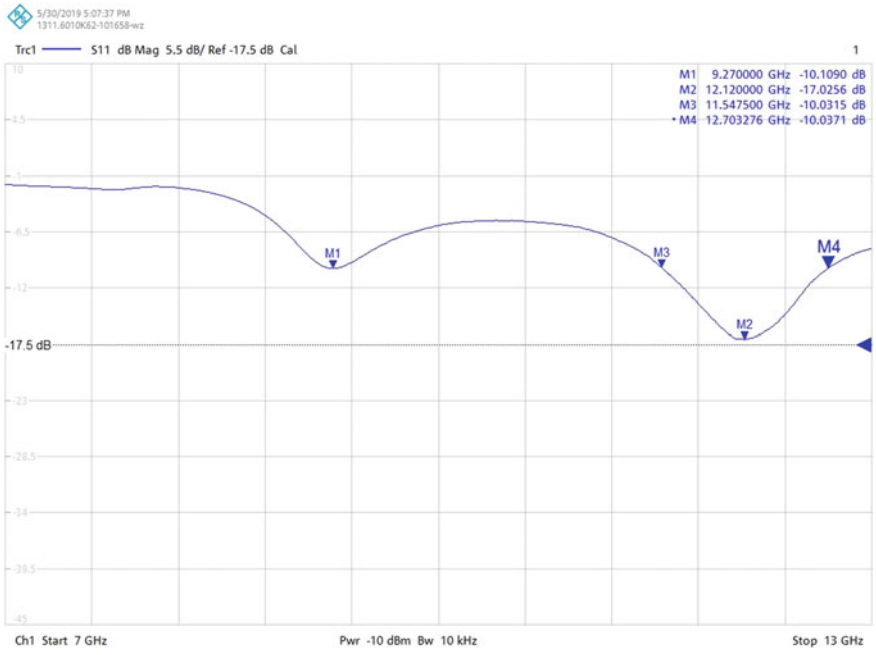


Fig. 11 Return Loss 1 × 2 antenna array with 5 mM Au-deposited ZnO nanorods on patch

14, respectively and in this too, an improvement in directivity is seen compared to conventional Cu-based array.

Table 2 provides a comparison of the measured parameters like return loss, bandwidth, and beamwidth of all the fabricated antennas. From the table, it is observed that the resonant frequency of fabricated patch antennas is different with different material. This is due to the electric length of that material. Again, it can be observed

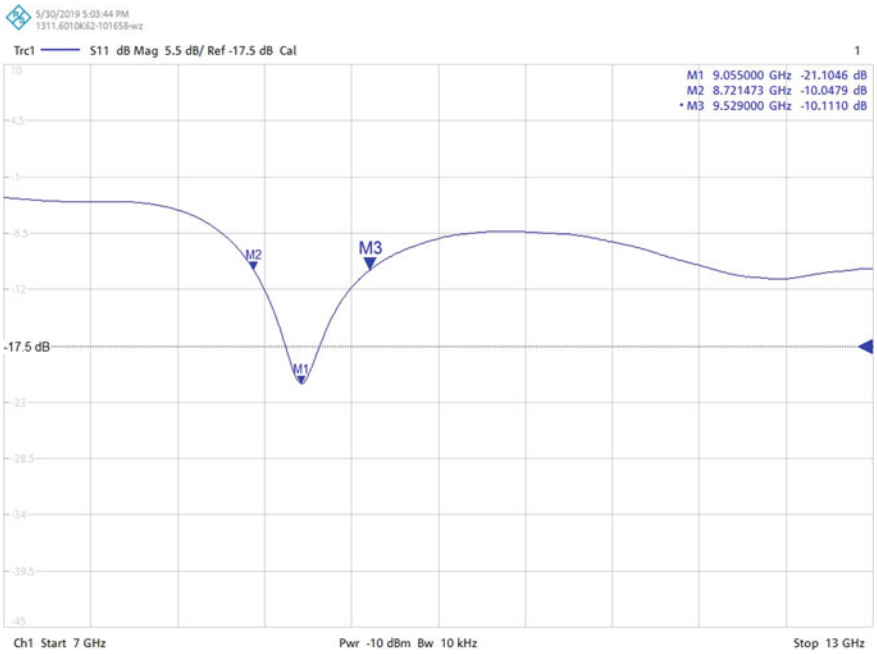


Fig. 12 Return Loss 1 × 2 antenna array with 10 mM Au-deposited ZnO nanorods on patch

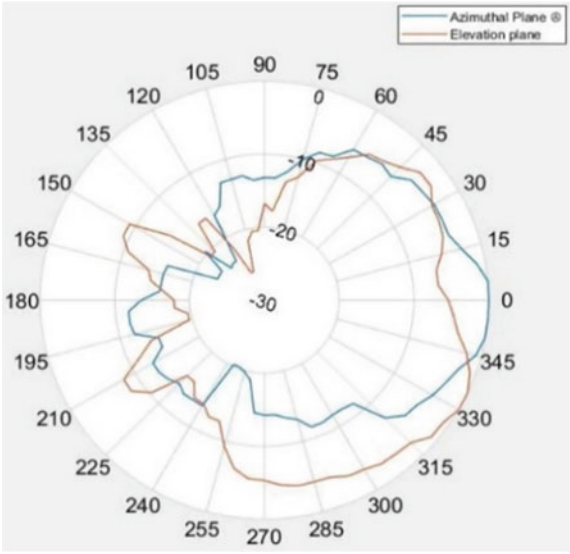


Fig. 13 Radiation Pattern of 1 × 2 antenna array with 5 mM Au deposited ZnO nanorods on patch

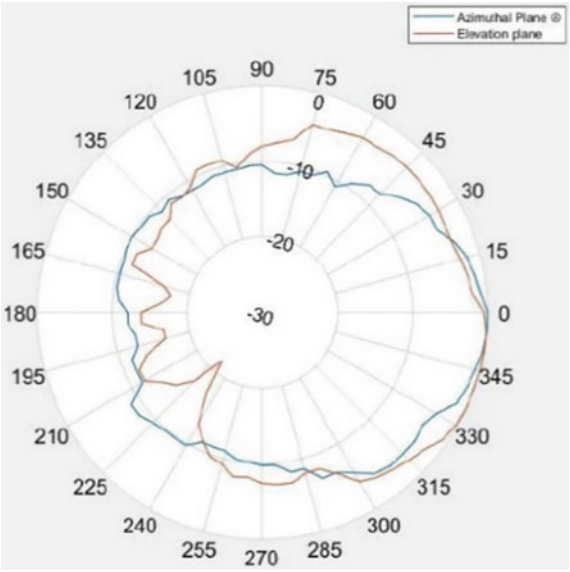


Fig. 14 Radiation Pattern 1 × 2 antenna array with 10 mM Au deposited nanorods on patch

Table 2 Comparison of measured performance parameters for different patch material

Patch with different materials	Operating frequency (GHz)	Return Loss (dB)	Bandwidth (GHz)	Beamwidth (degree)
Copper	11.181	−15.79	0.47	45
5 mM ZnO nanorods with gold	12.12	−17.02	1.16	40
10 mM ZnO nanorods with gold	9.05	−21.10	0.805	50

that the antenna arrays with nanomaterials have narrow beamwidth. With narrow beamwidth, the antenna arrays provide high directivity.

5 Conclusion

The use of nanomaterials on patch antennas can provide a low return loss and wider bandwidth. So, in this paper a 1 × 2 patch antenna array with individual radiating elements using gold as metal nanoparticles deposited over semiconductor ZnO nanorods have been fabricated to study the effects of nanomaterials on patch antenna array. The measured results of metal-semiconductor patch antenna arrays

show an improvement of −1.23 dB for 5 mM and −5.31 dB in the return loss and an improvement of 0.69 GHz for 5 mM and 0.335 GHz for 10 mM in operating bandwidth compared to conventional Cu-based patch antenna array. The bandwidth of nanomaterial-based array is high due to the presence of the resistive part. Due to the resistive element Q value of the system decreases. As Q value decreases bandwidth of the patch array increases. The beamwidth of fabricated patch antenna arrays is also calculated. As compared to Cu-based patch array, the array with 5 mM concentration nanomaterial have narrow beamwidth which proves that the directivity is enhanced. Hence, the array with 5 mM concentration provides a better result compared to the 10 mM concentration array.

Acknowledgements The authors acknowledge the support received from Center of Excellence in Nanotechnology (CoEN), Assam Don Bosco University, and MHRD-TEQIP III.

References

1. Balanis CA (1997) Microstrip antenna theory and design'. Wiley, NewYork
2. Kasinathan A, Jayaraj V, Pachiyaannan M (2014) E-shape microstrip patch antenna design for wireless applications. IJISET-Int J Innov Sci Eng Technol 1
3. Kidre A (2019) Wide band dual beam U- slot microstrip antenna. IEEE Trans Antennas Wirel Propag Lett 61:1415–1419
4. Shankar S, Chaurasiya H (2015) Inset feed microstrip patch antenna. In: IEEE International conference on computer, communication and control
5. Parmar N, Saxena M, Nayak K (2014) Review of microstrip patch antenna for WLAN and Wi-MAX application. Int J Eng Res Appl 4:2248–9622
6. Das R, Goswami S, Borgohain R, Baruah S (2015) Study on sheet resistance variation in ZnO nanorod arrays upon exposure to LPG at room temperature. In: International conference on energy, power and environment: towards sustainable growth (ICEPE) (2015)
7. Bhardwaj S, Samii YR (2012) C-shaped, E- shaped and U- slotted patch antennas: size, bandwidth and cross polarization characterizations. In: 6th European conference on antennas and propagation, pp 1674–1677

Study of TiO_2 and ZnO Nanostructures as Wide Band Gap Semiconductor for the Application of Nanotechnology-Based Solar Cells

Karen Das, Akib Khan, Arnab Sarkar, Pronoy Baishya, Priyanka Kakoty, and Sunandan Baruah

Abstract Rapid dwindling in reserves of conventional energy sources has created urgency among scientists to explore efficient techniques for tapping alternate renewable energy sources. Solar energy is the most promising among renewable energy sources which can address the ever-increasing demand for energy. Various dyes and quantum dots are experimented as photosensitizers for the design of low-cost solar cells for solar energy harvesting. Various wide bandgap semiconducting materials are also studied as transportation layer for the design of solar cell. A comparative study of TiO_2 mesoporous film and ZnO nanorods as transportation layer with various natural dyes and CdS quantum dot is presented in this paper.

Keywords Renewable energy · Solar cell · Photoanode · Nanorods · Quantum dots · Dyes

1 Introduction

Energy generation is of utmost importance in the modern era to keep the pace with the increased demand due to technological shift as well as ever-increasing consumers due to growing population. New methods of energy generation are, therefore, been researched one after another to reduce the dependence on dwindling resources. Renewable energy-based methods are the most sought out ones to keep the pace of energy demand. The amount of energy, we get from the sun, in an hour is adequate to satisfy the yearly global need [1]. Moreover, considerable sunlight is showered in most parts of the earth and solar energy being clean, free, and unlimited, it can be most vital in fulfilling all the energy demand in inexpensive and safest possible ways. Therefore, innovation to provide electricity by tapping the freely available solar energy at minimal expenses is the need of the hour.

K. Das · A. Khan · A. Sarkar · S. Baruah (✉)
Assam Don Bosco University, Guwahati, Assam, India
e-mail: sunandan.baruah@dbuniveristy.ac.in

P. Baishya · P. Kakoty
Tezpur University, Tezpur, Assam, India

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_9

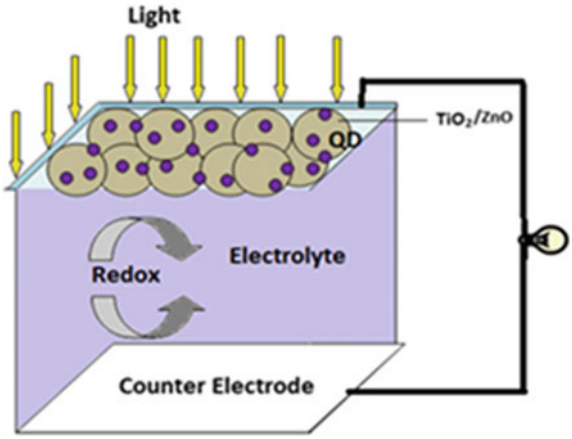


Fig. 1 Schematic of a QDSSC

Performance of solar cell (SC)s mainly depends upon light absorption as well as exposure of photon-sensitive material to light. Higher light absorption can be achieved using quantum dot (QD)s or different dye molecules adsorbed on nanostructured materials as nanostructures offer large surface to volume ratio. Bandgap engineering through size variations of the QDs can also be done to achieve absorption of various wavelengths of the visible spectrum of light. In SCs, electrons excited through photon absorption in the QDs/dyes, which are also called sensitizers, are transported to a wide semiconductor on which they are attached. Coupling the sensitizers with ZnO/TiO_2 nanostructures can result in transportation of a large number of electrons because of higher surface-to-volume ratios.

Typical construction schematic of such a SC is shown in Fig. 1. The structure of the SC is made of a photoanode and a counter electrode (CE) with a redox couple in between them [2–4]. A wide mesoporous semiconducting layer is attached to a conducting glass and sensitizers are adsorbed onto it to prepare the photoanode. The iodide/triiodide electrolyte is used in our work as redox couple which scavenges the holes generated by photon excitation at the junction of electrolyte and the photoanode and provides electrical equilibrium in the semiconducting layer. CdS QDs [5, 6] is one of the highly experimented sensitizers and the most popularly used wide bandgap semiconductor in QDSSCs is TiO_2 [7]. ZnOis also reported to be used as a mesoporous semiconducting layer in QDSSCs [8, 9].QDSSCs have evolved from dye-sensitized solar cells (DSSC) with only nominal difference between the-two. As the names suggest, the difference is only in the kind of sensitizers used. In this work, we present a comparative study of TiO_2 and ZnO as wideband gap semiconductor for the transportation of the photogenerated electrons for both QDSSCs and DSSCs. For DSSCs, we experimented with four natural dyes extracted from breadfruit (*Artocarpus Artilis*) bark, turmeric (*Curcuma Longa*) root, blackberry (*Rubus Fruticosus*) fruit and royal Poinciana (*Delonix Regia*) flower and CdS QDs are used for the design of QDSSC.

2 Experimental

2.1 Extraction of Dyes

A simple dye extraction process was used. 3–5 gm of the plant was crushed using mortar and pestle to form a paste. The paste was then dissolved in 50 ml of deionized water by continuous stirring at 60–70°C for one hour. The solution was then kept in dark condition for five hours. In the end, the solid residue was filtered out and a clear dye solution was obtained.

2.2 Preparation of Photoanode

Two types of photoanode were prepared using TiO_2 and ZnO for the experiments.

2.3 TiO_2-Based Photoanode

Fluorine doped tinoxide (FTO)-coated glass substrate was thoroughly cleaned with deionized water and acetone and a mesoporous TiO_2 thin film was deposited on the conducting side. 0.5 gm rutile TiO_2 nanopowder was then mixed with 5 to 6 drops of white vinegar to make a paste. Few drops of liquid detergent are added so that TiO_2 paste sticks to the conductive glass. Using duct tape, a rectangular area of 1 cm $\times$ 1 cm of the conducting side of the FTO glass is made available for deposition of TiO_2 mesoporous film. Few drops of the paste were deposited onto the exposed area of the FTO glass through doctor blading. After some time, the duct tape is removed and the paste-coated glass is heated at 400°C for half an hour. A brownish mesoporous film of TiO_2 was thus deposited on the FTO. TiO_2 coated glass was then dipped into the dye or the CdS QD solution for one hour so that the mesoporous film absorbs the dye QDs. After that, the substrate is taken out from the solution and washed with deionized water to remove additional material. The photoanode was then dried at room temperature in a desiccator.

2.4 ZnO-Based Photoanode

ZnO nanorods are grown on a rectangular area of 1 cm $\times$ 1 cm of the conducting side of the FTO glass using the method described in [10]. Then the electrodes are dipped in the dyes for one hour and then dried.

2.5 Preparation of Redox Couple

Iodide/Triiodide electrolyte is selected as redox couple as it has good solubility and suitable redox potential. It also provides rapid dye regeneration and absorbs small amount of light [11]. To prepare the redox couple, 127 mg of crystal iodine was mixed with 830 mg of potassium iodide (include formula) in a beaker. Almost instantly, Potassium Iodide reacts with Iodine to form Potassium Triiodide as depicted by changing color of the mixture to yellowish brown. 10 ml Ethylene Glycol solution was then added to the mixture under continuous stirring. The entire solution turns yellowish brown.

2.6 Preparation of CE and Fabrication of the SC

FTO coated glass was used as CE on to which silver NPs were deposited to achieve plasmonic effect. For the synthesis of silver, NPs deionized water was used as solvent. A Silver Nitrate ($AgNO_3$) solution was reduced using stoichiometric amount of trisodium citrate ($Na_3C_6H_5O_7$) resulting in nucleation and successive formation of Ag NPs through Ostwald Ripening. Then a film of silver NPs was grown over the conductive side of the FTO glass by dip and dry method to get the desired CE. After preparation of the photoanode, redox couple and the CE, the SC was assembled as shown in Fig. 1.

3 Results and Discussion

Optical absorbance of the sensitizer is a very important parameter for SCs. This is because higher is the absorption; better is the possibility of electron generation which in turn gives chance of better efficiency. UV-Vis spectroscopy was carried out to observe the optical absorption of the extracted dyes and CdS QDs and is shown in Fig. 2. Significant absorption of photons in the visible range by the sensitizers to be used can be interpreted from Fig. 2.

To understand the physical characteristics of the mesoporous film, a few physical characterizations were done as shown in Fig. 2. The SEM image, shown in Fig. 3(a), confirms deposition of a mesoporous film on the substrate. Figure 3(b) shows the TEM image of the TiO_2 nanoparticles and Fig. 3c shows the SEM image of the ZnO mesoporous film deposited on the substrate. Formation of ZnO nanorods is seen in Fig. 3c which is very much desirable for the transportation of photosensitized electrons. Figure 4 shows a comparative analysis of V-I characteristics between ZnO and TiO_2 SCs. From Fig. 4, it can be concluded that TiO_2-based SCs produced higher voltage and current as compared to ZnO-based ones.

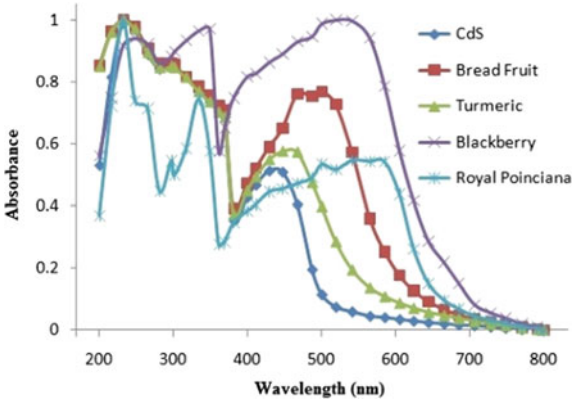


Fig. 2 UV-Vis spectra of CdS QD and the extracted dyes

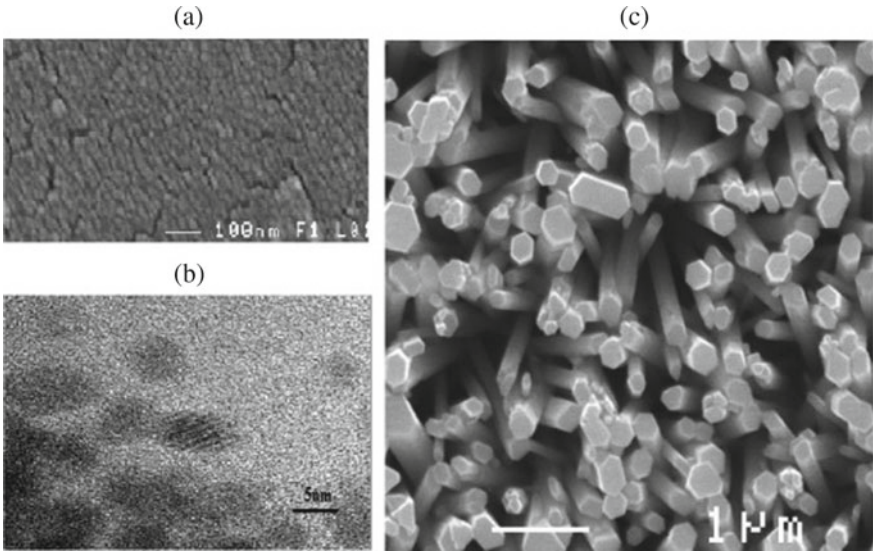


Fig. 3 physical characterization: **a** SEM image of the TiO_2 mesoporous film, **b** TEM image of the TiO_2 nanoparticles, **c** SEM image of ZnO nanorod mesoporous film

Figure 5a shows the fill factors of various SCs and nominal variation was observed in fill factors with varying mesoporous films of TiO_2 or ZnO. However, it is seen that SC with Blackberry and Breadfruit dyes and CdS Qds on TiO_2displayed improved fill factors, and for the rest, ZnO showed better fill factor. Figure 5b shows efficiency comparison of the various SCs. It is observed that in all cases, efficiency is found to be better with TiO_2 as compared to ZnO. We found highest efficiency of 1.92 with Blackberry dyes using TiO_2 mesoporous film while maximum efficiency of 0.9 is obtained with Blackberry dyes using ZnO mesoporous films.

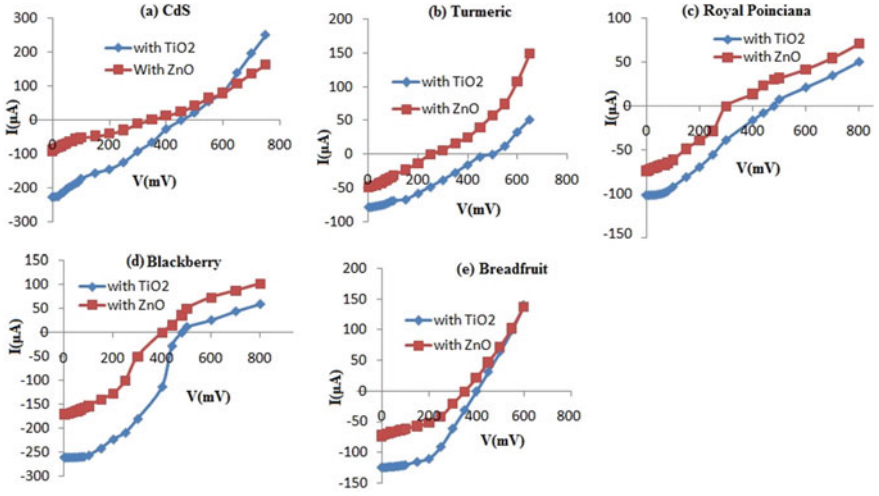


Fig. 4 V-I characteristics of various SCs with TiO_2 and ZnO mesoporous films

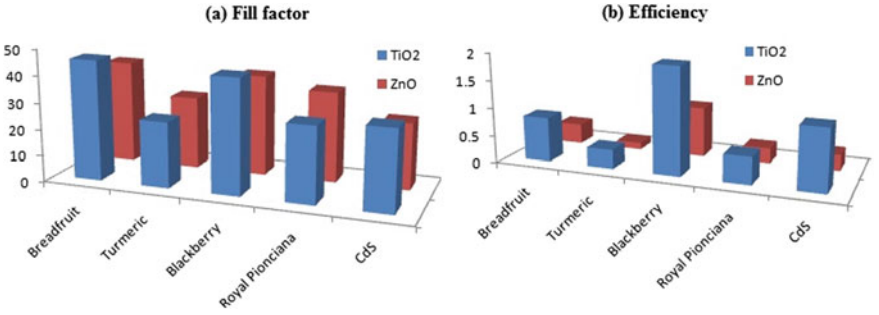


Fig. 5 Fill factors and Efficiencies of the designed SCs

4 Conclusion

Design of low-cost and environmentally benign solar cell has been studied extensively in the recent era. Various dyes and QDs are studied for possible use as sensitizers. Selection of wide bandgap semiconductor is also crucial for performance of the solar cell as it is the layer responsible for the transportation of photogenerated electrons. A comparative study of TiO_2 mesoporous films and ZnO nanorods as wide transportation layer is presented here. It is found that, in case of fill factor, both TiO_2 mesoporous films and ZnO nanorods produced quite similar performance with TiO_2 films slightly ahead of ZnO nanorods. However, it is found that TiO_2 is better as far as efficiency is concerned as compared to ZnO nanorods.

References

1. Zhang Q, Cao G (2011) Nanostructured photoelectrodes for dye-sensitized solar cells. Nano Today 6(1):91–109
2. Das K, Baruah S (2018) Quantum dots for solar energy harvesting. Curr Sci 115:659
3. Grätzel M (2004) Conversion of sunlight to electric power by nanocrystalline dye-sensitized solar cells. J Photochem Photobiol A: Chem 164:3–14
4. Halim MA (2012) Harnessing sun's energy with quantum dots based next generation solar cell. Nanomaterials 3:22–47
5. Meng K., Surolia P. K., Byrne O., Thampi K. R.: Efficient CdS quantum dot sensitized solar cells made using novel Cu2S counter electrode.Journal of Power Sources 248, 218–223(2014)
6. Sfaelou S et al (2014) Effect of the nature of cadmium salts on the effectiveness of CdS SILAR deposition and its consequences on the performance of sensitized solar cells. J Phys Chem C 118:22873–22880
7. Sudhagar P et al (2014) Quantum dot-sensitized solar cells. Low-cost nanomaterials. Springer, London, pp 89–136
8. Perera VPS (1999) An efficient dye-sensitized photoelectrochemical solar cell made from oxides of tin and zinc. Chem Commun 1:15–16
9. Leschkies KS et al (2007) Photosensitization of ZnO nanowires with CdSe quantum dots for photovoltaic devices. Nano Lett 7:1793–1798
10. Baruah S, Dutta J (2009) Hydrothermal growth of ZnO nanostructures. Sci Technol Adv Mater 10(1):01001
11. Boschloo G., Hagfeldt A (2009) Characteristics of the iodide/triiodide redox mediator in dye-sensitized solar cells. Acc Chem Res 42:1819–1826

AGC of Hybrid Solar-Hydro-Thermal System with GWO-based Conventional Secondary Controllers

Asadur Rahman, Lalit Chandra Saikia, and Yatin Sharma

Abstract Automatic generation control is exhibited for an unequal three-area hybrid solar-hydro-thermal power system. The hybrid power system comprises of conventional thermal and solar thermal system in area1. Conventional thermal system and hydro system with electric governor are positioned in area2 and area3, respectively. The conventional thermal systems are provided with single reheat turbine. Suitable generation rate constraint is provided in both hydro and thermal systems. The system dynamics with and without solar thermal are evaluated separately, considering conventional secondary controllers. The evolutionary computational technique called grey wolf optimizer algorithm is applied for simultaneous optimization of controller gains. Analyses reveal that system responses, with and without solar thermal, for PID controller are improved in comparison to other controllers. Investigation infers that the responses with reference to the hybrid system are enhanced with faster settling time and lesser deviation as compared to the system in absence of solar thermal.

Keywords Automatic generation control (AGC) · Grey wolf optimizer (GWO) · Solar-thermal (ST) system

1 Introduction

Efficient power system operation requires maintained frequency with limited tie-line power exchange. Deviations in these parameters occur due to change in load, other factors and are expected to make zero. "Automatic generation control (AGC)" holds the "frequency and tie line power exchange" stable to their respective nominal values.

A. Rahman (✉)
Department of Electrical Engineering, National Institute of Technology Srinagar, Srinagar, India
e-mail: asadur2003@yahoo.in

L. C. Saikia
Department of Electrical Engineering, National Institute of Technology Silchar, Silchar, Assam, India

Y. Sharma
Power Grid Corporation of India Ltd. (POWERGRID), Gurgaon, India

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_10

Several researches on AGC are reported in literature [1, 2]. Two-area AGC studies are presented by R.K. Sahu et al. [1]. Saikia et al. [2] reported studies on several other multi-area systems. The reported studies are demonstrated with the conventional sources of thermal, hydro, etc. According to recent studies, alternate energy sources such as solar and wind have the potential to become future energy sources. Modeling of integrated renewable energy sources for AGC studies is reported in literature [3–8]. The integration of solar photovoltaic system in AGC is presented in [3] by H. Asano et al. B. S Kumar in his paper [4] presented the AGC with respect to a distributed system. Rahman et al. [5] has incorporated the dish-Stirling solar thermal system for AGC of thermal system. This is been carried out with "biogeography based optimized" higher "degree of freedom PID controller". Bevrani et al. [6, 7] proposed modified area control error (ACE) concept for AGC incorporating solar photovoltaic (PV) system. However, the concept of solar-thermal (ST) system in comparison to solar PV system is growing rapidly. Further, the integration of these ST systems for AGC studies has been conceptualized from [5, 8, 9]. Sharma et al. incorporated "solar thermal power plant" [9] for grid-connected AGC studies but limiting to thermal system only. Thus, further studies on multi-area AGC system incorporating ST and other energy sources are important.

The governor has two controls with faster "primary control" and comparatively slower secondary control. In the present scenario, all research pertaining to AGC studies are done for the design and introduction of secondary control [1, 2, 9, 10]. Generally, conventional secondary controllers comprise of "I, PI and PID controllers". A performance of such controllers is studied in multi-area AGC with ST system [9], but is limited to thermal system only. Thus the current research needs to investigate the performance of conventional controllers for system with ST and other energy sources.

Techniques such as Bacterial foraging (BF) technique [2], etc. are available in literature for "AGC studies". Moreover, "fuzzy logic controller (FLC) [10] and supervised artificial neural network (ANN) controller" [11] requires higher computational time. Classical technique sometimes gives suboptimal result and is a trial and error technique [2]. To overcome local optimum, BF algorithm is implemented in AGC [2]. A "meta-heuristic algorithm" called "Grey Wolf Optimizer (GWO)" based on grey wolves' behavior [10] is established by Seyedali Mirjalili et al. [11]. The advantages of GWO over many other algorithms are presented in [11]. Its effectiveness is functional in solving AGC of thermal systems [9]. This can be checked for hybrid system as well. Thus, following are the main objectives of the present work:

(a) Integration of ST in a three-area hydro-thermal AGC system.
(b) Controller gains optimization with GWO algorithm.
(c) Comparison of system dynamic responses with and without solar thermal having conventional controllers, to find the optimal one in each case.

2 System Considered

An unequal three-area hydro-thermal system having conventional "thermal system" in area1 and area2, with hydropower plant in area3 is considered for investigation. For further investigation, the solar-thermal (ST) is included in area1 in addition to the "thermal system" and this combined system is termed as hybrid system. The area capacity ratio is 1:2:4. Conventional thermal system in area1-area2 is reflected with "generation rate constraint (GRC) of 3%/minute". GRC for hydro system is considered in area3 with "270%/minute for raising and 360%/minute for lowering generation". The base values for "hydro and thermal systems" are taken from [2] and for ST, from [5, 8, 9]. Area1 is provided with "1% step load perturbation (SLP)" for obtaining the system dynamics. Conventional controllers such as "I, PI and PID" are considered separately as secondary controllers for the investigation. The transfer function Simulink model of the hybrid system with ST in area1 is shown in Fig. 1. "Integral of squared error (ISE)" is used as "performance index (J)" and is given by (1). Here, i, j is the area number, for i = 1, 2, 3 and j = 2, 3 (j ≠ i).

$$J = \int_0^T \left\{ (\Delta f_i)^2 + \left(\Delta P_{tie\,i-j}\right)^2 \right\} dt \tag{1}$$

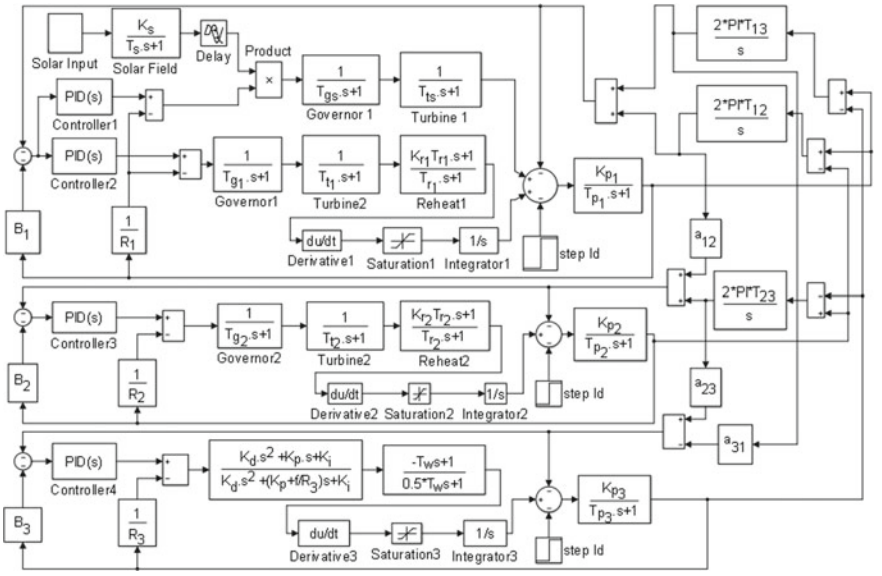


Fig. 1 Simulink model of the hybrid power system with ST in area1

3 Solar-Thermal (ST) System

The huge solar energy potential can be harnessed in producing the electrical power using the available technology, such as "solar photovoltaic (PV) and concentrated solar power (CSP)". Solar-thermal (ST) system falls under CSP technology and is appropriate for utility-based applications. The technology works on the principle of storing thermal energy from direct normal solar irradiance. This stored thermal energy can be retrieved instantly or at a later stage to produce electrical power. CSP is getting global importance due to its bulk energy storage facility and it comprises of a large collector area. Figure 2 demonstrates the block diagram representation of a solar-thermal (ST) system. It comprises of a solar collector area with heat exchanger, where thermal energy is stored. The other portion of the ST system includes the governor and a steam-turbine. A comparative study is reported in [12] for ST system with different types of collectors. The collectors concentrate the solar radiation to a central line/point to produce thermal energy and hence steam. Some modified steam generation process is developed and reported in [13]. A hybrid solar-fossil fuel power plant is proposed in [14]. J.B et al. [15] have suggested the modeling of the solar field. The output temperature of ST system changes at a rate as provided by Eq. (2) and the Laplace transform is as shown in Eq. (3).

$$\frac{dT_0(t)}{dt} = \frac{A\eta_0}{C} I(t) - \frac{U_L A}{C}[T_a(t) - T_e(t)] + \frac{v(t)}{V}[T_i(t) - T_0(t)] \tag{2}$$

$$T_0(s) = \frac{T_s.}{T_s.s+1}\frac{A\eta_0}{C} I(s) + \frac{T_s}{T_s.s+1}\left[\frac{v}{V} - \frac{U_L A}{2C}\right] T_i(s) + \frac{T_s}{T_s.s+1}\frac{U_L A}{C} T_e(s) \tag{3}$$

The changes in "inlet and environment temperature" are trivial and the transfer function of solar field w.r.t solar irradiance is applied as shown in Eq. (4), where K_s is the solar field gain. The steam produced in the heat exchanger drives the turbine.

$$G(s) = \frac{K_s}{1 + T_s.s} \tag{4}$$

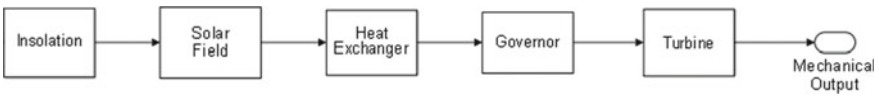


Fig. 2 Block diagram representation of solar thermal (ST) system

4 Grey Wolf Optimizer (GWO) Algorithm

This algorithm is based on the behavior of grey wolves [10], proposed by Mirjalili et al. [11]. Algorithm imitates grey wolves' behavior in pointing and chasing their target. "Grey wolves" have firm social prominent hierarchy. They live in group and are termed as "pack". The pack has four different levels namely, "alpha (α), beta (β), delta (δ) and omega (ω)". Decision-making is the responsibility of alpha and is the first level of pack. Beta helps in decision-making and is the second level. Delta is the next level having "scouts, sentinels, hunters, and caretakers". Scouts watch the boundary of the territory and issue warning at the time of danger. Pack's safety is the sentinel's responsibility. Alpha and beta get help from the hunters. Omega is the lowest level in the pack and has to submit to all acting as scapegoat. Muro et al. explain the hunting process of grey wolves in their paper [10]. The technique is explicitly explained in [9, 11]. The number of search agent (n) in GWO is tuned and considered as 5. The flowchart for GWO algorithm is as given below.

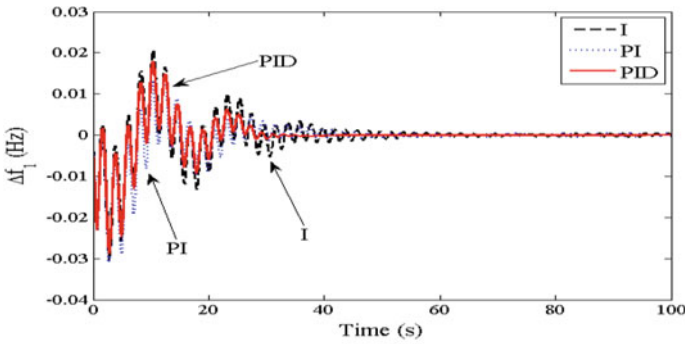
```
Begin
   Initialize population of 'n' grey wolf
   Evaluate fitness of every search agent for an objective function
   X_α X_β X_δ are the best, second, and third search agent
     While loop (Limit or Stop criterion);
           For every search agent
               Current search agent's position update
           End for
        Update the parameters 'a, A and C'
        Calculate search agent's fitness;
        Update the search agents X_α X_β X_δ ;
        Increase the count t;
     End while loop
  Return the best search agent X_α ;
End
```

5 Result Analysis and Discussion

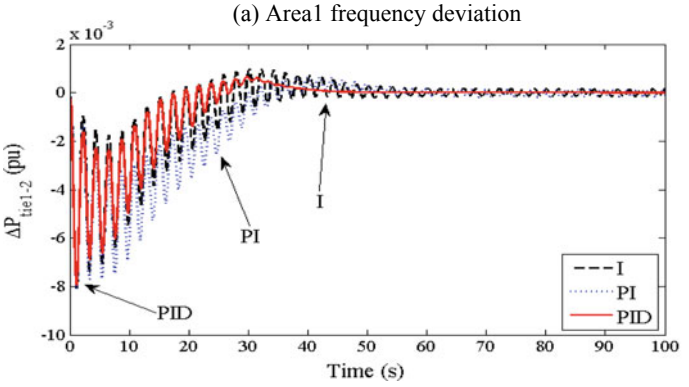
5.1 Optimization of Classical Secondary Controller Gains for Conventional System

Conventional power system is the basic hydro-thermal system as mentioned in Sect. 2. Classical secondary controllers such as I, PI, and PID are considered separately for investigation. The controller gains, in each controller case, are optimized simultaneously using GWO algorithm. Following are the optimum values of controller gains for the conventional system (without considering ST). The optimal gains with I controller

are, $K_{I1}* = 0.62913$, $K_{I2}* = 0.37382$, $K_{I3}* = 0.0041584$. The PI controller gains for the same system are computed and are $K_{P1}* = 0.007853$, $K_{P2}* = 0.0063624$, $K_{P3}* = 0.006127$, $K_{I1}* = 0.32858$, $K_{I2}* = 0.019018$, $K_{I3}* = 0.00095993$. Similarly, optimal PID controller gains are $K_{P1}* = 0.05373354$, $K_{P2}* = 0.00647322$, $K_{P3}* = 0.05531121$, $K_{I1}* = 0.9091905$, $K_{I2}* = 0.3177515$, $K_{I3}* = 0.000093765$, $K_{D1}* = 0.1355624$, $K_{D2}* = 0.0002108743$, $K_{D3}* = 0.100752$. The comparison of the responses pertaining to different controllers is demonstrated in Fig. 3. It is evident that PID controller is outperforming the other controllers. The responses depict better dynamics for settling time, peak overshoot, and oscillations. The deviation in area1 frequency and tie-line power exchange is reported in Fig. 3 for statement validation.



(a) Area1 frequency deviation



(b) Deviation in Tie-line power exchange for area1 to area2

Fig. 3 Comparison of the responses for the conventional system with I, PI, and PID as secondary controllers

5.2 Optimization of Classical Secondary Controller Gains for Hybrid System

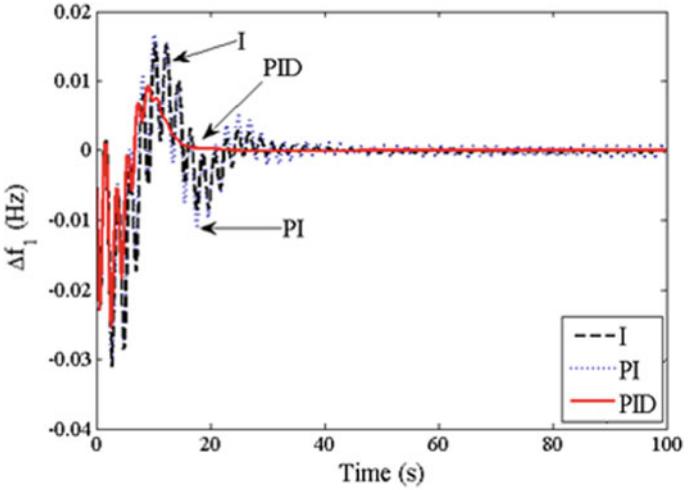
Hybrid system resembles the power system mentioned in Sect. 5.1 with additional solar-thermal integrated in area1. Classical controllers such as "I, PI, and PID" are considered individually as secondary controllers. GWO algorithm is applied to simultaneously optimize the controller gains. The optimal gains with I controller are, $K_{IST}^* = 0.030436$, $K_{I1}^* = 0.07912$, $K_{I2}^* = 0.029027$, $K_{I3}^* = 0.00002$. Similarly, the controller gains for PI are, $K_{PST}^* = 0.062878$, $K_{P1}^* = 0.0018044$, $K_{P2}^* = 0.0011514$, $K_{P3}^* = 0.065708$, $K_{IST}^* = 0.032429$, $K_{I1}^* = 0.093458$, $K_{I2}^* = 0.0000726$, $K_{I3}^* = 0.000654$ and for PID controller the optimal gains are $K_{PST}^* = 0.7968192$, $K_{P1}^* = 0.4892932$, $K_{P2}^* = 0.71400958$, $K_{P3}^* = 0.288658$, $K_{IST}^* = 0.09114182$, $K_{I1}^* = 0.5114575$, $K_{I2}^* = 0.1311596$, $K_{I3}^* = 0.008815331$, $K_{DST}^* = 1.0$, $K_{D1}^* = 0.42462$, $K_{D2}^* = 0.01272677$, and $K_{D3}^* = 0.1169604$. The comparison of the responses pertaining to different controllers for hybrid system is demonstrated in Fig. 4. It is evident that PID controller is outperforming the other controllers. The responses depict better dynamics with negligible oscillations and much lesser settling time, peak overshoot respectively. The deviation in area1 frequency and tie-line power exchange is reported in Fig. 4 for statement validation.

5.3 Comparison of Dynamic Responses for Conventional and Hybrid Power System

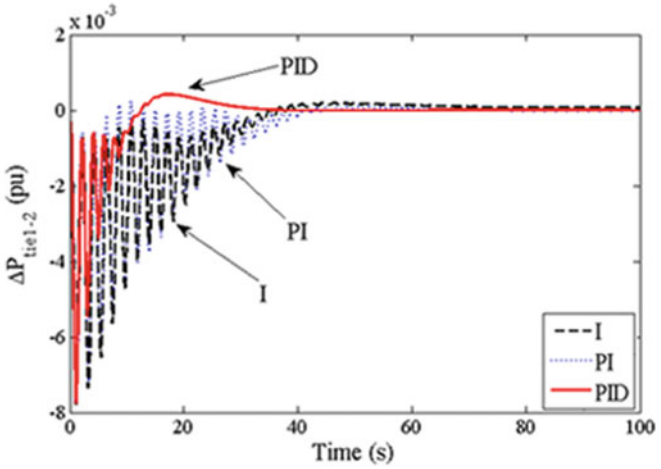
Figure 5 compares the system dynamic responses with PID controller for the conventional and hybrid system as depicted in Fig. 5 and Fig. 4, respectively. Figure 5 highlights the improvement in settling time and overshoots for hybrid system with respect to the conventional system. It also attracts the reduced oscillations for hybrid system. The observed values are reported in Table 1 for consideration. The incremental power generation in area1 with individual contribution from conventional thermal system and solar-thermal system is presented in Fig. 6. This reflects the power-sharing among the available energy sources to meet the area load demand (1 p.u).

6 Conclusion

Successful implementation of an unequal three area hybrid solar-hydro-thermal AGC system is presented. Investigation is carried out for three-area AGC system, conventional hydrothermal and hybrid system incorporating solar-thermal (ST) in area1, respectively. Meta-heuristic optimization technique, grey wolf optimizer (GWO), is



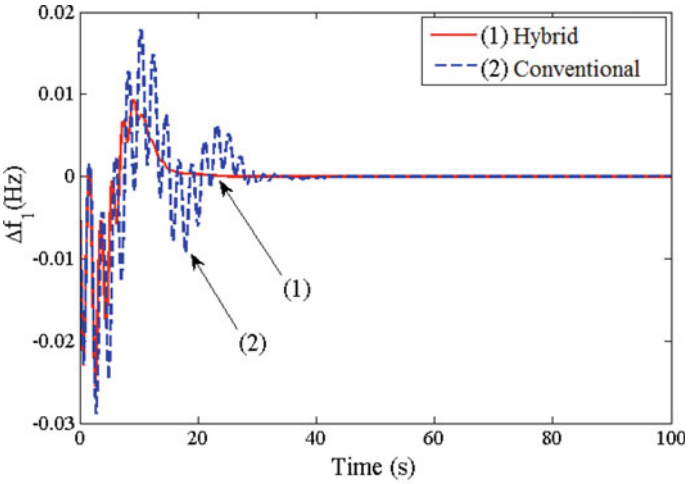
(a) Area1 frequency deviation



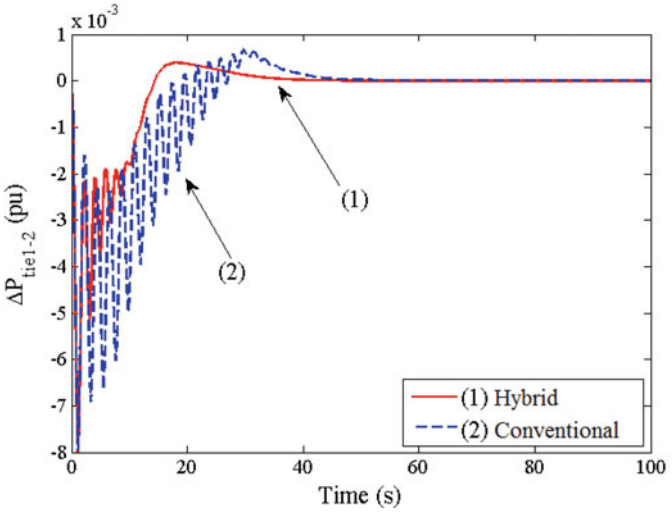
(b) Deviation in Tie-line power exchange for area1 to area2

Fig. 4 Comparison of the responses for the hybrid system with I, PI, and PID as secondary controllers

successfully applied in AGC of both the considered system for concurrent optimization of controller gains. Investigation reveals that PID controller, optimized with GWO algorithm, for the hybrid system gives improved responses. The responses are enhanced with reduced settling time, overshoots, and minor oscillations. Inclusion of solar-thermal for the hybrid system makes significant development in the system performance.



(a) Area1 frequency deviation



(b) Deviation in Tie-line power exchange for area1 to area2

Fig. 5 Comparison of system performances, Conventional vs. Hybrid system, with PID controller

Table 1 Observed values from Fig. 5

Fig. No.	Settling time(s)		Peak overshoot		Undershoot	
	Conventional system	Hybrid system	Conventional system	Hybrid system	Conventional system	Hybrid system
Figure 5(a)	42.28	24.95	0.01789	0.009315	0.02852	0.02566
Figure 5(b)	52.72	40.51	0.0006507	0.0003904	0.0079	0.007948

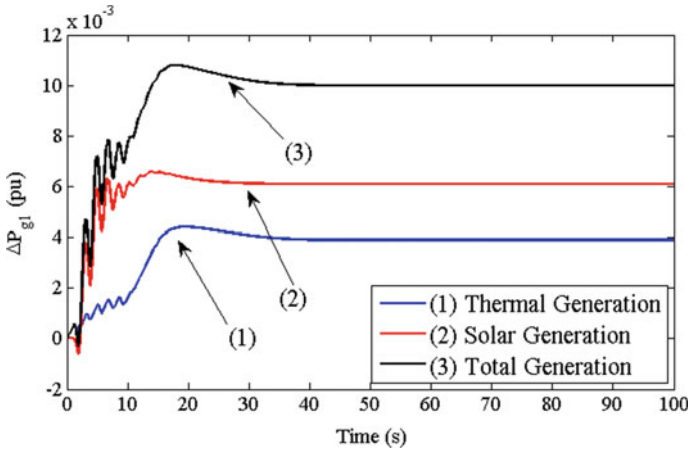


Fig. 6 Incremental power generation in area1 for hybrid system

Appendix

See Appendix Table.

Nominal parameters of the system are:							
Frequency, f: 60 Hz;	T_{gs}: 1.0 s;	T_{g1}, T_{g2}: 0.08 s;	T_{ts}: 3.0 s;				
T_{t1}, T_{t2}: 0.3 s;	T_{ri}: 10 s;	T_w: 1.0 s;		K_{ri}: 0.5;	K_{pi}: 120 Hz/pu MW; T_{pi}: 20 s;	$T_{12} = T_{23} = T_{13}$: 0.086 pu MW/rad;	H_i: 5 s;
D_i: 8.33 x 10^{-3} pu MW/Hz;	$B_i = \beta_i$: 0.425 pu MW/Hz;						
R_i: 2.4 Hz/pu MW;	loading: 50%;	K_s: 1.8;	T_s: 1.8 s;				
SLP: 1% in area1.							

References

1. Sahu RK, Panda S, Rout UK (2013) DE optimized parallel 2-DOF PID controller for load frequency control of power system with governor dead-band nonlinearity. Electr Power Energy Syst 49:19–33

2. Saikia LC, Nanda J, Mishra S (2011) Performance comparison of several classical controllers in AGC for multi-area interconnected thermal system. IJEPES 33(3):394–401
3. Asano H, Yajima K, Kaya Y (1996) Influence of photovoltaic power generation on required capacity for load frequency control. IEEE Trans Energy Convers 11(1):188–193
4. Kumar BS, Mishra S, Senroy N (2008) AGC for distributed generation. In: Proceedings of the international conference on sustainable energy technologies, pp 89–94
5. Rahman A, Saikia LC, Sinha N (2016) AGC of dish-Stirling solar thermal integrated thermal system with biogeography based optimised three degree of freedom PID controller. IET Renew Power Gen 10(8):1161–1170
6. Bevrani H, Ghosh A, Ledwich G (2010) Renewable energy sources and frequency regulation: survey and new perspectives. IET Renew Power Gen 4(5):438–457
7. Bevrani H (2009) Robust power system frequency control. Springer
8. Rahman A, Saikia LC, Sinha N (2017) Automatic generation control of an interconnected two-area hybrid thermal system considering dish-stirling solar thermal and wind turbine system. Elsevier Renew Energy 105:41–54
9. Sharma Y, Saikia LC (2015) Automatic generation control of a multi-area ST – Thermal power system using Grey Wolf Optimizer algorithm based classical controllers. IJEPES 73:853–862
10. Muro C, Escobedo R, Spector L, Coppinger R (2011) Wolf-pack (Canis lupus) hunting strategies emerge from simple rules in computational simulations. Behav Process 88:192–197
11. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey Wolf optimizer. Adv Eng Softw 69:46–61
12. Reddy VS, Kaushik SC, Ranjan KR, Tyagi SK (2013) State-of-the-art of solar thermal power plants—A review. Renew Sustain Energy Rev 27:258–273
13. Valenzuela L, Zarza E, Berenguel M, Camacho EF (2004) Direct steam generation in solar boilers. IEEE Control Syst 24(2):15–29
14. Fontalvo A, Garcia J, Sanjuan M, Padilla RV (2014) Automatic control strategies for hybrid solar-fossil fuel power plants. Renew Energy 62:424–431
15. Buzás J, Kicsiny R (2014) Transfer functions of solar collectors for dynamical analysis and control design. Renew Energy 68:146–155

Study on Substrate Dependency of Graphene-Based Patch Antennas for Gigahertz and Terahertz Applications

Dilruba Khanam, Kaustubh Bhattacharyya, Kandarpa Kumar Sarma, and Sunandan Baruah

Abstract Graphene is a thin sheet of carbon atoms in which carbon atoms are arranged in a hexagonal honeycomb lattice structure. It has very high electrical conductivity. Properties like high electron mobility and ability to support Surface Plasmon Polariton (SPP) waves make graphene suitable for THz communication. In this paper, antennas are designed and investigated with graphene patch at different substrates employing finite element method-based High Frequency Simulation Software (HFSS). Performance of the antennas are investigated for five different substrate materials such as FR-4, Bakelite, Rogers R04003, RT Duroid 6010 and Taconic TLC. Further, the performances are examined on the basis of return loss, voltage standing wave ratio (VSWR), gain, directivity and bandwidth of the graphene-based patch antennas at Gigahertz (GHz) and Terahertz (THz) range. Finally, it is proved that effect of substrate material is retained while linear scaling graphene-based patch antenna from GHz frequency range to THz frequency range.

Keywords HFSS · Microstrip patch antenna · Terahertz · Graphene · Substrate material · Return loss

D. Khanam · K. K. Sarma
Department of Electronics and Communication Engineering, Gauhati University, Gauhati, India
e-mail: ruba.onair@gmail.com

K. K. Sarma
e-mail: kandarpaks@gmail.com

K. Bhattacharyya (✉) · S. Baruah
Department of Electronics and Communication Engineering, Assam Don Bosco University, Assam, India
e-mail: kaustubh.bhattacharyya@dbuniversity.ac.in

S. Baruah
e-mail: sunandan.baruah@dbuniversity.ac.in

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_11

1 Introduction

Microstrip patch antennas are low profile, easy to fabricate, easy to feed and easy to incorporate with other microstrip circuit elements and integrate into systems. With the advancement of technology, the demand for higher wireless communication has increased day by day. The need for high-speed wireless communication has encouraged the researchers to study Terahertz (THz) waves [1]. THz wave occupies the electromagnetic wave with frequencies lying between 0.1THz and 10 THz [2, 3]. Prior to 1980, electromagnetic properties of THz wave are completely unknown. With the emersion of new technologies, THz wave is researched and found to be suitable for different applications. Properties like wave-particle duality, penetrability, and resolving power of spectrum make it suitable for applications in security, Medical, Military, etc. [4, 5]. THz band can be applied for local area networks and broadband mobile communications as it has high frequency and wide bandwidth. Spectrum at THz frequency band is not yet licensed and hence, there is huge scope of applications in this band [6]. Graphene-based nanoantennas will enable wireless communication at the nanoscale. The behavior of electron in graphene is being researched and graphene appears to be a feasible basis for antennas [7]. Graphene has some unique properties, which make it appropriate for use in miniaturized antennas. Properties like high electron mobility and ability to support SPP waves make it an appropriate choice for THz communication [8]. For Graphene patch antenna substrate materials act as a performance regulator [9]. It is reported that substrate material controls the properties of graphene patch and choice of good dielectric material can improve the quality. Transport properties of graphene and resonant properties of antenna are also influenced by substrate material [10]. A lot of research is undergoing to find suitable new materials that can be placed as substrates for the graphene patch antenna [11, 12]. Microstrip antennas normally have dimensions in mm. When similar structures of dimension in micrometers are designed, the antennas operating in THz frequencies [13]. A linear scaling technique is used to design THz antenna from a GHz antenna by reducing all dimension of the GHz antenna by a factor of 1000. In antenna designing, the primary step is to choose an appropriate substrate as it is needed for the mechanical support of the antenna. The resonance frequency of a nanoantenna is a function of length and width of substrate and patch. Resonance frequency of a graphene patch antenna further depends on location of patch [8, 14].

This paper presents a graphene-based patch antenna and comparative study for different substrate material. The performance of the antenna is evaluated on the basis of return loss, voltage standing wave ratio (VSWR), gain, bandwidth, and radiation efficiency.

2 Microstrip Patch Antenna at GHz Frequency

Microstrip patch antennas consist of a metallic patch which is mounted on a grounded substrate. Out of the variety of shapes of patch, square and rectangular shapes are mostly preferred because of the ease of analysis and fabrication, Here, antennas are designed and simulated using graphene as patch material. Simulation is carried out by using the software Ansoft HFSS. Further study was carried out by changing the substrate material for proposed antenna. Figure 1 shows the geometry of the proposed antenna. In this work, FR-4 epoxy substrate with dielectric constant = 4.4 and a thickness of 1.5 mm are used.

At first, antennas are designed with graphene as patch material and performances such as return loss, VSWR, gain, directivity and bandwidth are calculated. The design equations used for calculation are shown below:

The length of the patch,

$$L_p = \frac{c}{2f_r\sqrt{\varepsilon_r}} - 2 \times \Delta L \tag{1}$$

where, f_r is the operating frequency, c is the velocity of light in free space, ε_r is dielectric constant and the difference in length, ΔL which can be further defined as

$$\Delta L = 0.412 \times h \times \frac{(\varepsilon_{eff} + 1)}{(\varepsilon_{eff} - 0.258)} \frac{(\frac{w}{h} + 0.264)}{(\frac{w}{h} + 0.8)} \tag{2}$$

where h is the height of the substrate and ε_{eff} is the effective dielectric constant.

The width of the patch is given by

$$W_p = \frac{1}{2f_r\sqrt{\varepsilon_0\mu_0}}\sqrt{\frac{2}{\varepsilon_r + 1}} \tag{3}$$

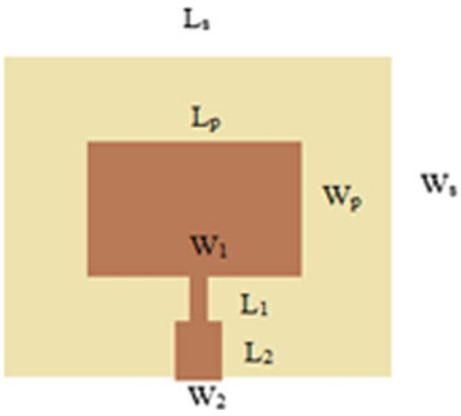


Fig. 1 Geometry of the proposed Antenna

Table 1 **1** Physical dimension of patch Antenna at GHz

Parameters	Symbol	Value
Operating frequency	f_r	(5–15) GHz
Patch length and width	L_p, W_p	6.4899 mm, 9.1287 mm
Dielectric substrate length, width, and thickness	L_s, W_s, h	15.48 mm, 18.12 mm, 1.5 mm
Microstrip line length	L_1, L_2	3 mm, 2.5 mm
Microstrip line width	W_1, W_2	0.5 mm, 1 mm

The length of the substrate

$$L_s = L_P + 6 \times h \tag{4}$$

The width of the substrate

$$W_s = W_P + 6 \times h \tag{5}$$

The detail dimensions taken into consideration are shown in Table 1.

The designed antennas with FR-4 substrate are first simulated and then analyzed for return loss, gain, directivity, radiation pattern, bandwidth, and VSWR. For further improvement in performance, the patch was investigated with five different substrate materials like FR-4, Bakelite, Rogers R04003, RT Duroid, and Taconic TLC.

3 Linear Scaling Technique

A linear scaling technique is used to design a THz antenna from GHz antenna [15]. In this technique, every dimension of GHz antenna are scaled down by a factor of 1000, and investigated whether the reduced sized antenna works at almost same magnitude at THz or not. Dimensions for simulation at THz range are taken in micrometres. Detailed dimensions are shown in Table 2.

Table 2 Physical dimension of patch Antenna at THz

Parameters	Symbol	Value
Operating frequency	f_r	(5–15) THz
Patch length and width	L_p, W_p	6.4899 μm, 9.1287 μm
Dielectric substrate length, width and thickness	L_s,W_s, h	15.48 μm, 18.12 μm, 1.5 μm
Microstrip line length	L_1, L_2	3 μm, 2.5 μm
Microstrip line width	W_1, W_2	0.5 μm, 1 μm

For further improvement in performance, the patch was investigated with five different substrate materials like FR-4, Bakelite, Rogers R04003, RT Duroid, and Taconic TLC.

4 Results and Discussion

The designed microstrip patch antenna with graphene as a patch material at GHz frequency range was investigated with different substrate materials. Figure 2 shows the return loss graph for five different substrates of graphene-based patch antenna at GHz.

From the figure, it is seen that return loss is minimum for Rogers R04003 and maximum for Bakelite.

Table 3 shows the measured results of graphene-based patch Antenna for different substrate material.

Graphene-based patch antenna attained a return loss of −41.70 dB with Rogers R04003 substrate which is a tremendous return loss in comparison to return loss obtained with other substrate material. Again considering the performance parameters—gain, bandwidth, and directivity, it is observed that RT Duroid 6010 exhibit good result with a huge bandwidth of 880.5 MHz at 12.4 GHz, although return loss is greater but below −10 dB which is quite acceptable. As Rogers RO4003 shows best result, return loss, VSWR, gain, and radiation pattern are also shown in Fig. 3.

Dimensions of the proposed GHz antenna is scaled down by a factor of 1000 to design an antenna operating in the THz frequency range and its radiation characteristics are investigated at 10 THz, utilizing the HFSS simulation tool. Again antenna parameters are investigated for different substrate material and the results are tabulated in Table 4.

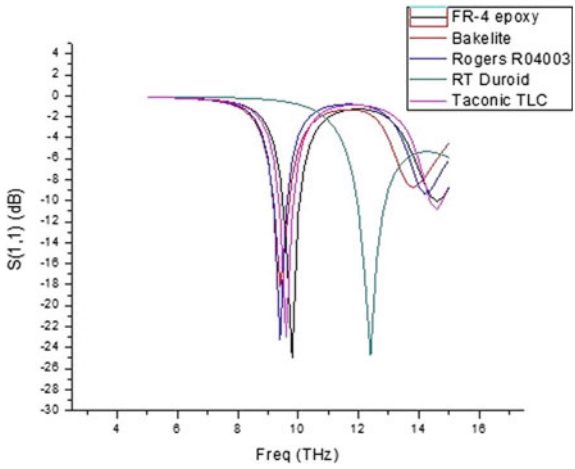


Fig. 2 Return Loss for different substrate of graphene-based antenna at 10 GHz

Table 3 Results of the graphene-based patch antenna at GHz for different substrate material

Substrate	Dielectric constant	Parameter	Graphene-based Antenna
FR-4	4.4	S11 (dB)	−33.87
		VSWR	1.04
		Gain (dB)	5.784
		Directivity (dB)	6.71
		Bandwidth (MHz)	6.71
		Resonating frequency (GHz)	9.6
Bakelite	4.78	S11 (dB)	−14
		VSWR	1.49
		Gain (dB)	6.68
		Directivity (dB)	6.57
		Bandwidth (MHz)	416
		Resonating frequency (GHz)	9.4
Rogers RO4003	3.4	S11 (dB)	−41.70
		VSWR	1.01
		Gain (dB)	7.03
		Directivity (dB)	6.94
		Bandwidth (MHz)	451.6
		Resonating frequency (GHz)	9.4
RT Duroid 6010	2.2	S11 (dB)	−17.73
		VSWR	1.63
		Gain (dB)	8.43
		Directivity (dB)	8.07
		Bandwidth (MHz)	880.5
		Resonating frequency (GHz)	12.4
Taconic TLC	3.2	S11 (dB)	−25.62
		VSWR	0.99
		Gain (dB)	7.38
		Directivity (dB)	7.22
		Bandwidth (MHz)	482.4
		Resonating frequency (GHz)	9.6

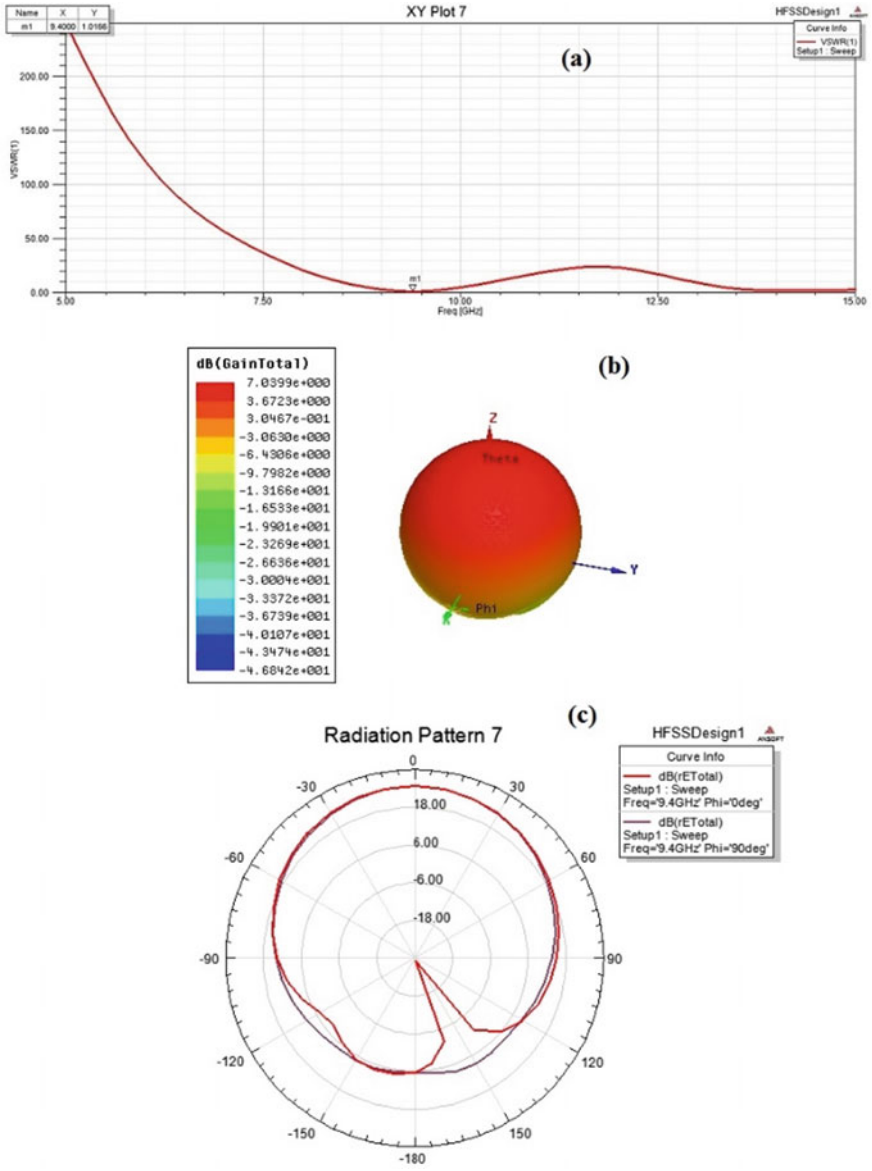


Fig. 3 **a** VSWR, **b** Gain, **c** Radiation pattern of graphene-based antenna with Rogers RO4003 substrate at 10 GHz

Table 4 Results of the graphene-based patch antenna at THz for different substrate material

Substrate	Dielectric constant	Parameter	Graphene-based Antenna
FR-4	4.4	S11 (dB)	−24
		VSWR	0.98
		Gain (dB)	6.07
		Directivity (dB)	6.63
		Bandwidth (GHz)	576.6
		Resonating frequency (THz)	9.8
Bakelite	4.78	S11 (dB)	−18
		VSWR	2.15
		Gain (dB)	5.94
		Directivity (dB)	6.52
		Bandwidth (GHz)	530
		Resonating frequency (THz)	9.4
Rogers RO4003	3.4	S11(dB)	−23.29
		VSWR	1.19
		Gain(dB)	6.47
		Directivity (dB)	7.01
		Bandwidth (GHz)	476.7
		Resonating frequency (THz)	9.4
RT Duroid 6010	2.2	S11 (dB)	−24.75
		VSWR	1.005
		Gain (dB)	8.04
		Directivity (dB)	8.12
		Bandwidth (GHz)	952.9
		Resonating frequency (THz)	12.4
Taconic TLC	3.2	S11 (dB)	−22.97
		VSWR	1.23
		Gain (dB)	6.56
		Directivity (dB)	7.07
		Bandwidth (GHz)	500
		Resonating frequency (THz)	9.6

Figure 4 shows the return loss graph for five different substrates of graphene-based patch antenna at THz range. It can be seen that return loss is minimum of −24.75 dB for RT Duroid substrate and maximum of −18 dB for Bakelite substrate.

From Table 4, it is seen that all the substrate materials attained gain more than 5 dB and return loss less than −10 dB at THz frequency range. Considering the performance parameters—return loss, VSWR, gain, bandwidth, and directivity, it is observed that RT Duroid 6010 exhibit good result with a huge bandwidth of

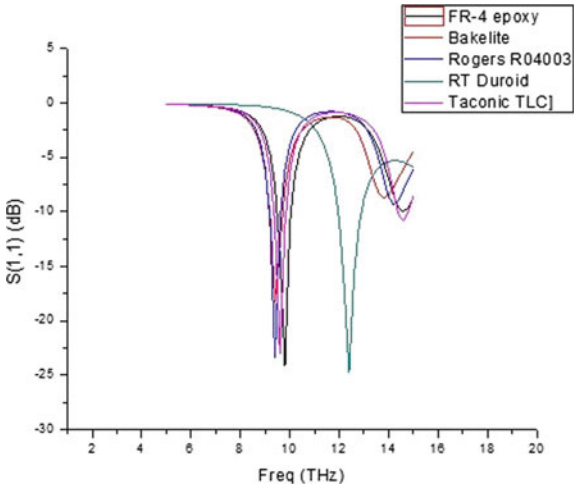


Fig. 4 Return Loss for different substrate of graphene-based antenna at 10 THz

952.9 GHz at 12.4 THz. Also, the effect of Rogers RO4003 material is retained at THz frequency range. It has been observed that the linear scaling technique performs well to design a THz antenna from its equivalent GHz antenna and experimental results validate its effectiveness. VSWR, gain, and radiation pattern of the proposed antenna with Rogers RO4003 are shown in Fig. 5.

5 Conclusion

In this paper, graphene-based patch antennas with different substrates are designed and simulated in the GHz and THz frequency range. Performance of the antennas was compared on the basis of return loss, gain, VSWR, directivity, radiation pattern, and bandwidth. Graphene-based Antenna achieved minimum return loss of −41.70 dB with a bandwidth of 451 MHz at 9.4 GHz with substrate Rogers RO4003, and after linear scaling, it shows a return loss of −23.29 dB with bandwidth 476 GHz at 9. 4 THz. All the substrate materials attained gain more than 5 dB and return loss less than −10 dB and can be used as a substrate material for graphene-based patch antennas. But for a higher frequency, RT Duroid would be the best substrate as its resonating frequency gets shifted toward higher frequency. Effect of substrate material is retained after scaling down by a factor of 1000 as graphene-based patch antenna attained good result with Rogers RO4003 at both GHz and THz without a shift in frequency.

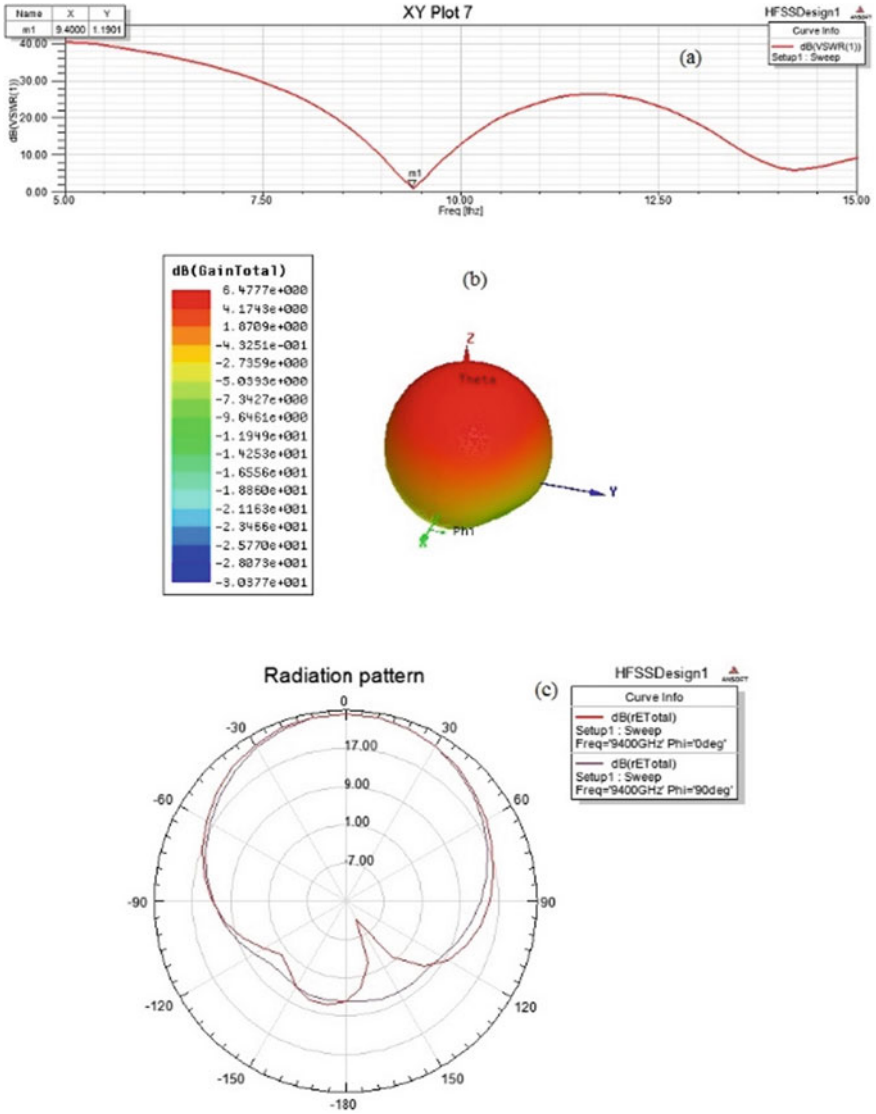


Fig. 5 **a** VSWR, **b** Gain, **c** Radiation pattern graphene-based antenna with Rogers RO4003 substrate at 10 THz frequency

References

1. Adnan S, Goni O (2015) Graphene nanoribbon based antenna for terahertz band communication. In: Proceedings of international conference on electrical information and communication technology (EICT 2015)

2. Sharma A, Singh G (2009) Rectangular microstrip patch antenna design at THz frequency for short distance wireless communication systems. J Infrared Millim Terahertz Waves 30(1):1–7
3. Llatser KC, Albert CA, Jornet JM, Alarcon E, Chigrin DN (2012) Graphene-based nano-patch antenna for terahertz radiation. Photon Nanostruct-Fund Appl 10(4):353–358
4. Anand S, Sriram Kumar D, Wu RJ, Chavali M (2014) Analysis and design of optically transparent antenna on photonic band gap structures. Optik 125:2835–2839
5. Llatser KC, Chigrin DN, Josep JM, Lemme MC, Albert C-A (2012) Alarcon Eduard: characterization of graphene-based nano-antennas in the terahertzband. In: 6th European conference on IEEE
6. Llatser CK, Cabellos-Aparicio A, Jornet J, Alarcon E, Chigrin D (2011) Scattering of terahertz radiation on a graphene-based nano-antenna. In: AIP Conference proceeding, 4th international conference on theoretical and nanophotonics, Germany, pp 144–147
7. Llatser I, Kremers C, Chigrin D, Jornet J, Lemme M, Cabellos-Aparicio A et al (2012) Characterization of graphene-based nano-antennas in the terahertz band. In: 6th European conference on antennas and propagation (EUCAP), pp. 194–198
8. Anand S, Sriram Kumar D, Jang Wu R, Chavali M (2014) Graphene nanoribbon based terahertz antenna on polyimide substrate. Optik 125:5546–5549
9. Zhu B, Chen Y, Deng K, Hu W, Yao ZS (2009) Terahertz science and technology and applications. In: PIERS Proceedings, Beijing, China
10. Zhou B, Yakup F, Du L, Dai JL (2010) Volakis: Polymer–carbon nanotube sheets for conformal load bearing antennas. IEEE Trans Antennas Propag 58:2169–2175
11. Balanis CA (2012) Antenna theory: analysis and design. Wiley
12. Llatser C, Kremers A, Cabellos Aparicio JM, Jornet EA, Chigrin DN (2011) Scattering of terahertz radiation on a graphene-based nano-antenna. In: AIP Conference proceedings, vol 1398, pp 144–146
13. Sharma A, Singh G (2009) Rectangular microstrip patch antenna design at THz frequency for short distance wireless communication systems. J Infrared, Millim Terahertz Waves, Springer 30(1):1–7
14. Kumar A, Marwaha S, Marwaha A, Kalsi NS (2010) Magnetic field analysis of induction motor for optimal cooling duct design. Simul Model Pract Theory, Elsevier Science 18(2):157–164
15. Bhattacharyya K, Goswami S, Sarmah K, Baruah S (2019) A linear-scaling technique for designing a THz antenna from a GHz microstrip antenna or slot antenna. Optik-Int J Light Electron Opt, Elsevier 199:1–8

Novel Approach to Reduce Rate of False Fire Alarms Through Multivariate Characterization of Fire

Pranjal Kumar, Subham Naskar, and Arpit Raj

Abstract Fire in residential or office premises can be a serious threat to life and property. There are many existing technologies for detecting fire and thereby raising alarms to alert the residents and appropriate authorities to take the required action. But, a false alarm is a major issue in these technologies. It can cause unwanted panic injuries and even stampede during the evacuation. With public budget cuts, some fire services have seen themselves forced to redefine response procedures. The new measures may not necessarily be in the best interest of society. E.g., Fire departments of certain countries prefer eye witness verification of fire before alerting the authorities. Delay in alerting fire extinguishing services might cost life and could lead to extensive damage to property. This motivated the authors to develop an approach to reduce the percentage of false alarms. The main focus of this study is to detect fire effectively and reduce the rate of false alarms by parameter-based distinction of any type of combustion occurring in a room.

Keywords Fire detection · False alarms · Disaster management

P. Kumar · S. Naskar · A. Raj (✉)
KIIT—Deemed to Be University, Bhubaneswar, India
e-mail: arpitraj.ece@gmail.com

P. Kumar
e-mail: pranjalcse30@gmail.com

S. Naskar
e-mail: subhamnaskar671@gmail.com

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_12

1 Introduction

False fire alarms are a major problem as it is associated with a high cost in terms of finance, time and manpower deployment. One study highlights that the "Fire and Rescue Service Authorities" of the United Kingdom claims the false alarms associated costs can rise to 1 billion pounds per year [1]. To substantiate, in 2011–2012 about 53 percent of alarms were, in fact, false positives [1]. Some studies conducted in the 90 s in Europe and the U.S., state even higher percentage. In some reports, the percentage of real alarms is as low as 11 percent [2]. Due to this pressing problem, in some countries visual verification is required before preventive and rescue forces are deployed. Studies conducted in Sweden and England have shown that verification of an alarm delays necessary resource allocation thereby increasing the overall cost for society [3].

Generally, it is observed that the large number of fire alarms is because of smoke detectors being incapable to distinguish between smoke particles released in a fire-related scenario and other events like cooking food, smoking cigarette, etc. Other factors responsible for false alarms are deceptive phenomena, equipment failure, human error, malice, good intent, etc. [3].

Fire detection using gas sensors has been recognized as a promising approach since the 1990s. As this approach can provide faster recognition of fire because during a fire, chemicals are released first then the visible smoke gets released, also additional security can be provided by monitoring chemical levels in the environment.

Autonomous fire monitoring and alert system performing with high precision can ensure extensive protection of life and property indoors. In this paper, the authors propose an approach to detect fire by monitoring the changes that occur in the environment in an event of a fire. Due to the variation of multiple factors caused in the environment during a fire, its detection becomes quite cumbersome [3]. Therefore, it would not be prudent enough to design a system based on one or two parameters only, as are many contemporary fire detection alarms.

Our proposed system takes into account multiple environmental factors such as gases released, temperature change, and humidity. Additionally, an algorithmic approach has been adopted which includes common materials found in a residential or official setting.

2 Concept Blocks

Authors have experimentally observed and analyzed the combustion behavior of various materials (especially found at home or office setting), which is explained in this section. This approach enabled the right selection of sensors for the system. Thereafter, the nature of the sensor readings during the combustion is examined, in an attempt to understand the relative significance of every sensor involved in the system.

2.1 Polyethylene

Polyethylene, a polymer comprising of carbon and hydrogen. When polyethylene undergoes combustion in abundant oxygen supply, it yields carbon dioxide and water. Whereas, carbon particles and carbon monoxide is produced when oxygen supply is constrained [4, 5].

2.2 Wood and Cellulose

When wood goes through combustion it produces gases like CH_2O [5, 6]. Thereafter, CH_2O reacts with surrounding oxygen produces CO_2, CO, H_2 as per the chemical equation:

$$6C_{10}H_{15}O_7 + \text{Heat} \rightarrow C_{50}H_{10}O + 10CH_2O$$
$$CH_2O + O_2 \rightarrow H_2O + CO_2 + CO + C + H_2$$

Carbon monoxide (CO), even in low concentration could adversely affect the human respiration system and various vital organs, leading to premature death [6].

2.3 Acrylic Fibers (Fibers, Rubbers, Adhesives, and Clothes)

Some synthetic fibers have about 80 to 85 percent content of acrylonitrile substance. These materials are made up of carbon, hydrogen, and nitrogen particles, and during their combustion, they produce hydrogen cyanide (HCN), an exceptionally harmful gas, in addition to CO [7–9].

Considering the gases released (covered in Sects. 1–3), MQ2 and MQ135 sensors have been selected as they are capable of detecting a wide range of NH_3, NO_x, alcohol, benzene, smoke, and CO_2.

2.4 Analysis of Temperature Variation

The temperature of the room rises significantly and can even reach greater than 500 degrees centigrade during fire, as the temperature increases the humidity present in the atmosphere varies sharply (depending upon the combustible materials) indicating a probable fire in the vicinity [10].

With a capability of detecting temperature over a wide range (-40 to 80 degrees centigrade) and measuring humidity from 0 to 100% with almost ± 1% accuracy, the DHT22 module proves to be appropriate for this purpose. Additionally, a flame sensor is also used because it is sensitive to various lightning conditions for its IR sensitive capabilities and it also enables better decision making by the system.

3 Experimental Setup

A setup was created to perform controlled combustion of materials and record the signals from various sensors of the system. The setup consisted of a rectangular enclosure of dimensions 8.5 × 7.5 cm^2. It had an opening of diameter 4 cm, which acted as ventilation. The distance between the edge of the enclosure and the measurement system is 4 cm. It was ensured that the materials were ignited for 10 s, and the time between the two readings was kept 2 s. 60 sets of readings were collected for each material. The measurement system and its constituent blocks are shown in Figs. 1 and 2.

The materials used for the experimental observation (along with their dimensions) were: cloth (8 x 8 cm^2), paper (10 x 10 cm^2), plastic (10 x 10 cm^2), rubber (2 × 2 cm^2), and wood (6 x 2 cm^2).

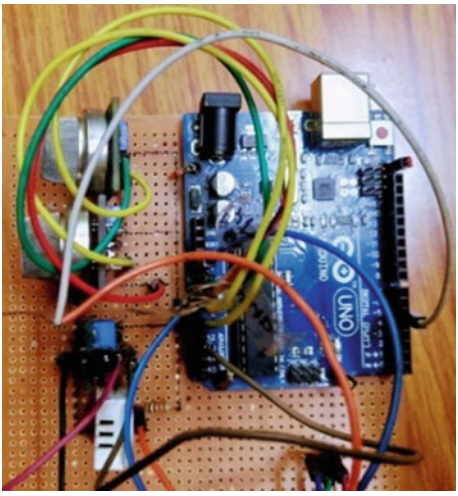


Fig. 1 Measurement system

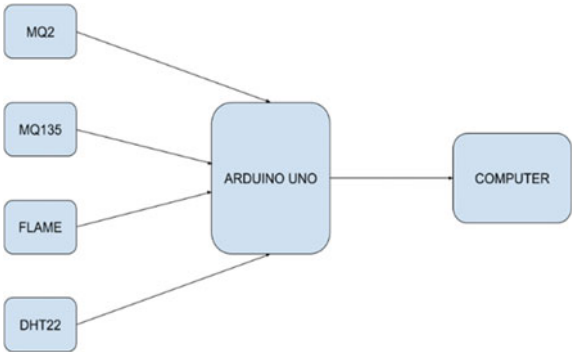


Fig. 2 Block diagram of the proposed system

3.1 Hardware and Software Specifications

- The Arduino UNO microcontroller R-3 board had been used for taking inputs from the sensors and for processing the output.
- MQ2 is a flammable gas and smoke sensor to detect and measure concentrations of flammable gas between 300 to 10,000 ppm [11].
- MQ135 is an air quality sensor for sensing a wide range of gases, including NH3, NOx, alcohol, benzene, smoke, and CO2. It has a high sensitivity to ammonia, sulfide, benzene steam, and smoke [11, 12].
- Flame sensor contains YG1006 sensor which is NPN silicon phototransistor which can detect a wide range of wavelength between 760 nm and 1100 nm [13]. It is sensitive to variation in lighting conditions, and because of its infrared-sensitive capabilities, it is beneficial under smoky conditions.
- DHT22 is humidity and temperature sensor which can measure temperature from −40 °C to 80 °C and humidity from 0% to 100% with an accuracy of ± 1 °C and ± 1% [14].
- Arduino IDE was used to develop and upload the program that would Arduino board.
- Python was used for recording the data from the Arduino microcontroller-based board into CSV (comma-separated value) file. It was also used for the analysis of data and plotting the graphs.
- VS Code was used as a text editor for writing python code.

4 Algorithm

```
Initialize
```

```
Data collected for individual sensors in
normal conditions and by burning
different materials.
Now every sensor will have two sets of data:
Fire
NonFire
M1 = Mean (Fire)
N1 = Mean (NonFire)
CHNG1 = abs(M1 - N1)/((M1 + N1)/2)
Similarly, we have, CHNG2, CHNG3,
  CHNG4, CHNG5 for all the sensors
```

```
Start Loop
C1 = Reading_Sensor1()
C2 = Reading_Sensor2()
C3 = Reading_Sensor3()
C4 = Reading_Sensor4()
C5 = Reading_Sensor5()
D1 = C1 > M1
D2 = C2 > M2
D3 = C3 > M3
D4 = C4 > M4
D5 = C5 > M5
Decision = (D1*CHNG1 + D2*CHNG2 + D3*CHNG3 +
D4*CHNG4 + D5*CHNG5)
If (Decision > 0)
               There is Fire
Else
               There is no fire
```

Here, *CHNG[i]* (where i = {1, 2, 3, 4, 5}) represents the change in the readings that have occurred in the individual sensors. By calculating it we get to know which sensor is the most sensitive one. Higher the *CHNG[i]* value, the more sensitive the sensor is and hence it has more priority in decision making. The *CHNG[i]* value is multiplied by *D[i]* which has a value of either 1 or 0 based on whether the sensor determines it is a situation of fire or not respectively.

5 Experimental Results and Discussion

Different materials have their unique combustion chemical signature and hence the sensors produce distinctive readings for each material. Figures 3, 4, 5, 6 and 7 shows the plot of various sensors for different materials under experimentation.

For performing the experiments for each test case (materials), the enclosure was kept in an optimal temperature condition with proper ventilation. The materials were ignited and kept in the enclosure for a fixed amount of time of two minutes. After obtaining the data for each test case the enclosure lid was opened to restore the internal environment to normal conditions. The above process was repeated for each test case, multiple times to ensure the consistency of the data collected.

The objective of collecting data from a diverse set of materials is to have a less biased dataset and thereby enhancing the actual fire prediction capability of the system. From Fig. 3 it is observed that paper shows the highest temperature increase,

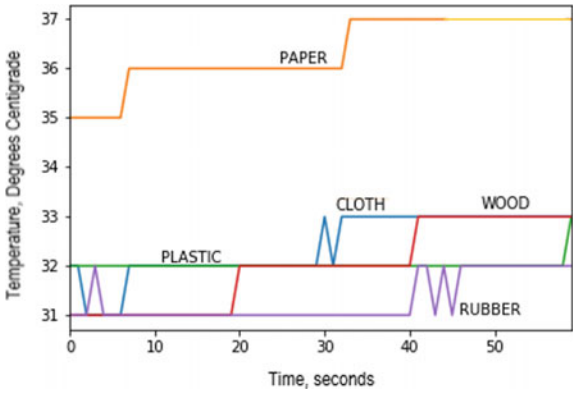


Fig. 3 Data from DHT22 sensor (Temperature)

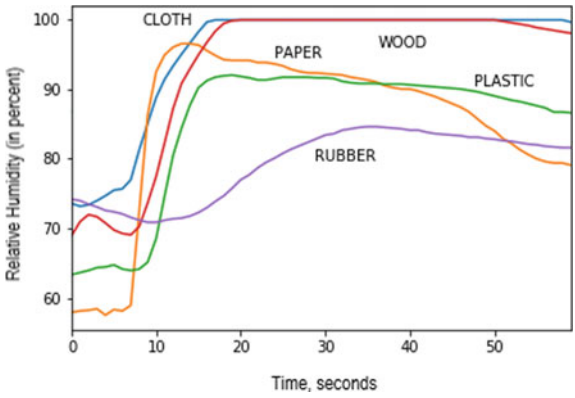


Fig. 4 Data from DHT22 sensor (Humidity)

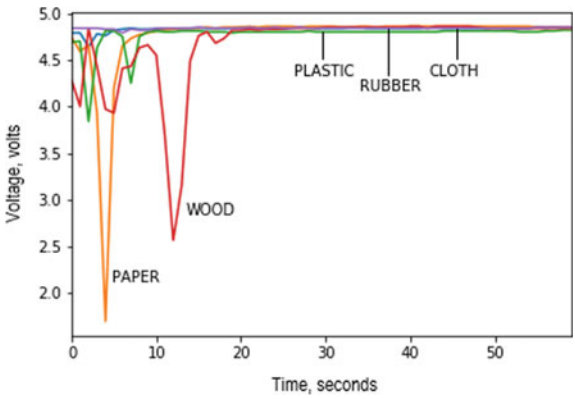


Fig. 5 Data from Flame sensor

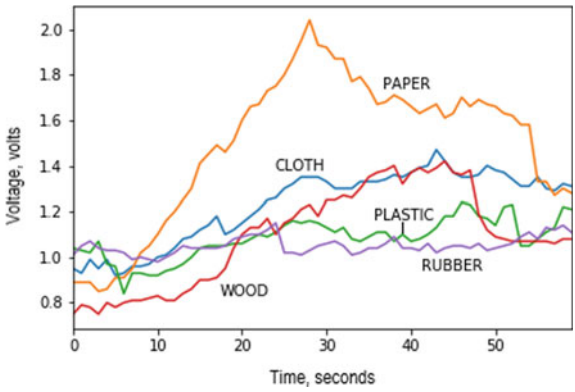


Fig. 6 Data from MQ135 gas sensor

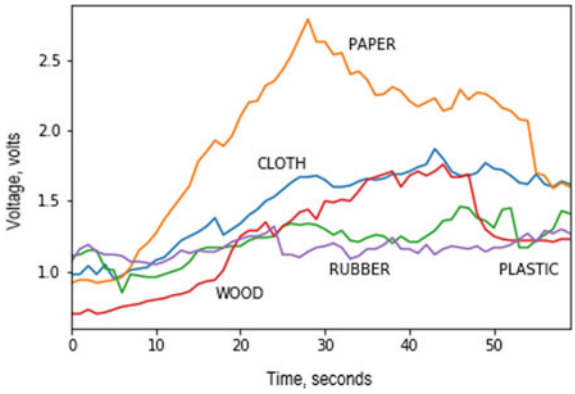


Fig. 7 Data from MQ2 gas sensor

because of complete combustion. Whereas, for a piece of cloth, wood, rubber and, plastic, change is not significant due to partial combustion.

From Fig. 4 it is seen wood initially take time to burn due to the moisture content. But, once ignited it burns steadily, which is evident from Fig. 5 As water vapor is one of the by-products of combustion for many materials, if there is no sufficient ventilation, humidity increases observed in Fig. 4 Wood and piece of cloth show the highest variations in relative humidity reading because they possess comparably higher moisture content among the others. Plastic and piece of cloth undergo slow combustion generating various gases and smoke as evident from Figs. 6 and 7.

Thus from the nature of graphs, and detecting the fire based on commonly available materials at home or office space, the proposed system can provide better and precise detection capability.

6 Conclusion

The main focus of this paper is to distinguish between a fire and non-fire scenario with high precision thereby decreasing the rate of false alarms. To fulfill this motive, parameter-based decision-making abilities of the various sensor is used for monitoring purpose. From the proposed methodology, combustion profile of commonly available materials at home or office space, such as wood, paper, cloth, rubber, and plastic are recorded. This is fed into the algorithm which enables fire detection with greater accuracy. The experimental setup system was built using low-cost sensors which are widely popular in the embedded system design, thereby keeping the setup cost-effective.

Authors are presently working to bring in more functionality and reliability by displaying warnings to users when the sensors get dirty due to the deposition of soot. To further improve the accuracy of the system an approach to classify the results using statistical techniques is currently under development.

References

1. Chagger R., Smith D (2014) The causes of false fire alarms in buildings. BRE Global Ltd., Watford, UK. Briefing Paper
2. Podrzaj P, Hashimoto H (2006) Intelligent space as a fire detection system. In: Proceedings of the IEEE international conference on systems, man, and cybernetics; Taipei, Taiwan. 8–11 October 2006, pp 2240–2244
3. Lance R (2014) Senior Manager Industry Affairs, Siemens Switzerland Ltd. REDUCING FALSE FIRE ALARMS: A Study of selected European Countries, 07 March 2014
4. Autian J (1970) Toxicologic aspects of flammability and combustion of polymeric materials. J Fire Flammab 1(239)
5. Sumi K, Tsuchiya Y (1971) Toxic combustion products of wood and polystyrene. National Research Council of Canada, Division of Building Research, BRN 76, Sept. 1971
6. Blumenthal I (2001) Carbon monoxide poisoning. J R Soc Med 94(6):270–272. https://doi.org/10.1177/014107680109400604
7. Henderson Y, Haggard HW (1927) Noxious gases and the principles of respiration influencing their action. Chem Catalog Co. Inc., 212 p
8. Autian J (1970) Toxicologic aspects of flammability and combustion of polymeric materials. J Fire Flammabil 1(239)
9. Documentation of the Threshold Limit Values. American Conference of Governmental Industrial Hygienists. Third Edition, 1971, 286 p
10. Essentials of Fire Fighting and Fire Department Operations 5th edn
11. MQ-2, MQ-135 Gas Sensor Technical Data, Hanwei Electronics Co. Ltd, 2002
12. The BOCA Basic Building Code/1970. Building Officials Conference of America, Inc., 5th edn, 413 p
13. Flame sensor, Phi Robotics Research Pvt. Ltd
14. DHT22 Technical Data Aosong (Guangzhou) Electronics Co. Ltd

A Graph-Based Band Selection Method for Hyperspectral Images Using Correlation Matrix

Jintu Kumar Das, Christopher D. Tholou, Alok Anand Minz, and Sonia Sarmah

Abstract Images are classified by analyzing the numerical properties of image features and then are organized into categories. Classification algorithms are typically performed in two phases i.e., training and testing. The number of training samples needed to design a classifier increases with the dimension of feature vector and it is challenging to determine if all features are necessary for the classifier. Therefore, we need a way to reduce the dimension of the feature vector without losing any important information. Remotely sensed hyperspectral images contain hundreds of spectral bands which provide detailed information about the objects. But this increased dimension at the same time also increases the computational complexity of classification. In this paper, we present a graph-based approach for band selection using correlation matrix and mutual information for dimension reduction of hyperspectral images so as to decrease the computational complexity during classification.

Keywords Band selection · Hyperspectral dataset · Correlation matrix · Mutual information · Feature reduction

1 Introduction

Hyperspectral remote sensing is an emerging and multidisciplinary field with many applications such as geology, ecology, atmospheric science, and forensic science. It provides spatial and spectral information simultaneously. The hyperspectral images are represented in a three-dimensional data cube (x, y, λ) for processing and analysis, where x and y represent two spatial dimensions of the scene, and λ represents the spectral dimension [2]. The hyperspectral imaging covers an extensive spectral range providing high potential for discrimination of subtle differences in ground covers. However, due to this high dimensionality, the classification performance for hyperspectral images decreases and may suffer from the curse of dimensionality [9]. As a result, we need to reduce the dimensionality of the such images without losing the

J. K. Das (✉) · C. D. Tholou · A. A. Minz · S. Sarmah
School of Technology, Assam Don Bosco University Azara, Guwahati 781017, Assam, India
e-mail: jintukumardas@gmail.com

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_13

original information. Feature reduction is the transformation that maps the data from a high order dimension to a low order dimension [4]. Feature reduction can be implemented with feature selection or feature extraction. Different techniques are already introduced in the past for band selection [3, 4, 6, 7, 11] to find crucial and significant bands present in a hyperspectral image. One of the techniques introduced a supervised feature extraction method based on the discriminant analysis (DA) [4] which uses the first principal component (PC1) to weight the scatter matrices. A graph-based feature reduction method was proposed in [11] which uses super-pixels as input to the proposed method and Simple Linear Iterative Clustering (SLIC) is performed followed by Laplacian Eigenmaps (LE). In [6] authors proposed a feature extraction method where the input hyperspectral images were segregated into multiple subsets containing adjacent bands. Later the bands were merged together by averaging. In subsequent steps, these merged bands were further processed with recursive filtering giving the resulting feature for classification. Local binary pattern (LBF) [6] of the extracted features are formed thereby increasing classification accuracy.

In this paper, we have discussed a graph-based feature reduction method based on the concept of mutual information and a correlation matrix. In the proposed method, band correlation is calculated considering each band as a vertex. Edges are created between bands having equal or greater correlation values than a predefined threshold value. In the next phase, connected components of the graph have been extracted and from each component, the band having the highest mutual information with respect to the ground truth is selected. The process eventually results in a reduced dataset comprising of only significant bands.

2 Background

2.1 Correlation

Correlation is the measure of similarity between two signals [8]. Correlation between two variables X and Y can be found using the formula,

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{X})^2 \sum(y_i - \bar{Y})^2}} \tag{1}$$

where, r_{xy} is the correlation coefficient of the variables X and Y. x_i and y_i represent the *i*th value of X and Y respectively in corresponding samples. $\bar{X}$ and $\bar{Y}$ are the mean of the values of X and Y respectively.

2.2 Mutual Information

Mutual Information is the measure of how much a random variable is related to another [10]. The formal definition of mutual information of two random variables X and Y is given by,

$$I(X;Y) = \sum_{y \int Y} \sum_{x \int X} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \tag{2}$$

where, p(x,y) is the joint distribution of X and Y.

3 Graph and Connected Components

Graph—A graph can be defined as G = (V, E) where V represents the set of vertices v_1, v_2, … etc and E represents the set of edges e1, e2, … etc. Each edge is a pair between two vertices (v_i, v_j).

Connected Components—In graph theory, connected components of an undi rected graph is a subgraph where any two pair of vertices are connected by at least one path [5] (Fig. 1).

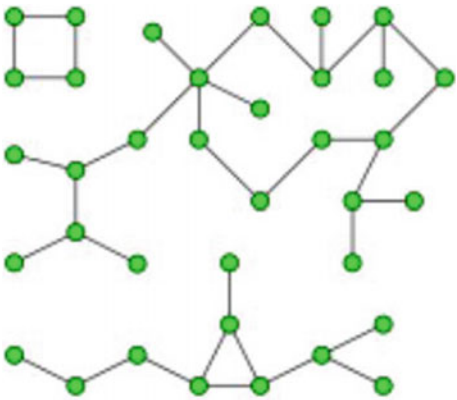


Fig. 1 A graph with 3 connected components

4 Proposed Methodology

In this proposed method, we have introduced a simple graph-based feature selection method using mutual information and a correlation matrix to select the significant bands from a hyperspectral image. The algorithm consists of two phases-graph construction and band selection. The overall time complexity is O(X * Y * Z) + O(V + E) where x, y, z are the height, width, the number of bands present in the image cube and V, E represents the vertices (the bands) and E represents the edges present in the graph. The two phases are discussed in detail in the following sub sections.

4.1 Construction of the Graph

Initially for the construction of the graph the hyperspectral dataset of dimension (x, y, λ), corresponding ground truth and a predefined threshold value of correlation coefficient are taken as input. Then a graph G is constructed considering each band of the input hyperspectral dataset as a vertex. Thus the number of vertices in G is equal to the number of bands in the hyperspectral i.e. λ. An edge is added between a pair of vertices(bands) in G if the correlation coefficient between those two vertices is greater than the input threshold.

4.2 Finding the Connected Components and Band Selection

In this phase, from graph G (constructed in Sect. 1), the connected components are extracted. From each connected component, we select the vertex(band) having the highest mutual information score with ground truth. So, if there are k($k \leq \lambda$) connected components in the graph then the total number of selected bands is also 0k0. Finally, a reduced dataset is constructed considering only the selected bands (Fig. 2).

5 Experimental Setup

5.1 Dataset Description

For carrying out the experiment, we have taken 3 datasets acquired by various sensors namely- Indian pines(corrected), Pavia University and Salinas-A. The Indian Pines(corrected) scene was gathered by AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) sensor and consists of 145 × 145 pixels, 200 spectral bands and 16 identified classes. Salinas-A dataset consists of 86*83 pixels and 204 spectral bands

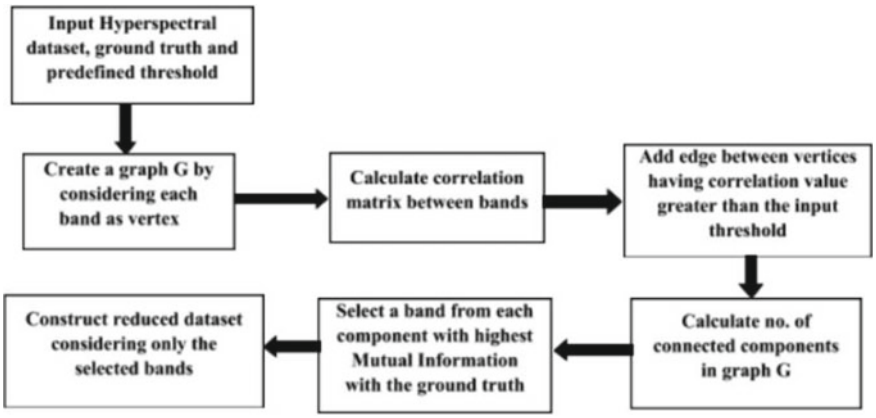


Fig. 2 Proposed method block diagram

and includes 6 classes. Pavia University Scene contains 103 spectral bands, 610*340 pixels and 9 classes [1]. For each dataset a series of experiments with three different threshold values of correlation coefficient—0.95, 0.97, 0.98.

6 Classifier Used

For classification purpose, two classifiers namely Support Vector Machine and Convolutional Neural Network were used separately.

In case of SVM, multiclass SVM with one against all strategy and linear kernel was used for training and testing purpose. The C parameter was set to 1.0 and gamma was set to auto.

A hybrid CNN classifier was also used for the classification. There were 10 hidden layers, 1 input and 1 output layer. The kernel size was taken as (3,3) and number of epoch chosen was 10. For the fully connected layers, the number of neurons were 256 and 128.

7 Evaluation Metrics

For training and testing the SVM classifier tenfold cross-validation method was used. Cross-validation is used to estimate the skill of machine learning model4.2. For CNN the training and testing data was divided into 70% and 30%, respectively. The following evaluation metrics were used to evaluate the performance of the classifiers:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{3}$$

with TP, FP, TN, FN being number of true positives, false positives, true negatives and false negatives, respectively.

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{4}$$

where κ represents the Cohen Kappa Score, p_0 is the empirical probability of agreement on the label assigned to any sample (the observed agreement ratio), and p_e is the expected agreement when both annotators assign labels randomly. p_e is estimated using a per-annotator empirical prior over the class labels.

8 Results and Discussion

From the series of experiments on the aforementioned datasets, we may observe that using only a small number of bands selected by the proposed methodology, adequate accuracy could be achieved using the SVM and CNN classifer. Table 1 presents the accuracy achieved by using all the bands of the datasets while Tables 2, 3 and 4 present the detailed classification result achieved by using only the selected bands for Indian Pines, Salinas-A and Pavia University dataset. The classification results which are improvement over using all the bands are shown in bold in the respective tables.

For the Indian Pines dataset the number of bands selected for the thresholds were 53, 67 and 96, respectively. From Table 1 it can be seen that using all the 200 bands of Indian Pines the obtained accuracies were 84.39% and 99.40% for SVM and CNN respectively. Table 2 shows that using 96 selected bands SVM gave classification accuracy of 74.92%. But with CNN classifier accuracy increased significantly to 95.32% with the same set of selected bands.

Table 1 Classification results considering all the bands

Dataset	SVM	CNN
Indian Pines	84.39	99.40
Salinas-A	99.92	98.33
Pavia University	91.64	98.81

Table 2 Classification results of Indian Pines on reduced dataset

τ	Bands Retained	SVM		CNN	
		Accuracy	Kappa	Accuracy	Kappa
0.95	53	71.23	73.24	89.48	90.23
0.97	67	73.33	75.13	91.10	91.85
0.98	96	74.92	76.97	95.32	96.32

Table 3 Classification results of Salinas-A on reduced dataset

τ	Bands retained	SVM		CNN	
		Accuracy	Kappa	Accuracy	Kappa
0.95	16	98.95	98.88	98.65	98.10
0.97	20	99.23	98.96	98.83	98.95
0.98	32	99.29	99.15	98.85	98.99

Table 4 Classification results of Pavia University on reduced dataset

τ	Bands retained	SVM		CNN	
		Accuracy	Kappa	Accuracy	Kappa
0.95	3	68.58	56.09	98.78	98.53
0.97	4	69.20	58.04	98.36	98.10
0.98	8	76.58	70.43	98.92	99.23

For Salinas-A dataset both SVM and CNN gave outstanding results with only limited number of selected bands. The obtained accuracy with all the 204 bands of Salinas-A were 99.92 and 98.33%, respectively for SVM and CNN. From Table 3, it may be observed that with only 16 selected bands(which is only .08% of the original number of bands) SVM gave an accuracy of 98.95% and CNN gave an accuracy of 98.65%. Similarly for the Pavia University with all the bands accuracies of 91.64% and 98.81% were achieved by using SVM and CNN respectively. However, for Pavia University dataset, as we can observe from Table 4, SVM performed moderately. But with CNN and using relatively very small number of bands, 8 in our case, high accuracy of 98.92% could be achieved, which was an improvement compared to the same using all the bands.

From the experimental results, it may be observed that using only a small number of bands selected by the proposed methodology, adequate accuracy could be achieved using SVM and CNN classifier. For Indian pines dataset using all the bands and SVM, the obtained accuracy is 84.39%. Using 96 selected bands accuracy of upto 74.92% could be achieved. But using CNN classifier the accuracy of upto 95.32% could be achieved. For Pavia University dataset, as we can observe from Table 4, SVM gives moderate results but with CNN and using relatively very small number of bands, 8 in our case, high accuracy could be achieved. For Salinas-A dataset both SVM and CNN gives outstanding results with only 16 number of bands.

9 Conclusion

In this work, We have proposed an algorithm for graph-based feature reduction, which tackles the challenges posed due to the high computational complexity involvement

while processing the hyperspectral dataset having hundreds or even thousands of bands. We have experimented the proposed method over three different hyperspectral datasets using two classifiers and found that using hybrid CNN classifier the selected bands give close or higher accuracy than using all bands. Our future work will concentrate on the tuning of the hyper-parameters and testing the proposed method on various other large datasets.

References

1. Hyperspectral remote sensing scenes (2019) http://www.ehu.eus/ccwintco/index.php/Hyperspectral Remote Sensing Scenes
2. Eismann MT (2012) Hyperspectral remote sensing. SPIE Bellingham
3. Hinton GE (2006) Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. Science 313(5786):504–507
4. Imani M, Ghassemian H (2015) Feature reduction of hyperspectral images: discriminant analysis and the first principal component. J AI Data Mining 3(1):1–9
5. Jain R, Kasturi R, Schunck BG (1995) Machine vision, vol 5. McGraw-Hill, New York
6. MS, D'souza D (2016) Feature extraction of hyperspectral images based on lbp and rf feature extraction techniques. Int J Sci Res (IJSR) 5(5):1977–1979. https://doi.org/10.21275/v5i5.nov163861
7. Reshma R, Sowmya V, Soman K (2016) Dimensionality reduction using band selection technique for kernel based hyperspectral image classification. Proc Comput Sci 93:396–402
8. Sarmah S, Kalita SK (2016) A correlation based band selection approach for hyperspectral image classification. In: 2016 IEEE 6th international conference on advanced computing (IACC). IEEE, pp 271–274
9. Steinbach M, Erto¨z L, Kumar V (2004) The challenges of clustering high dimensional data. In: New directions in statistical physics. Springer, pp 273–309
10. Wang B, Wang X, Chen Z (2012) Spatial entropy based mutual information in hyperspectral band selection for supervised classification. Int J Numer Anal Model 9(2)
11. Zhang X, Chew SE, Xu Z, Cahill ND (2015) Slic superpixels for efficient graph-based dimensionality reduction of hyperspectral imagery. In: Algorithms and technologies for multispectral, hyperspectral, and ultraspectral imagery XXI. vol 9472. International Society for Optics and Photonics, p 947209

A Study on Using Deep Learning for Segmentation of Medical Image

Lal Omega Boro and Gypsy Nandi

Abstract Segmentation of medical images using deep learning has provided state-of-the-art performances in this area of work. With the availability of large digital datasets and access to powerful GPUs, deep learning has transformed our world. We are now able to make computers mimic and replicate the functions of the human mind simply by providing enough data and computing the problem. Deep learning has a huge potential for medical image analysis and now it has been firmly established as a robust tool in image segmentation. This paper addresses the six popular methods that have employed deep-learning techniques for the segmentation of medical images which play a massive impact in the medical healthcare industry and in turn make a contributing role towards the concept of smart cities. A comparative study on these deep learning-based segmentation techniques will provide a researcher working in the field of medical imaging to explore further in this area for higher accuracy and better results.

Keywords Segmentation · Medical images · Deep learning · U-Net · V-Net · Mask R-CNN

1 Introduction

Artificial Intelligence plays a huge role in medical health care industries especially when it comes to medical imaging. It has great potential to address various challenges caused by excessive urbanization. Soon the future of smart healthcare in smart cities will be driven by AI doctors which can outperform human physicians in diagnosing diseases that need quick judgemental calls. There are numerous papers regarding medical image analysis ranging from segmentation, classification, detection, etc. that have proved the efficiency and effectiveness of deep learning which is a subset

L. O. Boro (✉) · G. Nandi
School of Technology, Assam Don Bosco University, Guwahati 781017, Assam, India
e-mail: omegaboro7@gmail.com

G. Nandi
e-mail: gypsy.nandi@gmail.com

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_14

of machine learning and in turn, a subset of Artificial Intelligence [1, 2]. Organ and substructure segmentation is considered as one of the most challenging tasks for analyzing clinical parameters like volume and shape as well as to identify the region of interest to locate a tumor, lesion, and other abnormalities due to high variability in medical images. A lot of research has been carried out in an attempt to develop an architecture for automated segmentation. The most popular architecture is U-Net proposed by Ronneger et al. [3] discussed in Sect. 3. It is built upon the fully convolutional network architecture and it is modified in such a way that it works with very few images and yet yields more precise segmentation. Furthermore, some researchers contributed to improving the architecture that can be seen from the work of Cicek et al. [4] and Military et al. [5] discussed in Sect. 3.

Segmentation is a process of partitioning images into a set of segments with similar properties to locate objects or boundaries. Segmentation is the first and most essential step in medical image analysis. There are many applications of segmentation. For example, during the medical diagnosis of cancer, the severity can be determined from the shape of the cancerous cell. For this, segmenting the image into multiple regions is required to filter out those regions that represent the characteristic of cancer cells. In this type of application areas the quality of the final output largely depends on the quality of segmented output [6]. The importance of segmentation in medical images are varied such as for:

- Locating the region of interest.
- Learning the anatomical structure of the human body.
- Measuring tissue volume to determine the size of the tumor and other abnormalities.
- Helping in treatment planning like differentiating cancerous cells from non-cancerous cells before undergoing radiation therapy.

This paper consists of three main sections which are structured as followed. Section 2 gives a brief overview of deep learning. Section 3 discusses the different state-of-the-art architectures of deep learning for the segmentation of medical images. Section 4 describes some of the challenges of deep learning in the medical healthcare industry. Section 5 concludes the chapter.

2 Deep Learning Overview

Deep learning is a subfield of machine learning which in turn is a subfield of artificial intelligence. Artificial intelligence is a technique to mimic human behavior, machine learning is a technique to achieve AI through an algorithm that is trained with data and deep learning is a type of machine learning inspired by the structure of the human brain. In terms of deep learning, this structure is called an Artificial Neural Network (ANN). Deep learning is powered by a huge amount of data. The “Big Data Era” will open up a lot of opportunities and innovations in deep learning. It has remarkably

improved the results for many problems like speech to text conversion, object recognition, self-driving cars, robotics, and medical healthcare. Deep learning has a huge potential for medical image analysis. The breakthrough came in the year 2012 [7] in an event where a deep convolutional neural network gave better accuracy of results and outperformed other established models and also won an overwhelming victory in the worldwide computer vision competition called the *ImageNet Classification.*

Traditional machine learning techniques for pattern recognition relied heavily on the human-engineered features extraction process. The process required careful engineering and domain expertise about a given problem to reduce the complexity of the data and to improve the model accuracy. Moreover, if erroneous features are extracted, the model is inherently limited in performance. Deep learning, on the other hand, execute feature engineering on its own and enable faster learning without explicitly told to do so. This eliminates the need for domain expertise and hardcore feature extraction.

The existence of deep learning can be traced back to 1943 [8] when Warren McCulloch and Walter Pitts proposes the first mathematical model for a neural network in their seminar paper titled "A Logical Calculus of Ideas Immanent in Nervous Activity". Warren McCulloch and Walter Pitts created a computational model called threshold logic to mimic the human thought process. Deep learning architectures are based on the framework of Artificial Neural Network (ANN) which is inspired by the biological neural network. ANN consists of multiple nodes (neurons) that are interconnected with each other from one layer to another as shown in Fig. 1. These artificial neurons are connected by synapses that contain weighted values and like synapses in the biological brain, they can transmit a signal to other neurons. Each artificial neuron outputs a decision signal based on the weighted sum. The weighted values of the neurons are optimized with loss function by matching the actual value and the predicted value. The optimized weights are then updated using gradient descent.

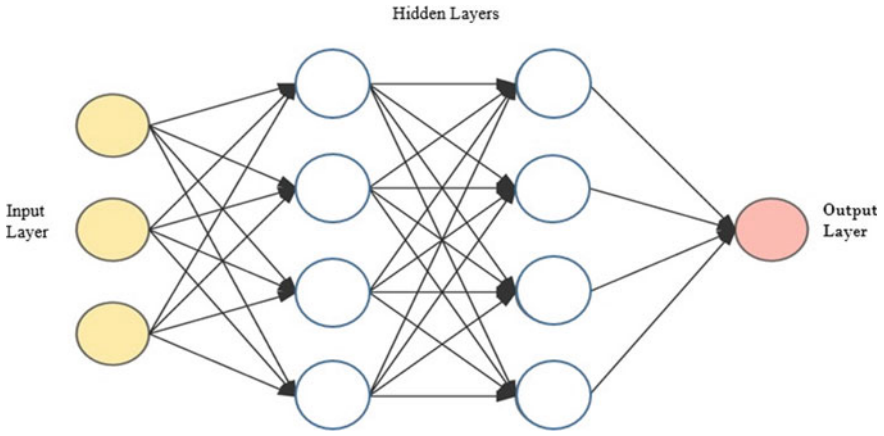


Fig. 1 Artificial Neural Network with 1 input layer, 1 output layer, and 2 hidden layers

There are two types of ANN, FeedForward and Feed-Back ANN. In FeedForward, the flow of information is uni-directional. Its main goal is to approximate some functions. For example a regression function $y = f^*(x)$ maps an input x to a value y. A feedforward network defines a mapping $y = f(x; \theta)$ and learns the value of the parameters θ that result in the best function approximation. Here x is used to evaluate the intermediate computation in the hidden layer which defines f which in turn is used to calculate y. In FeedBack ANN, there is a feedback loop in which the output of the model is fed back to itself. The combination of FeedForward and FeedBack ANN is called Recurrent Neural Network.

3 Basic Principles of Segmentation Approaches/Network Structures

There are many segmentation techniques that are applicable for medical imaging. However, the algorithm developed for one class of images may not always be applicable to other classes of images. This section briefly describes some of the most popular and successful deep learning-based segmentation architectures.

3.1 2D U-Net

One of the most well-known CNN architecture for medical image segmentation is U-Net which was initially proposed by Ronneberger et al. [3]. The U-Net architecture was built upon a more elegant architecture, the Fully Convolutional Network and it was modified in such a way that it yields precise segmentation with very fewer training images. The architecture looks like "U" and it basically consists of two parts a "contracting path" and an "expansive path". The contracting path follows the structure of a convolutional network which consists of two 3×3 convolutions and each followed by ReLU (rectified linear unit) and a 2×2 max pool with stride 2 for downsampling. The expansive path is similar to the contracting path however each upsample is appended with high-resolution features from the opposite side of the "U", the contracting path. The whole thing ends with a 1×1 convolution mapped all the component feature vector to the desired number of classes.

Figure 2 illustrates the 2D U-Net architecture for 32×32 pixels resolution where the blue boxes correspond to a multi-channel feature map and white boxes correspond to the copied feature map. The arrows indicate the different operations. The number of channels is given on top of the boxes.

The U-Net architecture is tested on three different segmentation tasks. It achieves a warping error of 0.0003529 and a rand-error of 0.0382 on the segmentation of neuronal structures in electron microscopic recordings. The architecture is further tested on two data set for cell segmentation tasks in light microscopic images. It

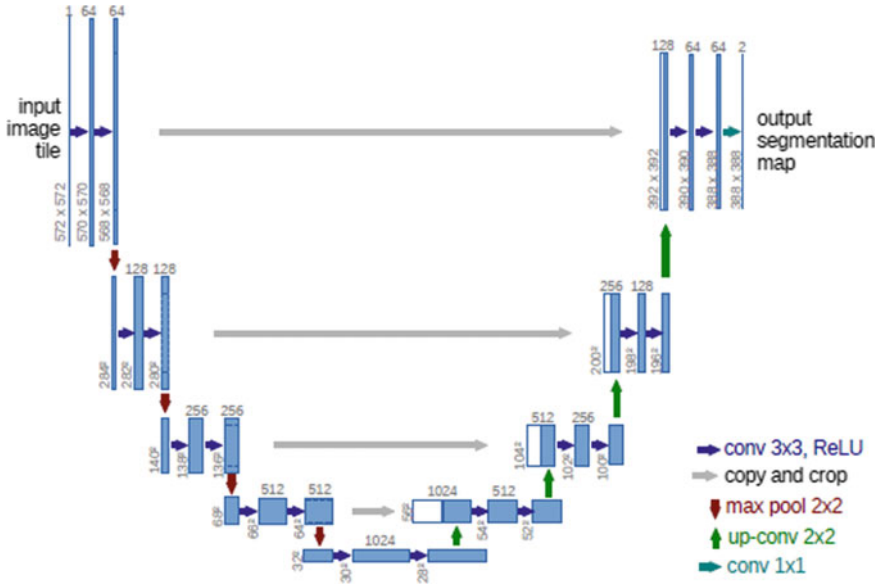


Fig. 2 The structure of 2D U-Net architecture [3]

achieves an average IOU of 92 percent on the first data set "PhC-U373" which is provided by DR. Sanjay Kumar, Department of Bioengineering University of California at Berkeley CA (USA), and an average IOU of 77.5 percent on the second data set "DIC − HeLa" which is provided by Dr. Gert van Cappellen Erasmus Medical Center, Rotterdam, Netherland.

3.2 3D U-Net

In an attempt to improve the U-Net architecture with richer spatial information, Cicek et al. [4] present a 3D U-Net architecture for volumetric segmentation that requires some annotated 2D slices for training. The proposed architecture is quite similar to the early work of Ronnebeger [3]. They redesign the architecture to perform a 3D operation. The proposed network was able to generate dense volumetric segmentation from some 2D annotated slices. Key to their model is weighted loss function and special data augmentation which enable them to train the network with only a few manually annotated slices, i.e., from sparsely annotated training data.

Figure 3 illustrates the 3D U-Net architecture. The structure is similar to the standard 2D U-Net structure. It has an analysis path and a synthesis path. The analysis path contains two 3×3x3 convolutions in each layer and a max-pooling 2×2x2 in each dimension with strides two. The synthesis path contains 3×3x3 convolutions and $2 \times 2 \times 2$ up-convolution in each layer with strides of two in each dimension.

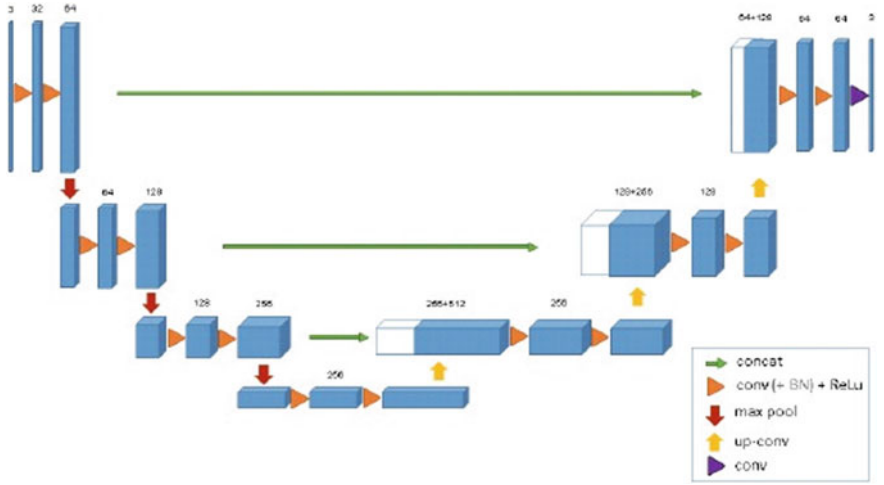


Fig. 3 The structure of 3D U-Net architecture [4]

The architecture experimented into two setups. In semi-automated setup, the user annotates few slices in the volume to be segmented and the network generates a dense 3D segmentation. The average IOU achieve in this setup is 0.863 experimented in threefold cross-validation. In a fully-automated setup, the performance gain of the 3D architecture is demonstrated to an equivalent 2D implementation.

3.3 V-Net

Furthermore, the most popular derivations of U-Net is the V-Net proposed by Millitari et al. [5]. Figure 4 illustrates the structure of V-Net architecture. The left part of the network consists of a compression path and the right part decompresses the signal until its original size is reached. They applied convolution at each stage both for extracting features from the data and reducing its resolution with 2x2x2 voxels wide kernel applied with stride 2. In their approach, pooling operations have been replaced with convolutional which helps to have smaller memory footprints during training. PReLU was used as a non-linearity activation function throughout the network.

The right part of the network extracts features and expands the spatial support of the lower resolution maps to output two-channel volumetric segmentation. These two output feature maps are the probabilistic segmentation of the foreground and background regions by applying soft-max voxelwise.

The architecture is trained on 50 MRI volumes and the relative manual ground truth annotation which is obtained from the "PROMISE2012" [9] challenge dataset. The performance is evaluated in terms of Dice overlap and Hausdorff distance between

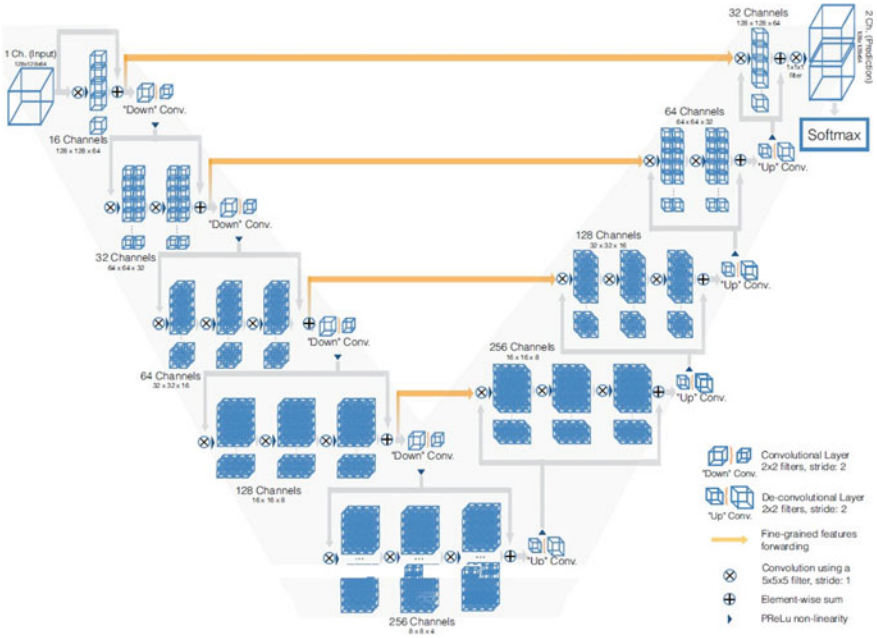


Fig. 4 Schematic representation of V-Net architecture [5]

the predicted delineation and the ground truth annotation and the obtained challenge score from "PROMISE2012".

3.4 Fully Convolutional Network (FCNs)

Long et al. [10] proposed fully convolutional networks that take input of the arbitrary size and produce an output of the corresponding(resample) spatial dimensions. They cast ILSVRC classifiers into FCN and augmented for dense prediction using in-network upsampling and a pixel-wise loss. Training for segmentation is carried out by fine-tuning which is done by back-propagation on the entire network. Skips are added between layers to fuse coarse and to refine the semantics and spatial precisions of the output.

The architecture is evaluated on the test set of NYUDv2 [11], PASCAL VOC [12], and SIFT Flow which gives 20 percent relative improvement to 62.2 percent mean IU in 2012 and reduces inference time in less than one-fifth of a second for a typical image.

3.5 DeepLab

It was first proposed by Liang-Chieh et al. [13] with a motive to perform semantic segmentation with deep learning. They make three main contributions that experimentally shown to have substantial practical merits. First, they introduced convolutions with upsampled filters for dense prediction tasks. Second, they propose atrous spatial pyramid pooling (ASPP) for segmenting objects at multiple scales. Third, they improved the localization of object boundaries by using DCNNs. They tackle the main challenges faced in [14] such as the repeated combination of max-pooling and downsampling which significantly reduces the spatial resolution of the resulting features map.

The proposed architecture achieves the new state-of-the-art at the PASCAL VOC-2012 [15] semantic image segmentation reaching 79.7 mIOU in the test set. During the experimental phase, the model weights of the Imagenet pre-trained VGG-16 or ResNet-101 were finetuned and the 1000-way Imagenet classifier in the last layer is replace with a classifier having the same targets number as that of semantic classes. DCNN and CRF training stages were decoupled and the proposed architecture is evaluated on four datasets PASCAL VOC 2012, PASCAL-Context [16], Pascal-Person-Part [17], and Cityscapes [18].

3.6 Mask R-CNN

Kaiming He et al. [19] propose a Mask R-CNN, a framework for instance segmentation of an object. The framework extends Faster R-CNN by adding a branch for predicting segmentation masks on each of Region of Interest (ROI). Segmentation is determined in parallel with predicting the class and box offset. The framework consists of two stages, the first one is Region Proposals and the second one is classifying the proposal and generating mask and bounding boxes. The authors indicate the convolutional backbone architecture using the term network-depth-features for feature extraction over the entire image. The additional branch for segmentation adds only small overheads and the entire network processes at 5 fps.

Some of the use cases of Mask R-CNN in medical image segmentation can be seen from [20] where they developed a technique for lesion detection and classification between benign and malignant. They achieve 0.75 mean average precision for detection and segmentation and 85 percent for differentiation between benign and malignant. Another recent application of Mask R-CNN can be seen from [21] where they achieve a weighted score of 0.906 for classifying MR images for automatic knee meniscus tear detection and meniscal tear orientation.

Table 1 shows a comparative analysis of the basic principles of segmentation network structures. The table contains a brief summary of the different architectures that have been discussed here to give an idea of how one architecture differs from another. This may help future researchers to explore the best-suited architectures for

Table 1 Comparative analysis of the basic principles of segmentation network structures

Architectures	Remarks
2D U-Net	It consists of a contracting path to capture context and expanding path that allows precise localization. The network has 23 convolutional layers in total. It is fast and required a few training images to give better results
3D U-Net	It comprises of two paths, the analysis path, and the synthesis path. Each path consists of specific layers followed by ReLU. Batch normalization ("BN") is used before each ReLU. It has 19069955 parameters in total
V-Net	It is a 3D image segmentation based on a volumetric convolutional neural network. A novel objective function based on the Dice Coefficient is introduced. PReLU is used as the non-linearity activation function
Fully Convolutional Network (FCN)	Classification networks such as AlexNet, the VGG net, and GoogLeNET are contemporary adapted into FCN and transfer their learned features by fine-tuning to segmentation task
DeepLab	It is a state-of-the-art semantic segmentation that uses atrous convolution and atrous spatial pyramid pooling. It also combines methods from DCNN and fully-connected conditional random fields to improve the localization of object boundaries
Mask R-CNN	It is a state-of-the-art instance segmentation which is an extension of Faster R-CNN. It uses ResNet 101 architecture to extract features from the images. It uses a pixel-to-pixel alignment

their problems. Table 2 review some of the application of the state-of-the-art model in medical image segmentation.

Other than the deep learning techniques mentioned in this section, several techniques such as the Convolutional Residual Networks (CRNs), Recurrent Neural Networks (RNNs), Ensemble learning, SegNet Neural Network has also been used in segmenting medical images.

4 Challenges of Deep Learning in Medical Healthcare

No doubt deep learning has a huge impact in medical healthcare industries. Its application is widely spread across several domains ranging from drug discovery, medical imaging, natural language processing, to understanding genomics. However, despite the advancement, it also has its limitations and challenges.

Table 2 Application of the state-of-the-art model for medical image segmentation

Paper Title	Method	References
Automatic Breast and Fibroglandular Tissue segmentation in Breast MRI using Deep Learning by a Fully Convolutional Residual Neural Network U-Net	U-Net	[22]
Organs at risk segmentation in Head and Neck CT Image Using a Two-Stage segmentation Framework Based on 3D U-Net	3D U-Net	[23]
Learning-based automatic segmentation of arteriovenous malformations on contrast CT images in brain stereotactic radiosurgery	V-Net	[24]
Fully Convolutional Multi-scale Residual DenseNets for Cardiac Segmentation and Automated Cardiac Diagnosis Using Ensemble of Classifiers	DenseNet based FCN architecture	[25]
A two-stage approach for automatic liver segmentation with Faster R-CNN and DeepLab	Faster R-CNN DeepLab	[26]
Automatic knee meniscus tear detection and orientation classification with Mask R-CNN	Mask R-CNN	[21]

- The entire deep learning model is often considered as "black box" because it is not easily interpretable. Consequently, in most situations, it is impossible to explain the technical and logical base to explain why and how it gives an excellent result.
- The high dependency on the quality and amount of training data as well as the tendency of overfitting should be considered [27]. To train a reliable deep learning model effectively, a large set of training data is required. This may not seem the problem in today's world especially with the explosion of medical healthcare data recently but specific disease data is limited. A common problem of training the model with limited data is overfitting. In this case, the model will not be able to generalize to new samples that it has not observed. Therefore, a particularly rare disease is not well suited with deep learning [28].
- There could be legal, ethical, and privacy issues on the basis of data usage. Since the data used consisted of human medical imaging and their corresponding meta-information, therefore its use may subject to specific regulations for the commercial development of a deep learning-based system.
- Adopting a deep learning system in medical image analysis would raise questions of ethics, responsibility, and oversight. As we know any system cannot be a hundred percent perfect. Therefore, it is difficult to assign responsibility to an individual researcher in case of life-threatening mistakes that may cause harm to patients.

5 Conclusion

This paper began with a brief overview of deep learning and highlighted its advantages over traditional machine learning. Six deep learning-based architectures have been discussed in this paper which is mostly built upon the convolutional neural networks. Some of the challenges of deep learning when applying to medical healthcare have also been discussed. From this review paper, one can see and learn the different architectures for the segmentation of medical images. This paper may help researchers to get a brief introduction of the recent state-of-the-art architectures relating to segmentation and also be concerned and well informed of the possible challenges in the medical healthcare industry.

References

1. Buetti-Dinh A, Galli V, Bellenberg S, Ilie O, Herold M, Christel S, Boretska M, Pivkin IV, Wilmes P, Sand W et al (2019) Deep neural networks outperform human expert's capacity in characterizing bioleaching bacterial biofilm composition. Biotechnol Rep 22:e00321
2. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K et al (2017) Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv:1711.05225
3. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 234–241
4. Çiçek O, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O (2016) 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 424–432
5. Milletari F, Navab N, Ahmadi SA (2016) V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth international conference on 3D Vision (3DV). IEEE, pp 565–571
6. Pal NR, Pal SK (1993) A review on image segmentation techniques. Pattern Recognit 26(9):1277–1294
7. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
8. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5(4):115–133
9. Litjens G et al (2014) Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge. Med Image Anal 18(2):359–373
10. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3431–3440
11. Silberman N et al (2012) Indoor segmentation and support inference from rgbd images. In: European conference on computer vision. Springer, Berlin, Heidelberg
12. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A (2011) The PASCAL Visual Object Classes Challenge 2011 (VOC2011) Results. http://www.pascalnetwork.org/challenges/VOC/voc2011/workshop/index.html
13. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2017) Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans Pattern Anal Mach Intell 40(4):834–848

14. Sermanet P, Eigen D, Zhang X, Mathieu M (2013) Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229, 2013
15. Everingham M et al (2015) The pascal visual object classes challenge: A retrospective. Int J Comput Vis 111(1):98–136
16. Mottaghi R et al (2014) The role of context for object detection and semantic segmentation in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition
17. Chen X et al (2014) Detect what you can: Detecting and representing objects using holistic models and body parts. In: Proceedings of the IEEE conference on computer vision and pattern recognition
18. Cordts M et al (2016) The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE conference on computer vision and pattern recognition
19. He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp 2961–2969
20. Chiao JY, Chen KY, Liao KYK, Hsieh PH, Zhang G, Huang TC (2019) Detection and classification the breast tumors using mask r-cnn on sonograms. Medicine 98(19)
21. Couteaux V, Si-Mohamed S, Nempont O, Lefevre T, Popoff A, Pizaine G, Villain N, Bloch I, Cotten A, Boussel L (2019) Automatic knee meniscus tear detection and orientation classification with mask-rcnn. Diagnost Intervent Imaging 100(4):235–242
22. Zhang Y et al (2019) Automatic breast and fibroglandular tissue segmentation in breast MRI using deep learning by a fully-convolutional residual neural network U-net. Acad Radiol 26(11):1526–1535
23. Wang Y, Zhao L, Wang M, Song Z (2019) Organ at risk segmentation in head and neck ct images using a two-stage segmentation framework based on 3d u-net. IEEE Access 7:144591–144602
24. Wang T, Lei Y, Tian S, Jiang X, Zhou J, Liu T, Dresser S, Curran WJ, Shu HK, Yang X (2019) Learning-based automatic segmentation of arteriovenous malformations on contrast ct images in brain stereotactic radiosurgery. Med Phys 46(7):3133–3141
25. Khened M, Alex Kollerathu V, Krishnamurthi G (2019) Fully convolutional multi-scale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. Med Image Anal 51:21–45
26. Tang W, Zou D, Yang S, Shi J, Dan J, Song G (2019) A two-stage approach for automatic liver segmentation with faster r-cnn and deeplab. Neural Comput Appl:1–10
27. Lee JG, Jun S, Won Cho YW, Lee H, Kim GB, Seo JB, Kim N (2017) Deep learning in medical imaging: general overview. Korean J Radiol 18(4):570–584
28. Ravì D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, Yang GZ (2016) Deep learning for health in- formatics. IEEE J Biomed Health Inf 21(1):4–21

Attention Recognition System in Online Learning Platform Using EEG Signals

Swadha Gupta and Parteek Kumar

Abstract The current scenario of education has been transformed from paper to paperless world. With the advancement in technology, smart education concept, and internet access, an adaptive learning environment has been created due to which participation of online learners across the smaller cities has been increased. In the online learning platform, there is a need to identify students' mental state so as to motivate them and to help them deepen their learning performance. The current stage of the online education system does not focus on providing immediate feedback and thus results in a lower performance of the user. The immediate feedback from student-to-teacher and teacher-to-student is not possible in online learning environments, unlike traditional classroom environments. To help online course teachers to effectively observe students' state of attention in online learning environments. This paper proposes a solution based on the EEG feedback approach. While watching online course video, a single electrode EEG device is worn by the user will keep on analyzing the brain activity to measure the attention level of the student. The interesting results for student learning assessments produced by the analysis of the raw data from the EEG device. This will ensure that the online user is able to receive constant feedback in real time while watching online course videos, thus improving the user's overall learning experience.

Keywords E-learning · EEG · MOOCs · Classroom education · ICT

S. Gupta (✉) · P. Kumar
Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology, Patiala, Punjab, India
e-mail: swadhagupta15@gmail.com

P. Kumar
e-mail: parteek.bhatia@thapar.edu

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_15

1 Introduction

The development of ICT (Information and Communication Technology) [1, 2] has had a global and profound influence on educational activities in a new era. ICT has enabled the learner to learn anytime, anywhere, and anyplace and has changed the concept of space and time. The primitive man's hominoid was known as Ape and now we are in the generation of the App, where change is speeding. The use of laptops, mobile devices, etc. has generated the idea that the context and place of learning are not very important. Stepping forward into the online learning world which is paperless, today's education technology is making yesterday's heavy textbook academic institutions redundant.

A teacher can help students to learn more from their learning by using advanced learning methods i.e. virtual learning environment, smartphones, online virtual classrooms, cloud servers, etc. In the digital age, the term Smart Education is used to describe modern age learning [3]. Smart education's aim is to meet the demands of life and work in the twenty-first century. In order to create which is known as "smart learning", attempts are being made to merge technology and pedagogy. Now, MOOC [4] is considered as the buzzword especially for higher education and some regarded it as the tsunami in the ocean of education. Over the last few decades, e-learning has taken many forms and shapes. From PowerPoint presentations to virtual reality, it has come a long way. By using modern technology, the aim of smart learning is to provide students with holistic learning so that they adapt and prepare themselves well for a fast-changing world.

In a traditional classroom environment, by observing student's movements or body signals, a teacher can pass judgment on the learning aspect of the learner and can alter the content of the corresponding subject according to the learning pace of an online student [5]. However, in web-based courses, the teacher is unable to assess whether or not a student actually understands the content in web-based courses. Moreover, there is a lack of teacher and student interaction in an online education environment. In this way, considering the intelligent tutoring system is becoming important, where it can evaluate how deeply students are learning while watching the online video content. To confirm that the learner is on the right track, the correct feedback is very necessary. As most of the time when learner faces some challenges and difficulties, they prefer not to complete their course. Such easily distracted and attention-deficient behavior is of high concern for the future of online education. To resolve the Attention and Feedback issue in an online learning environment, cognitive load assessment using EEG-based devices is the solution. Cognitive load assessment acts to improve student learning and it is based on cognitive load theory (CLT) [6]. The current research in the field of low power and portable EEG devices is still in its nascent stage. There has been very little development in the field of providing accurate EEG data using low-power devices. At present, there is no way to provide real-time feedback to the learner who is learning through online video courses. A lot of research has been done in the field of collecting brainwave data but much of it has been done by bulky non-portable BCI devices.

To monitor and improve the attention levels of users, a single channel electrode Headset named Neurosky, Mindwave, compatible with Android, Windows, and MAC has been used in our proposed work. In the field of measuring the attention of people [5, 7, 8], an enormous amount of work has been done. Therefore, there is scope for incorporating new techniques to improve the attention levels of an online student.

2 Electroencephalography (EEG)-Based BCI and Its Brainwaves

The brain is an electrochemical organ and generates limited electrical power but very useful. Electroencephalogram (EEG) is a method to record or detect the electrical activity generated within the brain by using electrodes (small, metal disc), which are placed on the surface of the scalp. The cluster of neurons gives enough signal to measure an electrical signal from the scalp by using an EEG device. The cells of the brain remain active all the time whether sleeping, relaxing, or doing any activity and it communicates by sending electrical impulses. The collection of EEG signals that is generated by the brain is amplified and digitized. The digitized signals are then sent to a mobile device or computer system for data processing and storage purpose. The EEG data are analyzed, which is an exceptional way to cognitive processes study.

The electrical activity arises from the brain that has shown in the form of brainwaves. Brainwaves are generated through synchronized electrical pulses from neuron masses that communicate with each other. These brainwaves are displayed in the form of frequencies, which is divided into different bandwidth to represent different functions. The changes in the brainwaves are according to what action the brain is performing. The slower frequency brainwaves describe the feeling of slow, sluggish, tired, or dreamy. The higher frequencies describe the feeling of hyperactive or weird. The speed of brainwave is measured in the form of Hertz and is dived into different bands defining fast, moderate, and slow waves. The four categories of brainwaves ranging from least activity to most activity are shown in Fig. 1:

1. Delta waves (frequency range up to 4 Hz)

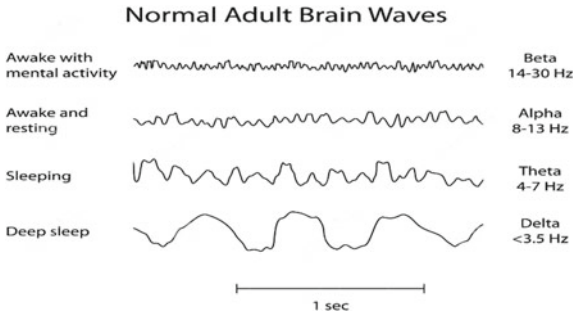


Fig. 1 Brainwave Frequency values

2. Theta waves (frequency range from 4 Hz to 7 Hz)
3. Alpha waves (frequency range from 7 Hz to 13 Hz)
4. Beta waves (frequency range from 14 Hz to about 30 Hz)

2.1 Cognitive Load Theory (CLT)

The concept of cognitive load theory (CLT) is based on how the human information system deals with information storing, handling, and processing. Figure 2 shows three main building blocks for the human information processing system. The three memories are sensory memory, working memory, and long-term memory. First, information is captured by sensory memory. This memory grabs the relevant information and forwards it to the next working memory.

The working memory processes the information received from sensory memory. The working memory buffers five to nine chunks of information at a time. This is central to CLT. Later, the brain receives this information and starts processing it. During processing, the brain categorizes information and passes to long-term memory. In long-term memory, the information is stored in a knowledge structure denoted as schemas. Based on schemas, human habits develop.

3 Related Work

Different works have investigated the effectiveness of EEG in predicting cognitive load in various real-world tasks. The concept of EEG is used in predicting the cognitive load in solving complex real-world problems. EEG signals are recorded and then after various predictive approaches are used to modularize and solving the problem. Kohlmorgen et al. [1] focused on the automobile simulation problem. In this problem, the information is recorded during driving to solve the automatic driving simulation

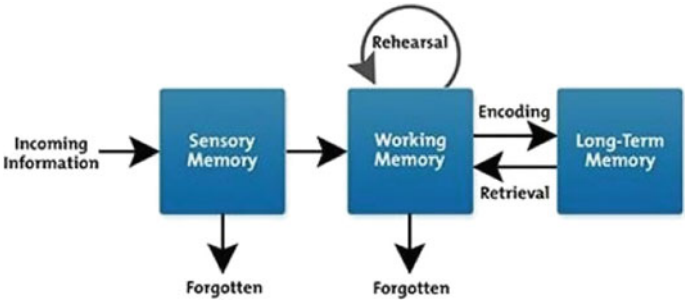


Fig. 2 Cognitive Load Theory (CLT)

system. The information is considered as a lower load related to active responsiveness. Deviations in cognitive load can be observed in monitoring and memorizing activities as well as in solving arithmetic issues, which have been shown in other studies [9, 10].

Cognitive load has also performed well in handwritten or pattern recognition problems [2, 11]. CLT deals with the cognitive load that examined a vital impact on educational problems and gives an acceptable solution to it. Traditionally, EEG equipment has major problems such as scalability, cost, nature of electrodes, etc. Nowadays, advancement in the EEG technology provides: lower-cost equipment, fewer number electrodes, nature of electrodes is dry [3, 12], which makes it is easy to handle and make it a well-calibrated form of equipment. Using EEG-based study in learning concepts gives impeccable results [4, 7]. For instance, EEG signals information in the form of alpha, beta, gamma, and theta waves that show correlated during playing a serious game [4] as well as solving true–false questions [8].

In other fields such as student's responses during study or reading [5], EEG-based information is recorded and tried to process with various machine learning techniques. As there are no cross-validation results present, thus it becomes difficult to get an actual solution. The overall study is overfitting to this problem. To overcome this problem, recent studies detect some close evidence that EEG signals with cognitive load are related to the difficulty of an instrument as well as complexity in signal information [6].

EEG signals with CLT help in solving student response problem, the regression analysis gives notable results. Another study deals with EEG-based data on reading easy and difficult passages with an Intelligent Tutoring System. Each passage has 61–83 words. EEG data with cognitive load process and provide feedback of the reader based on signals and categorized into a good candidate or average candidate or bad candidate. This type of problem can be solved using machine learning approaches that can deal with this complex type of arithmetic operations and provides accurate results [1, 13]. Vanitha and Krishnan [14] used Electroencephalography as a non-invasive procedure for the identification of stress in the students. As EEG is based on the feedback as received from the stress hormones, it is a reliable and effective method for the identification of stress level with the help of brain waves, Soman [15] used EEG signals for capturing and identification of the neural responses for a known and unknown language learning task. The experimental results revealed that the time-based frequency features obtained by the EEG signals have significant language discriminatory information.

In recent studies of brain–computer interface (BCI) studies, it collects the learner's brain signal information and process to identify the brain's attention level, anxiety level, and relaxation level [16, 17]. Further, the researcher used non-embedded EEG detectors to record EEG brain waves to analyze the learner's various attention and behavior levels [16–21].

4 Motivation

In the field of education, the rise of the Internet has opened new horizons over the years. Now the information is not only limited to classrooms but also by surfing the internet one can get quick and easy access to information. Maintaining optimum attention levels has adversely been affected due to the availability of information from many sources. To perform a cognitive task with a high degree of accuracy, it is important to maintain an optimum level of attention. Hence, when the attention level drops too low, there is a need for a real-time attention-based neurofeedback mechanism. The attention system can be used by the online learner to improve and monitor their attention level while watching the online course video

We consider an experiment on online learners while watching video content. A group of 80 students participated in the experiment. The preparation involves watching various course videos with a certain degree of attentiveness. The online learner can check their attentiveness from their attentive level values. But they can not keep track of their attentiveness throughout the whole course video. They also can not identify in which section of the video their performance and attentiveness fell and thus it leads to the hindrance of their preparation and performance. So, it becomes important to have a mechanism that can continuously monitor the performance of an online learner by keeping track of their attention throughout the learning process.

5 Our Contribution

The attention-based real-time neurofeedback is used in our experiment to tackle with the problem of varying attention. A novel approach is proposed in which an online learner can monitor and analyze their attention level on an individual basis. The participants are made to watch online course video [5], with and without the attention-based neurofeedback. In this paper, we focus on using the attention-based neurofeedback for the online learner that provides instant feedback to both online users and instructors and can be applied to various online courses. Our feedback-based model solves the problem of unavailability of a real-time feedback system for online courses. The experiment requires students to wear wireless portable Neurosky headset. The Neurosky headset records and monitors the attention level of the students while watching the online educational course. The difference between the previous approaches and the proposed approach is that it provides real-time instant feedback regarding the attentiveness. The online learner can see when their attention is falling. The values that are received from the EEG headset can further be analyzed in different subsections to see student's performance.

6 Proposed System

The EEG-based BCI system is proposed to monitor the attention level of the online learner while watching the online course video. The BCI-based non-invasion device is used to access the brain activity of the online learner [22, 23]. The wireless EEG device is used, which is worn by the learner. The EEG device gives the raw EEG data, which is in the form of brainwave frequency. The brainwave frequency helps to determine the attention level of the online learner watching the course video, which gives the feedback based on the attention level of the online learner [24, 25]. In this proposed system, the experiment is performed on the selected subjects consisting of males and females. They watched the online course video with EEG headset taking the readings of each subject. Based on the outcomes of the EEG device, the EEG-based attention system gives feedback when the attention level falls. In this way, the learning performance of the online learner increases and learner's real-time data of mental efforts are captured for analyzing purpose.

7 Experiment Design

In this study, a real-time feedback and attention recognition method is proposed, which is shown in Fig. 3 for the online learning platform that uses EEG brainwave frequencies to access the attention state of the learner.

7.1 Block Diagram of the System

The design of the system shown in Fig. 3 is categorized into four parts: Initiating, Planning, Executing, and Closing, which is given below:

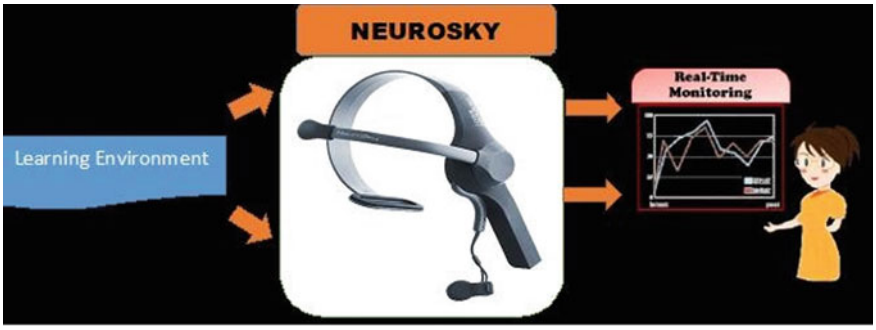


Fig. 3 A framework of the proposed system

1. Initiating (First Step)

In the first step, the research team decides its aims. After discussion among the research team, teachers, and educational experts, based on the requirement of the real-time feedback system, the ideas of the team are accepted formally.

2. Planning (Second Step)

The development of the project management plan is done after the identification of the system requirements. The plan summarizes the necessary development steps required to complete the project schedule, theory support of the scope document.

3. Executing (Third Step)

Based on the visual sensor and brainwaves monitoring, the attention state of the user is identified while watching the online course video.

4. Closing (Fourth Step)

At last, to accomplish the objectives as mentioned above, usability testing, stability testing, and expert evaluation have been carried out.

7.2 Feedback System/Attention Recognition

Neurosky brainwave Mobile2 sensor is used in this study. The EEG device consists of three main parts, including the data processing chip, the earlobe sensor, and the frontal lobe sensor. The device detected the electrical signals before translated into meditation and attention values. The attention and meditation values can be seen using visualizer and through different emoji for both the values. Meanwhile, it also provides an interface to illustrate the brainwave attention of the online student.

8 Experiment

Neurosky brainwave Mobile2 sensor is used in this study. The EEG device consists of three main parts, including the data processing chip, the earlobe sensor, and the frontal lobe sensor. The device detected the electrical signals before translated into meditation and attention values. The attention and meditation values can be seen using visualizer and through different emoji for both the values. Meanwhile, it also provides an interface to illustrate the brainwave attention of the online student.

Fig. 4 Student using our EEG headset

8.1 Subjects

A group of 80 students participated in the attention recognition experiment consisting of 40 males and 40 females aged 21–26 in the training phase. Each subject takes six sessions consisting of mental tasks and watched the education-related video with a length of 10 min can be seen in Fig. 4.

8.2 Experimental Procedure

The subjects are seated in a quiet room and made them watch a video for a duration of 10 min. When the student wore the EEG device, the start button is clicked after 1 min of watching the video. The 1-minute delay is taken so that the student's mind gets relaxed and chances of errors reduce. On the click of the start button, the device starts recording the EEG signals from the subject. When the subject starts learning through online courses, the human brain emits some brainwaves, thus system activates the function for the detection of attention signals by clicking the button. The system collects the brainwave signals during learning through an online course. Figure 4 shows an example of how this experimentation was conducted.

8.3 Experimental Analysis

In this section, we discuss the performance evaluation of the user, which is as follows:

8.3.1 Data

The EEG brainwaves data recorded by the application are shown in Figs. 5 and 6. The raw data are retrieved from the device, which is in the form of different frequency ranges and that raw data are preprocessed to remove the artifacts. The important features are extracted from it by using the feature extraction algorithm and classification is performed on it. The classification algorithm classifies the different attention levels of the online learner by analyzing the alpha, theta, beta, and gamma frequency that represents the different mental state of the learner. These data are first converted into required numerical form and then processed and manipulated by application for further conversion into visualization format.

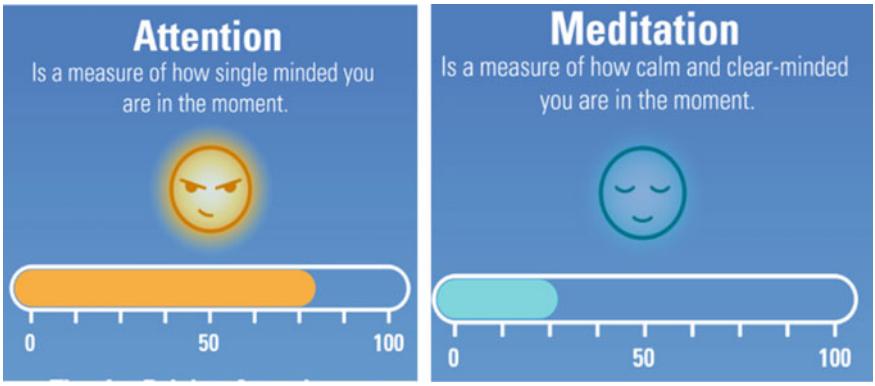


Fig. 5 Display of attention and meditation values in mobile application

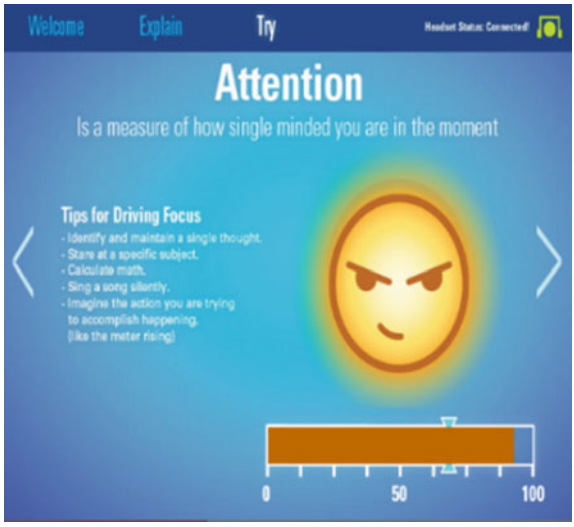


Fig. 6 Display of attention and meditation values in web application

8.3.2 Performance Parameters

Following are the main performance parameters of the proposed system:

- Time Delay. The time difference in notification alerts. Lower the value of the time delays better the performance of the system.
- Accuracy. This is one of the important parameters that specify the evaluation of the performance of user data recorded if only validated data are taken under consideration.
- Ease of use. The system is benefitted directly to the end-user. It is a simple click-based application easy to handle and modify.

8.4 Experimental Results

In this section, we will present our experimental results. The application performed as per targeted results. Figures 7 and 8 represent the actual application view and readings.

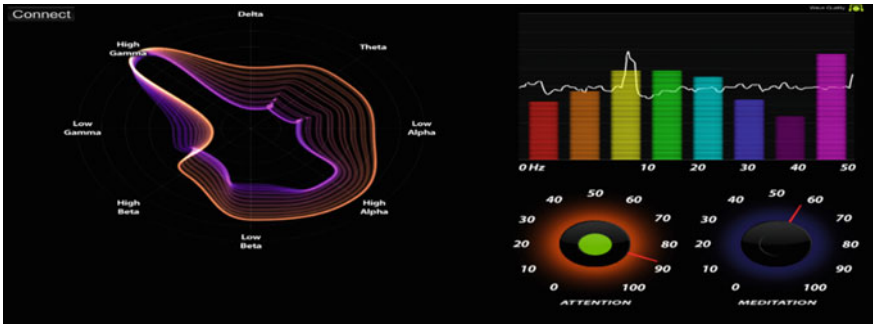


Fig. 7 Visual display of attention and meditation values in signal form

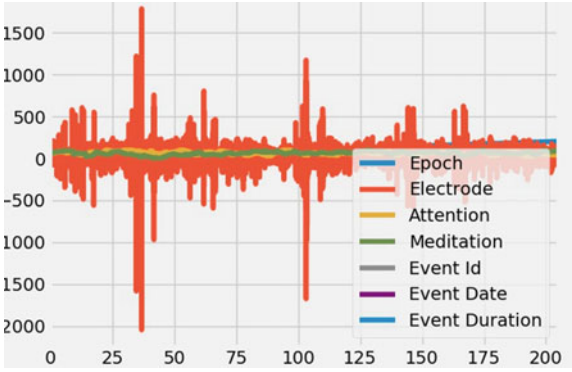


Fig. 8 Real-time interface of report

As illustrated in Fig. 8, the attention index falls below 40 and persists for more than 10 s when the learner cannot maintain his/her focus. The higher value defines higher attention, which is provided by beta frequency range as higher attention means higher beta value. The lower value defines lower attention value, which means the alpha frequency is dominant in this case. The program will automatically provide voice input for the student at this moment. For this participant, we can find out that when the system gives feedback, the attention value of the subject has improved significantly. This helps in improving the learning performance of online learners. This system will also help the content creator by sharing with them the analysis of their created content if there are changes required to improve their content or not.

9 Conclusion and Future Directions

The proposed system aims to use EEG technology to determine the mental states of online leaner while watching the online course video. The mental state's data are captured to make dynamic adjustment of video contents within e-learning contexts just like in the traditional classroom environment. This helps to make the online learning process efficient and effective with immediate real-time feedback. However, there are numerous gaps between online learning platforms and in-class teaching environments. Dis-similar to in-class teaching, where an educator can pass judgment if the students are understanding the materials through verbal requests or their non-verbal communication but in long-distance education, this is not possible. While watching MOOC, it is really important that students get his/her attention recognition on his/her performance and also get a prompt if she or he is not concentrating or understanding the concepts which are being taught. We try to solve this problem by using EEG input to analyze students' mental states. In this paper, we developed an EEG-based system to detect the attention level of the student within an online learning environment. This study's main focus is to increase the learning performance of online students by enhancing the concentration and attention level. The immediate feedback to the online students is provided by the values extracted from the raw data of the wireless EEG device. The parametric value was obtained after the analysis of the raw data. In the way, the system helps to enhance the student's attention while learning in an online environment. The future work will be including the function having dynamic adjustments of the attention feedback. The EEG-based smart system will be developed using machine learning algorithms for more efficient results.

References

1. Kohlmorgen J et al (2007) Improving human performance in a real operating environment through real-time mental workload detection. In: Dorhage G et al (eds) Toward brain-computer interfacing. MIT Press, Cambridge, MA, pp 409–422

2. Chaouachi M, Jraidi I, Frasson C (2011) Modelling mental workload using EEG features for intelligent systems. In: Proceedings of the international conference on user modeling, adaptation and personalization. Springer, Berlin, Heidelberg, pp 50–61
3. Matthews R et al (2007) Novel hybrid bioelectrodes for ambulatory zero-prep EEG measurements using multi-channel wireless EEG system. In: Proceedings of the 9th International conference on foundations of augmented cognition. Springer, Berlin, Heidelberg, pp 137–146
4. Derbali L, Frasson C (2010) Players' motivation and eeg waves patterns in a serious game environment. In: Proceeding of the 10th international conference on intelligent tutoring systems. Springer, Berlin, Heidelberg, pp 297–299
5. Heraz A, Frasson C (2007) Predicting the three major dimensions of the learner's emotions from brainwaves. Int J Comput Sci 2(3):187–193
6. Chaouachi M, Frasson C (2010) Exploring the relationship between learner EEG mental engagement and affect. In: Proceedings of the 10th international conference on intelligent tutoring systems. Springer, Berlin, Heidelberg, pp 291–293
7. Heraz A, Frasson C (2007). Predicting the three major dimensions of the learner's emotions from brainwaves. Int J Comput Sci 2(3):187–193
8. Chaouachi M, Frasson C (2012) Mental workload, engagement and emotions: an exploratory study for intelligent tutoring systems. In: Proceedings of 11th International conference on intelligent tutoring systems. Springer, Berlin, Heidelberg, pp 65–71
9. Berka C et al (2007) EEG correlates of task engagement and mental workload in vigilance, learning and memory tasks. Aviat Space Environ Med 78(5):231–244
10. Gevins A et al (1998) Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. Hum Factors: J Hum Factors Ergonom Soc 40(1):79–91
11. Cirett Galán F, Beal CR (2012) EEG estimates of engagement and cognitive workload predict math problem solving outcomes. In: Proceedings of the 20th International conference on user modeling, adaptation, and personalization. Springer, Berlin, Heidelberg, pp 51–62
12. Matthews R et al (2005) The invisible electrode–zero prep time, ultra low capacitive sensing. In: Proceedings of the 11th International conference on human-computer interaction. Lawrence Erlbaum Associates, Hillsdale, NJ, pp 22–27
13. Sun JCY, Yeh KPC (2017) The effects of attention monitoring with EEG biofeedback on university students' attention and self-efficacy: the case of anti-phishing instructional materials. Comput Educ 106:73–82
14. Vanitha V, Krishnan P (2016) Real time stress detection system based on EEG signals. Biomedical Research (2016) Computational Life Sciences and Smarter Technological Advancement, pp S271–S275
15. Soman A, Madhavan CR, Sarkar K, Ganapathy S (2019) An EEG study on the brain representations in language learning. Biomed Phys Eng Express 5(2):1–20
16. Wang C-C, Hsu M-C (2014) An exploratory study using inexpensive electroencephalography (EEG) to understand flow experience in computer-based instruction. Inf Manag 51(7):912–923
17. So WK, Wong SW, Mak JN, Chan RH (2017) An evaluation of mental workload with frontal EEG. PLoS ONE 12(4):
18. Chen C-M, Wu C-H (2015) Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. Comput Educ 80:108–121
19. Shadiev R, Wu T-T, Huang Y-M (2017) Enhancing learning performance, attention, and meditation using a speech-to-text recognition application: evidence from multiple data sources. Interact Learn Environ 25(2):249–261
20. Chen CM, Huang SH (2014) Webbased reading annotation system with an attention based self-regulated learning mechanism for promoting reading performance. Br J Edu Technol 45(5):959–980
21. Sun JC-Y (2014) Influence of polling technologies on student engagement: An analysis of student motivation, academic performance, and brainwave data. Comput Educ 72:80–89
22. Chang K, Nelson J, Pant U, Mostow J (2013) Toward exploiting EEG input in a reading tutor. Int J Artif Intell Educ 22(1–2):19–38

23. Chen CM, Wang JY, Yu CM (2015) Assessing the attention levels of students by using a novel attention aware system based on n brainwave signals. Br J Educ Technol
24. Sun JCY, Katherine PCY (2016) The effects of attention monitoring with EEG biofeedback on university students' attention and self-efficacy. UMEDIA-2016
25. Sharma A et al (2019) Smart learning system based on EEG signals. International conference on advances in computing and data sciences. Springer, Singapore

Comparative Study of Jute Fiber and PET Fiber-Reinforced Concrete

Rony Kutum, Piyush Singh, and Anirban Saha

Abstract The reinforcement of fibers into the concrete is gaining more popularity due to its application of waste as a fiber. Out of several industrial waste products such as polyethylene terephthalate (PET) bottles can be used for the production of fiber-reinforced concrete. This study investigates the effect of PET and jute fibers on concrete through a compressive strength test. The PET bottle-reinforced concrete shows better compressive results in a short duration of curing time (7 days) compared with both the jute fiber-reinforced concrete and concrete without reinforcement.

Keywords Fiber-reinforced concrete (FRC) · PET · Plastic waste · Jute · Reuse

1 Introduction

Concrete is composed of materials like coarse aggregate and fine aggregate that are bonded together with water and cement [1]. It is one of the most widely used materials on earth after water. They are brittle material and have a very low tensile strength. Plain concrete has a limited ductility and little resistance to cracking. The addition of uniformly dispersed and closely spaced fibers into concrete could act as a crack arrester and it also improves the properties of concrete [2]. PET and natural fibers like jute can be an effective material to reinforce concrete. Fiber-reinforced concrete (FRC) contains fibrous material. PET is one of the most used plastics in the world especially for the manufacturing of bottles, containers, etc. [3]. As per the estimate, about 12,342 tonnes of plastic waste is generated in India [4] and some 485 billion PET bottles were produced worldwide [3]. These PET and natural fibers can be used as a reinforcement in the concrete blocks to increased fracture toughness after the concrete hardening process. The recycled PET fibers improve the compressive and tensile strength and ductility of the concrete [5]. Pelisser et al. [6] studied recycled PET fibers in concrete measuring 10, 15, and 20 mm in length with 0.18% and 0.3% of fiber content.

R. Kutum (✉) · P. Singh · A. Saha
Department of Mechanical Engineering, Assam Engineering College, Guwahati, Assam, India
e-mail: rony.kutum1@gmail.com

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_16

2 Materials and Methods

2.1 Portland Pozzolana Cement (PPC)

Portland Pozzolana Cement (IS 1489-1) is used for the preparation of specimens that is most known and available everywhere, the pozzolanic materials commonly used are

- Volcanic Ash
- Calcined Clay
- Fly Ash
- Silica Fumes.

2.2 Aggregate

The fine aggregate of size less than or equal to 0.5 mm and coarse aggregate of size 10–12.5 mm was used for preparing the specimen.

2.3 Water

The water plays an important role in the hardened concrete. The impurities in the water may affect the setting of the cement and can affect the strength of the concrete. Fresh clean water without any impurities and oil has to be used for casting and curing of the specimen. The water–cement ratio of 0.5 is taken.

2.4 Fiber Reinforcement

Plastic and natural fibers were used as reinforcement. PET fiber is a waste material that is obtained from industries. The PET bottles are collected from houses and restaurants. The bottles were cut after removing the top and bottom parts. The length of the PET fiber was kept 100 mm and a width of 50 mm. Jute fibers are obtained from bags, the length of the fibers was kept 100 mm.

3 Methodology

3.1 Mixed Design

Mix design carried out for M20 grade, having mix proportion of 1:1.5:3 with water–cement ratio of 50%.

3.2 Compressive Strength Test

Concrete blocks are prepared with different percentages of fibers and cured under normal conditions for 7 days. A compressive strength test was done for different concrete specimens. The following is the formula for compressive strength test:

$$Compressive\ strength = \frac{Compressive\ load}{Cross-sectional\ area\ of\ the\ specimen} (1)$$

In this research, an experimental investigation was conducted to improve the strength of fiber-reinforced concrete provided with PET and jute fibers. PET fibers and jute fibers are extracted from the wasted bottles and bags collected from the household wastes and industries. The PET bottles and jute fibers are cut into a uniform length of 100 mm (Fig. 1).

A total of three types of concrete specimens of size 150 mm × 150 mm × 120 mm were used for the experiment (no-fiber, PET bottles, and jute fiber). Concrete blocks were cast with a design mix ratio of 1:1.5:3 with a water–cement ratio of 50%. A slump test was done to check the workability of the concrete mix as shown in Fig. 2. The mix was distributed and filled into the mold. The fibers are then placed over it in three layers as shown in Fig. 3. A total of three concrete samples are prepared for each fiber and non-fiber type.

(a) PET-fibers (b) Jute-fibers

Fig. 1 Reinforced materials used in concrete

Fig. 2 Slump test to check the workability of concrete

(a) PET layer

(b) Jute fiber layer

Fig. 3 Layering of fibers on mix

After curing for 7 days, the concrete blocks were tested as shown in Fig. 4, Table 1 shows different parameters for the concrete.

4 Results and Discussion

The concrete block with PET fiber mesh shows higher compressive strength as compared with the jute-reinforced and control concrete block. The jute fiber block shows lower compressive strength as it required more time to get higher compressive

(a) PET-bottle reinforced concrete blocks (b) Jute fiber reinforced blocks

Fig. 4 Fiber-reinforced blocks after curing

strength. The crack propagation rate was very less in PET fiber-reinforced concrete. The failure mode of control concrete blocks and fiber (jute and PET) concrete blocks is shown in Fig. 5 below. Table 2 shows the compressive strength test for cubes after 7 days (Fig. 6).

5 Conclusion

In this paper, the behavior of concretes reinforced with PET bottle fiber, jute fiber, and zero fiber has been investigated using compressive strength tests. A series of compressive tests with and without fiber concrete have been performed. Obtaining the following conclusions:

1. Higher compressive strength was achieved on PET fiber concrete 26.66 N/mm^2 as compared with the jute fiber 12.59 N/mm^2 and without fiber concrete blocks 16.73 N/mm^2.
2. PET fiber-reinforced concrete shows 59.35% increase in compressive strength as compared with no-fiber concrete blocks.
3. The PET fiber-reinforced concrete shows the better compressive result in short duration of time (7 days).
4. Using waste PET bottles and jute fiber as reinforcing material contributes to generating a benefit to environmental preservation.

Table 1 Different parameters used for concrete (with and without fibers)

Sl. no	W/C ratio	Sample		Size (mm)	Cement (% vol)	Fine aggregate (% vol)	Coarse aggregate (% vol)	Fiber content (% vol)	Fiber size (mm)	Fiber types
1	50%	S1	a	150 × 150 × 120	20	30	50	0	0	Without fiber
			b	150 × 150 × 120	20	30	50	0	0	
			c	150 × 150 × 120	20	30	50	0	0	
2	50%	S2	a	150 × 150 × 120	20	30	40	10	100	PET bottle fiber
			b	150 × 150 × 120	20	30	40	10	100	
			c	150 × 150 × 120	20	30	40	10	100	
3	50%	S3	a	150 × 150 × 120	20	30	40	10	100	Jute fiber
			b	150 × 150 × 120	20	30	40	10	100	
			c	150 × 150 × 120	20	30	40	10	100	

Fig. 5 Compressive test of fiber-reinforced concrete

Table 2 Compressive test for cube after 7 days

Sl.no	% of fiber	Technical replicates	Load KN	Average load KN	Compressive strength (N/mm^2)	Average compression (N/mm^2)	Fiber types
1	0	T1	360	376.66	16	16.73	Without fiber
		T2	300		13.33		
		T3	470		20.88		
2	10	T1	610	600	27.11	26.66	PET bottle fiber
		T2	590		26.22		
		T3	600		26.66		
3	10	T1	250	283.33	11.11	12.59	Jute fiber
		T2	300		13.33		
		T3	300		13.33		

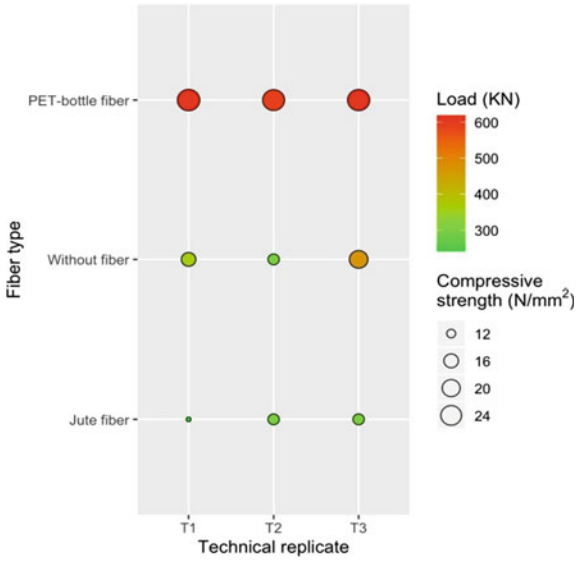


Fig. 6 Bubble heat map showing load and compressive strength in different fiber types

References

1. Haridass M, Gunasekaran R, Vijayakumar V, Vijayaraghavan P (2017) Experimental study on concrete with plastic plate fibre reinforcement. Int J ChemTech Res ISSN: 0974-4290, ISSN(Online):2455-9555
2. Zakaria M (2018) A comparative study of the mechanical properties of jute fiber and yarn reinforced concrete composites. J Nat Fibers
3. What a waste: an updated look into the future of solid waste management. https://www.worldbank.org/en/news/immersive-story/2018/09/20/what-a-waste-an-updated-look-into-the-future-of-solid-waste-management
4. ZadgaonkarA (2015) Plasticfuel. http://www.ipiindia.org/recycling/item/plastic-fuel-alka-zadgaonkar-pdf
5. Fraternali F, Ciancia V, Chechile R, Rizzano G, Feo L, Incarnato L (2011) Experimental study of the thermo-mechanical properties of recycled PET fiber-reinforced concrete. Compos Struct 93:2368–2374
6. Pelissera F, Montedoa ORK, Gleizeb PJP (2012) Romanb HR mechanical properties of recycled PET Fibers in concrete. Mater Res 15(4):679–686

Reused-Based Replacement Policy for Last-Level Cache with Minimum Hardware Cost

Purnendu Das and Bishwa Ranjan Roy

Abstract In multicore processor, the demand for cache memory is higher to bring frequently accessed data close to processor. Off-chip miss reduces the performance of the processor. Nowadays, multilevel cache architecture is used to increase the capacity of the cache memory. Upper level caches are dedicated to particular core act as private cache, whereas the Last-Level Cache is shared by the multiple core of the processor. The size of Last-Level Cache is bigger than the upper level of cache. So, the chance of large number of dead block in Last-Level Caches is higher. Efficient replacement policy is required to handle Shared Last-level Cache with minimum hardware cost. Traditional LRU replacement policy fails to remove dead block early in the large-sized cache. Many dead block detection algorithms have been proposed with higher hardware cost. In this paper, we have proposed a replacement policy that utilizes the reused locality instead of temporal locality. Also, the proposed technique reduces the hardware cost significantly. The proposed technique has been experimentally compared with two different baseline systems. It is observed that the proposed technique is capable of detecting dead block early and reduces the memory access time of the system.

Keywords Cache architecture · Last-Level cache · Replacement policies · Dead block · Hardware overhead

1 Introduction

Cache memory plays an important role in the performance of computer. In modern data-intensive applications, it is very challenging to reduce off-chip memory access. Multilevel cache memory is used to increase the size of on-chip memory. Last-Level Cache (LLC) is usually shared by multiple core of the Processing Unit. The capacity

P. Das · B. R. Roy (✉)
Department of Computer Science, Assam University Silchar, Silchar, India
e-mail: brroy88@gmail.com

P. Das
e-mail: purnen1982@gmail.com

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_17

of LLC is required to be increased to avoid the off-chip memory access. Hardware cost is also a critical component in On-Chip Multicore Processor. So, effective cache architecture is required to increase the capacity of cache memory. Many architectures have been proposed to replace SRAM by DRAM to reduce the space as DRAM is 8 times denser than SRAM [1].

The purpose of cache memory is to hold frequently accessed data by the processor. So, efficient cache replacement policy is required to identify and replace a block that is no longer required, by a block which is likely to access near future. Belady [2] proposed a replacement algorithm with the assumption that the future sequence of reference is known to the system, which is purely theoretical. Many practical algorithms [3, 4] have been proposed to detect and replace the dead block [5, 6] (which will not be accessed before eviction) at the earliest. But still there is a huge performance gap between Belady's optimal replacement policy and other practically implemented replacement policies.

In this paper, we have discussed the need and challenges of large cache memory in the first section. In the next section, we have discussed different approaches related to our works. Then, we have proposed a hardware cost-effective replacement algorithm based on re-used distance. Finally, performance evaluation has been done by comparing the experimental results of the proposed technique with two different baselines.

2 Background

In set-associative cache, the size of working set may become larger than the size of individual set during the execution of data-intensive applications, which causes conflict misses. So, it is required to increase the associativity of the cache to improve the cache hit. Increase in the cache associativity results into two different issues: 1) hardware cost to represent the sets and 2) increased number of dead blocks. The structure of typical set-associative cache is shown in Fig. 1. The most popular LRU replacement policy requires $n \times \log_2 n$ bits to represent each set, where n is the size

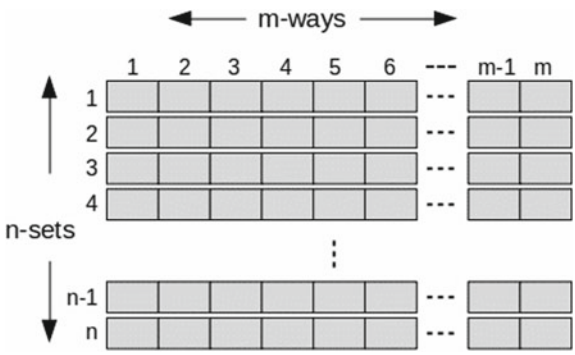


Fig. 1 Structure of m-ways set-associative cache

of the set. Many other algorithms like MRU, DIP, LRR required the same additional storage cost [7, 8]. Many effective dead block predictors [9–11] have been proposed to remove the never re-accessed block at the earliest but compromises the hardware cost. Purnendu et al. have proposed a hardware cost-effective cache architecture that reduces storage cost significantly [12]. They have used a hybrid replacement policy to select victim line. Basically, replacement policy takes the advantages of temporal locality to decide whether a cache line is to be retained or replaced. In hierarchical cache architecture, the upper level private cache consumes the effect of temporal locality. For example, if a line is frequently accessed by the processor after being loaded into the cache, the line will reside in the MRU (most recently used) position of private cache. But in shared Last-Level Cache, the line will not be accessed again until it is reloaded into the private cache. So, it is observed that the Shared LLC is influenced by reuse locality instead of temporal locality [3, 13, 14]. The principle of reuse locality is that if a line is reused at least once, likely to be reused more in future. LRR algorithm exploits the reuse locality to manage the SLLC with the same cost of LRU ($\log_2 n$).

2.1 *SplitWays*

This technique reduces the hardware cost significantly by splitting the set-associative cache vertically into multiple wayGroups [12]. If n is the size of set and m is the number of wayGroups, the size of wayGroup is n/m as shown in Fig. 2. Let the set-associative cache is C and wayGroups are G_1, G_2,…, $G_{\text{m-1}}$, G_{m}. Then

$$C = G_1 \cup G_2 \cup \ldots G_{m-1} \cup G_m. \tag{1}$$

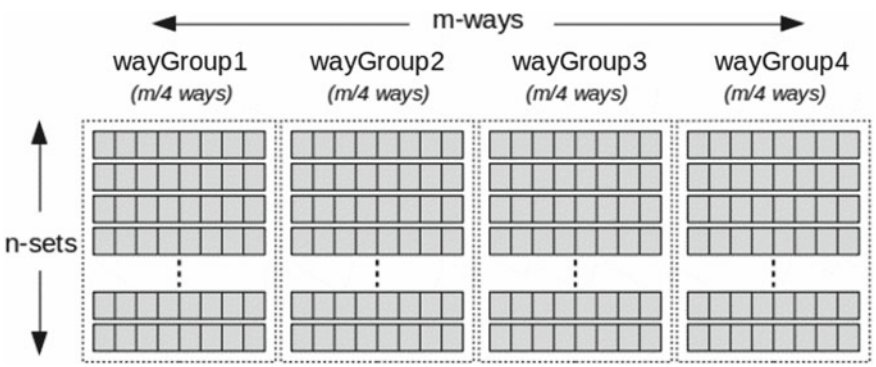


Fig. 2 Representation of m-ways set-associative cache in splitWays technique

To select a victim line, following steps are used: 1) LRU line is selected from each wayGroups to form a set of LRU lines, *S*. 2) Then victim line is selected from the set *S* randomly.

In this technique, the bits required to implement LRU algorithm in each set is $(n/m) \times (m \times \log_2 m)$, whereas without splitWays, it is $n \times \log_2 n$. SplitWays reduces hardware cost by 17–66% over LRU depending on different combination of associativity and number of wayGroups. In case of dead block detection, splitWays performs better than LRU. By splitting the large cache into multiple wayGroups, the chances of early dead block detection become higher.

2.2 Least Recently Re-Used (LRR)

This technique uses the re-used locality to obtain the cache line replacement order [13]. The technique is based on the assumption that a line present in the private cache is important as it is being used and a line with reuse locality has a chance to access again. The victim line is selected in the following order: 1) randomly selects a line neither in private cache nor reused. 2) select the line with longest reused distance if not present in private cache. 3) LRU line in private cache. This technique is capable of detecting dead block quickly in SLLC. The hardware cost of this algorithm is $n \times \log_2 n$ same as LRU.

3 Motivation

The main challenges of cache replacement algorithm are to detect dead block as early as possible with minimizing hardware cost. The base algorithms discussed in background section are splitWays and LRR. The splitWays is really effective approach in terms of hardware cost. These techniques perform better compared with traditional LRU algorithm, but it suffers a large performance gap with OPT [2] algorithm. SplitWays performs poor in case of applications where the cache lines show longer reuse interval. On the other hand, LRR reduces cache miss by removing dead block quickly, but the hardware cost is same as the traditional LRU algorithm. So, we have proposed an efficient dead-block predictor with reduced hardware cost.

4 Proposed Technique

In this technique, set-associative cache is equally partitioned into multiple subsets of ways called wayGroups. An additional set of LRR lines selected from different wayGroups is maintained. The proposed replacement policy works in two phases. In the first phase, LRR line of individual wayGroups is determined. In the second

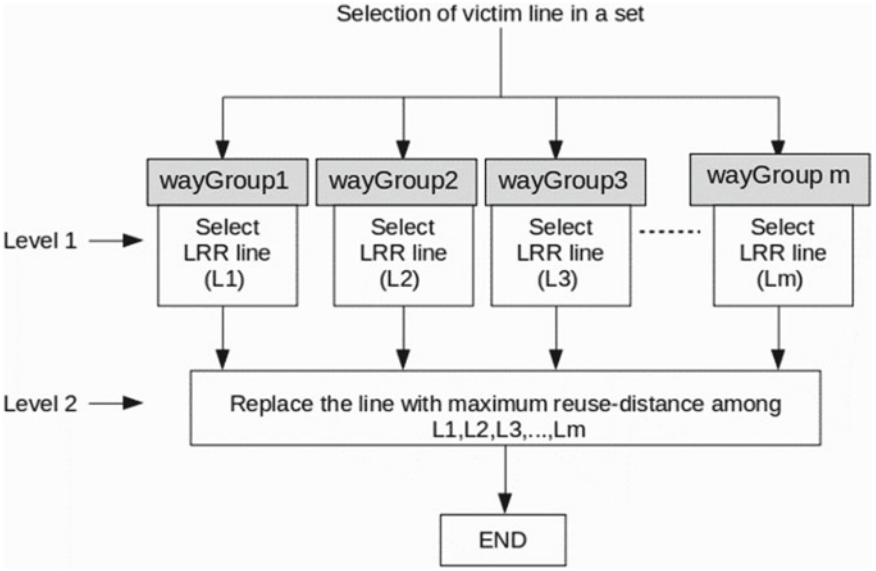


Fig. 3 Flow diagram of the proposed replacement policy

Phase, LRU wayGroup is identified and the LRR line belongs to that wayGroup is selected as victim. The flow diagram of this algorithm is shown in Fig. 3. The steps followed by the proposed are given below:

Phase I: selection of LRR line in each wayGroups

(a) Cache hit on line, say P
1. Set maximum value to the counter of line P
2. Set being-used bit of line P to 1.
3. Decrement the counter values of lines other than P and line with counter value = 0

(b) Cache miss
1. Set being-used bit of line that is not present in private cache to 0.
2. Pass the LRR line to phase II

Phase II: Replacement of victim block

(a) Cache hits on wayGroup G
1. Move the wayGroup G in the most recently used position

(b) On miss
1. Replace the LRR line of the waygroup in the LRU position
2. Initialize the counter value of new line in that wayGroup with 0 and being used bit with 1.

[Note: line with counter value 0 signifies the recently loaded line and line with maximum counter value signifies the most recently reused line. Being-used bit is used to specify if the block is present in the private cache]

4.1 Hardware Cost Analysis

SplitWays approach requires $(n/m) \times (m \times \log_2 m)$ bits to implement LRU replacement policy in n-way set associative cache with m wayGroups. In the proposed technique, an additional set of LRR line of different wayGroups is used to find the recency order of waygroups. So a total of $(n/m) \times (m \times \log_2 m) + (n/m) \times \log_2(n/m)$ bits required to select victim line. These additional bits are negligible compared with the hardware cost $n \log_2 n$ of traditional LRU. Table 1 shows the reduction of bit in splitWays and the proposed technique with respect to traditional LRU replacement policy. Figure 4 shows that the proposed technique is considerably cost-effective compared with different reused-based dead block prediction algorithms.

Table 1 Comparison of hardware cost among LRU, splitWays and proposed technique

Associativity	WayGroups	LRU (bits required)	SplitWays (bits required)	Proposed (bits required)	SplitWays (bit reduction in %)	Proposed (bit reduction in %)
16	2	64	48	50	25	21
16	4	64	32	40	50	37
32	2	160	128	130	20	18
32	4	160	96	104	40	35
32	8	160	64	88	60	45
64	2	384	320	322	17	16
64	4	384	256	264	33	31
64	8	384	192	216	50	43
64	16	384	128	192	66	50

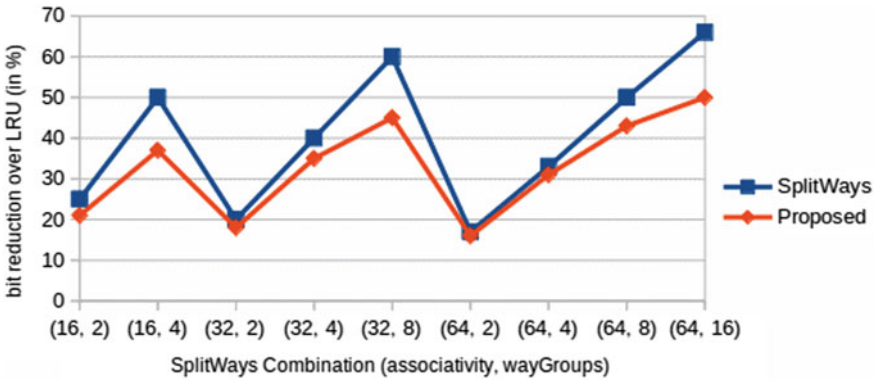


Fig. 4 Comparison of bit reduction by splitWays and proposed techniques over LRU

5 Performance Evaluation

The proposed technique is implemented in system emulation mode (SE) mode of gem5 [15, 16]. We have developed target machine using ALPHA architecture with MESI CMP protocol. The system consists of four core processor with private cache (L1 & L2) and shared cache (LLC). The proposed cache architecture and replacement algorithm are developed using special module of gem5 simulator called RUBY. The specification of the target machine is given in Table 2. We have executed seven parsec benchmark [17] applications in the target machine to observe the performance of the system in terms of Miss-Per-Kilo-Instruction (MKPI) and Cycle Per Instruction(CPI). All the seven benchmarks viz. *dedup, body, swaption, vips, fluid, ferret* and $\times$ *264* are executed for 200 million cycles for the analysis. The performance of the proposed technique is compared with the performance of LRU and splitWays replace replacement policy separately.

5.1 Results of Baseline Systems

It is observed from the results that the parameter MKPI for the proposed technique is reduced by 7% compared with baseline 1 while 12.5% in case of baseline 2, shown in Fig. 5. MKPI depends on the number of dead blocks present in the cache. The baseline-1 fails to evict dead block at the earliest in the lower level cache. The performance of splitWays is better compared with LRU because the chances of early detection of dead block increase due to splitting of ways in cache. The dynamic behavior of splitWays to detect the victim from any wayGropus increases the performance of the cache memory. The proposed technique utilizes the advantages of splitWays as well as the reuse-locality to make dead block prediction effective. The reduction in the cache miss reduces the memory access time. So, number of cycle per instruction (CPI) decreases with the decrease in the MKPI parameter. Figure 6 shows

Table 2 Specification of simulated system

Specification	Values
No. of cores	4
Levels of hierarchical cache memory	3
Private Cache	L1 and L2
Shared Last-Level Cache	L3
L3 cache	8 MB, 32-way set-associative
L2 cache	256 KB, 4-way set-associative
L1 cache	64 KB, 2-way set-associative
Cache block size	64B
No. of wayGroups	4

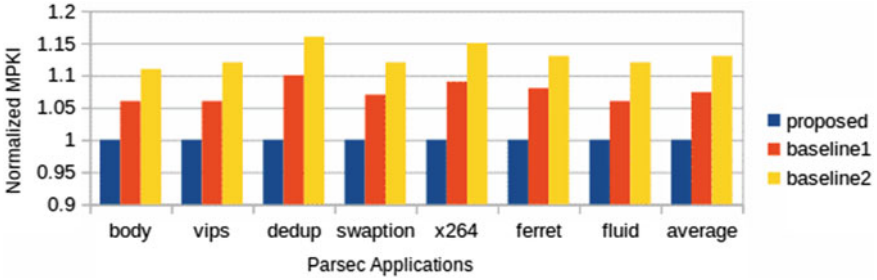


Fig. 5 Normalized comparison of the proposed technique with the baseline systems over MPKI

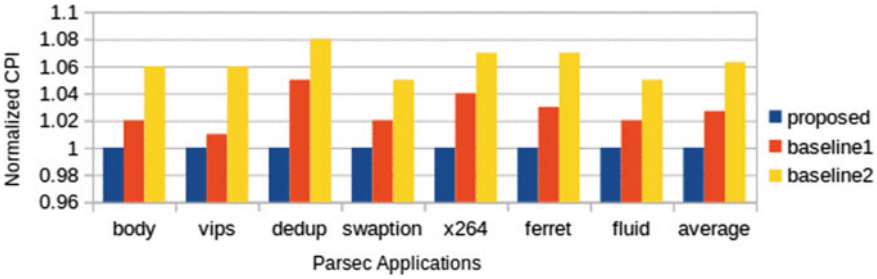


Fig. 6 Normalized comparison of the proposed technique with the baseline systems over CPI

the significant improvement of CPI in the proposed technique over the splitWays replacement technique. The increase in the hardware cost in the proposed technique is negligible compared with the improvement achieved on the overall performance of the system.

6 Conclusions

The presence of dead block increases with the increase in the size of cache. So, early dead block detection in the large Last-Level Cache is the challenging task of the replacement of policy. Hardware cost increases to make replacement policy efficient to detect the dead block quickly. We have proposed a hardware cost-effective replacement policy capable of detecting dead block in the large-sized cache. The proposed policy reduces the hardware cost by splitting the cache into multiple wayGroups and increases the chance of dead block detection by considering the reused locality.

References

1. Mittal S, Vetter JS, Li D (2015) A survey of architectural approaches for managing embedded dram and non-volatile on-chip caches. IEEE Trans Parallel Distrib Syst 26(6):1524–1537
2. Belady LA (1966) A study of replacement algorithms for a virtual-storage computer. IBM Syst J 5(2):78–101
3. Balasubramonian R, Jouppi NP, Muralimanohar N (2011) Multi-core cache Hi- rarchies. Morgan and Claypool Publishers
4. Das S, Polavarapu N, Halwe PD, Kapoor HK (2013) Random-LRU: a re- placement policy for chip multiprocessors. In: Proceedings of the international symposium on VLSI design and test (VDAT)
5. Liu H, Ferdman M, Huh J, Burger D (2008) Cache bursts: A new approach for eliminating dead blocks and increasing cache efficiency. In: 2008 41st IEEE/ACM International symposium on microarchitecture, pp 222–233
6. Khan SM, Jim´enez DA, Burger D, Falsafi B (2010) Using dead blocks as a virtual victim cache. In: Proceedings of the 19th international conference on parallel architectures and compilation techniques, ser. PACT '10, 2010, pp 489–500
7. Tian G, Liebelt M (2014) An effectiveness-based adaptive cache replacement policy. Micro-process Microsyst 38(1):98–111
8. Qureshi MK, Jaleel A, Patt YN, Steely SC, Emer J (2007) Adaptive insertion policies for high performance caching. In: Proceedings of the 34th annual international symposium on computer architecture, June 09–13, 2007, San Diego, California, USA
9. Wong WA, Baer JL (2000) Modified LRU policies for improving second-level cachebehavior. In: Proceedings of the international symposium on high performance computer architecture (HPCA), pp 49–60
10. Kaxiras S, Hu Z, Martonosi M (2001) Cache decay: exploiting generational behaviorto reduce cache leakage power. In: Proceedings of the International symposium in computer architecture (ISCA), pp 240–251
11. Liu H, Ferdman M, Huh J, Burger D (2008) Cache bursts: a new approach foreliminating dead blocks and increasing cache efficient. In: Proceedings of the International symposium on Microarchitecture (MICRO), 2008, pp. 222–233
12. Das P, Roy B (2019) SplitWays: an efficient replacement policy for larger sized cache memory. Int J Eng Adv Technol
13. Albericio J, Ibanez P, Vinals V, Llaberıa JM (2013) Exploiting reuse locality on inclusive sharedlast-level caches. ACM Trans Architect Code Optim 9(4) Article 38:19 p
14. Das S, Kapoor HK (2017) Latency aware block replacement for l1 caches in chip multi-processor. In: 2017 IEEE computer society annual symposium on VLSI (ISVLSI), pp 182–187
15. Binkert N, Beckmann B, Black G, Reinhardt SK, Saidi A, Basu A, Hestness J, Hower DR, Krishna T, Sardashti S et al (2011) The gem5 simulator. ACM SIGARCH Comput Architect News 39(2):1–7
16. Martin MMK, Sorin DJ, Beckmann BM, Marty MR, Xu M, Alameldeen R, Moore KE, Hill MD, Wood DA (2005) Multifacet's general execution-driven multiprocessor simulator (gems) toolset. SIGARCH Comput Architect News 33(4):92–99: http://www.cs.wisc.edu/gems/
17. Bienia C (2011) Benchmarking modern multiprocessors. Ph.D. dissertation, Princeton University. http://parsec.cs.princeton.edu/

Dr. Purnendu Das is an assistant professor of the Department of Computer Science, Assam University Silchar. He has pursued Ph.D. Degree from Tripura University. He has published research papers in many reputed journals.

Bishwa Ranjan Roy is an assistant professor of the Department of Computer Science, Assam University Silchar. He has done M. Tech degree at NIT Silchar. He has many publication in many reputed journals.

Revealing Structural and Optoelectronic Properties for Bi-Doped $CuGaS_2$ Chalcopyrite: A Density Functional Investigation

Karina Khan, Aditi Gaur, Amit Soni, U. Ahuja, and J. Sahariya

Abstract Electronic and optical analysis of a ternary chalcopyrite compound, $CuGaS_2$, with Bi doping has been performed using the first principle investigation based on the density functional theory calculations. These properties have been computed using Trans-Blaha modified Becke Johnson (TB-mBJ) as the exchange and correlation potential as embedded in Wien2k code. The basis of full potential linear augmented wave (FP-LAPW) has been followed in order to attain accurate and efficient results. The band gap reported through the calculation is 0.7 eV with 3% doping of Bi in $CuGaS_2$. The platform of optical and electronic analysis is decided on the basis of band structure, density of states (DOS), refractivity, dielectric tensor's components, reflectivity and integrated absorption coefficient. The parallel and perpendicular component of the optical spectra depicts an isotropic/anisotropic nature explained through their approximate coincidence.

Keywords Density functional theory · Supercell · Solar cell

K. Khan
Department of Physics, Manipal University Jaipur, Jaipur 303007, Rajasthan, India

A. Gaur · A. Soni (✉)
Department of Electrical Engineering, Manipal University Jaipur, Jaipur 303007, Rajasthan, India

U. Ahuja
Department of Electrical Engineering, NMIMS, Mukesh Patel School of Technology Management & Engineering, Mumbai 400056, Maharashtra, India

J. Sahariya
Department of Physics, National Institute of Technology, Uttarakhand, Srinagar (Garhwal) 246174, India

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_18

1 Introduction

The density functional theory (DFT) methods are applied to the system that explains the Kohn and Sham (KS) demonstration of the many electron system and exact ground state density [1–3]. The calculations follow the methodology stated as the most apt and standard method, i.e. FP-LAPW, which provides computational efficiency and result accuracy [4]. These approaches are centered on a set of basis expansion of KS orbitals [5]. The substitution of two atoms in the place of one metal atom of Zinc Blende yields a chalcopyrite compound. The yielded structure is in the form of tetragonal super-structure [6] exhibiting quasi cubic nature, i.e. c/a $\approx$ 2. The categorical division of the chalcopyrite family has been done into two subclasses of compounds formed by the elements like I, III, and VI or II, IV and V [7]. Chalcopyrite semiconductors have direct energy gaps that offer large value of absorption coefficients above the given band gap [8]. These materials can then be quoted in the form of absorber materials of thin-film solar cells. Thermal, optical, structural, magnetic, etc. are the properties exhibited by these compounds in order to state their usage in different application fields. Copper-based chalcopyrite compounds are suitable for optoelectronic device applications. Some of the copper series chalcopyrite compounds are known for their wide band gaps. The variety offered in terms of optical energy gaps and mobilities of carrier atoms provided by such ternary compounds gave its significant contribution to the formation of technological devices [9]. Theoretical and experimental investigation of the different properties for $CuGaS_2$ ternary chalcopyrites have been shown in the works carried out by many researchers [10–12]. The mixing of impurity atoms like 3d transition metals and other elements has led to the improvisation of the energy gap of $CuGaS_2$ as quoted in ref [13–15]. Role of addition of some impurity in $CuGaS_2$ has also depicted its use as an intermediate band solar cell [16–18]. The Bi-based doping of $CuGaS_2$ is synthesized through experimental method, i.e. solvothermal method [15].

2 Models and Methods

First, the unit cell structure of $CuGaS_2$ and its supercell have been presented in Fig. 1 (a) and 1 (b). In this work, the Ga atoms belonging to the $CuGaS_2$ compound are replaced with Bi atoms. The total atoms modeled in the 2 $\times$ 2x2 supercell are 31 for Bi-doped $CuGaS_2$ semiconductor with 3% of doping. The structure belongs to the tetragonal 122-I4-2d space group that gets transformed into 81-P4 tetragonal primitive symmetry. The structure is framed in a manner in which lattice parameters and atomic coordinates with the use of DFT are incorporated in WIEN2K code. The FP-LAPW program implements the PBE-GGA and TB-mBJ exchange correlation explaining the calculation of the compound's properties with doping of Bi. The cutoff energy input is -6 Ry. The sphere's atomic radius termed as muffin tin radius is reported as 2.38 a.u, 2.22 a.u and 1.95 a.u for Cu, Ga and S, respectively. The

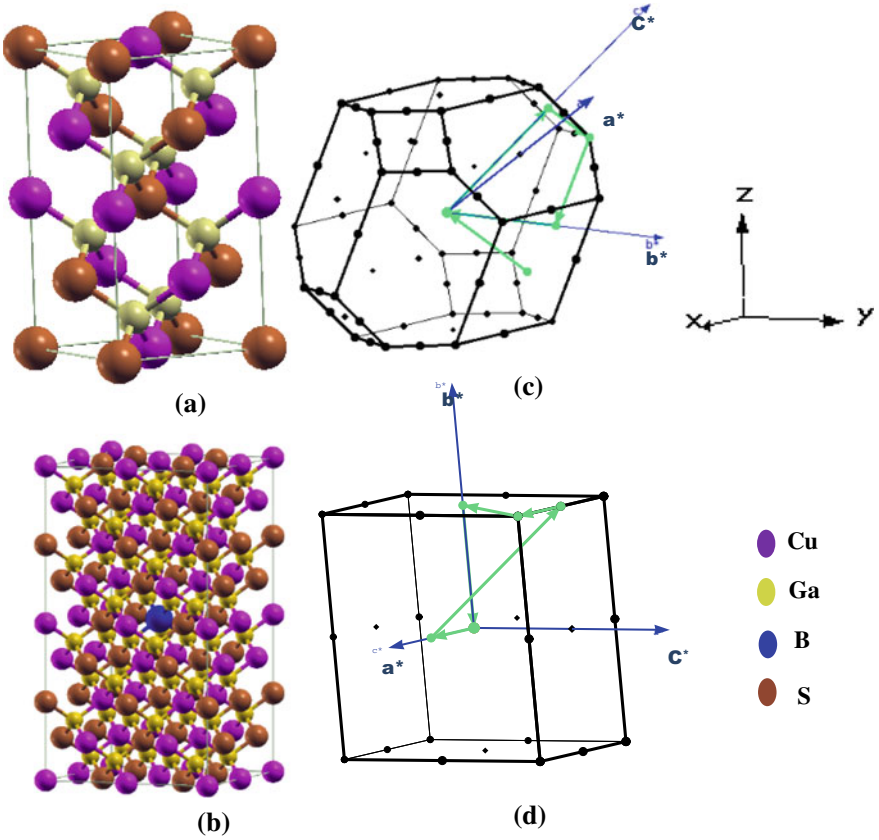


Fig. 1 A Crystal Structure of **a** $CuGaS_2$ and **b** $CuGa_{0.97}Bi_{0.03}S_2$ and Brillouin zone along high symmetry direction of **c** $CuGaS_2$ and **d** $CuGa_{0.97}Bi_{0.03}S_2$

following values of RMT Kmax, lmax, Gmax are 7, 10 and 12, respectively, depicting the brillouin zone's directions of high symmetry. These are expressed along the k points that are taken to be 20. The lattice parameters used to define the structure are a = b = 5.263 Å and c = 10.378 Å. The atomic positions assigned to each element, i.e. Cu (0, 0, 0), Ga (0, 0, 0.5) and S (0.25, 0.25, 0.125) [19].

2.1 Electronic Structure Analysis

The pure $CuGaS_2$ crystal's high symmetrical and directional representation of Brillouin zone along with the state density plots around the energy gap is demonstrated in Fig. 2 (a) and (c), respectively. The value of optical energy gap present between the minima of valence band and maxima of conduction band is reported as 1.12 eV

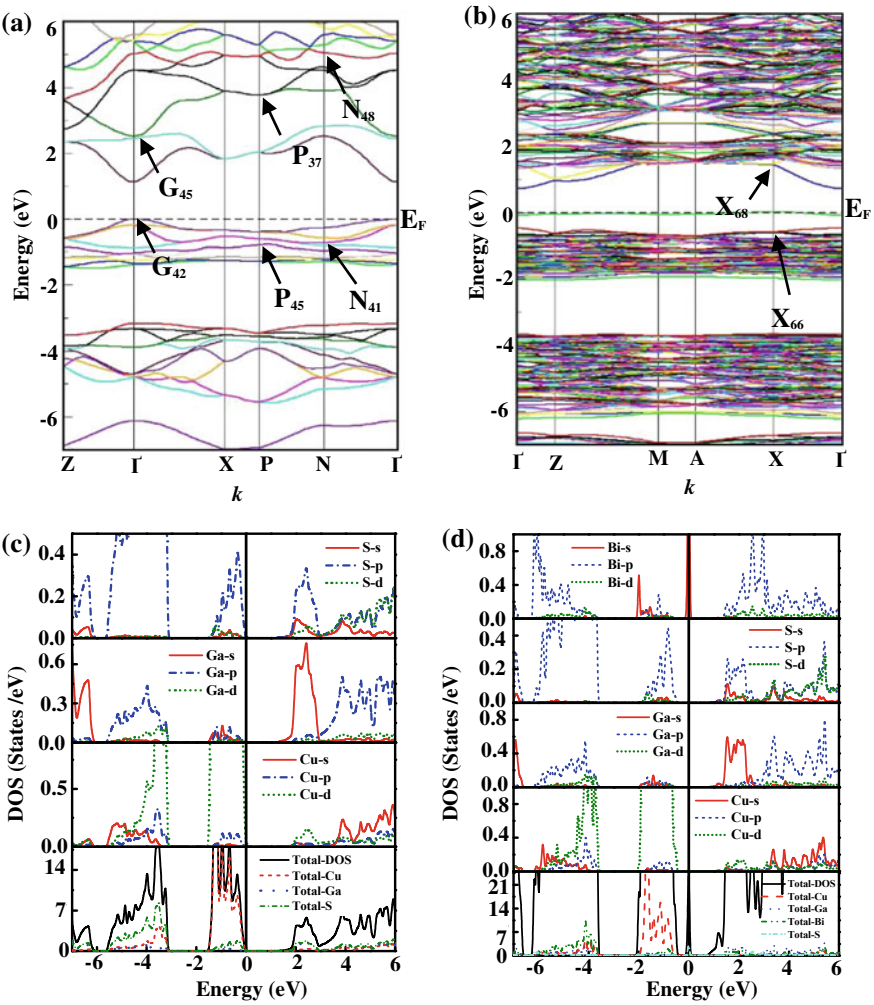


Fig. 2 Band Structure of **a** $CuGaS_2$ and **b** $CuGa_{0.97}Bi_{0.03}S_2$ and DOS spectra of **c** $CuGaS_2$ and **d** $CuGa_{0.97}Bi_{0.03}S_2$

and 0.71 eV for pure and doped $CuGaS_2$, respectively, exhibiting direct band gap at Γ-Γ point. This stands out to be in accordance with the experimental data. The band structure representation of Bi-doped $CuGaS_2$ is given in Fig. 2(b) along with density of states diagram given in Fig. 2 (d). Through these DFT calculations, we can observe the underestimation of band gap in case of doping with Bi. Due to this reason, the earlier reported experimental band gap turns out to be higher than the calculated one [15]. The interband transitions are given in Table 1. Through the analytical study, the DOS spectra reveal the upper-most valence band s, p, and d states of Cu, Ga and S, respectively. Most of the contribution is due to the 3d, 4 s and 3p of Cu, Ga

Table 1 Interband transition of $CuGaS_2$ and $CuGa_{0.97}Bi_{0.03}S_2$

Samples	Peaks	Peak Position (eV)	Dominant Transitions
$CuGaS_2$	A	2.57	$\Gamma'_{42} \rightarrow \Gamma'_{45}$
	B	4.51	$P_{37} \rightarrow P_{45}$
	C	5.62	$N_{41} \rightarrow N_{48}$
$CuGa_{0.97}Bi_{0.03}S_2$	D	2.74	$X_{664} \rightarrow X_{684}$

and S atoms, respectively. $CuGa_{0.97}Bi_{0.03}S_2$ shows transition in three phases with direct band gap as 0.7 eV with two energy sub-bands of 0.389 and 0.007 eV. Thus, $CuGa_{0.97}Bi_{0.03}S_2$ possesses intermediate band gap that arises due to the Bi doping in $CuGaS_2$. The DOS spectra showed that this happens due to main contribution of the 6 s-state of Bi.sub-bands of 0.389 and 0.007 eV. Thus, $CuGa_{0.97}Bi_{0.03}S_2$ possesses intermediate band gap that arises due to the Bi doping in $CuGaS_2$. The DOS spectra showed that this happens due to main contribution of the 6 s-state of Bi.

3 Optical Property Analysis

The transitions seen in electronic structure of solid lead to the study of optical properties using WIEN2k program. The optical absorption value is obtained by the study of band structure of energy. Based on the electronic structural calculation, real component of the dielectric function, integrated absorption coefficient for the parallel component, the absolute value of refractive index of $CuGaS_2$ and $CuGa_{0.97}Bi_{0.3}S_2$, respectively, have been performed and calculated in Table 2, also the description of all these optical parameters is given in Fig. 3. Figure 3 (b) explains the transition peaks in the imaginary component curve which in turn explains the electronic transitions that take places from jumping of electrons from valence to conduction band. Figure 3 (c) explains the absorption coefficient of the pure and Bi-doped $CuGaS_2$ by the analysis of the area under the curve process.

Absorption coefficient is an important parameter as it defines the optical band gap of a sample given by the Tauc law and is studied by the interband transitions near band gap. The value of IAC is decreased by insertion of impurity atom. Figure 3 (d) and 3 (e) explains the reflectivity and refractivity of the pure and Bi-doped $CuGaS_2$. The knowledge of optical constant like refractive index can also

Table 2 Value of optical constants: dielectric constant ε1 (0), refractive index η (0), Integrated Absorption Coefficient (IAC) and reflectivity of $CuGaS_2$ and $CuGa_{0.97}Bi_{0.03}S_2$

Samples	$\varepsilon_1(0)$	η (0)	IAC ($x10^4$ eV/cm)
$CuGaS_2$	6.830	2.601	84.828
$CuGa_{0.97}Bi_{0.03}S_2$	6.149	2.484	81.964

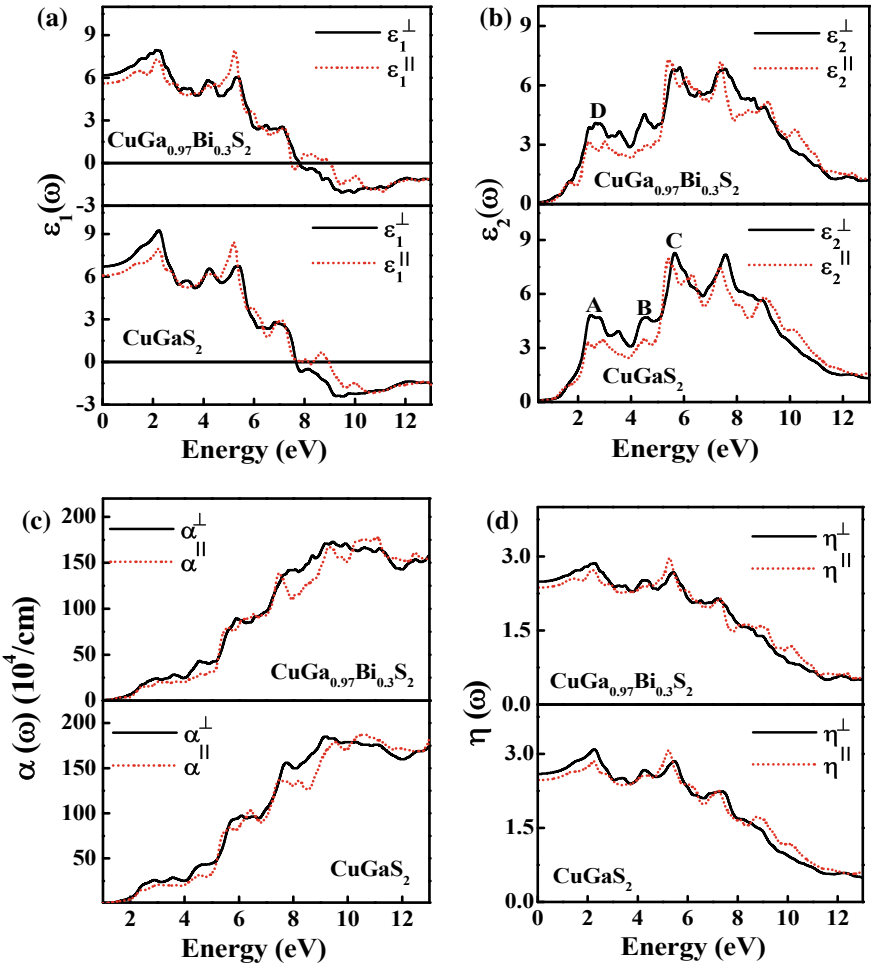


Fig. 3 **a** Real component of Dielectric function; **b** Imaginary component of Dielectric function; **c** Absorption curve; **d** Refractive index plot and of $CuGaS_2$ and $CuGa_{0.97}Bi_{0.03}S_2$

be calculated explaining the system response in respect to system propagation of electromagnetic wave through it.

4 Conclusion

Analysis of the structural, electronic and optical properties of Bi-doped $CuGaS_2$ has been done using the WIEN2k package. The exchange correlation method followed is TB-mBJ providing accuracy in results with band gap as 0.7 eV at 3% of doping.

The calculated result depicts that when we increase the doping of Bi, the band gap is found to be reducing, further carrying forward the conclusive results of one of the experimental works.

Acknowledgements We feel highly obliged by the help provided by Prof. Blaha through the Wien2k code and financial aid by DST-SERB, New Delhi under the project vide grant number EMR/2017/005534.

References

1. Ventura ON, Kieninger M, Irving K (1997) Density functional theory: a useful tool for the study of free radicals. Adv Quant Chem 28:293–309
2. Schwarz K, Blaha P, Madsen GKH (2002) Electronic structure calculations of solids using the WIEN2k package for material sciences. Comput Phys Commun 147:71–76
3. Amusia MY, Shaginyanr VR (2000) Calculations of single particle spectra in density functional theory. Phys Lett A 269:337–342
4. Aray Y, Rodriguez J, Vega D (2002) An implementation of the atoms in molecules theory to the FPLAPW method. Comput Phys Commun 143:199–212
5. Singh DJ (1994) Plane waves, Pseudo potentials and the LAPW Method, 2nd edn. Kluwer Academic, Boston
6. Romeo N (1980) Solar cells made by chalcopyrite materials. Jpn J Appl Phys 19:5–13
7. Arnulf J-W (2012) Practical handbook of photovoltaics. Academic Press
8. Unold T, Kaufmann CA (2012) Comprehensive renewable energy. Elsevier, Oxford, UK
9. Jaffe JE, Zunger A (1983) Electronic structure of the ternary chalcopyrite semiconductors $CuAlS_2$, $CuGaS_2$, $CuInS_2$, $CuAlSe_2$, $CuGaSe_2$ and $CuInSe_2$. Phys Rev B 28:5822–5847
10. Ringeissen J, Regolini JL, Lewonczuk S (1973) Optical properties of CuGaS2. Surf Sci 37:777–785
11. Liuyang Y (2015) 3rd International conference on machinery, materials and information technology applications (ICMMITA 2015), pp 941–945
12. Soni A, Gupta V, Arora CM, Dashora A, Ahuja BL (2010) Electronic structure and optical properties of $CuGaS_2$ and $CuInS_2$ solar cell materials. Sol Energy 84:1481–1489
13. Zongyan Z, Dacheng Z, Juan Y (2014) Analysis of the electronic structures of 3d transition metals doped $CuGaS_2$ based on DFT calculations. J Semicond 35:013002-1-10
14. Zalewski W, Bacewicz R, Antonowicz J, Schorr S, Streeck C, Korzun B (2007) Local order in Mn doped $CuGaS_2$. Science annual report, pp 1037–1038
15. Marcos A, Mascaro LH (2018) Bismuth doping on CuGaS2 thin films: structural and optical properties". MRS Commun 8:504–508
16. Han MM, Zhang XL, Zeng Z (2016) Sn doping induced intermediate band in CuGaS2". RSC Adv 6:110511–110516
17. Marsen B, Klemz S, Unold T, Schock HW (2011) Investigation of the sub-bandgap photoresponse in CuGaS2: Fe for intermediate band solar cells. Prog Photovolt: Res. Appl 20:625–629
18. Chen P, Qin M, Chen H, Yang C, Wang Y, Huang F (2013) Cr incorporation in CuGaS2 chalcopyrite: a new intermediate-band photovoltaic material with wide-spectrum solar absorption. Phys Status Solidi A 210:1098–1102
19. Laksari S, Chahed A, Abbouni N, Benhelal O, Abbar B (2006) First- principles calculations of the structural, electronic and optical properties of $CuGaS_2$ and $AgGaS_2$. Comput Mater Sci 38:223–230

Designing Lightweight S-Box Using Simplified Finite Field Inversion Mapping

Mebanjop Kharjana, Fabiola Hazel Pohrmen, and Goutam Saha

Abstract Emerging areas like the IoT, etc., have computing environments that consist of numerous resource-constraint devices that are interconnected and communicated to each other. These devices need to operate in a secured environment; however, conventional cryptography is not suitable as they have low computational and memory resources. Security for such devices can be ensured by using lightweight cryptography instead. In this paper, a technique was proposed to design smaller S-boxes that can be used in lightweight block ciphers, hash functions, etc. The design technique used in the AES S-box was adopted and simplified in order to make these smaller S-boxes. The proposed S-boxes were compared with those used in the PRESENT cipher and the LUFFA hash function in terms of the different cryptographic properties and parameters. In addition, a change in the nonlinearity value of the proposed S-box was also calculated with reference to that of an AES S-box.

Keywords AES · Lightweight S-box · Inversion mapping · Internet of things

1 Introduction

Lightweight cryptographies are algorithms or protocols that aim to provide security solutions to resource-constraint environments like the Internet of Things (IoT) by using minimal computing and memory resources. These algorithms and protocols use simpler cryptographic components, which make them faster to execute than those of conventional cryptography.

A substitution-box (S-box) is one such cryptographic element, whose main function is to perform substitution. An $m \times n$ S-box provides a nonlinear mapping for input of m bits into output of n bits.

One way to describe such S-box is to view it as a vectorial Boolean function that is mapping from F_2^m to F_2^n. Since a single Boolean function is a mapping from F_2^m to F_2, an S-box can be visualized as a collection of n Boolean functions [1]. These

M. Kharjana (✉) · F. H. Pohrmen · G. Saha
North-Eastern Hill University, Shillong 793022, India
e-mail: meban@nehu.ac.in

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_19

collections of n Boolean functions of m variables are called the coordinates of the S-box [2].

Though resistance to linear and differential cryptanalysis is the main design criteria for a cryptographic S-box, there are other properties [3] that they should also exhibit. These properties are: (i) Balanced, (ii) Bijective, (iii) Strict avalanche criterion (SAC) and (iv) Bit independence criterion (BIC).

2 AES S-Box

An AES S-box is a nonlinear transformation that substitutes an input byte with another byte. It is an 8×8 S-box that involves calculating the multiplicative inverse of the input over GF (2^8). It is designed to resist linear and differential cryptanalysis and also prevent algebraic attacks such as interpolation attacks [4].

It is implemented as a function SubBytes and consists of the following two important steps:

- **Multiplicative Inverse** The AES S-box was constructed based on the concept of inversion mapping under modulus of an irreducible polynomial. An irreducible polynomial is used so that all the elements will have a corresponding multiplicative inverse. The modular reduction also ensures that the result can be easily represented by a byte, whose binary polynomial representation has a degree lower than that of the irreducible polynomial. There are 30 irreducible polynomials under GF(2^8) that can be used as the modulus and the designers selected ($\alpha^8 + \alpha^4 + \alpha^3 + \alpha + 1$).
- **Affine Transformation** Inverse mapping has a very simple algebraic expression, which makes it susceptible to algebraic manipulations. Therefore, an additional invertible step is added by performing an affine transformation in order to strengthen the existing inverse mapping. The affine transformation on the multiplicative inverse i^{-1} is given by the following expression:

$$(\alpha^7 + \alpha^6 + \alpha^2 + \alpha) + i^{-1} \times \left(\alpha^7 + \alpha^6 + \alpha^5 + \alpha^4 + 1\right) mod\left(\alpha^8 + 1\right) \quad (1)$$

Polynomial ($\alpha^8 + 1$) is selected since it is the simplest modulus possible with a degree 8. The multiplication polynomial ($\alpha^7 + \alpha^6 + \alpha^5 + \alpha^4 + 1$) belongs to the set of polynomials that are co-prime to the modulus and is chosen as it has the simplest description. The constant polynomial ($\alpha^7 + \alpha^6 + \alpha^2 + \alpha$) is selected such that there are no fixed point, i.e., $SBOX(x) \neq x$, and no opposite fixed point, i.e., $SBOX(x) \neq \widehat{x}$.

The paper is organized into sections as follows. In Sect. 2, a brief description of the related works is given. Section 3 gives an insight into the methodology and design criteria for the proposed lightweight S-boxes. Their security analysis is also

given in this section. In Sect. 4, the results are discussed. The conclusion of the paper is given in Sect. 5.

3 Related Works

Lightweight S-boxes are simpler S-boxes compared with those used in conventional cryptography. They have a simple description with minimal memory requirements. PRESENT [5], a lightweight block cipher and LUFFA [6], which is a family of hash functions, are examples of cryptographic algorithms that use lightweight S-boxes.

In both cases, the S-boxes are of the size 4 × 4 and perform mapping from F_2^4 to F_2^4. These S-boxes are given in hexadecimal notation in Table 1.

The S-box proposed in [7] is another lightweight S-box that is designed using the same technique as that of AES by performing multiplicative inversion combined with an affine transformation. It is a 4 × 4 S-box that transforms between the finite field GF(2^4) and composite field GF((2^2)^2). The multiplicative inverse uses Euclidean algorithm in the composite field GF((2^2)^2), whereas the affine transformation is done in the finite field GF(2^4). The transformation between the finite field GF(2^4) and composite field GF((2^2)^2) is isomorphic in nature.

The cipher described in ref [8] is a simplified version of the Rijndael encryption algorithm. This version uses a lightweight S-box and defines its algebraic finite field as GF(16) = GF(2)[α]/($\alpha^4 + \alpha + 1$). No particular reason was cited for the selection of ($\alpha^4 + \alpha + 1$) as the modulus.

3.1 Proposed Lightweight S-Box

3.1.1 Design Criteria

The proposed S-box is of the size 4 × 4 that performs multiplicative inverse and affine transformation in the finite algebraic field of GF(2^4). GF(2^4) consists of only 16 polynomials, whose degree is less than 4 and hence can be represented by a nibble. Addition of coefficients is done under modulo-2, whereas multiplication is more complicated and is performed under modulo of a polynomial of degree 4, so that the result lies in GF(2^4). The proposed S-box will be substituting nibble with another nibble and consists of the following two steps:

Table 1 Sample 4 × 4 S-boxes

Input	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
Present	C	5	6	B	9	0	A	D	3	E	F	8	4	7	1	2
LUFFA	D	E	0	1	5	A	7	6	B	3	9	C	F	8	2	4

Multiplicative inverse: The multiplicative inverse of an input nibble is calculated in the finite field GF(2^4). The following are the only available irreducible polynomials under GF(2^4):

$$\alpha^4 + \alpha + 1$$
$$\alpha^4 + \alpha^3 + 1$$
$$\alpha^4 + \alpha^3 + \alpha^2 + \alpha + 1$$

Affine transformation: Affine transformation is a linear mapping in the finite field GF(2^4) that is applied on a multiplicative inverse i^{-1}. The following expression is the representation of this transformation:

$$\left[i^{-1} \times q(\alpha)\right] mod\, r(\alpha) + p(\alpha) \tag{2}$$

where p(α), q(α) and r(α) are polynomial expressions, whose descriptions are given as follows:

- **Modular polynomial r(α)**: The simplest polynomial possible with a degree equal to 4 is selected as the reduction polynomial r(α). It need not necessarily be an irreducible polynomial, i.e.,

$$r(\alpha) = \alpha^4 + 1 \tag{3}$$

- **Multiplication polynomial q(α):** The multiplication polynomial q(α) is selected depending on the modulus r(α), since q(α) should be co-prime to r(α). It is known that multiplication by α can be implemented at the byte level, as a left shift and a subsequent conditional bitwise XOR with 1B [4]. Multiplication by higher powers of α can be implemented by repeat application of the above steps. Therefore, to simplify the multiplication process at the hardware level, q(α) is selected with a simple polynomial as follows:

$$q(\alpha) = \alpha \tag{4}$$

- **Translation polynomial p(α)**: The constant polynomial p(α) performs linear translation. Since the result of the linear translation is a polynomial that belongs to the finite algebraic field GF(2^4), translation is analyzed by starting with the smallest value possible from this finite field and choose the one that causes no fixed point and no opposite fixed point. As shown in Table 2, the required result is given by ($\alpha^4 + \alpha^3 + \alpha^2 + \alpha + 1$) when $p(\alpha) = 1$ and by ($\alpha^4 + \alpha^3 + 1$) when $p(\alpha) = \alpha^2 + \alpha$. However, ($\alpha^4 + \alpha + 1$) will always cause a fixed point and (or) opposite fixed point problem for all possible translation values.

Table 2 Analyzing translation polynomial p(α

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
$\alpha^4 + \alpha^3 + \alpha^2 + \alpha$+1	1	3	E	4	0	D	B	2	9	F	7	C	A	8	6	5
$\alpha^4 + \alpha^3 + 1$	6	4	F	7	A	9	E	B	0	D	1	3	2	5	8	C
$\alpha^4 + \alpha + 1$	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

For each irreducible polynomial, the multiplicative inverse and translation for a given input nibble have an equivalent expression of $[i^{-1} \times \alpha] \bmod (\alpha^4 + 1) + p(\alpha)$ and are calculated using SageMath [9], which is a free and open-source mathematics software system.

Note that though same-size finite fields are isomorphic, whereby the choice of the irreducible polynomial is irrelevant [10], it does make a difference when an affine transformation is involved.

Consequently, only two S-boxes are chosen and are expressed in polynomial form as shown in Eqs. 5 and 6.

$$S1(\alpha) = \left[i^{-1} \times \alpha\right] mod\left(\alpha^4 + 1\right) + 1 \tag{5}$$

$$S2(\alpha) = \left[i^{-1} \times \alpha\right] mod\left(\alpha^4 + 1\right) + \left(\alpha^2 + \alpha\right) \tag{6}$$

3.2 Security Analysis

The proposed S-boxes S1 and S2 are evaluated and compared with the 4 × 4 S-boxes of PRESENT and LUFFA, in terms of different cryptographic properties and parameters as shown in Tables 3 and 4.

Similarly, the proposed S-boxes are also compared with the 8 × 8 AES S-box to get an idea about their differences in terms of the different cryptographic properties and parameters.

Table 3 Comparing cryptographic properties

	S1	S2	PRESENT	LUFFA	AES
Is balanced?	√	√	√	√	√
Is bent?	×	×	×	×	×
Have linear structures?	×	×	√	√	×
Is a permutation?	√	√	√	√	√
Is almost bent?	×	×	×	×	×
Is APN function?	×	×	×	×	×
Is monomial?	×	×	×	×	×

Table 4 Comparing cryptographic parameters

	S1	S2	PRESENT	LUFFA	AES
Linear Branch Number	2	2	2	2	2
Relative Maximal Linear Bias	0.25	0.25	0.25	0.25	0.0625
Differential Branch Number	2	2	3	2	2
Differential Uniformity	4	4	4	4	4
Maximal Difference Probability	0.25	0.25	0.25	0.25	0.015625
Nonlinearity	4	4	4	4	112

4 Results and Discussion

On comparing, the proposed S-boxes, S1 and S2, with other 4 × 4 S-boxes of PRESENT and LUFFA, it is observed that they have cryptographic properties and parameter values that are similar to each other as shown in Tables 3 and 4. The only exception is that the PRESENT S-box has a higher differential branch number value of 3 compared with other S-boxes with value 2. This implies that the proposed S-boxes have lower diffusion power than that of the PRESENT S-box.

On comparing the proposed 4 × 4 S-boxes with the 8 × 8 AES S-box, it is interesting to note they have similar cryptographic properties and even have the same value for a differential branch number. However, there is a significant reduction in their nonlinearity value, which is a 96.42% drop from that of the AES S-box as shown in Table 4.

Consequently, there is an increase in the values of the relative maximal linear bias and maximal difference probability of S1 and S2 from that of the AES S-box. The resistance of S1 and S2 to linear and differential cryptanalysis is, therefore, considerably lowered compare with that of the AES S-box.

Resource-constraint environments like the IoT can improve their efficiency by using simplified cryptographic S-boxes. The memory footprint of lightweight ciphers and hash functions can be minimized by using such components. The simplified S-boxes have lower resistance to linear and differential cryptanalysis as compared with that of the conventional S-boxes. However, it is a compromise that has to be made when they are used in a resource-constraint and real-time environments in order to achieve better computing efficiency.

5 Conclusion

Based on the comparison among the 4 × 4 S-boxes, the proposed S-boxes are as secured as the PRESENT and LUFFA S-boxes. This is validated by the different cryptographic properties and parameters, with the only exception being a minor difference in the differential branch number value.

Based on the comparison of the proposed 4 × 4 S-boxes with that of the 8 × 8 AES S-box, it is observed that they all exhibit cryptographic properties that are similar to each other. However, the decrease in their nonlinearity value by 96.42% is expected considering the huge difference in their sizes.

Therefore, inversion mapping technique can be simplified for designing lightweight S-boxes that are simple, practical and that can be used in resource-constraint environments like IoT.

References

1. Bejo A, Adji TB (2017) AES S-box construction using different irreducible polynomial and constant 8-bit vector. In: IEEE conference on dependable and secure computing. IEEE,, pp 366–369
2. Canteaut A (2016) Lecture notes on cryptographic Boolean functions. Inria, Paris, France
3. Cheung JM (2010) The design of S-boxes. Diss. Sciences
4. Daemen J, Rijmen V (1999) AES Proposal: Rijndael http://csrc.nist.gov/archive/aes/rijndael Rijndael-ammended.pdf
5. Bogdanov A, Knudsen LR, Leander G, Paar C, Poschmann A, Robshaw MJ, Vikkelsoe C (2007) PRESENT: an ultra-lightweight block cipher. In: International workshop on cryptographic hardware and embedded systems. Springer, Berlin, Heidelberg, pp 450–466
6. De Cannière C, Sato H, Watanabe D (2008) Hash function Luffa: specification. Sub-mission to NIST SHA-3 Competition
7. Prathiba A, Bhaaskaran VS (2018) Lightweight S-box architecture for secure Internet of Things. Information 9(1):13. https://doi.org/10.3390/info9010013
8. Musa M, Schaefer EF, Wedig S (2002) A simplified Rijndael algorithm and its linear and differential cryptanalyses. Santa Clara University, Santa Clara
9. SageMath-Open-Source Mathematical Software System. http://www.sagemath.org/. Accessed 09 Dec 2019
10. Aslan B, Sakalli MT, Bulus E (2008) Classifying 8-bit to 8-bit S-boxes based on power mappings from the point of DDT and LAT distributions. In: International workshop on the arithmetic of finite fields. Springer, Berlin, Heidelberg, pp 123–133

Review of Performance of a Single Basin Passive Solar Still

Ranbir Kalita, Parimal Bakul Barua, and Deva Kanta Rabha

Abstract Solar stills are simple devices that can convert impure or brackish water into pure potable water through the process of distillation. It utilizes the thermal energy received in the form of radiation from the sun for carrying out the distillation process. Hence, it is a device that entirely works on the solar energy and as such, solar stills are simple and environment-friendly devices. In this paper, the effects of different parameters on the performance of a single basin passive solar still are discussed along with the ways of improving its efficiency.

Keywords Solar still · Water distillation · Water purification

1 Introduction

The term salinity of water indicates the amount of dissolved salts, like chlorine, fluorine, sodium, potassium, arsenic, etc. in a definite volume of water. Salinity is expressed either in the units of parts per thousand (ppt) or parts per million (ppm). As per IS 10500—2012 (Second revision), the salinity of potable water should not be more than 500 ppm.

The conventional ceramic filters can remove solid impurities from water. But they are actually not capable of decreasing the salinity of water fully. The only possible way of decreasing the salinity of water effectively is by reverse osmosis (RO). As the name indicates it is a reverse of the natural osmosis process, where the solvent moves through a selectively permeable membrane from a higher solute concentration region to a lower one till the concentration of solute on both sides becomes equal.

The osmosis process proceeds spontaneously and does not require any external work input. But in the RO process, the solvent is forcefully moved through the selectively permeable membrane and as such the process of RO requires external work input. In the commercially available RO filters, this external work input is provided by work-consuming devices that work on electrical energy.

R. Kalita (✉) · P. B. Barua · D. K. Rabha
Mechanical Engineering Department, Jorhat Engineering College, Jorhat 785007, Assam, India
e-mail: ranbirkalita@gmail.com

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_20

Non-RO filters available in the market though can decrease the salinity of water to an appreciable limit, yet they cannot make the water germ free. To kill the microorganisms present in the water, the output water from these filters is passed through another device in which ultraviolet rays are made to fall on the water running through the device. This ultraviolet radiation actually kills the microorganisms present in the water. This process is commercially called as e-boiling.

But there is an alternate way of desalinating water and killing the microorganisms at the same time. It is called the distillation process, in which the water is heated above the saturation temperature to change its phase from liquid to vapour. The generated water vapour is then condensed to get the purified liquid water.

The heat energy required for distillation can be provided to the water mass by burning fossil fuel. But this can also be accomplished by utilizing the thermal energy of the sun received in the form of radiation, i.e. the solar thermal energy.

Though water is abundantly available on the earth, yet approximately 3 percent is available as the usable water. Remaining 97 to 98 percent are either present in the oceans as saline water or as ice burgs in the arctic regions. The groundwater that is available in the Sea Islands or coastal regions is also saline in nature and it may be a bit less saline than the seawater. In these regions, desalinating the water is a must to make it potable. Also for non-coastal regions, if the groundwater is polluted, it may be due to various reasons, its purification is utmost necessary to make it potable and reduce health risks.

2 Working Principle of a Simple Solar Still

A simple solar still works just similar to the natural rain-water cycle. As displayed in Fig. 1, raw water is placed in the still basin. The solar radiation incident on the transparent glass cover passes through it and falls on the water contained in the basin, heating it up to the evaporation point. The vapour molecules leaving the water

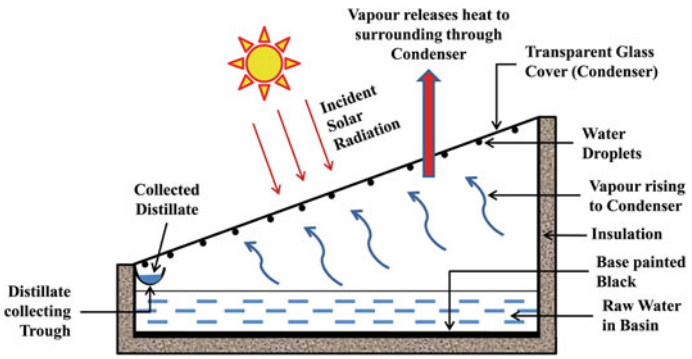


Fig. 1 A typical single basin passive solar still in operation

surface leave all the impurities behind in the basin, rise upward and touches the glass cover on its inner surface. As the outer surface of the glass cover is exposed to the ambient air, vapour molecules release their latent heat to the outside air, and tiny water droplets are formed that adheres to the glass surface. Due to the inclination of the glass cover at a certain angle, these tiny water droplets trickles down to the lowest end, and gets collected on a distillate collection trough, which is connected to an outlet pipe.

During the process of distillation, the water level in the basin depletes continuously. To compensate this, raw water can be fed into the basin, either manually at a definite interval of time or continuously by attaching water feed mechanism.

3 History of Solar Stills

As per the documental evidence, solar stills have been in use since sixteenth century. The first large size solar still plant was constructed for supplying drinking water to a mining community at Las Salinas in the Northern Chile in the year 1872 [14]. The basins of the stills were made of woods, having a total area of 51,000 sq. feet, and was found to work with an efficiency of 30% [23].

Until World War II, the solar distillation did not receive much attention from researchers. But during the WW-II, a handsome amount of research was done on the solar stills, to make them more efficient and portable in size, so that the sailors who live in the sea continuously for many days can make use of it and produce drinking water from the saline seawater. Approximately 2,00,000 inflatable solar stills were manufactured for US Navy and were kept in the life rafts for being used by the sailors at the time of emergency. The efficiency of these portable devices was found to be 50–60% [23].

4 Types of Solar Stills

Solar stills can be broadly classified as passive solar stills and active solar stills. If no external mean is used to accelerate the heat transfer processes taking place in the still, then the still is called as a passive solar still. On the other hand, to accelerate the heat transfer processes and as such to increase the rate of output, additional components can be attached to a solar still. In such a case, the still is called as active solar still [5]. .

As shown in Fig. 2, a Flat Plate Collector (FPC) has been attached to a single basin solar still to increase the heat receiving area of the still. Until an electric water pump is attached to the system, the solar still is called as passive solar still because the flow of water throughout the whole system is taking place solely because of the convection current. But as soon as the electric pump is attached to the system to

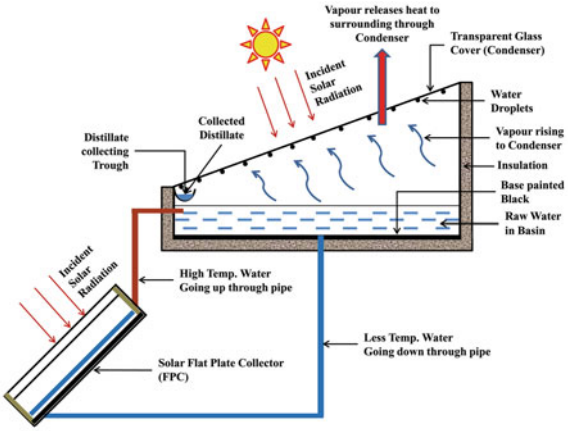


Fig. 2 Passive solar still

accelerate the water flow within the system, as shown in Fig. 3, the solar still is now called as an active solar still.

Solar stills can also be classified as single-effect solar stills and multi-effect solar stills. In single-effect solar stills, the water vapour during condensation releases the latent heat directly to the ambient air. But in multi-effect solar stills, also called as multi-stage solar stills, the released heat during condensation in one stage is utilized to increase the heat content of water in the next stage, and this process continues

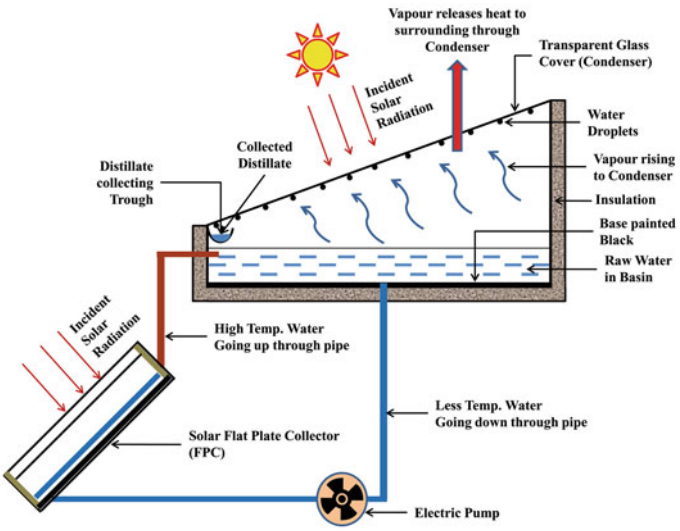


Fig. 3 Active solar still

till the last stage. This reutilization of the released heat during condensation makes the multi-effect solar stills more efficient than single-effect solar stills [17].

Hogan et al. [7] tried a technique called the Membrane Distillation (MD). MD plants are different than the conventional solar distillation plants because of their capacity to recover large portions of the latent heat of vaporization using conventional heat-exchange devices.

5 Constructional Elements of a Solar Still

Considering a single basin passive solar still, the main components of the still are:

(a) The Basin of the still
(b) Transparent Condensing cover
(c) Condensate collection trough

(a) The Basin of the Still
The basin of the still can be constructed by wood, ceramic material or sheet metal. Though woods are cheaper construction materials, yet the problem of using woods for constructing the basin is that it requires frequent maintenance and has a short life span. Besides that, water sipping through the joints is another main problem with the basins made of woods.
Sheet metal can also be used for constructing the basin. If the joints are properly welded, water will never percolate through the basin. Sheet metal requires less maintenance than the wooden constructions; just the inner and outer surfaces of the basin have to be painted to prevent corrosion. Sheet metal construction is a bit expensive mean than the woods. But this increase in cost is justified by an increase in the life span of the basin and less requirement of maintenance efforts.
Using ceramic materials is an alternate mean to construct the basin. An advantage with the ceramic materials is that it can be moulded to the required shape of the basin making it a continuous piece of construction. Also the ceramic materials are not prone to corrosion; hence require least maintenance efforts as compared to sheet metal and wooden constructions.

(b) Condensing Cover
The transparent condensing cover can be made either using glass or plastic. The refraction properties of the condensing cover should be good enough, so that maximum amount of solar radiation can pass through it with minimum loss of incident radiation due to reflection [13].

(c) Condensate Collection Trough
This trough is fitted below the lowest end of the condensing cover and collects the trickling condensed water droplets. As this trough is fitted within the still, it remains in continuous contact with water and exposed to high temperature during the period of operation. As such the material of the trough should be corrosion resistant and must be able to withstand high temperatures.

6 Factors Affecting the Performance of a Solar Still

The main factors that affect the performance of solar stills include:

1. Inside area of the basin.
2. Solar radiation absorption properties of the inner basin lining.
3. Loss of heat through the side walls and bottom of the basin.
4. Depth or height of water in the basin.
5. Area of the condensing cover.
6. Adhesion properties of the condensing cover
7. Inclination of the condensing cover.
8. Thickness of the condensing cover.
9. Reflection losses of incident solar radiation due to reflection at the surface of the condensing cover.
10. Gap between the free surface of water in the basin and condensing cover.
11. Loss of heat due to leakage of water vapour from the still.
12. Facing direction of the solar still.

It is an obvious fact that more the heat collection area of the basin more is the amount heat received. As such, the inner area of the basin has to be large enough to collect more incident solar radiation. But the selection of the basin area is limited by the size of the still and other economic considerations. To increase the heat absorption, the inner lining of the basin is always painted in black. The selection of the paint should be made keeping in mind that the paint should not react chemically with the water or the salts present in the water. Earlier researches reveal that even a complete black basin lining cannot prevent 4% loss of the total incident energy [23]. Rajvanshi [18] has tried to increase absorptivity of solar radiation of the basin water by adding water-soluble dies of black, green and red colour whose boiling points are higher than the water in the basin. An increase in output has been observed and maximum output is obtained when black dye is added to the water in the basin. Okeke et al. [14] have studied the effects of adding charcoal and coal into the basin water of a solar still. Charcoal and coal both increases the solar energy absorption in the basin water and thus minimizes the reflection losses. As such the output and efficiency of the still are found to increase. Besides reducing the reflection losses, addition of coal and charcoal also increases the surface area for evaporation.

As most of the heat is absorbed by the inner basin lining, heat loss through the bottom of the basin is obviously maximum as compared to that through the side walls of the basin. For this reason, the critical thickness of insulation at the bottom of the basin should be more than that of the side walls. Tenthani et al. [24] have experimented by painting the inner surface of the side walls with white paint, so that the amount of solar radiation falling on the inner surface of the side walls gets reflected back to the water in the basin. They have observed an increase in the amount of distillate output and confirmed the results through statistical analysis.

One of the main parameters that controls the rate of condensation of the vapour molecules is the surface area of the condensing cover. Obviously, larger the area

available for condensation more will be the amount of condensate produced per unit time. Various shapes of the condensing cover have been tested by many researchers like semi-circular or oval, elliptical, etc. But the most effective one is found to be a flat area inclined at some angle, also called the angle of tilt of the condensing cover [9]. Some researchers have tried to minimize the cost of construction of the still by using cheap plastic materials as condensing cover. But, the problem with plastic materials is that, plastic deforms when exposed to high temperature for a period of long duration. Also the phenomenon of fogging occurring at the inner surface of the plastic cover prevents the solar radiation to pass through it due to increased reflection and scattering losses [23] and as such the performance of the still decreases.

Glass is less susceptible to deformation and can withstand higher temperature than cheap plastic materials [9]. Even the window grade glass materials show good adhesion properties than the plastic materials of the same price range. The amount of condensed water droplets falling back to the basin is also less in case of glass than in the case of plastic materials.

Earlier researches have revealed that lower depth of water in the basin shows more rate of evaporation during the peak sun shine hours but decreases rapidly till the end of the day. But higher depth of water in the basin, though initially shows less rate of evaporation, yet the phenomenon of evaporation continues to take place till the end of the day due to heat storage effect in the water mass [8, 10, 16].

The inclination of the flat condensing cover is also important in increasing the efficiency of the still. As found by some researchers, the angle of tilt equal to the latitude of the geographic location increases the performance of the still. Others believe that tilt angle approximately equal to 10 degrees gives an appreciable performance irrespective of the geographic location [3, 9]. It is also found in the literature that the tilt angle should always be less than 50° [23]. Some authors also believe that decrease in tilt angle increases the performance of the still, whereas others have found that the angle of inclination of the glass plate has no effect on the output [22]. The optimum tilt angle (δ) with respect to a geographic location can also be given by the following relation [19]:

$$\delta = \text{latitude} \pm 15^{\circ}$$

where 15° is to be added to the latitude if the still design is intended for use during the winter months and subtracted from the latitude if the still is intended for use during the summer months.

Since a large amount of heat transfer is taking place through the condensing cover during the process of condensation, its thickness should be selected carefully. Thinner is better, but thinner glass covers are more susceptible to form cracks at higher operating temperatures. To avoid such damage to the condensing cover, toughened glass could be a better option when high operating temperatures are expected. Also the glass material selected for the condensing cover should have good refraction properties; otherwise some of the incident radiation may be lost due to the reflection

effect at the cover surface. This kind of reflection loss is less in thin glass covers as compared to the thicker ones.

Some researchers have also found that more is the gap between the free water surface in the basin and the inner surface of the glass cover, more will be the rate of evaporation due to higher temperature gradient between the glass cover and water surface as compared to a lesser gap between the two [5, 8, 16, 22]. Some authors have also tried to increase the output of the still by blowing ambient air inside the still. Kianifar et al. [10] have observed an increase of 15%–20% in the daily output from the still when the still is fitted with a fan to blow ambient air within the still [8, 10]. Deniz [4] has advised to use Shade Black Cloth Wick spread over the entire area of the basin to increase the distillate output and achieve higher thermal efficiency in case of an inclined system for water distillation.

One of the main factors that decreases the efficiency of a solar still is the loss of heat with the leaking water vapour through the gaps between the glass cover and basin. Hence, the entire still should be made airtight as far as possible using gaskets capable of withstanding high temperatures.

Unlike solar parabolic trough collectors, solar stills do not always require continuous tracking of the sun. For appreciable results, the still should be placed facing the North-South direction during the whole day of operation [25].

Mahdi and Smith [12] have noticed that the distillate production rate is higher in Winter than that in Summer. They thought the reason to be lower proportion of diffuse radiation on clear winter days than on clear summer days.

7 Means to Improve the Performance of a Single Basin Passive Solar Still

The efficiency of a simple passive solar still can be improved by:

1. Increasing the heat collection area
2. Increasing the heat transfer rate during condensation
3. Decreasing the loss of heat through side walls and bottom of the basin, pipings and minimizing vapour leakage.

The constructed size of the still imposes a constraint on the available basin area for heat collection. This area for heat collection can be increased by attaching an external solar FPC to the still through proper piping arrangement [25]. The water from the lower end of basin enters the FPC and after receiving the thermal energy, the water leaves the collector and enters the still basin again. The flow of water between the FPC and the still takes place spontaneously due to convection current when the still is placed at a higher level from the ground than that of the FPC. In case of active solar stills, an electric pump is attached to the system to accelerate the water flow within the system.

To enhance the heat transfer during condensation, either air or water can be made to flow over the condensing cover on its outer surface [1, 15] using an air blower,

fan or an electric pump [1]. Some researchers claim that when air is blown over the condensing cover, the amount of output increases [1, 18], whereas others have found that this process has a negative effect on the output. Some other researchers even claim that wind velocity does not have any prominent effect on the output of the still [22]. Arunkumar et al. [1] have observed an increase in the output from the still when the condensing cover is cooled by flowing water over its outer surface. But as passive stills are concerned, work-consuming devices cannot be attached to it. Instead, the outer surface can be cleaned manually at definite interval of time to keep it dust free, so that the incident radiation can pass through it easily [2]. The ratio of the evaporating surface area to the solar energy collection area is also an important parameter. With an increase in this ratio, the output of distilled water also increases [11].

To reduce the negative effects of heat loss from the solar still, sufficient amount of insulation should be provided to the side walls and bottom of the basin and also to the piping arrangements. The joints in the still should be made airtight by using proper gaskets. But even with perfect insulation, it was earlier found that the efficiency of a single basin still working in passive mode cannot be more than 74% [23] and also that the efficiency of a solar still lies between 50% to 60% [22]. It is also found in the literature that the maximum efficiency of a simple basin type solar still equals nearly to 30% only [21].

Shashikanth et al. [20] have commented that PCM (Phase Change Materials) can be used as heat storage medium and it is a promising mean to increase the efficiency of the desalination process and is economically feasible too. Such increase in the yield from a solar still has also been observed by Gugulothu et al. [6].

8 Conclusion

A solar still could be a very effective mean to produce potable water if the factors affecting its performance are minimized and the parameters that determine its efficiency are set at the optimum levels with respect to the geographic location where the still is in operation. The initial installation cost of a solar still may seem to be high as compared to the daily output it gives, but it is justified by the fact that, once the still is designed and installed properly, it does not require any further investment for its operation, apart from occasional maintenance and cleaning. Besides the factors already mentioned, the number of total sunny days in a year and the intensity of solar radiation on a geographical region are obviously the main factors that determine the performance of a solar still in that region. But despite this fact, the performance of a solar still can be evaluated under local environmental conditions and design parameters can be changed to increase its performance further.

Acknowledgements The authors are thankful to the TEQIP-III cell of Jorhat Engineering College for timely funding and kind co-operation received time-to-time.

References

1. Arunkumar T, Jayaprakash R, Ahsan A, Denkenberger D, Okundamiya MS (2013) Effect of water and air flow on concentric tubular solar water desalting system. Appl Energy 103:109–115
2. Cooper PI (1973) The maximum efficiency of single-effect solar stills. Sol Energy 15:205–217
3. Cooper PI, Read WRW (1974) Design philosophy and operating experience for Australian solar stills. Sol Energy 16:1–8
4. Deniz E (2012) An investigation of some of the parameters involved in inclined solar distillation systems. Wiley Online Library, pp 1–5. https://doi.org/10.1002/ep.11612, (wileyonlinelibrary.com)
5. Gugulothu R, Somanchi NS, Reddy KVK, Gantha D (2015) A review on solar water distillation using sensible and latent heat. Proc Earth Planet Sci 11:354–360
6. Gugulothu R, Somanchi NS, Vilasagarapu D, Banoth HB (2015) Solar water distillation using three different phase change materials. In: Materials Today: Proceedings 2, 4th International conference on materials processing and characterization, pp 1868–1875
7. Hogan PA (1991) Desalination by solar heated membrane distillation. Desalination 81:81–90
8. Hollands KGT (1963) The regeneration of lithium chloride brine in a solar still for use in solar air conditioning. Commonwealth Sci Ind Res Org Eng Sect Melbourne, Aust 7(2):39–43
9. Howe ED, Tleimat BW (1974) Review paper twenty years of work on solar distillation at the University of California. Sol Energy 16:97–105
10. Kianifar A, Heris SZ, Mahian O (2012) Exergy and Economic Analysis of a Pyramid-shaped Solar Water Purification System: Active and Passive Cases. Energy 38:31–36
11. Kiatsiriroat T, Bhattacharya SC, Wibulswas P (1987) Performance analysis of multiple effect vertical still with flat plate solar collector. Solar Wind Technol 4(4):451–457
12. Mahdi JT, Smith BE (1994) Solar distillation of water using a V-trough solar concentrator with a wick-type solar still. Renew Energy 5(1):520–523
13. Moustafa SMA, Brusewitz GH, Farmer DM (1979) Direct use of solar energy for water distillation. Sol Energy 22:141–148
14. Okeke CE, Egarievwe SU, Anmalu AOE (1990) Effect's of coal and charcoal on solar-still performance. Energy 15(11):1071–1073
15. Pandey GC (1984) Effect of dried and forced air bubbling on the partial pressure of water vapour and the performance of solar still. Sol Energy 33(1):13–18
16. Patel P, Kumar R (2016) Comparative performance evaluation of modified passive solar still using sensible heat storage material and increased frontal height. Proc Technol 23:431–438
17. Prasad B, Tiwari GN (1996) Analysis of double effect active solar distillation. Energy Convers Manag 37(11):1647–1656
18. Rajvanshi AK (1981) Effect of various dyes on solar distillation. Sol Energy 27:51–65
19. Saikia A, Bhuyan NJ, Sarma TP, Uddin P, Sarma D, Barua PB (2018) Design, fabrication and performance evaluation of solar water still. Int J Eng Res Technol (IJERT) 7(9):152–157. ISSN:2278–0181, http://www.ijert.org
20. Shashikanth M, Khadka B, Lekhana Y, Kiran PMS, Alaparthi N, Veerammneni S (2015) Solar water distillation using energy storage material. Proc Earth Planet Sci 11:368–375
21. Sodha MS, Kumar A, Tiwari GN, Tyagi RC (1981) Simple Multiple Wick Solar Still: Analysis ad Performance. Sol Energy 26:127–131
22. Soliman SH (1972) Effect of wind on solar distillation. Sol Energy 13:403–415
23. Telkes M (1953) Fresh water from sea water by solar distillation. Ind Eng Chem 45(5):1108–1114
24. Tenthani C, Madhlopa A, Kimambo CZ (2012) Improved solar still for water purification. J Sustain Energy Environ 3:111–113
25. Zaki GM, Al-Turki A, Al-Fatani M (1992) Experimental investigation on concentrator-assisted solar-stills. Int J Solar Energy 11:193–199

Flat Plate Solar Thermal Collectors—A Review

Dhrupad Sarma, Parimal Bakul Barua, Deva Kanta Rabha, Nidhi Verma, Soumyajyoti Purkayastha, and Sudipta Das

Abstract Solar thermal flat plate collectors (STFPC) are the mainstay in modern household solar thermal applications and in industrial sectors requiring low-temperature applications. They are easy to design and manufacture and are available in many forms. STFPCs are used in water heating, crops drying, timber seasoning, space heating and solar absorption/adsorption refrigeration systems. It is one of the most widely used and studied solar collectors. In this paper, an attempt has been made to review research works on improving the thermal performance of the solar flat plate collector. Detailed discussions have been presented on the various losses of the STFPC and methodologies suggested by the researchers to reduce the losses as well as improving the thermal performance.

Keywords Solar thermal flat plate collectors · Front heat loss · Collector efficiency

1 Introduction

The sun is an unlimited and environmentally friendly source of energy. As per the World Radiation Centre (WRC), the solar energy incident on, outside the earth's atmosphere is 1367 W/m^2 with 1% uncertainty. Most of this radiation energy comes in the wavelength range of 0.3 to 3 micrometre [1]. A part of this radiation get scattered in the earth's atmosphere (diffused solar radiation) and the remaining part directly falls on the earth's surface as direct solar radiation. This abundant energy through periodic in nature can be trapped using solar collectors and is converted into usable forms of energy such as heat or electricity [2]. Solar collectors are noiseless, reliable and low maintenance systems that do not produce any toxic or radioactive waste [3].

The earliest examples of solar thermal energy harvesters are solar water heater, which was black colour-painted box filled with water. A major drawback was that it had inherently high thermal mass. Apart from heating water, solar thermal energy is

D. Sarma (✉) · P. B. Barua · D. K. Rabha · N. Verma · S. Purkayastha · S. Das
Mechanical Engineering Department, Jorhat Engineering College, Jorhat 785007, India
e-mail: thinkdpd@rediffmail.com

P. K. Bora et al. (eds.), *Emerging Technologies for Smart Cities*, Lecture Notes in Electrical Engineering 765, https://doi.org/10.1007/978-981-16-1550-4_21

also employed in space heating, water desalination, crops drying, power generation etc. However, in high-temperature applications such as solar thermal power generation, the application of solar thermal flat plate collector (STFPC) is limited because of its low output temperature. Gaur and Tiwari [4] developed two different types of solar distillation systems: passive solar still and active solar still. In the former, solar radiation heats the water directly and in the later one, additional thermal energy is used to heat the water in the basin. Ramani, et al. [5] designed a counter flow solar air collector with double-pass having porous material in the second air passage and observed that the thermal performance is 25% higher than that of without porous material and 35% higher than that of a single-pass collector. Michaelides and Eleftheriou [6] presented the performance characteristics of a solar water heating system of 3 m^2 flat plate collector area with a 68 L capacity tank and observations were done over a duration of 2 years under real weather conditions. It was also observed that the maximum exergy efficiency is 6.27% in a double-pass corrugated plate solar air collector having optimum values at 12.2 mm corrugation height, 1.79 m^2 heater area and 0.005 kg/s air flow [7].

To convert solar radiation directly into electricity, photovoltaic (PV) cell collectors are used. Change in ambient temperature affects the efficiency of PV collectors. As temperature increases, the performance of PV collectors decreases [8]. Huang et al. [9] reported that 80% of the absorbed energy is discarded to the environment after the electric energy conversion in PV collectors and suggested that a better thermal efficiency is obtainable by using PVT (photovoltaic–thermal) collectors made from corrugated polycarbonate panel. It is also found that the performance of PV increases as with the working fluid mass flow [10]. Thus, to improve the overall thermal efficiency using the unused solar radiation, hybrid solar collectors have been designed by merging the photovoltaic (PV) cell and thermal collector, which is also termed as PVT (photovoltaic–thermal) collector. This arrangement has an added benefit of providing a cooling effect to the PV cells, further improving its efficiency. Chauhan et al. [11] designed a hybrid solar air collector having an absorber plate, a transparent glass glazing cover, PV module, air duct and DC fan. Omrany and Marsono [12] demonstrated the use of passive strategies especially in the building sector enhancing sustainability measures by reducing building's negative impacts besides optimizing its energy performance. It is shown that applying laminate glazing with a spectrally selective coating reduces radiative heat loss and improves the thermal output of glazed PVT collectors. However, due to the lower solar radiation for PV, electricity generation slightly get reduced [13]. Moreover, the efficiency of liquid-based PV collectors is found to be in the range of 5 to 15% more than traditional ones [14]. Balaji et al. [15] designed a solar linear Fresnel reflector of an area of 154 m^2 and analyzed it using two different profiles of secondary concentrator. Kraemer et al. [16] developed a new type of thermal collector called solar thermal electric generators with the efficiency of 4.6% under 1 kW/m^2 solar radiation. It is also reported that, non-imaging solar collectors, with efficiencies ranging between 9 and 40%, can be used for low and medium heat applications [17].

Solar thermal systems generally consist of two subsystems: Solar collectors and thermal energy storage components [18]. In low-temperature solar thermal systems,

the STFPC are most commonly used for its low cost and easy operability. Ayompe and Duffy [19] designed a 4 m^2 flat plate collector and the investigators monitored the collector for 1 year under various temperate climatic conditions. The researchers reported 32.2% average solar fraction, 45.6% collector efficiency and 37.8% system efficiency. However, certain factors including dust and bird droppings hinder solar insulation [20]. Yang, et al. [21] experimentally determined that decreasing thermal resistance in the air flow channel is the most effective method to enhance thermal efficiency. Juanico et al. [22] found that the STFPC can be improved using thick plastic hoses with large diameters because of higher thermal inertia for more storage capacity. It was suggested to keep the storage tank at an elevation of around 30–60 cm higher than the collector top, to prevent reverse circulation in absence of sunshine [23]. The evacuated tube solar collector is also shown to have improved performance over the STFPC [24]. Leon and Kumar [25] found that solar absorptivity, collector pitch and air flow rate have the strongest effect on the collector thermal effectiveness and efficiency. Addition of passes also improves the performance of collectors in solar air heating systems. Prajapati et al. [26] found that the efficiency of double-pass solar air heaters is 34–45% higher than that of the single-pass solar heaters. In case of the building mounted collectors, Pacheco et al. [27] found that certain parameters such as tilt and building orientation affect energy efficiency. In addition to that, gap between absorber collectors, number of passes, area of the collector covered and solar cells also influences the collector efficiency [3].

2 Components of the Solar Thermal Flat Plate Collector

STFPC is made up of the following components: top glazing cover, the absorber/collector plate, fluid circulating tubes and back and side insulation. The components of a typical solar water heater STFPC and its various heat losses are shown in Fig. 1.

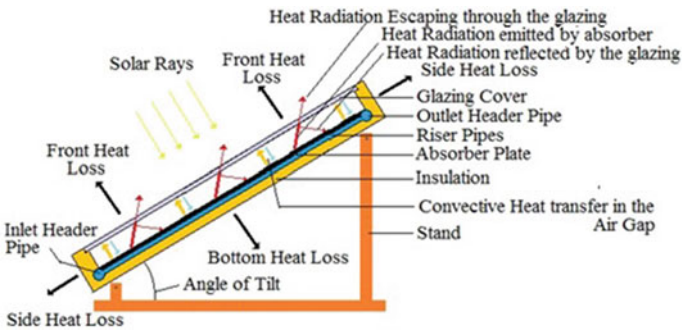


Fig. 1 Components of a typical solar water heater STFPC and its various heat losses

2.1 Top Glazing Cover

The function of the top gazing cover is to allow solar radiation to come in but to serve as an opaque medium for the heat radiated by the hot absorber plate. It also reduces the heat loss by convection from the absorber acting as a barrier between the cold atmospheric air and the hot air above the absorber. The most widely used glazing material is translucent low iron glass [28]. The distance between the glazing cover and absorber plate is termed as air gap that is found to be optimum around 20–30 mm [13].

2.2 Absorber Plate

The function of the absorber plate is to absorb the incoming solar radiation and convert it into heat. In general, flat plate collectors have $2m^2$of collector area and thickness of absorber plate is around 8 mm [29, 30]. The material used for the collector plate is aluminium or copper [31]. Generally the absorber plate is painted black to enable it to absorb incoming solar radiation.

2.3 Fluid Circulating Tubes

These tubes are attached longitudinally to the backside of the absorber that transfers heat to the fluid flowing through it. These tubes are also referred as riser tubes in solar water heating systems. The tubes have an outer diameter of 18 mm and inner diameter of 16 mm and the distance between the consecutive tubes is 100 mm [32]. The ends of these tubes are connected to header pipe at both ends. The header pipe diameter varies from 20 mm to 25 mm [33]. The tubes are mostly made of copper due to its high conductivity [28].

2.4 Back and Side Insulation

Insulation is important for minimizing heat losses. STFPCs are insulated to minimize the heat loss by conduction and convection through the back and side of the box. The insulation must be stable in all respects within the working temperature of the collector. Glass wool and mineral wool are the most commonly used insulation materials [19, 34, 35]. The thickness of back and the side insulation are kept around 80 mm and 40 mm, respectively [36].

3 Heat Losses in a Solar Thermal Flat Plate Collector

Heat loss from an STFPC can be divided into three groups as described below:

3.1 Top Heat Loss

Heat loss through the top or front side of an STFPC is the largest among the other two heat losses. It constitutes around 75% of the total heat loss of a single-glazed STFPC [37]. So it has a very significant effect on STFPC performance. The top heat loss occurs partly due to the convection in the air gap between the absorber plate and the glazing, and partly due to the heat transfer by conduction through the air inside the air gaps. The conduction heat transfer also occurs at the glazing and its frame. For a single glazing cover, the heat loss is in the range of 5 to 10 W/m^2 K, which can be reduced further to about 4 W/m^2 K using two glazing covers [38]. Honeycomb materials also play a role in reducing the convective heat loss in the air gap. The optimum bottom gap thickness of Honeycomb material is found to be 3 mm [39]. Chan et al. [35] suggested the use of antireflective glass covers. It was found that TiO_2, SiO_2 and Al_2O_3 allowed good control and improvement in the optical properties of a thin film combined with a high rate of deposition [40]. Dudita et al. [41] also suggested coloured TiO_2 thin films, which increases architectural acceptance. It is also found that the convective heat transfer through the air gap can be minimized by evacuating and optimizing the air gap between the absorber and the glazing cover. In case of double-diffusive natural convection in a triangular-shaped solar collector, with an increase in Buoyancy ratio and Rayleigh number, the heat and mass transfer increase [42].

Another important contributor to this heat loss is the radiation heat loss by the absorber. It can be reduced by applying selective coating having high absorbance to emittance ratio. Thermal losses were found to be reduced with the use of spectrally selective and transparent low emissivity coatings [43]. The solar absorber material may be a transition metal oxide such as aluminium coated with *eta* plus with 95% absorbance and 5% emittance [19]. Chan et al. [35] suggested transition metal Cu and Al as absorber, with selective Tinox coating. However, when copolymer absorbers were tested, it was found to have good UV light protection, physical strength and chemical stability [44]. Low emittance-coatings based on silver ($\epsilon = 0.13$ and $\tau = 0.79$) is also used for flat type, liquid PV/T collectors [43].

3.2 Bottom Heat Loss

The heat leakage through the backside of the STFPC is termed as the bottom heat loss, which is mostly conduction heat transfer through the insulating materials. There

is no air gap between the absorber plate and back insulation in a well-constructed STFPC; therefore, the convection heat transfer is negligible. The Bottom Heat Loss increases with an increase in the temperature difference between the absorber and atmosphere. For a well-insulated STFPC the bottom heat loss coefficient is around 0.69 W/m^2 K [38].

3.3 Side Heat Loss

The heat lost by the side of the STFPC also is due to the conduction heat transfer by the insulation materials. The side heat loss coefficient is also around 0.69 W/m^2 K. As the side area of a STFPC is much less than the bottom area, the total side heat loss is not very significant.

All these heat losses are also governed by the absorber temperature. The higher the difference between the absorber and atmospheric temperature, the higher are the heat losses and consequently the lower is the system efficiency. The absorber temperature depends largely on the efficiency of the fluid flow system that carries out heat from the absorber to the storage. Thus, the circulating fluid in the tubes plays an upmost important role in the heat transfer. The back wall temperature can be reduced by increasing the mass flow rate [45]. Cardinale et al. [46] recommended the use of water and 30% mono-propylenic glycol for better heat transfer. Nanofluids are also found to be helpful in enhancement of the thermal and optical properties [47]. Al_2O_3 water nanofluid is found to have better performance than the pure water [48]. Similar improvements are also found by using CuO-based nanofluid [49]. Nasrin and Alim [50] studied the behaviour of water-based nanofluid on free convective boundary layer inside a solar collector and found that the heat and mass transfer rates are best enhanced by Ag/CuO nanoparticles with highest Nr (buoyancy ratio) and Sc (Schmidt number). The investigators reported that, when the water is replaced with Single-Walled Carbon Nanotube Technology (SWCNT)-based nanofluid, the heat transfer coefficient improves by15.33% and the entropy generation reduces by 4.34% [51].

In summary, all the three types of heat losses as described above are shown with their contributing factors in the accompanied Ishikawa cause–effect diagram (Fig. 2).

A comprehensive comparison of the efficiency obtained by various researchers and their test setup is presented in Table 1. Table 2 shows comprehensive comparison of the improvement in reduction of heat loss and other factors by various researchers.

4 Performance Equations Related to STFPC

The following equations are used for performance evaluation of a STFPC:

The amount of solar radiation received by the collector is calculated using Eq. (1) [61],

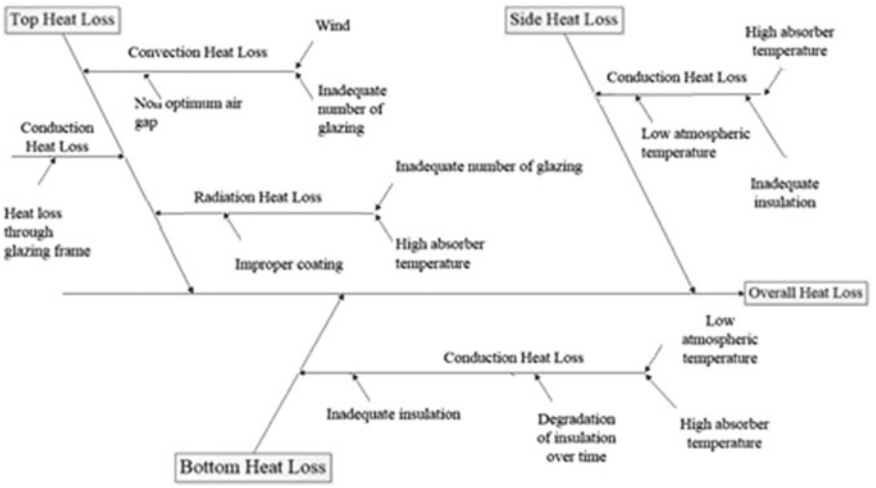


Fig. 2 Ishikawa cause–effect diagram showing the types of heat loss from STFPC with their contributing factors

$$Q_I = IA \tag{1}$$

where Q_I is the amount of solar radiation received by the collector (W), I is the intensity of solar radiation (W/m^2) and A is the collector area (m^2)

The amount of the useful energy gained by the collector after considering absorption and transmittance by the glazing cover and the heat lost to the surroundings is calculated using Eq. (2) [61],

$$Q_u = Q_i - Q_o \tag{2}$$

$$Q_i = I(\tau\alpha)A \tag{3}$$

$$Q_o = U_L A(T_c - T_a) \tag{4}$$

where Q_i is the amount of solar radiation received by the collector after considering absorption and transmittance by the glazing cover (W), Q_o is the rate of heat loss by the collector to its surroundings (W), α is the absorption coefficient of the absorber plate, τ is the transmission coefficient of the glazing material, U_L is the collector overall heat loss coefficient (W/m^2), T_c and T_a are the collector average temperature and ambient temperature (°C), respectively.

The rate of extraction of heat (Q_u) from the collector by the fluid passing through it is calculated employing Eq. (5) [36, 61, 62],

$$Q_u = mc_p(T_{out} - T_{in}) \tag{5}$$

Table 1 A comprehensive comparison of the efficiency and test setup

S/No.	Authors	Test Setup and Improvement in Efficiency
1	Bhowmik and Amin [52]	Using reflector with the collector, the overall efficiency increases by 10%
2	Hedayatizadeh, et al. [7]	Glazed double pass v-corrugated plate solar air heater with 12.2 mm corrugation height, 1.79 m^2 heater area and 0.005 kg/s air flow rate, the maximum exergy efficiency obtained 6.27%
3	Said et al. [53]	With TiO_2-H_2O (0.1vol.% and flow rate 0.5 kg/min) nanofluid was used as the working fluid, the energy efficiency increases by 76.6% and the exergy efficiency was 16.9%
4	Shojaeizadeh, et al. [54]	Using Al_2O_3-water nanofluid as base fluid, the maximum collector exergy efficiency increases by 0.72% with corresponding decrease in the mass flow rate by 67.8% and the collector inlet fluid temperature by 1.9%
5	Abad et al. [55]	By using Cu-water nanofluid over water, the collector efficiency increases by 24%
6	Jafarkazemi and Ahmadifard [56]	With 40 °C inlet water temperature and low mass flow rate, the maximum energy efficiency and exergy efficiencies were 80% and 8%, respectively
7	Chen, et al. [35]	The efficiency of the collector with ETFE foil is 2–3% higher than without ETFE foil
8	El-Sawi, et al. [57]	The thermal efficiency of chevron pattern is 10% higher than v-grooved and 20% higher than that of the flat absorbers
9	Ramani, et al. [5]	The thermal performance of the porous absorbing material-based double pass solar air collector is 25% higher than the double pass solar air collector without porous absorbing material and 35% higher than the single-pass collector
10	Charalambous, et al. [58]	The thermal efficiency was 42% using air as the working fluid
11	Tiwari and Sodha [59]	In IPVTS system, an overall thermal efficiency increases from 24% to 58% due to the additional thermal energy collection by the water flow

where m is the mass flow rate of cooling fluid (Kg/s), Cp is the heat capacity of cooling fluid at constant pressure (J/K), T_{out} and T_{in} are the collector outlet and inlet temperature (°C), respectively.

The collector efficiency (η_c) is defined as the ratio of the useful energy gained by the collector to the solar energy incident on it over a particular time period and can be calculated using Eq. (6) [61],

Table 2 A comprehensive comparison of the Improvement in reduction of Heat loss and other factors by various researchers

S/No	Authors	Improvement in reduction of Heat loss and other factors
1	Hedayatizadeh, et al. [7]	Variation in the temperature between sun and absorber accounted 63.57% of the whole exergy loss at the point corresponding to the maximum exergy efficiency
2	Said, et al. [51]	Use of SWCNT nanofluids over water as the absorbing fluid reduce entropy generation and enhance the heat transfer coefficient by 4.34% and 15.33%, respectively
3	Singh, et al. [60]	Using selective coating on absorbers over the ordinary black paint reduced the overall heat transfer loss coefficient by 20–30% whereas the use of double glass reduced the overall heat transfer loss coefficient by 10–15%
4	Agbo and Okoroigwe [23]	The use of double glazing cover minimizes the overall heat loss coefficient by 44%. For an absorber emissivity varying from 0.86 to 0.10, the collector loss coefficient is reduced by 47% and the overall collector heat loss coefficient increases by 0.1 with 10 K rise in ambient temperature
5	Leon and Kumar [25]	Unglazed transpired solar collector shows good performance in the temperature range 45–55 °C

$$\eta_c = \frac{\int Q_u dt}{A \int I dt}. \tag{6}$$

The collector efficiency (η_c) also can be calculated by Eq. (7) [63],

$$\eta_c = \frac{mc_p(T_{out} - T_{in})}{IA}. \tag{7}$$

The thermal efficiency (η_{th}) can be calculated by Eq. (8) [61],

$$\eta_{th} = \frac{Q_u}{IA} \tag{8}$$

5 Conclusions

Due to the low cost, easy operability and less maintenance requirements, solar thermal flat plate collectors (FPC) are still in use, especially in households or as a passive preheater where the output temperature requirement is low. The modern FPCs have a good thermal efficiency yet retaining their simplicity and ruggedness in comparison to other solar collector designs. The following points can be summarized regarding the current status of research in this field:

- Advent of evacuated tube collectors certainly is a lip forward in the field of solar collector, but it is very prone to thermal shock that is not much seen in general FPC design.
- Hybrid solar collectors or PVT collectors show improved overall efficiency over non-hybrid FPC designs.
- In solar thermal Flat Plate Collectors, a major heat loss occurs from the front side.
- The front heat loss problem is being solved in three ways. First, the heat loss can be reduced by reducing the convection and conduction losses through the glazing cover by optimizing the type of glazing material, number of glazing, using honeycomb structures and sometimes, by evacuating the air gap. Second, some researchers are working to reduce radiation heat loss from the absorber by applying selective absorber coating, special spectral coatings on the underside of the glazing cover that reflects back the thermal radiation more efficiently. Third, the heat loss is reduced by lowering the absorber temperature using nanofluids in the riser tubes.

All these researches are certainly improving the performance characteristics of the overall FPC design, but the performance of any collector system is also largely dependent upon the local environmental variables. Thus, the design of FPC should be optimized keeping in view the local environmental variables.

Acknowledgements The authors wish to acknowledge with a deep sense of gratitude the TEQIP-III cell, Jorhat Engineering College for providing us with necessary arrangements and help during our research work, which enable us to carry out our research and publish our results.

References

1. Duffie JA, Beckman WA (2013) Solar engineering of thermal processes, 4th edn. Wiley, USA, pp 5–6
2. Amri A, Jiang ZT, Pryor T, Yin CY, Djordjevic S (2014) Developments in the synthesis of flat plate solar selective absorber materials via sol-gel methods: a review. Renew Sustain Energy Rev 36:316–328
3. Ibrahim A, Othman MY, Ruslan MH, Mat S, Sopian K (2011) Recent advances in flat plate photovoltaic/thermal (PV/T) solar collectors. Renew Sustain Energy Rev 15(1):352–365
4. Gaur MK, Tiwari GN (2009) Optimization of number of collectors for integrated PV/T hybrid active solar still. Appl Energy 87(5):1763–1772
5. Ramani BM, Gupta A, Kumar R (2010) Performance of a double pass solar air collector. Sol Energy 84(11):1929–1937
6. Michaelides IM, Eleftheriou PC (2011) An experimental investigation of the performance boundaries of a solar water heating system. Exp Therm Fluid Sci 35(6):1002–1009
7. Hedayatizadeh M, Sarhaddi F, Safavinejad A, Ranjbar F, Chaji H (2016) Exergy loss based efficiency optimization of a double pass/glazed v-corrugated plate solar air collector. Energ 94:799–810
8. Skoplaki E, Palyvos JA (2009) On the temperature dependence of photovoltaic module electrical performance: a Review of Efficiency/Power correlations. Sol Energy 83(5):614–624
9. Huang BJ, Lin TH, Hung WC, Sun FS (2001) Performance evaluation of solar photovoltaic/thermal systems. Sol Energy 70(5):443–448

10. Han J, Ji J, Chow TT, Yi H, Lu J, He W, Sun W (2006) Effect of fluid flow and packing factor on energy performance of a wall-mounted hybrid photovoltaic/water-heating collector system. Energy Build 38(12):1380–1387
11. Chauhan D, Shishodia YS, Agarwal S (2015) Performance of hybrid air collector under different conditions. Eur J Adv Eng Technol 2(3):69–75
12. Omrany H, Marsono AK (2016) Optimization of building energy performance through passive design strategies. Br J Appl Sci Technol 13(6):1–16
13. Pokorny N, Matuska T (2016) Performance analysis of glazed liquid photovoltaic thermal collector with use of detail model. Sol Energy 83(12):2157–2164
14. Aste N, Leonforte F, Pero CD (2012) Optimization of solar thermal fraction in PVT systems. Energy Proc 30:8–18
15. Balaji S, Reddy KS, Sundararajan T (2016) Optical modelling and performance analysis of a solar LFR receiver system with parabolic and involute secondary reflectors. Appl Energy 179:1138–1151
16. Kraemer D, Poudel B, Feng HP, Caylor JC, Yu B, Yan X, Ma Y, Wang X, Wang D, Muto A, McEnaney K, Chiesa M, Ren Z, Chen G (2011) High performance flat panel solar thermoelectric generators with high thermal concentration. Nat Mater 10(7):532–538
17. Reddy KS, Vikram TS, Mallick TK (2018) Experimental performance investigations of an elliptical hyperbolic non-imaging solar concentrator with trapezoidal surface receiver for process heat applications. J Clean Prod 192:735–750
18. Tian Y, Zhao CY (2013) A review of solar collectors and thermal energy storage in solar thermal applications. Appl Energy 104:538–553
19. Ayompe LM, Duffy A (2013) Analysis of the thermal performance of a solar water heating system with flat plate collectors in a temperate climate. Appl Therm Eng 58(1–2):447–454
20. Mani M, Pillai R (2010) Impact of dust on solar photovoltaic(PV) performance: research status, challenges and recommendations. Renew Sustain Energy Rev 14(9):3124–3131
21. Yang M, Wang P, Yang X, Shan M (2012) Experimental analysis on thermal performance of a solar air collector with a single pass. Build Environ 56:361–369
22. Juanico L, Dilalla N (2014) Optimization of the hose-based low cost solar collector. Int J Renew Energy Biofuels 2014
23. Agbo SN, Okoroigwe EC (2007) Analysis of thermal loses in flat plate collector of a Thermosyphon Solar water heater. Res J Phys 1(1):35–41
24. Reddy PM, Venkataramaiah P, Sairam P (2012) Optimization of process parameters of a solar parabolic trough in winter using Grey-Taguchi approach. Int J Eng Res Appl (IJERA) 2:816–821
25. Leon MA, Kumar S (2007) Mathematical modeling and thermal performance analysis of unglazed transpired solar collectors. Sol Energy 81(1):62–75
26. Prajapati V, Soni U, Devi M (2016) A review to optimize the parameter of solar air heater. Int J Adv Eng Res Dev 3(5):71–82
27. Pacheco R, Ordonez J, Martinez G (2012) Energy efficient design of building: a review. Renew Sustain Energy Rev 16(6):3559–3573
28. Abubakar GB, Egbo G (2014) Performance evaluation of flat plate solar collector (Model Te39) in Bauchi. Am J Eng Res (AJER) 3(10):34–40
29. Zima W, Dziewa P (2010) Modelling of liquid flat plate solar collector operation in transient states. J Power Energ 225(1):53–62
30. Raghuraman P, Hendrie SD (2009) Analytical predictions of liquid, air photovoltaic/thermal flat plate collector performance. J Sol Energy Eng 103(4):291–298
31. Pena JLDL, Aguilar R (2014) Polymer solar collectors: a better alternative to heat water in Mexican homes. Energy Proc 57:2205–2210
32. Rehim ZSA (1998) A new design of solar water heater. Proc Indian Acad Sci (Chem Sci) 110(3):373–384
33. Sukhatme SP (1997) Solar energy: principles of thermal collection and storage, 2nd edn. Tata McGraw Hill Publishing Company Ltd., India

34. Amrutkar SK, Ghodke S, Patil KN (2012) Solar flat plate collector analysis. IOSR J Eng (IOSRJEN) 2(2):207–213
35. Chen Z, Furbo S, Perers B, Fan J, Andersen E (2012) Efficiencies of flat plate solar collectors at different flow rates. Energy Proc 30:65–72
36. Farahat S, Sarahaddi F, Ajam H (2008) Exergetic optimization of flat plate solar collectors. Renew Energy 34(4):1169–1174
37. Matuska T, Zmrhal V, Metzger J (2009) Detailed modelling of solar flat-plate collectors with design tool kolektor 2.2. In: Eleventh International IBPSA conference, Glasgow, Scotland, 27–30 July 2009
38. Rai GD (2005) Non- conventional energy sources, 4th edn. Khanna Publishers, India, pp 91–93
39. Ghoneim AA (2005) Performance optimization of solar collector equipped with different arrangements of square-celled honeycomb. Int J Therm Sci 44(1):95–105
40. Boudaden J, Oelhafen P, Schuler A, Roecker C, Scartezzini JL (2005) Multilayered Al_2O_3/SiO_2 and TiO_2/SiO_2 coatings for glazed coloured solar thermal collectors. Sol Energy Mater Sol Cells 89(2–3):209–218
41. Dudita M, Manceriu LM, Anastasescu M, Nicolescu M, Gartner M, Duta A (2013) Coloured TiO_2 based glazing obtained by spray pyrolysis for solar thermal applications. Ceram Int 40(3):3903–3911
42. Rahman MM, Öztop HF, Ahsan A, Kalam MA, Varol Y (2012) Double-diffusive natural convection in a triangular solar collector. Int Commun Heat Mass Transf 39(2):264–269
43. La¨mmle M, Kroyer T, Fortuin S, Wiese M, Hermann M (2016) Development and modelling of highly-efficient PVT collectors with low-emissivity coatings. Sol Energy 130:161–173
44. Chow TT (2010) A review on PV/T hybrid technology. Appl Energy 87(2):365–379
45. Alfegi EMA, Sopian K, Othman MY, Yatim B (2009) The effect of flow rates on the performance of finned single pass, double duct photovoltaic thermal solar air heaters. Eur J Sci Res 25(2):339–344
46. Cardinale N, Piccininni F, Stefanizzi P (2013) Economic optimization of low-flow solar domestic hot water plants. Renew Energy 28(12):1899–1914
47. Taylor RA, Otanicar TP, Adrian R, Prasher R, Phelan PE (2011) Nano fluid optical property characterization towards efficient direct absorption solar collector. Nanoscale Res Lett 6(1):225
48. Khanjari Y, Kasaeian AB, Pourfayaz F (2017) Evaluating the environmental parameters affecting the performance of photovoltaic thermal system using nanofluid. Appl Therm Eng 115:178–187
49. Nasrin R, Alim MA, Chamkha AJ (2012) Combined convection flow in triangular wavy chamber filled with water-CuO nanofluid: effect of viscosity models. Int Commun Heat Mass Transf 39(8):1226–1236
50. Nasrin R, Alim MA (2014) Modeling of a solar water collector with water-based nanofluid using Nano particle. Heat Transf-Asian Res 43(3):270–287
51. Said Z, Saidur R, Rahim NA, Alim MA (2014) Analysis of exergy efficiency and pumping power for a conventional flat plate solar collector using SWCNTs based nanofluid. Energy Build 78:1–9
52. Bhowmik H, Amin R (2017) Efficiency improvement of flat plate solar collector using reflector. Energy Rep 3:119–123
53. Said Z, Sabiha MA, Saidur R, Hepbasli A, Rahim NA, Mekhilef S, Ward TA (2015) Performance enhancement of a flat plate solar collector using titanium dioxide nanofluid and polyethylene glycol dispersant. J Clean Prod 92(2015):343–353
54. Shojaeizadeh E, Veysi F, Kamandi A (2015) Exergy efficiency investigation and optimization of an Al_2O_3-water nanofluid based flat-plate solar collector. Energy Build 101(2015):12–23
55. Abad MTJ, Zamzamian A, Imani E, Mansouri M. Experimental study of the performance of a flat-plate collector using Cu-Water nanofluid. J Thermophys Heat Transf https://doi.org/10.2514/1.t4074
56. Jafarkazemi F, Ahmadifard E (2013) Energetic and Exergetic evaluation of flat plate solar collectors. Renew Energy 56(2013):55–63

57. El-Sawi AM, Wifi AS, Younan MY, Elsayed EA, Basily BB (2010) Application of folded sheet metal in flat bed solar air collectors. Appl Therm Eng 30:864–871
58. Charalambous PG, Kalogirou SA, Maidment GG, Yiakoumetti K (2007) PV/T collectors-a review. Appl Therm Eng 27(2–3):275–286
59. Tiwari A, Sodha MS (2006) Performance evaluation of solar PV/T system: an experimental validation. Sol Energy 80:751–759
60. Singh PL, Sarviya RM, Bhagoria JL (2010) Heat loss study of trapezoidal cavity absorbers for linear solar concentrating collector. Energy Convers Manag 51(2010):329–337
61. Struckmann F (2008) Analysis of a Flat-plate Solar Collector. Heat Mass Transp.1–4. https://scholar.google.co.in/scholar?hl=en&as_sdt=0%2C5&q=F.+Struckmann%2C+Analysis+of+a+Flat-plate+Solar+Collector&btnG. Accessed 15 Nov 2019
62. Singh B, Othman MY (2009) A review on photovoltaic thermal collectors. J Renew Sustain Energ 1(6). https://doi.org/10.1063/1.3266963
63. Otanicar TP, Phelan PE, Tyagi H, Taylor RA (2011) Spatially varying coefficient for direct absorption solar thermal collector optimization. J Sol Energ Eng 133(2). https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1087.9648&rep=rep1&type=pdf. Accessed 15 Nov 2019