Transactions on Computer Systems and Networks

Amrita Rai
Dinesh Kumar Singh
Amit Sehgal
Korhan Cengiz *Editors*

Paradigms of Smart and Intelligent Communication, 5G and Beyond

Springer

Transactions on Computer Systems and Networks

Series Editor

Amlan Chakrabarti, Director and Professor, A. K. Choudhury School of Information Technology, Kolkata, West Bengal, India

Transactions on Computer Systems and Networks is a unique series that aims to capture advances in evolution of computer hardware and software systems and progress in computer networks. Computing Systems in present world span from miniature IoT nodes and embedded computing systems to large-scale cloud infrastructures, which necessitates developing systems architecture, storage infrastructure and process management to work at various scales. Present day networking technologies provide pervasive global coverage on a scale and enable multitude of transformative technologies. The new landscape of computing comprises of self-aware autonomous systems, which are built upon a software-hardware collaborative framework. These systems are designed to execute critical and non-critical tasks involving a variety of processing resources like multi-core CPUs, reconfigurable hardware, GPUs and TPUs which are managed through virtualisation, real-time process management and fault-tolerance. While AI, Machine Learning and Deep Learning tasks are predominantly increasing in the application space the computing system research aim towards efficient means of data processing, memory management, real-time task scheduling, scalable, secured and energy aware computing. The paradigm of computer networks also extends it support to this evolving application scenario through various advanced protocols, architectures and services. This series aims to present leading works on advances in theory, design, behaviour and applications in computing systems and networks. The Series accepts research monographs, introductory and advanced textbooks, professional books, reference works, and select conference proceedings.

Amrita Rai · Dinesh Kumar Singh · Amit Sehgal ·
Korhan Cengiz
Editors

Paradigms of Smart and Intelligent Communication, 5G and Beyond

Springer

Editors
Amrita Rai
G L Bajaj Institute of Technology and Management
Greater Noida, Uttar Pradesh, India

Dinesh Kumar Singh
G L Bajaj Institute of Technology and Management
Greater Noida, Uttar Pradesh, India

Amit Sehgal
Electrical Electronics and Communication Engineering (EECE)
Sharda University
Greater Noida, Uttar Pradesh, India

Korhan Cengiz
Electrical and Electronics Engineering Department
Trakya University Faculty of Engineering
Trakya, Turkey

ISSN 2730-7484 ISSN 2730-7492 (electronic)
Transactions on Computer Systems and Networks
ISBN 978-981-99-0108-1 ISBN 978-981-99-0109-8 (eBook)
https://doi.org/10.1007/978-981-99-0109-8

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

Dedicated to the Almighty

Preface

Paradigms of Smart and Intelligent Communication, 5G and Beyond is a book about the concept and application of Intelligent communication and networking. Recent rapid development in the field of intelligent communication and network have resulted in opening the doors to numerous information services and applications which were marginally explored earlier. This book endeavors to highlight the untapped potential of Smart communication for the innovation and expansion of the 5G and beyond.

We wrote this book because the advent of the wireless communications systems augurs new cutting-edge technologies, including self-driving vehicles, unmanned aerial systems, autonomous robots, the Internet of Things, and virtual reality. These technologies require high data rates, ultra-low latency, and high reliability, all of which are promised by the fifth generation of wireless communication systems (5G). Academics and industries are already discussing beyond 5G wireless systems which will be the sixth generation (6G) of the evolution. These technologies have high impact on the society and the way people are leading their lives. Thus, it is highly desirable to develop smart technologies for communication to improve the overall services and management of wireless communication and 5G networking. Integration of Artificial Intelligence (AI) with 5G wireless networks, Machine learning, and IoT have converged to give some ice-breaking solutions in many aspects of 5G wireless communication design including radio resource allocation, network management, and cyber-security. Therefore, the utilization of artificial intelligence, machine learning techniques, and IoT will lead to better performance of Smart Communication and wireless systems Networking.

The book can be used as an introductory text on communication and networking with their advancement using AI/ML and applications for undergraduate or graduate students in Electronics and communication engineering and computer science or related disciplines such as computer engineering, cognitive science, and wireless networking.

The references given at the end of each chapter are not meant to be a comprehensive list. Rather, we have referenced the works that have been directly used and works that we think provide good overview of the related or prerequisite literature, by referencing both classic works and more recent surveys. We hope that no researchers feel slighted by their omission, and we are happy to have feedback where someone feels that an idea has been misattributed.

Greater Noida, India — Amrita Rai
Greater Noida, India — Dinesh Kumar Singh
Greater Noida, India — Amit Sehgal
Trakya, Turkey — Korhan Cengiz

Contents

1 Artificial Cognitive Computing for Smart Communications, 5G and Beyond 1
Amsini, Uma Rani, and Amrita Rai

2 Green IoT Networks Using Machine Learning, Deep Learning for 5G Networks 17
Neha Jain, Ritesh Pratap Singh, Harsh Arora, and Krishanu Kundu

3 Integration of IoT and 5G Communication 43
Seema Nayak, Pankaj Jha, and Manoj Nayak

4 Role of IoT and Antenna Array in Smart Communication and 5G 57
Krishanu Kundu, Amrita Rai, and Ritesh Pratap Singh

5 Machine Learning and Deep Reinforcement Learning in Wireless Networks and Communication Applications 83
Om Prakash, Prabina Pattanayak, Amrita Rai, and Korhan Cengiz

6 Detection of Consumption of Alcohol Using Artificial Intelligence 103
Anubhav Thapliyal, Arunendra P. Singh, and Rashmi Priyadarshini

7 Application of Machine Learning in Finger Vein Pattern Recognition 113
Rahul Dev, Krishanu Kundu, Astha Sharma, and Shiv Narain Gupta

8 Machine Learning Techniques for Anomaly Detection Application Domains 129
Alka Chaudhary and Reshu Agarwal

9 Application of AI & ML in 5G Communication 149
Lipsa Das, Biswa Mohan Sahoo, Ajay Rana, Khushi Dadhich, Smita Sharma, and Suman Avdhesh Yadav

10 Software Defined Network-Based Management Architecture for 5G Network 171
Catherine Nayer Tadros, Bassem Mokhtar, and Mohamed R. M. Rizk

11 Reversible Logic Based Single Layer Flip Flops and Shift Registers in QCA Framework for the Application of Nano-communication 197
Rupali Singh and Pankaj Singh

12 Machine Learning Technique for Few-Mode Fiber Design with Inverse Modelling for 5G and Beyond 221
Bhagyalaxmi Behera, Abhishek Das, and Mihir Narayan Mohanty

13 IoT for Landslides: Applications, Technologies and Challenges 245
Amrita Mohan, Ramji Dwivedi, and Basant Kumar

14 A Review: Dust Cleaning Approach of Solar Photovoltaic System Using IOT & ML 261
Kunwar Babar Ali, Zaiba Ishrat, and Seema Nayak

15 Prediction of Heart Disease Using Hybrid Machine Learning Technique 277
Nagaraj M. Lutimath, Chandra Mouli, B. K. Byre Gowda, and K. Sunitha

About the Editors

Amrita Rai received her B.Tech. degree from Nagpur University, M.Tech. degree from Maharishi Dayanand University, Rohtak, and a Ph.D. degree from Thapar University, Patiala, in 1998, 2007, and 2015, respectively. She has over 21 years of teaching, research, and administrative experience. She had published several reputed international/SCI papers and attended various national and international conferences. She has guided more than 20 postgraduate and 100 undergraduate theses. She has written a book and 6 book chapters. She has filed six patents and delivered expert lectures on smart sensors transforming healthcare systems, IoT and sensors, and adaptive signal processing in many universities and colleges.

Dinesh Kumar Singh received B.E. in Electronics and Communication Engineering from Kumaon Engineering College, India in 2003. He pursued his M.Tech. in Digital Communication from RGPV University, India, and his Ph.D. from the Indian Institute of Technology (ISM), Dhanbad, India. His area of interest is microwave engineering. He is currently an Associate Professor in the Electronics and Communication Engineering Department at the G. L. Bajaj Institute of Technology and Management, India. His further research areas are designing high gain, compact, re-configurable, fractal-shape, circularly polarized microstrip antennas, substrate integrated waveguide, and magneto-electric dipole antennas for a modern communication system. He has been credited with publishing several research papers in reputed international journals and conferences.

Amit Sehgal is Ph.D. in Electronics Engineering from Dr. A. P. J. Abdul Kalam Technical University, India. He completed his M.Tech. from Guru Nanak Dev Engineering College, India, and his B.E. from C. R. State College of Engineering. He has over 20 years of experience in the industry, academia, and research. He has contributed to several research and development projects in the field of wireless networks, health informatics, and IoT, including corporate and international academic collaboration. Several projects guided by him have received project grants under various state and central government schemes. He is a voluntary startup mentor for Atal Innovation Mission Incubators and Atal Tinkering Labs. He has been contributing to the startup ecosystem by mentoring innovators and early-age startups. On completion of his Cambridge Certification for International Teachers and Trainers and Dale Carnegie Certificate for High Impact Teaching Skills, he made a significant contribution to the pedagogic uplift of engineering education by serving as Master Trainer for Mission 10X—a faculty development initiative by Wipro Technologies, for over two years. He has more than 35 research papers in journals and conferences. His research interests include wireless networks, remote monitoring systems, and IoT.

Korhan Cengiz is an Assistant Professor in the Department of Engineering at Trakya University, Turkey. Dr. Cengiz completed his Ph.D. at the Department of Electronics Engineering at Kadir Has University. He is the author of more than 40 papers in the field of communications. He participated as a keynote speaker at several conferences. He has numerous referee and editor positions in several journals and conferences.

Chapter 1
Artificial Cognitive Computing for Smart Communications, 5G and Beyond

Amsini, Uma Rani, and Amrita Rai

Abstract Artificial Intelligence (AI) is computer intelligence that manifests itself in "cognitive" capabilities that people identify with other brains. AI employs various technologies including Deep Learning, Machine Learning and Natural Language Processing. The self-learning systems are utilizing pattern recognition, natural language processing and data mining to replicate the person's brain functions are called cognitive computing. Cloud-based communication has bolstered this by delivering vital communication services. However, due to restricted capacities and a need for low latency, high reliability, and a good user experience, providing a cloud-based environment and intensive data processing algorithms are insufficient. Cognitive computing is considered a branch of computer science that simulates human cognitive processes. As a result, when cognitive science skills are combined with communications and existing systems may be improved, resulting in higher accuracy and lower latency. We have gone through cognition-based communications in depth in this study, which blends smart communication technologies and intelligent computing based on AI. Following is an overview of the cognitive computing and its evolution. Then, combining networking, analytics, and cloud computing, a systematic and comprehensive framework for using cognition in communication is provided.

Keywords Artificial intelligence · Cognitive · Smart communications

1.1 Introduction

Since the early 1990s, computing has been established to be a real game–changer with a source of fascination for academics. Before time computers were compiling strategy

Amsini (✉) · U. Rani
Sri Sarada College for Women (Autonomous), Salem-16, India
e-mail: amsini1621992@gmail.com

A. Rai
GL Bajaj Institute of Technology and Management, Plot No 2, Knowledge Park III, Greater Noida, Uttar Pradesh 201306, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
A. Rai et al. (eds.), *Paradigms of Smart and Intelligent Communication, 5G and Beyond*, Transactions on Computer Systems and Networks,
https://doi.org/10.1007/978-981-99-0109-8_1

such as calculators where decreased the intricacy required for human computing. Then, there was come the application era, because everything was handled and operated by microprocessor-equipped gadgets. The cognitive computing era is about discussing data and making sense of it with statistics. The large scope of requirements for increasing the client experience at the time coverage of wide sort, massive connections using a low-power, high capacity hub, and better reliability in system communications, which leads to network survival and prosperity with wireless communication. The research was conducted based on the architectures of network and enablement management optimization, flexible configuration, and quick response time. This paper discuss about the components and architecture of cognitive computing, smart communication, cognitive computing and the impact of COVID-19 and 5G network (Dash et al. 2021). Cognitive computing is a branch of research that seeks to replicate human cognitive process (Gudivada 2016). The evaluation of cognitive computing elaborate (Fig. 1.1).

Tabulating machines were popular in the late nineteenth century, allowing advancements such as the census report and the Social Welfare System in the United States. Programmable computing first appeared in the 1940s and has since allowed all

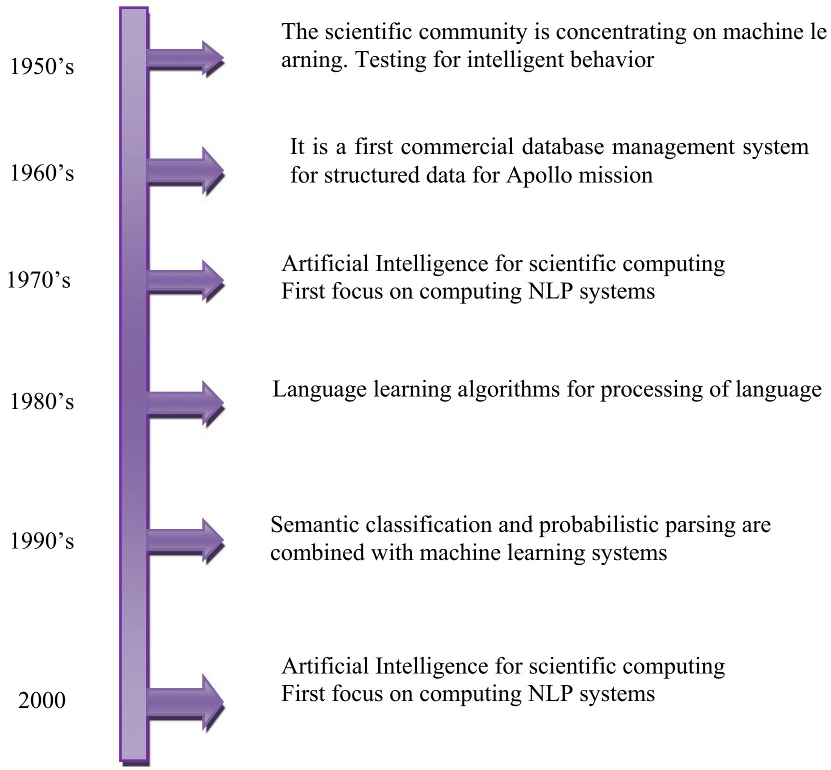


Fig. 1.1 The evaluation of cognitive computing (The previous few decades)

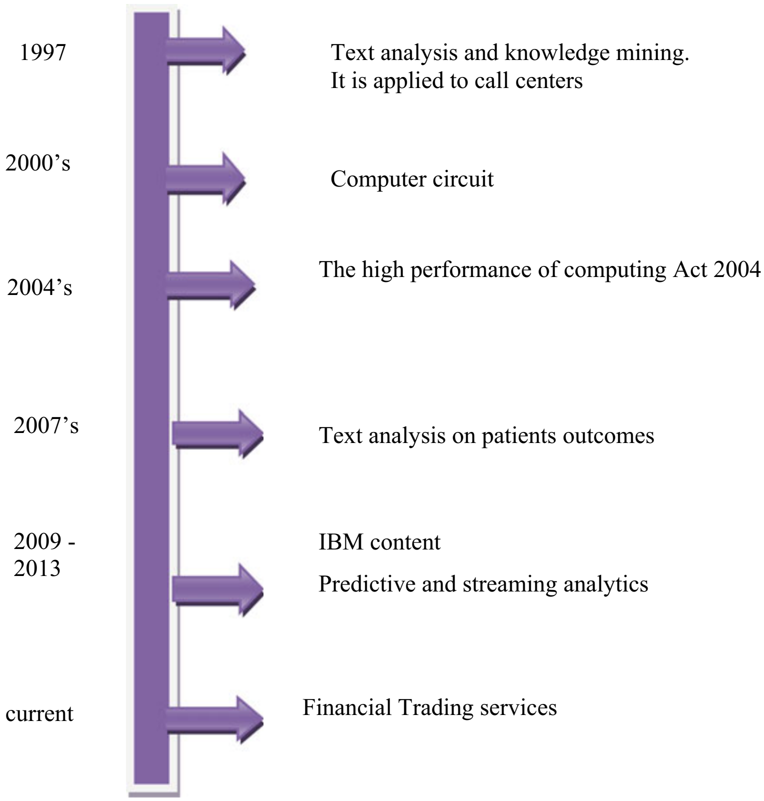


Fig. 1.2 The evaluation of cognitive computing (The last ten years with current year)

from space exploration towards the Internet. The evaluation of cognitive computing in the last ten years with the current year is shown in (Fig. 1.2).

After tabulating machines and programmable computers, the cognitive computing age may be regarded as the third most disruptive paradigm in the history of computing. The purpose of artificial intelligence is to allow the construction of self-aware intelligent systems. The cognitive systems identified data from the text and speeches to understand the systematic and semantic relations. In general, the 90's were early evidence of successful reinforcement (Du et al. 2020; Elsayed 2021) learning in AI, NLP, machine learning, reasoning, and data mining. Nowadays, Google Deep mind and IBM Watson are at the forefront of cognitive computing.

As a result, cognitive science encompasses a wide range of themes and study fields, including anthropology, linguistics, artificial intelligence, philosophy, psychology, and neuroscience (Wang 2003). In a sense, the accomplishments of a cognitive science scientist as much as this point are inextricably linked toward multidisciplinary research methodologies. Cognitive computing systems combine the finest of several technologies, including NLP, real-time computing, AI (Artificial Intelligence), and machine learning.

A cognitive computing system can evaluate massive amounts of semi- and unstructured data utilizing these technologies. the advantages of the cognitive computing are better than data analysis, it also enables business to identify opportunities and expose challenges in real time process them to respond faster and more efficiently and then improved the degree of consumer contact (Hwang and Chen 2017).

1.2 The Cognitive Computing Components

The Cognitive computing is characterized as a collection of autonomous and cognitive knowledge processing theories and technologies that mimic the brain's mechanisms beyond conventional forced data processing. Cognitive computing conceptual and behavioral models have been established. To address the design and implementation of cognitive computing systems, a strong mathematical technique called as label mathematics has been proposed. A wide range of cognitive computing applications, such as autonomous agent systems and intelligent search engines, have been identified.

The complete cognitive figure architecture comprises the following components:

Algorithms

- Algorithms are rules to follow to define approaches that are being used for solving certain problems.
- AI refers to a machine sensing or generally seeing its own state and then figuring out how to achieve the goals it was designed to achieve.
- Initial AI was progressively basic, for such, knowledge retrieval, and cognitive registration is often seen as a primitive sort of AI.

Reasoning and Decision Automation

- This is a strategy used by an intellectual data structure to apply its knowledge of where to accomplish.
- Goal is referred to as reasons. It's not much like human cognition, but it's meant to mimic it.
- The outcome of a thinking technique might well be decision mechanization, wherein the program autonomously generates and implements a solution to a problem (Tarafdar and Beath 2018).

Sentimental Intellectual Ability

- For a long time, psychological figuring shied away from striving to reflect passionate understanding; however, fascinating operations.
- For instance, the MIT innovation Affectiva, are trying toward processing the frameworks that were constructed which preserve a fathom sense of humanity through these signposts as good looks and then generate responses.

- The objective of intellectual figuring frameworks, which are remarkable in resemblance to a human in their ability toward read passionate cues (Tarafdar and Beath 2018).

The Speech Recognition Processing and Natural Language processing (NLP)

- Natural language processing is the use of registered techniques to comprehend and generate answers to human languages in their natural—that is, generally both spoken and written structures.
- Natural Language processing is divided into two distinct levels such as Natural Language Generation (NLG) in addition to natural language understanding (NLU).
- Multimodal acknowledgment is a comparable mechanism that translates contributed conversation into language suited for NLP (Tarafdar and Beath 2018).
- Deep learning calculations are used in visual recognition to analyze photos and differentiate things, such as individuals.

Emotional Intelligence

- For a long time, psychological beginning to figure avoided desperately attempting to reflect an enthusiastic viewpoint; however, fascinating operations. For example, the MIT innovation Affective, is working on developing frameworks is an understanding sense of humanity by means of indicators as external manifestations and then generate responses.
- The objective is to generate cognitive thinking structures that are startlingly mortal owing to their ability to read impassioned cues (Tarafdar and Beath 2018).

1.3 The Architecture of Cognitive

Figure 1.3 depicts a cognitive computing system architecture that makes use of cloud-based solutions, tensor flow, and numerous database tools to develop various cognitive programs. There are some technological challenges available for human computer interaction with emotional cognition. The communication area focuses on data transmission, whereas the computer realm focuses on information consumption.

In real-world cognitive computing applications, knowledge is mostly expressed through data, which includes both unstructured and semi-structured data. The Internet of Things (IOT) (Chen et al. 2017a) collects real-time valuable information in the material reality. It forms a massive network via the Web, and realizes intercommunication among huge wearable sensors in order to create a founder in between the physical and the digital world (Chen et al. 2017a; Sheth 2016).

Cognitive robots were created in the mid-twentieth century. Over the course spanning more than a half-century, human production practices have been substantially changed by robot technology and living styles (Sheth 2016). This technique is now a significant marker for measuring a country's stage of methodical and technical

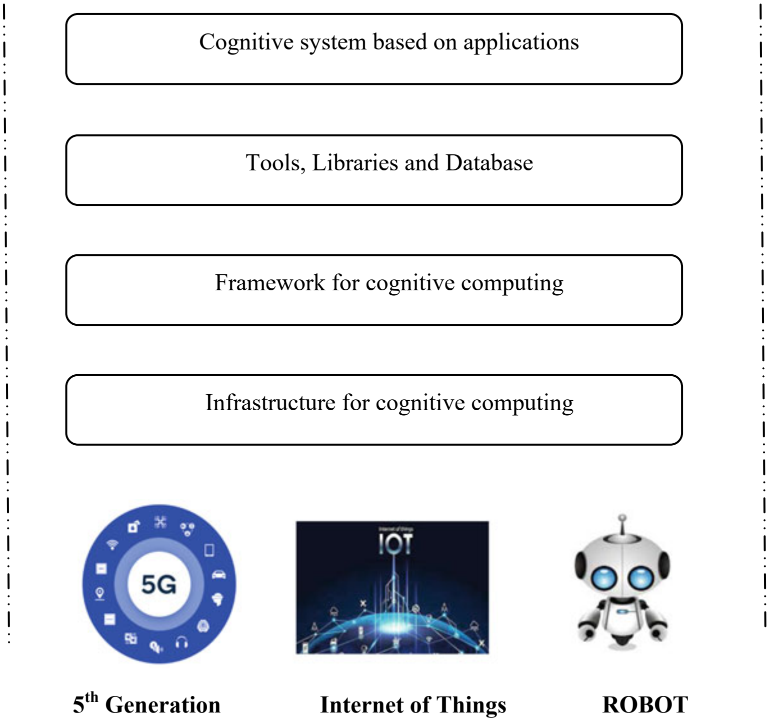


Fig. 1.3 The system architecture of cognitive

modernization and production of high-end. The new trend in community relationships demonstrates that the next era of robot systems will visualize human beings in more ways than one. For instance, particularly in the form of a partnership relationship in which humans and machines coexist in unity and make their corresponding advantages amenable to one another. The ability to coexist with humans is a key trait for the next generation of robots. Cognitive computing system construction is offered were comprised of different segments such as Internet of Things, cloud computing and big data analysis.

Cognitive applications such as audio and video processing data were based on structure and then unstructured, Machine learning analytics, Machine vision, game theory, and patterns were predictions and visualization, then cost savings at productivity and security.

An important use of the cognitive computing is emotional cognition. In addition, it is a novel human mechanism interface technique. At present, contemporary human–machine interaction systems frequently imply human–machine interaction supported in a physical distance context. IBM's (International Business Machines) intellectual service for language and Google's cognitive computing application stress the recognition of brain resembling cognition along with judgment through employing a service of cloud paradigm toward delivering correct decision making support. Cloud computing and internet of things provide the hardware and software foundation for cognitive computing, whereas providing big data analysis techniques furthermore ideas for determining and identifying novel ideas and challenges of worth data statistics.

1.4 Cognitive Computing for Smart Communications

As network technology and AI techniques progress, applications of smart information and various services are becoming increasingly popular throughout the world. The cloud based communication is facilitating to providing some powerful communication capabilities. Nevertheless, owing to limited capacity and the necessity for little latency, elevated reliability, and a pleasant user familiarity, just providing a cloud-based infrastructure and intensive data processing system is insufficient.

The study of mimicking cognitive functions is known as cognitive computing. The development and completion of 5G smart networks is particularly demanding since resource optimization is required (Chen et al. 2017b). Cognitive computing, combined with artificial intelligence, administers wireless system services and then cloud computing applications. There are some characteristic of cognitive computing is shown in (Fig. 1.4).

The cognitive systems continued to learn and received additional information and update it, making it dynamic training. Continuously interdependent and probabilistic in nature since it connects several modules and makes decisions based on the past historical events. The system can assign a weightage to each conceivable option and determine the best solution to tools.

1.4.1 The Cognitive Computing for the Society—Use Cases

Smart information apps and services are becoming more popular throughout the world as network technology and artificial intelligence approaches to advance. Cloud communication has also contributed to this by offering robust communication capabilities. However, just offering a cloud-based environment and intense data processing techniques is insufficient due to restricted capacity and the need for low latency, high dependability, and better user experience. Cognitive computing is the study of replicating human cognitive processes. Thus, when cognitive science skills are combined with communications, they can improve existing systems by providing higher accuracy and lower latency (Wang 2002).

Computing is applied in various departments, including education, health care, finance, and retail. In healthcare fields, build a cognitive based system on research data, clinical, biomedical, and biomedical. The cognitive computing requires skills such as planning, reasoning and learning (Wang 2003). Whatever applications are needed for cognitive process were as product, process and insight (Wang 2007b).

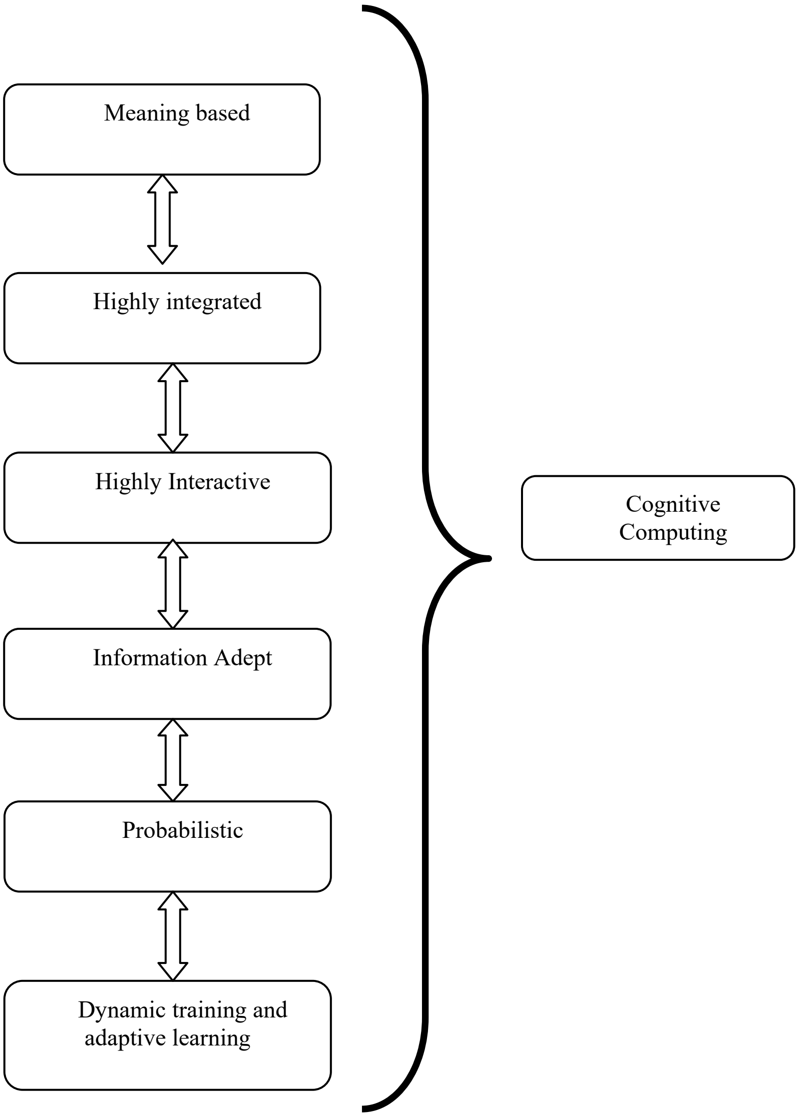


Fig. 1.4 The characteristic of cognitive computing

In the financial sector, cognitive computing is used for fraud detection. Retail-based systems, designed for mining financial data, also aid in the detection of fraud. The student learning styles are recorded as the date and mined in the education industry. The advantages of computing are the reduction of human error, digital assistance, prevention of wrong decisions by machine and helping in repetitive jobs.

This cognitive computing represents the self learning systems for utilizing machine learning models to mimic the way of brain works. These products are based on the field of artificial intelligence. The cognitive is acquiring knowledge from the data to understand human interaction and provide the solution to the mining problems. It used existing knowledge to get or generate new knowledge. The basis of cognitive such as big data analytics, machine learning and cloud computing.

- Big data analytics is the process of examining large data sets containing a variety of data, and analyze the big data with sophisticated tools.
- Machine learning is the use of algorithms to enable computers to analyze data and make predictions based on the information fed to them. Artificial intelligence tools enable the cognitive analytics to follow and understand the content, apply reasoning and then learn continuously.
- The cloud computing is applied to analyze huge amounts of data in real time and it need extensive computing power.
- Some cognitive computing applications like computer vision or speech recognition require good storage and computing infrastructure. Enterprise can now elastically scale their storage and processing infrastructure with bid data platforms like hadoop, and cloud computing platforms like Azure and Google cloud. Cloud based computing also provides powerful communication (Modha et al. 2011).
- Cognitive computing uses in education field for helping institutions to save on resources and deliver better administrative services. There is no doubt that if teachers and support teams use cognitive services to better support their students, educational services can be improved.
- Lot of fundamental actions based on computing such as internet of things (IOT), research in image processing sensors, robotics and then communication.

Cognitive Informatics is defined as a disciplinary investigation of computer science, information science, cognitive science, and intelligence science into the internal information processing mechanisms and processes of the brain and natural intelligence, as well as their engineering applications in cognitive computing.

1.4.2 The Cognitive Analytics as Parts of Cognitive Computing

The cognitive analytics is one of the intelligent technologies to analyze large sets of unstructured data with human like intelligence. This might involve comprehending the context and meaning of a statement or recognizing certain items in an image given a great quantity of information. Cognitive analytics frequently use artificial intelligence algorithms and machine learning, allowing a cognitive application to develop over time. Cognitive analytics exposes patterns with relationships to traditional analytics cannot. Real life applications for cognitive analytics are given below,

- The medical sector is already beginning to apply cognitive analytics to match patients with the best available therapies. Microsoft's Cortana, Apple's Siri, and IBM's Watson are examples of cognitive analytics in use today.
- Organizations are utilizing cognitive analytics to access unstructured data sources such as photos, emails, text documents, and social media postings.

Though cognitive analytics is still in its infancy, it may be the explanation for discovering real-time solutions for enormous volumes of heterogeneous data and ushering in a paradigm change from traditional analytics (Modha et al. 2011). The cognitive analytics paradigm is shown in (Fig. 1.5) with brief capabilities. The cognitive platforms which combine the necessary elements for data analytics and decision making at effective manner.

- Different types of users—The platform supports different views for data scientists and simple end users. A data scientist can train and compare AI models. The end user can monitor the model's outcome that was previously trained by a data scientist.

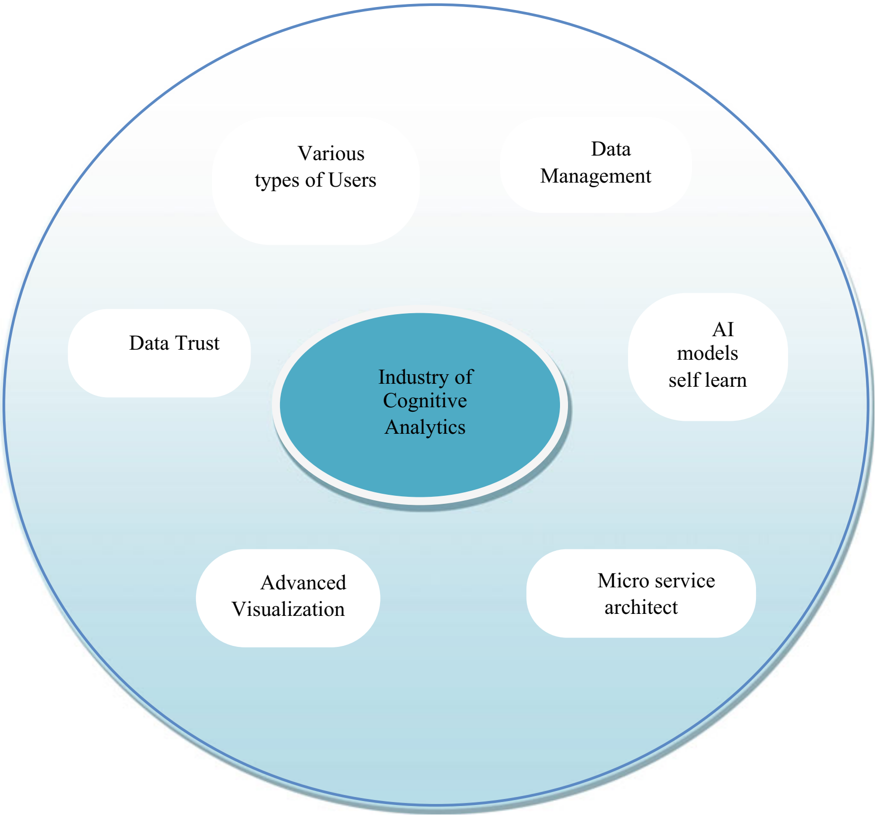


Fig. 1.5 Cognitive analytics platform

- The microservice architecture—It is consisting of a wide range of machine learning algorithms. It will be expanding the learning algorithms and developing artificial learning methods.
- Self learned AI methods—To learn the self learning of AI algorithms such as supervised and unsupervised learning models. It is applied for various data like industries.
- The data driven model—It is providing a continuous learning system which provide a high accuracy compared to manual learning.
- Data Management—Data storing, management, cleansing and pre-processing functionalities were provided for data management.
- Data Trust—the internet development system connectors are used by the platform in order to provide connectivity, interoperability, data transition and security.

The cognitive analytics use cases such as Microsoft Cortana, Apple's Siri and IBM's Watson. Organizations are using the cognitive analytics to tap into unstructured data sources such as images, emails, text documents and then social posts. Cognitive computing can generate data in any form, organized or unstructured, from many sources, allowing for risk mitigation while delivering value, a great user experience, and increased security. Cognitive analytics use case scenarios of banking, health care and then retail.

- Banks and financial institutions employ cognitive analytics to improve customer experience, operational efficiency, and income generation. Cognitive analytics has the potential to transform the banking industry in three fundamental areas. Improve your analytical skills, Contextual involvement that is stronger and deeper and then the organization's whole digital transition.
- There are several application cases for cognitive analytics in a healthcare company. One of the most important applications for medical professionals in the healthcare sector is to enhance patient treatment, which leads to better patient outcomes.
- Marketers may collect a large amount of customer data based on their purchasing patterns using cognitive analytics. This enables businesses to increase shop efficiency and make smarter decisions. Cognitive analytics may also be connected with e-commerce platforms to extract information about consumers, their purchasing cycles, and the various sorts of things they purchase, and so on.

1.5 Impact of Covid-19 on Cognitive Computing Market 6 Cooperative and Cognitive Network for 5G Network

Artificial intelligence (AI) is crucial for cognitive wireless communications under the new service paradigm to overcome many technical challenges. In complex wireless networking environments, application capabilities include complex decision making, wireless network management, resource optimization, and deep information discovery (Wang 2007a). Also, wireless communication and network ecosystems need to be upgraded with new capabilities such as Adaptive Wireless Resource

Management Based on Cognitive Power, and so on. Innovative information services and applications are spreading globally due to the fast development of wireless communication and networking technology. Advanced communications and networks have significantly enhanced the user experience and have had a significant influence on all parts of people's lifestyles at home, at work, in social interactions, and economically.

Therefore, developing smart wireless communication and networking technologies to support optimal management, dynamic configuration, and fast service organization is a major challenge. Derived from cognitive science and data analytics, cognitive computing can mimic or augment human intelligence, and exhibits great potential to power smart wireless communication networks (Wang 2002, 2007a).

Recently, communication and network technology have advanced at a rapid rate of growth. Several information applications and services have been established globally. This technology has a significant influence on society and individual lives have improved. The previous generations of wireless networking advancements were constrained and unable to meet the user needs. But 5G networks are expected to transform wireless networks. This network is intended to provide a higher quality of service, scalability, low latency, and faster speeds (Qin et al. 2018). 5G network technology combined with machine learning provides high intelligence scope in the network era (Akhtar et al. 2021).

Cognitive computing help medical practitioners diagnose illness from the massive amount of COVID-19 data and produces the result of smart recommendations (Dash et al. 2021). This data is dealt global wise so early detection and identification are improving the patient survival rate. Huge quantity rate of data is processed by deep learning with AI. The COVID-19 data is collected from the chest X-ray image (Abdul Salam and Taha 2021). This data is very huge which is handled by deep learning with IOT (Internet of Things).

Federated deep learning-based COVID-19 identification methods are based on internet of things and edge based computing is designed on Net architecture (Liu et al. 2020). The COVID-Net is helped to improve the testing and training performance with improved sensitivity values and testing sets from cloud central server respectively. The sample cognitive computing based identification of COVID-19 based on internet of things is shown in (Fig. 1.6). Several sectors have been devastated by the COVID-19 epidemic, but the cognitive cloud computing market has benefited from the plague. Furthermore, the NLP techniques in pharmaceutical and healthcare companies assist scientists and healthcare practitioners in various ways. During the crisis, these reasons drove the expansion of the cognitive cloud computing industry. The market has been divided into many groups based on technology, business size, industrial sector and then geography.

Currently, the information and communication sector is confronting a global problem, such as the creation of new services with improved QoS and immediate reductions in environmental impact. Wireless demand and supply are not balanced (Qin et al. 2018).

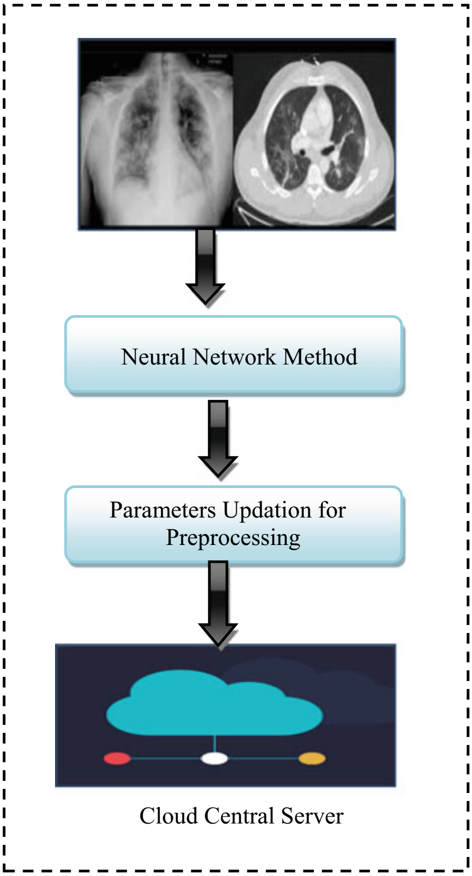


Fig. 1.6 Cognitive Computing-based COVID-19 identification on internet of things

1.6 Challenges and Future Aspects of Cognitive Computation on 5G and Communication

Global 5G installation has recently begun. The consistent process has a variety of features, including ultra-reliability, mass connection, and tailored low latency. However, 5G won't be enough to support all of the demands of technology in 2030 and beyond (Fourati et al. 2021). The development of new information and communication technologies has a significant impact on the recruitment of technical personnel, businesses, and researchers. Regarding 5G networks, sixth-generation (6G) CR networks are predicted to familiarize users with cutting-edge use cases and performance metrics, including providing global coverage, cost effectiveness, increased spectrum, energy intelligence, and safety (Ari et al. 2019; Muhammad Muzamil Aslam et al. 2021; Yu et al. 2020). Currently, the information and communication sector is confronting a global problem, such as the creation of new services with improved QoS and immediate reductions in environmental impact. Wireless demand and supply are not balanced (Miah et al. 2020). There seems to be concern about an impending spectrum crisis where the demand for smart mobile devices will soon outpace wireless capacity. Lack of fresh spectrum in wireless data carriers is a problem. Smartphone data traffic is growing so quickly that it will eventually fill the available spectrum if nothing is done. This might damage not only phones but also other wireless devices that are in use (Ostovar et al. 2020).

Fifth-generation networks are mainly based on robust architectures and ML-based solutions in heterogeneous behavior of communication networks. Here, we mention the blocks of ML application (Wang et al. 2020) and also describes security and privacy challenges. Possible technologies for example: Microwave and tera-hertz band, RAN, NFV, network slicing, wireless, and cloud computing are also described to upgrade the upcoming 5G network architecture. Various challenges to completing 5G implementation lie ahead (Naeem et al. 2020). The Cognitive computation proposal was proposed to investigate the effectiveness of the spectrum utilization issue. This CC technology has the potential to cause issues for spectrum management (Xu et al. 2019). Multiple uses of the spectrum suggest that it is in use, indicating that the risk of complicated spectrum utilization at unanticipated geographic locations and at fixed frequencies is not very high (Al-Turjman et al. 2019).

Due to several issues, including the installation of spectrum sensing capabilities, sensing nodes with the current frequency band, and node responsibility selection for nodes responsible for spectrum sensing, the CR method is difficult and complex for WSNs (Lee et al. 2020). On the other hand, CR encompasses a diverse network and is not constrained by opportunistic spectrum access. Frequency bands cannot be allocated to different wireless access technologies in the current spectrum locations (Muwonge et al. 2020). Such technology may make a number of new uses and possibilities possible, including rural connection, city and university coverage, huge wireless hotspots, and content distribution networks.

1.7 Summary

In this chapter, we focused on providing a comprehensive description of cognitive computing architecture and smart communication using a 5G network. The architecture of cognitive computing system is comprised of three sections such as Internet of Things (IOT), big data analysis, and cloud computing. The cognitive computing is paired with artificial intelligence and cloud computing has a potential to handle the application of wireless system and provides services with ease. The cognitive computing is a one of the problem solving approach based on modern AI that is fully focused on understanding, learning, planning and reasoning. These technologies boost up the medical data testing and improve prognosis accuracy too. Finally, we also discussed the potential major difficulties in 5G CR network communication as well as important empowering technologies that are vital to the development of 5G and beyond. We have quantified the many technologies that might be useful for

5G CR network communication in addition to highlighting the visualization and objective of 6G CR network communications.

References

Abdul Salam M, Taha S (2021) Ramadan M : COVID-19 detection using federated machine learning. PLoS ONE 16(6):e0252573

Akhtar T, Tselios C, Politis I (2021) Radio resource management: Approaches and implementations from 4G to 5G and beyond. Wirel Netw 27:693–734

Al-Turjman F, Ever E, Bin Zikria Y, Kim SW, Elmahgoubi A (2019) SAHCI: scheduling approach for heterogeneous content-centric IoT applications. IEEE Access, 7, pp 80342–80349

Ari AAA, Gueroui A, Titouna C, Thiare O, Aliouat Z (2019) Resource allocation scheme for 5G C-RAN: A Swarm Intelligence based approach. Comput Netw 165:106957

Chen M, Miao Y, Hao Y, Hwang K (2017a) Narrow band internet of things. IEEE Access 5:20557–20577

Chen M, Yang J, Hao Y, Mao S, Kai H (2017b) A 5G cognitive system for healthcare, Big Data Cognit Comput, 1(1), pp 1–15

Dash S, Chakravarty S, Mohanty SN, Pattanaik CR, Jain S (2021) A deep learning method to forecast COVID-19 outbreak. New Gener Comput 1–25

Du Z, Deng Y, Guo W, Nallanathan A, Wu Q (2020) Green deep reinforcement learning for radio resource management: Architecture, algorithm compression, and challenges. IEEE Veh Technol Mag 16:29–39

Elsayed M (2021) Machine learning-enabled radio resource management for next-generation wireless networks. Ph.D. Thesis, Université d'Ottawa/University of Ottawa, Ottawa, ON, Canada

Fourati H, Maaloul R, Chaari L (2021) A survey of 5G network systems: Challenges and machine learning approaches. Int J Mach Learn Cybern 12:385–431

Gudivada VN (2016) Cognitive computing: Concepts, architectures, systems, and Applications, Handbook Stat., vol. 35, pp 3–38

Hwang K, Chen M (2017) Big-Data Analytics for Cloud, IoT and Cognitive Learning. Wiley, London, U.K.

Lee S, Youn J, Jung BC (2020) A cooperative phase-steering technique with on-off power control for spectrum sharingbased wireless sensor networks. Sensors, 20(7)

Liu B, Yan B, Zhou Y, Yang Y, Zhang Y (2020) Experiments of federated learning for covid-19 chest x-ray images

Miah MS, Ahmed KM, Islam MK, Mahmud MAR, Rahman MM, Yu H (2020) Enhanced sensing and sumrate analysis in a cognitive radio-based internet of things. Sensors (switzerland) 20(9):2525

Modha SD, Ananthanarayanan KR, Esser S, Nadirango A, Sherbondy JA, Singh R (2011) Cognitive computing. Commun, ACM 54(8):62–71

Muhammad Muzamil Aslam, Liping Du, Xiaoyan Zhang, Yueyun Chen, Zahoor Ahmed, Bushra Qureshi (2021) Sixth generation (6G) cognitive radio network (CRN) application, requirements, security issues, and key challenges. Wirel Commun Mob Comput, vol. 2021, Article ID 1331428, 18 pages. https://doi.org/10.1155/2021/1331428

Muwonge BS, Pei T, Otim JS, Mayambala F (2020) A joint power, delay and rate optimization model for secondary users in cognitive radio sensor networks. Sensors (switzerland) 20(17):4907–4918

Naeem MA, Ali R, Alazab M, Yhui M, Bin Zikria Y (2020) Enabling the content dissemination through caching in the state-of-the-art sustainable information and communication technologies. Sustain Cities Soc, vol. 61, article 102291

Ostovar A, Bin Zikria Y, Kim HS, Ali R, (2020) Optimization of resource allocation model with energy-efficient cooperative sensing in green cognitive radio networks, IEEE Access, vol. 8, pp 141594–141610

Qin M, Yang Q, Cheng N, Zhou H, Rao RR, Shen X (2018) Machine learning aided context-aware self-healing management for ultra-dense networks with QoS provisions. IEEE Trans Veh Technol 67:12339–12351

Sheth (2016) Internet of things to smart IoT through semantic, cognitive, and perceptual computing, IEEE Intell Syst, 31(2), pp 108–112

Tarafdar M, Beath CM (2018) Wipro Limited: Developing a cognitive DNA. In: Thirty ninth International Conference on Information Systems, San Francisco, vol. 3, pp 6–7

Wang Y (2002) Keynote: On cognitive informatics. In: Preceding 1st IEEE International Conference on Cognitive Informatics (ICCI'02), Calgary, Canada, IEEE CS Press, August, pp 34–42

Wang Y (2003) On cognitive informatics brain and mind: A Trans disciplinary Journal of Neuroscience and Neorophilisophy, 4(3), 151–167. Kluwer Academic Publishers

Wang Y (2007a) Keynote: Cognitive Informatics Foundations of Nature and Machine Intelligence. In: Preceding 6th IEEE International Conference on Cognitive Informatics (ICCI'07), Lake Tahoe,CA, USA, IEEE CS Press, pp 2–12

Wang, Y ((2007b)) The theoretical framework and cognitive process of learning. In: Preceding 6th International Conference on Cognitive Informatics (ICCI'07), (pp. 470–479). IEEE CS Press

Wang S, Chen M, Liu X, Yin C, Cui S, Poor HV (2020) A machine learning approach for task and resource allocation in mobile-edge computing-based networks. IEEE Internet Things J 8:1358–1372

Xu H, Gao H, Zhou C, Duan R, Zhou X, (2019) Resource allocation in cognitive radio wireless sensor networks with energy harvesting. Sensors, 19(23)

Yu H, Afzal MK, Zikria YB, Rachedi A, Fitzek FHP (2020) Tactile internet: technologies, test platforms, trials, and applications. Futur Gener Comput Syst 106:685–688

Chapter 2
Green IoT Networks Using Machine Learning, Deep Learning for 5G Networks

Neha Jain, Ritesh Pratap Singh, Harsh Arora, and Krishanu Kundu

Abstract Internet of Things (IoT) alludes to the enormous network interconnection of items frequently furnished with universal savvy utilized to offer savvy facilities to end clients. Although, one of the significant problems of IoT is the small power of IoT gadgets which are supposed to perform reliably for an extensive stretch of time irrespective to the battery substitution. Additionally, in search of universal IoT, quantity of IoT gadgets has detonated and tends to an enormous ascent in carbon impression of IoT networks. In such manner, power management of IoT and Green-IoT arose as exciting and alluring exploration points for both scholarly world and industry. We lead an extensive and a cutting-edge overview on novel power management procedures in IoT networks, in this chapter. Initially we introduce the difficulties of power absorption in Internet of Thing networks. After that, we will introduce novel and notable methods of power management for IoT that is yet emphasize the latest arrangements suggested in each method. Then, we will give a thorough overview of the latest power management methods for IoT environment. We will likewise introduce late patterns and new exploration points that may utilized for power preservation in IoT. At last, we will present some proposals on the most proficient method to utilization of the methods introduced in our overview to accomplish the IoT uses QoS necessities.

Keywords IoT · Machine learning · Deep learning · Power management

N. Jain
Department of Electronics and Communication Engineering, Shobhit Institute of Engineering and Technology, (Deemed-to-be-University), Modipuram, Meerut, UP 250110, India

R. P. Singh (✉)
School of Electrical and Computer Engineering, Haramaya Institute of Technology, Haramaya University, P.O. Box 138, Dire Dawa, Ethiopia
e-mail: ritesh.pratap@haramaya.edu.et

H. Arora
Institute of Innovation in Technology and Management, Janakpuri, New Delhi, UP 110058, India

K. Kundu
Department of Electronics and Communication Engineering, G L Bajaj Institute of Technology and Management, Greater Noida, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
A. Rai et al. (eds.), *Paradigms of Smart and Intelligent Communication, 5G and Beyond*, Transactions on Computer Systems and Networks,
https://doi.org/10.1007/978-981-99-0109-8_2

2.1 Introduction

A prototype which envisages the accomplishment of the daily targets and incorporating them to the web with μC (microcontroller) based system, comprising of transmitter–receiver section is known as IoT. The prototype encourages both researchers and enterprises and focuses on the computerization of many applications with maximum physical significance. In our day-to-day routine exercises, the involvement of IoT is complex, from intelligent transportation, which offers momentous advancement in transportation and administrative organization to savvy home that simplifies its resident's actions. IoT covers many fields, for example, savvy traffic, medical services, farming, industry 4.0 and might in fact be found in our ordinary items like pens, garments, and so on. Thus, the society has known a blooming development in the quantity of these gadgets. As per IHS Markit, up to year 2030, approximately 125 billion of IoT gadgets are supposed to be associated with one another (Howell 2017). This growth of 5G and exposure of new range frequencies profoundly impacts on these numbers (Li et al. 2018).

In addition, IoT gadgets are normally compact and battery restricted and due to high information exchange between these gadgets brings about high energy prerequisites. These prerequisites are frequently not upheld by IoT gadgets and can rapidly prompt battery consumption and the network loss. So, what are energy efficient ways out that won't just minimize the energy utilization in IoT networks and their carbon impressions yet in addition to enhance the network lifetime? Likewise, we see the development of the Green IoT which mainly emphasises on regulating power in IoT networks to diminish the power utilization and improve the life of IoT networks. In future, It will roll out huge improvements in our day to day life and will assist with understanding the vision of "green ambient intelligence" (Al Turjman et al. 2019).

In this chapter, we will introduce a wide scope of energy collecting and energy saving methods and create a new trend for Green IoT, for example, social green IoT, 5G green IoT, green network digitalization, backscatter communication, and so on. This chapter is planned as: In second part, we describe the literature on how power management in IoT works. In third part, we discuss difficulties of power utilization in networks. In fourth part, few energy farming and energy saving strategies introduced in literature of researchers are described. In fifth part, management of power for IoT ecosystem is discussed. In sixth part, we describe the new point of view of researchers and trends that can be used in management of energy for IoT. At last, we will finish up the chapter by giving a knowledge into prospective of Green IoT.

2.2 Recent Advances in 5G IoT Ecosystem

The topic of management of energy for IoT was broadly reviewed and many audits and studies has been introduced over time. Many researchers are keen on energy saving Wireless Sensor Networks as WSNs are broadly exploited and ubiquitous in Internet of Things. The researchers introduced battery-driven power preservation and power farming methods utilized to plan algos in WSNs, in Khan et al. (2015).

The researchers mainly reviewed the protocols utilized for the management of power of WSNs. In this chapter, sensors and various gadgets Internet of Things networks are discussed. Many reviewed power management and power efficient methods and power saving methods for Wireless Sensor Networks concentrating on management of energy transmission and various methods, in Singh et al. (2020). The researcher reviewed the power farming methods utilized in Wireless Sensor Networks that help to further develop Wireless Sensor Network lifetime. Various power saving and power farming methods for sensors as well as expensive assets gadgets are described in this chapter.

2.3 Green IoT Enabling Technologies

As in our daily life we are using IoT facilities, and carbon impression of these uses is expanding correspondingly with quantity of IoT gadgets. Gadgets with rich and restricted power assets collectively made IoT networks. The main purpose of Green IoT is to minimize the power utilization of both gadgets as well as performing their facilities and functions properly. Main and broad IoT advancements, for example, WSNs are made out of tiny and battery restricted sensor gadgets that accumulate data and construct effective and dependable data and communication devices. However these gadgets are normally utilized to operate observing and control jobs in IoT applications for prolonged period without any interference. Sensors may likewise be put in regions which are hard to get to (for example submerged networks, military organizations, risky conditions, and so forth) and may immediately exhaust of power. Therefore, changing batteries regularly turns into a burdensome and expensive affair. Power capacity and energy management ought to guarantee the steady accessibility and functionality of IoT gadgets since neither battery substitution nor cable-energy is particle choices in cruel circumstances.

Numerous power saving hardware and software methods have been created to minimize power utilization in IoT networks and service expense. Right now, the society encounters a shortage in power sources with the significant discharge of Carbon dioxide due to human action. Greening data and communication methods turn into a peremptory requirement with consideration of harms due to our current circumstance. Subsequently, power gathering methods have been concentrated to control IoT gadgets with sustainable and green power sources like solar power, wind power, mechanical power, and so forth. Actually, along with power restricted nature of IoT gadgets and their big number and power utilization, estimating and investigating the power utilization of IoT gadgets is critical to accomplish applications' QoS necessities. Because of the significant expense and enormous size of current commercial power estimation methods, different other options are present. But these options are generally restricted with their intricacy, estimation reach and precision (Dezfouli et al. 2018). IoT power estimation instruments must be normalized to have the option to help a wide range of gadgets with their various intricacies for a minimal price.

As obvious, while the breakaway of IoT changing our lives has begun, many obstacles actually required to get over concerning power utilization and ecological problems. Various expected methods in various sectors are being investigated to accomplish this objective that needs a ton of endeavors.

2.4 IoT Ecosystem Energy Management Techniques

To handle the power problems in IoT organizations, two methodologies are normally utilized: power saving and power gathering. Power saving plans mainly emphasize on creating protocols, algorithms, and also hardware resolution to limit network power utilization and expand its lifetime. While Power gathering can be portrayed as a procedure utilized to generate power from network surroundings to give a continuous power to network gadgets. According to these two classifications, Fig. 2.1 shows new power management methods to maximize the lifetime of IoT networks and minimize carbon impression of networks.

Powers saving methods are divided according to the technology they used. Power saving method generally uses two types of technologies; Deep Learning and Machine Learning to minimize power utilization in IoT networks. Power efficient data collection method and sleep/wake methods are two main technologies used for power saving.

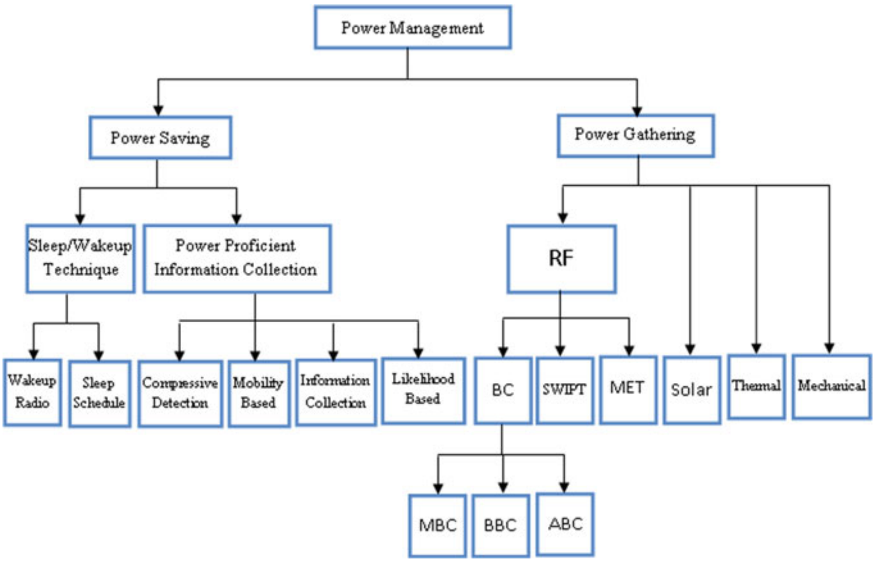


Fig. 2.1 Novel power management methods

Power gathering can be categorized into four segments: RF power gathering, Solar power, Mechanical power, and Thermal power.

2.4.1 Power Saving Techniques

2.4.1.1 Sleep/Wake-Up Technique

(i) Wakeup Radio: Wakeup Radio is a new hardware method developed to minimize the power utilization of gadgets and maximize the lifetime of networks. It comprises of a minimal energy radio connected to the gadget. When an incoming transmission is detected, the primary radio wakes it up. In IoT, Wake-up Radio is a μC module connected with a wakeup recipient and primary radio (Froytlog et al. 2019). The μC and primary radio are in profound sleep more often. As displayed in Fig. 2.2, when the recipient gets a call through a wakeup alert communicated by the external hub, it produces an interference wave to awaken the μC to execute the specific job and send the information back by means of its primary radio.

According to power utilization, there are three kinds of wakeup radios (Piyare et al. 2017): active wakeup radio, passive wakeup radio and semi active wakeup radios. Active wakeup radios need uninterrupted external power like batteries. In passive wakeup radios, there is no need of external power. They gather power from the communicated signal itself. While in semi active wakeup radio, few parts from the receiver end need constant energy from external supply and the RF front-end stays passive.

Wakeup radios may be used in smart offices, smart cities, smart homes, or in industry warehouses (Deng et al. 2019). But, because of wakeup and sleep wake-up characteristic of the wakeup radio, there is an increment in information gathering idleness. To design a wakeup radio in a huge climate various design viewpoints are considered (Piyare et al. 2017),

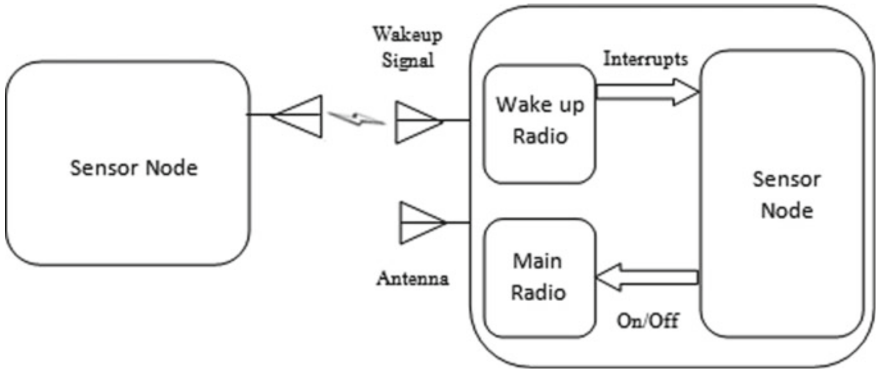


Fig. 2.2 Wakeup radio

(a) The hub's energy utilization cannot surpass sleep energy of the primary radio to maintain harmony between the power saved and utilized by the gadget.
(b) The time expected to wake up the gadget radio ought to be least to minimize the idleness and finally maximize the range of utilizations of the wakeup radios.
(c) Because of interruptions and incorrectly getting signals planned to different gadgets, a hub or many hubs could be unessentially initialized causing colossal power misspend. So, planers ought to track down the exact difference and track down the solution for difficulties without expanding the energy utilization.
(d) In the Wakeup Radio surrounding, the sensitivity of receiver hub extensively affects the energy requirement as there is need of high energy at the receiver due to high sensitivity and less sensitivity implies requiring high radiating energy at the transmitting end.
(e) The scope of the Wakeup Radio affects the kind of use it is utilized in, as well as on the energy utilization of system. Such as, for increment in the hub's density and energy utilization, multi-hop communication is needed in small range communication.
(f) Information rate: To enhance power efficiency and wakeup idleness, high information rate in wakeup radio is the best solution. Likewise, to increase the wakeup signal's reliability and communication range, maximum bit duration is needed.
(g) Wakeup Radio is ought to be financially savvy to be incorporated into IoT and current sensor hubs. The expense shouldn't surpass 5–10% of the total sensor hub expense.
(h) Wake-up Radios ought to be designed with frequency guidelines of the ISM band.

(ii) Sleep Schedule

Sleep schedule, also known as duty cycling, is viewed as the best known approach to saving power and increasing the lifetime of networks. Sleep schedule algorithms are utilized to decrease the hub's battery utilization by switching back and forth among wakeup and sleep modes. The hub is generally set in a small energy mode and awakens when required. These algorithms generally described under three protocols: MAC protocol, wakeup/sleep protocol, and topology control.

Asynchronous sleep schedule techniques go through extreme transmission power utilization when it waits for the receiver to wake up with various unicast transmissions which lead to network clog as well as utilization of power (Passos et al. 2019). There is a protocol to enhance the sleep schedule time span of small energy IoT gadgets furnished with power gartering circuits in IoT networks with Wi-Fi. In this technique, to foresee the harvested power, an ideal weighting moving-normal filter is used and estimates the ideal weighting variable to assess the IoT gadgets' span of off periods. The gadgets will take advantage of off periods for reaping sufficient power for information securing. There is one another novel sleep schedule proficient track designing algorithm for precision horticulture with IoT (Dhall and Agrawal 2018). This method is executed on the information transmission stage and track designing algorithms research on all tracks and picked the smallest track. This method is also successful for countless gadgets. But, performance of this method relies upon the groups, so as to minimize total power utilization that it is required to enhance the group association.

2.4.1.2 Power Proficient Information Collection

A huge number of IoT gadgets are utilized in smart applications such as savvy homes, savvy transport, etc. These gadgets create colossal amount of information. To convey its information to its adjacent and base station, an IoT gadget will utilize a lot of power. Besides, IoT networks are generally comprised of an enormous number of gadgets with restricted batteries. When exhausted or drained, it is not easy to change the batteries particularly when the gadgets are set in a segregated or hard to get to position. To minimize power expenses on transmission, gathering of the produced information and conveying it in an effective manner is a good solution. Many information gathering methods have arisen to decrease the power utilization, delay and expand the lifetime of the network (Chandnani and Khairnar 2020).

(i) Compressive Detection Based Method: In the compressive detection-based method, gathered information is compressed and the network's transmit information rate is minimized. Compressive detecting assists to minimize the size of the information matrix of the sensor to decrease data and communication repetitiveness. Compressive detection-based methods can be divided into three types: Distributed compressive detecting, Distributed source coding, and Conventional compression (Chandnani and Khairnar 2020). Information compressing methods are depicted by their portrayal proficiency, computational intricacy, memory management and compatibility. Information compressive methods ought to enhance these attributes to upgrade the productivity of IoT networks and minimize the gadgets' intricacy. There is one another power proficient compressive sensing-based protocol that is available known as the clustering routing method (Wang et al. 2019). This method evaluates the optimal cluster size, cluster heads distribution, and it also mitigates the "hot spot issue". This method minimizes the total power usage of the network very proficiently and expands the network's lifetime. There is a temporal relationship between the information which implies it may be fascinating to involve a predictive methodology to decrease power utilization due to spacial-temporal connection.

(ii) Mobility Based Information Collection: Nowadays, Mobility based information collection methods are generally utilized as a mode to foster power proficient IoT networks and expand the lifetime. This method permits battery-restricted gadgets to conserve the gadget's transmission power and subsequently broaden the lifetime. This can be accomplished by three ways.

First, technique comprises in utilizing the static sensor hubs and mobile sink. They collectively gather information from hubs when they are adequately near it. Second, techniques have mobile hubs and a static sink. As the hubs can move, they further enhance the coverage of the network and reduce required the hubs' quantity. Third, the technique comprises of mobile sensor hubs and a portable sink.

There is an internet of vehicles dependent method which, permits traffic information collecting design with an effective sink hub choice (Wang et al. 2018). This method shows specific vehicles and ways ought to be chosen to diminish network power utilization. Mobility based information collection is an extremely encouraging power saving method if carbon impressions are permitted by mobile sink and is given an enhanced direction.

(iii) Information Aggregation Method: Information Aggregation approach: This method comprises in minimizing the network transmission rate and eliminates excess and replicated transmission of information. Information aggregation is the method of one or various hubs that accumulate information from different hubs for a while, aggregate and transmit the outcome. The information aggregation methods can be divided into three types: tree-based information aggregation method, cluster-based information aggregation method and centralized based information aggregation method (Chandnani and Khairnar 2020). The tree-based information aggregation methods have low power utilization, huge network lifetime and extraordinary versatility. Cluster based information aggregation methods have low traffic and huge adaptation to internal failure. Centralized based information aggregation methods emphasize generally on security (Pourghebleh and Navimipour 2017). The tree is made out of hubs organized as a tree. A tree contains leaf hubs and sending hubs. This method uses less energy in comparison to other two. But, inadequacy of intermediate hub influences the entire topology and affects the network's operation (John and Jyotsna 2018). Cluster based information aggregation method utilizes a sensor-based group foundation algorithms to choose the μC and operate various kinds of sensors. This method permits the user to expand the lifetime of the network, amplify the inclusion and terminate repetitive information at μC. But these methods expand the network's weakness. These methods likewise consider transmission and sleep inertness (Song et al. 2017). Centralized based information aggregation method assembles information and transmits information to the centralized gateway of IoT. To guarantee the network's information consistency, these techniques have an effective sleep schedule strategy with information aggregation. The centralized based information aggregation method is safer than the other two but this method has low adaptation to internal failure (Ko et al. 2019).

(iv) Likelihood based method: The likelihood-based approach saves power by utilizing the estimate models to discover the relation between information. To anticipate the benefit of the sensed information is the main objective of likelihood-based methods. Bayesian method eludes the transmission of exceptionally connected sensor information and minimizes power utilization in IoT networks (Razafimandimby et al. 2018). This method was used on a hierarchical structure with savvy gadgets and information centers. This method minimizes the quantity of transmitted information with great data precision in three distinct situations. The devices used in likelihood-based methods are complex and the method does not consider high connection between various devices' information. Due to these reasons, this method is complex to execute.

2.4.2 Power Gathering Methods

Mainly four types of power gathering methods are available; RF Power Gathering Method, Solar Power Gathering Method, Thermal Power Gathering Method, and Mechanical Power Gathering Method. All These types of methods are explained below.

2.4.2.1 RF Power Gathering

RF power gathering method is also known as the wireless energy transfer method. These techniques mainly emphasize on the backscatter method as the backscatter method is a developing method in power gathering.

(i) **Backscatter Communications (BC)**

Since all IoT gadgets are battery dependent so the lifetime of IoT networks is very small and this is the main problem with IoT devices. As these gadgets work for a long period batteries easily depleted. This initiates a substitution expense. Also, the sensors are not always active and due to the battery's breakdown unnecessarily. In BC wireless signals convert into source of power and also used as a communication mode for battery less gadgets. A battery less gadget help to minimize power utilization, eliminate the requirement of regular battery substitution and deflects from undesirable execution because of power lack.

BC have three different models (Hoang et al. 2020) as shown in Fig. 2.3: Monostatic (MBC), Bistatic (BBC) and Ambient (ABC).

(a) Monostatic Backscatter Communication: MBC is made out of two parts: Reader and Backscatter transmitter, as shown in Fig. 2.3c. In MBC recipient and RF source are situated in the reader. A signal is sent by the RF source which initiates the tag that modulates signal and sends it back to the receiver. This method is mostly used for short range RFID as recipient and RF source are situated in the same gadget so modulated signals might experience from round trip track loss.

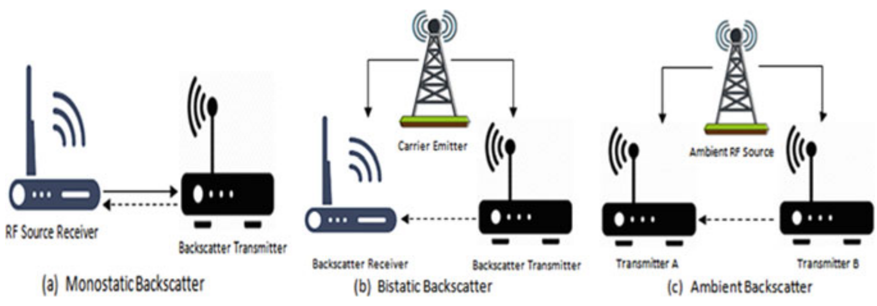


Fig. 2.3 Backscatter communication

(b) Bistatic Backscatter Communication: In BBC the RF source (carrier Emitter) and receiver are situated in different devices which minimize the round-trip track loss issue (as shown in Fig. 2.3b). Additionally, by ideally putting carrier emitters, the total communication range may be increased and the performance of the system is enhanced. However, development and maintenance expense in BBC is high but they have a simple circuit. Because of this BBCs are cheaper than MBCs.

(c) Ambient Backscatter Communication: Ambient backscatter communication is approximately same as BBC (Fig. 2.3c) as its carrier emitters are isolated from the backscatter recipient. Except that, it can have a connected RF source, for example, Wi Fi access or cell based node. An ABC can be a battery less gadget (sensor or tag) that may gather wireless power from RF and send a signal (either 0 or 1) to a sink hub (Zhang et al. 2019). The assembling expenses and power utilization of ABCs can be decreased by the utilization of minimal expense and low-power parts.

There is a cooperative ABC that permits a gadget to retrieve data from both the surrounding backscatter gadgets and RF (Yang et al. 2018). If RF signal time is smaller than backscatter gadget symbol time than this method improves the maximum likelihood detection capacity. This method may be enhanced to be used in savvy home and wearable uses. For complex IoT, different variables, (for example, the quantity of base stations, the quantity of backscatter gadgets and their location) should be enhanced to be power proficient and to adjust to IoT requirements. There are two techniques to achieve the above stated goals; Centralized MAC protocol (Feng et al. 2018), distributed MAC protocol (Ma et al. 2018; Cao et al. 2019). For huge IoT networks with temperamental network topologies and expensive maintenance, Distributed MAC is more appropriate between these two.

However, BC appear to be best the accomplish green IoT, however, it actually needs to conquer few difficulties, for example, the possibility of interruptions if there are countless hubs as well as, and the bit rate impediments as it isn't easy to carry out extremely complex regulations on a backscatter tag because of its restricted capacity in controlling the RF sign and executing baseband handling. Also, BC is required to conquer short range restriction and huge cost.

(ii) **Simultaneous Wireless Information and Power Transfer (SWIPT)**

It is most alluring and emerging innovation in power management. This empowers the synchronous exchange of data and energy wirelessly (Perera et al. 2017). Basically, this permits an IoT gadget to collect power and get RF data from a similar RF signal which conveys both power and data. So, an IoT gadget can be self-controlled by receiving information. However, generally, it is impossible to execute power gathering and data decoding at the same time from a similar signal. The data may be modified or lost by power gathering activity. So, various recipient models were intended to accomplish SWIPT: Antenna Switching, Time Switching (TS), Separate Receiver, and Power Splitting (PS) (Perera et al. 2017). In this method, energy and data required for IoT hub are transmitted simultaneously that retain the hub live and expand the IoT network's lifetime. By a double battery green power collecting design, SWIPT can achieve this (Liu and Ansari 2019). It comprises of a first battery to provide power to the gadget and second one to gather power to minimize the power deficiency likelihood. But this method is restricted to compact IoT networks having gadgets which may uphold a double battery. For big IoT networks, this method is complex and expensive. To fulfill the QoS prerequisites of every client, there is a method-based DL to limit all transmitted energy of a multi-transporter NOMA SWIPT (Luo et al. 2019). DL was additionally utilized with SWIPT (Park et al. 2020b). In this technique, there is a double mode SWIPT which utilizes a versatile mode changing dependent on DL to enhance both single tone SWIPT and multi tone SWIPT. However, DL is proficient for SWIPT-based device; it likewise needs a significant measure of assets that ought to be needed to construct green and maintainable IoT.

SWIPT is hypothetically an extremely proficient power collecting method. But this method has many difficulties. In SWIPT, data and power are transmitted through a similar signal. The signal should be effectively decoded to retrieve right data. This signal may change by the power gathering procedure or by interrupt. SWIPT is restricted in range as signal can debilitate while crossing significant distances. At last, signals transmitted to gadgets may be hacked. In this way, a degree of safety ought to be included to safe private and individual information.

(iii) **Mobile Energy Transfer**

MET is an aspiring power proficient method for IoT. It expands wireless energy transfer methods as utilization of traditional wireless power transfer methods isn't appropriate if hub is remotely placed, situated in regions which are hard to approach and doesn't uphold free charging. MET comprises in providing to a mobile hub (like a vehicle), the capacity to transfer power to gadgets inside its transmission range.

The power utilization of the mobile hub is a critical problem of MET empowered IoT. The different variables can affect the power utilization of MET-dependent IoT, for example, mobile base station areas and its speed, the quantity of gadgets and their average time for conveying information and power as well as the transmission energy (the energy required for power and information transmission) (Almasoud and Kamal 2017).

Adaptive Resonant Beam Charging (ARBC) is a MET method. As displayed in Fig. 2.4, it comprises of two specifically isolated parts: a sender and a collector (vehicle, TV, and so forth) and sometime, relay (UAV in Fig. 2.4) which is utilized to transfer the power between the sender and the collector.

2.4.2.2 Solar Power Gathering

Solar Power is reasonable and clean power source that can dispense the power deficiency issue in IoT. This method precisely changes sunlight in feasible energy which is helpful for area with huge light accessibility. Although, the solar power gathering method fails to emerge where light accessibility can't necessarily be ensured. Also, there is huge power wastage during transport of solar power from one place to another place.

Fig. 2.4 MET module

There is a smart solar power gathering dependent on maximum energy point chasing for wireless sensor hubs utilized in IoT, which likes to involve the solar energy and accepts Li battery as complementary under the state of insufficient enlightenment (Li and Shi 2015). In case of shortage of solar power nearby, this method gives great solidity of IoT. A power projection algorithm involves the intensity of light of the luminance lamp in the inner climate for wireless sensor hubs for IoT (Shin and Joe 2016). This method can be especially useful to accomplish green and feasible savvy home as lamps may be utilized to provide energy to sensors utilized in the savvy home.

2.4.2.3 Thermal Power Gathering

Thermal power is the power that generates from heat and changes in temperature. The main source of thermal power can be the climate and the human body. Such as, a variation in temperature may be generated through the human body and the climate relying upon the medical issue, the position of an individual, or body movement.

There are two methods that are utilized in thermal power gathering:

i. Thermometric—A method that right away changes temperature variation into useful power
ii. Pyroelectric—A method that produces electrical power with the variation in temperature.

It is essential to effectively produce electric power from thermal generator to generate the highest electric power to gadgets and utilize the maximum capability of this methodology. Thermal power gathering is very proficient for different IoT uses (Abdal-Kadhim and Leong 2018).

2.4.2.4 Mechanical Power Gathering

Mechanical power gathering comprises of gathering power that comes from motion, force, human venture or any kind of movement from an item. This method is an extremely proficient power gathering method as many and quicker the item moves more the power delivered to do much jobs. Mechanical power generates from different sources, for example, breeze and aqua movement or human movement and may be changed into electrical power and utilized to provide IoT gadgets. Such as, it tends to be fascinating to utilize the power generated from human movement to control a human temperature observing gadget that the client wears or driving adjoining houses utilizing power delivered by breeze and aqua movement. Mechanical power gathering gives a large number of chances so it is an optimal answer for making green and feasible applications.

Breeze power can be used in independent Wireless Sensor Networks and climate prediction to enhance the forecast of breeze condition in future (Jushi et al. 2016). This method utilizes a two-hour climate prediction window, if there is no breeze power which displays uncertain characteristics of breeze power making it unacceptable for IoT that needs consistent accessibility and real time executions. Kinetic power generated from human movement in wearables can also be used (Magno et al. 2016). The human movement can generate a huge quantity of power daily but, as the power generated can differ from one to other it may be challenging to foster a method that is bound to an enormous populace. Table 2.1 shows the basic idea and shortcoming of the power gathering methods.

Table 2.1 Power gathering methods and their shortcomings

Power gathering method	Idea	Shortcoming
Backscatter communication	1. Battery less gadgets 2. Utilize RF to code data and transmit it to receiver	1. Minimal range 2. Proficient data 3. Safety 4. High maintenance expense
SWIPT	1. Battery operated or battery less gadgets 2. Data and energy for gadget are transmit via same signal	1. Proficient data 2. Safety 3. Interrupt 4. Minimal range
MET	Utilize a mobile hub to control IoT gadgets	1. Power utilization of mobile hub 2. Interrupt
Solar power	Utilize solar power to control gadgets	1. Light availability 2. Power transport
Thermal power	Produce electric energy from difference in temperature	Proficient energy production
Breeze/aqua power	Produce electric energy from aqua and breeze motion	1. Difficult to maintain 2. Unforeseeable
Human movement	Produce electric energy from human motion	1. Low power 2. Changes from one person to another

2.5 Energy Management in IoT Cloud Computing Techniques

To accomplish Green-IoT, many novel methods for power management have been investigated by the specialists to minimize the power utilization of the various parts of the IoT from information centers to IoT hubs. These methods enhanced with the utilization of novel communication models such as 5G and the fast development of methods like AI, cloud, and so on. In this segment, we will introduce the new advances in power management for IoT networks environment.

2.5.1 Cloud Computing (CC)

The quantity of IoT gadgets and information gathered by those gadgets are increased with the revolution of IoT. As, IoT gadgets don't have the computational ability to handle the gathered information, cloud computing (CC) is an extremely encouraging protocol that gives very good performance computing capacity and high limit storage to help the universal requirement for information gathering and handling in IoT networks. This permits the battery-dependable IoT gadgets to offload different jobs to the cloud (for example information cleaning) and preserve their power. However, it may be easy to utilize this method, as it is obvious that cloud servers are exceptionally power utilizing. So, diminishing the cloud power utilization by utilizing its assets efficiently is necessary. Minimizing power utilization in CC may be accomplished by the utilization of low-energy gadgets and the virtualization of facilities.

The power proficient system for IoT permits the architecture to foresee the sleep timespan of sensor-dependent IoT with respect to their leftover battery and minimize the power utilization of sensor hubs and cloud assets (Kaur and Sood 2015). Although, this method isn't good for systems with real time needs, it is good for devices with wakeup/sleep methods. There are many methods for power saving with the help of service composition development in cloud computing. (Baker et al. 2017) The multi cloud Internet of Thing service composition method (E2C2) is the power saving method and to fulfill the client's necessities, it coordinates with the least number of IoT users. This method is better than the other used method: Smart Cloud, Base Cloud, All Cloud (Zou et al. 2010) and COM2 (Kurdi et al. 2015). To enhance this method, we have to consider the total power of the cloud source, yet additionally the power required for service performance. To minimize power utilization and performance time, there is a parceling method for a user with up loadable and non-up loadable components. This method uploads components on cloud but, it requires maximum network liableness to guarantee application well operation and accessibility.

Virtual Machine Placement (VMP) method is a power saving method for cloud source. This method is used to minimize the quantity of working physical machines and minimize their power utilization while additionally balancing the load between them (Gharehpasha et al. 2020). In any case if virtual machine (VM) is placed than it cannot be relocated to other physical machines. There is an architecture for dynamic VMP which minimizes the quantity of the working physical machine and decreases the total power utilization of the information center (Alharbi et al. 2019). In dynamic VMP, VMs may be relocated from one to the other. So, security procedures ought to be used during relocation to safe personal information from hackers.

2.5.2 Fog Computing (FC)

We can get a model and liberates business by CC and it also frees clients from the load of numerous parametric characteristics. But it doesn't comprise a successful answer for dormancy sensitive uses and frequently neglects to fulfill their delay necessities. Thus, Fog computing (FC) arose as novel model that comprises of an expansion of CC at the lead of the network instead to substitute it. FC is a deeply virtualized model that gives networking, storage, and computing facilities between conventional CC Information Centers and end gadgets (Bonomi et al. 2012). FC models are made to help the versatility, geo-distribution as well as low idleness needed in different IoT uses (for example industrial IoT). These parameters make the FC models appropriate to help the power limitations of Green-IoT uses.

FC models should design in a way that various applications can utilize. Such models ought to offer in-house help for various sorts of communication and information analytics structures to carry out the various characteristics of Fog-IoT models. There is an AI-dependent application placement method that directs idleness, asset usage and power utilization problems in fog empowered IoT networks (Nayeri et al. 2021). This method mainly emphasizes on accomplishing proficient load balancing techniques. ML is generally utilized and joined with developmental algorithms to solute the Fog application placement issue. Many algorithms may be joined to hide one and all's shortcomings and accomplish difficulties like safety, secrecy, accessibility, and so forth. To diminish end-gadgets power utilization for industrial IoT, there is a power proficient fog-dependent IoT model by decreasing their transmission (Peralta et al. 2017). This method utilizes a MQTT broker to foresee future information estimations through forecast methods, work as gateway and offload costly information handling from Cloud to Fog.

The linear FC protocol minimize the overall power utilization. However, these protocols are basic and helpful but they are just appropriate for IoT with small quantity of sensor hubs (Oma et al. 2018a). The tree-based FC protocol upholds an enormous quantity of IoT gadgets and minimizes the complete power utilization in comparison to CC (Oma et al. 2018b). A distributed learning protocol is developed by mining the gathered information that estimated the information nature and afterward sending just the updated protocol characteristics to a FC model. This method saves power in IoT gadgets by restraining nonessential uplink transmission (Lavassani et al. 2018).

2.5.3 Edge Computing (EC)

With the rising number of IoT gadgets, the information generated by those gadgets is likewise increasing rapidly. Handling proficiently those information turn into a need in Green-IoT. In this way, Edge Computing (EC) methods arise as an extremely alluring solution for power management in IoT because they overcome the gap between restricted capacity of low-powered gadgets and computational requests. EC permits the responsibility to be uploaded from cloud to an area nearer to the source of information that should be handled. It gives the chances to downstream information for cloud and upstream information for IoT. It subsequently has the capacity to increase the lifetime of the battery compelled IoT gadgets, decrease network traffic, extend BW, enhance secrecy and accomplish an impressive communication delay and power saving.

There are many methods to accomplish power saving in EC based IoT. Power proficiency utilizing EC may be accomplished via various plans: By uploading methods to information decrement plans as well as proficient management of IoT gadgets and applications, and by using a design with cloud, mobile EC and IoT, the power utilization and versatility issues can be handled (Sun and Ansari 2016). It incorporates a structure which creates a specific uploading plan to minimize power utilization of IoT gadgets and aerial of edge gadgets. Also, this design requires to defeat many difficulties related to hubs ID by VMs (particularly in the case of position change) also the hubs and VMs movement. As a matter of fact, if it changes its position, the model ought to find an effective method that can either relocate the VM or is not dependent on the QoS necessities. There is a novel reinforcement learning (RL) dependent uploading method for IoT gadgets with power collection (Min et al. 2019). In this method, IoT gadgets chose an edge gadget and their uploading rate as indicated by their present battery, the past radio transmission rate to every edge gadget, and the forecasted measure of collected power (Liu et al. 2019). To decrease power utilization in savvy home, EC with deep RL in an IoT-based power management. This method works in an online mode as well as offline mode.

In offline mode, Deep Neural Network (DNN) and a deep RL protocol dependent on Q-learning is fabricated and utilized in the online mode. This method minimizes the power expense and may be used to diminish power utilization of savvy homes in the savvy city scenario. Another method for great performance uses with great information generation to compress the information gathered through edge gadgets before transmission (Azar et al. 2019). This method depends on the error limited lossy compressor. By this, exact grouping precision is same as the compressed information.

2.6 Savvy Power Management Techniques for Internet of Things

Many actual gadgets, for example, automobiles, homes and some other things improved with sensors and software that empower data collection, dispersion and evaluation collectively known as the IoT network. But, because of the great quantity of Internet of Things gadgets with their little dimensions, gadgets don't have the analytical ability to handle all information gathered; thus, ML and big data innovations are productive ways to enhance smart solutions to gathered data management.

Big data innovations permit the assortment and evaluation of a lot of information while ML concentrates on algos and analytical models depend on prototype and intercession as to which system to utilize to fulfill their objectives (Machorro-Cano et al. 2020). Moreover, ML is an entirely dependable and powerful methodology, in this way; it is widely utilized in actual time uses. ML is observed best for erudition depended issues and equipped for distinguishing foundation and qualities of these issues to gain from ML and perform activities that depend on erudition procured and subsequently enhanced system executions.

Because of the great quantity of Internet of Thing gadgets, power used by these gadgets is also high. Likewise, the planet encounters a high ecological emergency along with environmental turnarounds and huge paces of water adulteration, air adulteration, and so forth, while diminishing energy utilization in IoT networks is the toughest test we could confront. Analysts can take advantage of ML and smart arrangements to anticipate energy utilization in Internet of Thing networks and uses.

Different arrangements are given to decrease power utilization of Internet of Thing gadgets. But these arrangements are normally stable, restricted to specific setting and cannot take savvy decisions. For example, various expansion techniques that have been introduced for energy management. But those techniques are relying upon a specific setting, so that, ML has arisen as a promising way to create savvy arrangements for power efficiency in savvy homes as these homes are capable, utilizing savvy meters information, to lessen power utilization as well as accustom to the decision taking relying upon the in-habitants' propensities. Indeed, it would be intriguing to take advantage of ML to change transport planning relying upon air adulteration or activity decontamination systems if sensors recognize a huge water adulteration rate, and so on. Consequently, numerous analysts show interest for accomplishing Green-IoT at various network layers utilizing savvy arrangements such as ML and DL (Deep Learning) that demonstrated their effectiveness in different fields.

ML and DL, two Artificial Intelligence (AI) approaches, are generally confounded as DL are known as a subsection of ML while, we think it is necessary to distinguish them as ML and DL as two ideas utilized in various settings and applied fields.

ML may be characterized as algos which gives the system capacity to recklessly grasp and further develop its potential. ML algorithms grasp from encounters and use algorithms which have grasped to take the right choice. They typically include individual intercession to provide input to algos. ML algorithms are generally divided as supervised learning and unsupervised learning. The learning gained from previous information is utilized as Supervised Learning and uses this data to new information utilizing labeled samples to anticipate subsequent activities or acts, but unsupervised learning utilizes unlabeled and unclassified information and assists the system with finding a wide range of unknown patterns in information. One more class of ML is semi-supervised ML which utilizes both previous explained learning. Normally, those algorithms utilize a modest quantity of labeled information and remaining in typically unlabeled information. Semi-supervised algos decline labeling exertion and increment learning exactness of the system. At last, reinforcement learning is a methodology wherein the learning strategies comprise in cooperating with the climate by making moves and getting rewards or errors. Inside a particular setting, this strategy expands the system's performance and consequently decides its optimal way of behaving.

DL algorithms are a development of ML algos as these algorithms utilize a layered and programmable formation of algorithms called ANN (Artificial Neural Network). ANN empowers systems to settle on precise choices without human mediation. DL varies from the other ML algorithms as it can achieve complex tasks utilizing unstructured information, process a lot of information and grasp proficiently without human mediation. In this chapter, we will introduce ML and DL-based arrangements utilized to lessen the energy utilization of IoT networks.

2.6.1 Machine Learning

Along with the development of Internet of Thing and the detonating quantity of Internet of Thing gadgets, the amount of information created by Internet of Thing systems likewise increments impressively. Therefore, ML methods have arisen to utilize the information produced by IoT gadgets and make IoT application savvier. But, because of the great number of IoT gadgets, the energy used by those gadgets has likewise expanded significantly, in this manner, power saving is turned into the need for Internet of Thing systems creators and end-clients. The idea propelled analysts to concentrate on utilization of ML to make green savvy and practical IoT-network applications.

The quantity of effort concerning the utilization of ML to accomplish energy saving is huge. To assist on decreasing the energy utilization, ML may be utilized in power saving Internet of Thing uses, for example, Smart premises (Zhao et al. 2018; Thomas and Cook 2016), Smart Energy (O'Dwyer et al. 2019; Sheikhi et al. 2015), Smart Water (Mounce et al. 2015). Different works emphasize saving energy by proficiently transmitting information and decreasing the quantity of transmitted bundles. While in Paris et al. (2019), the researchers mainly emphasized on power proficient mobility expectation with the use of ML. This shows that the information highlights are known and with the accessible datasets, ML algorithms end up being a fascinating arrangement that may be utilized and adjusted to various parts of power saving in Internet of Thing networks.

2.6.2 *Deep Learning*

DL is optimistic methodology for management of power in Internet of Thing networks. Table 2.2 shows that DL varies from ML as DL gives preferable operation along with high scale information. DL algos can be independent on information evaluation and consequently extricate novel elements for issues as contradiction to ML algorithms which rely upon the extricated and heretofore distinguished highlights.

DL has been utilized in a few works connected with energy saving. In Liu et al. (2019), deep support erudition-dependent power planning was suggested to manage vulnerability and irregularity of power reserves. Deep support erudition method was suggested to get management procedure of power for a melded electrically powered vehicle, in Hu et al. (2018), Liessner et al. 2018). Then again, in Wu et al. (2018), deep Q-erudition dependent management procedure of power for energy split melded electrically powered buses is suggested. As described in studies already, DL may utilize in various power saving uses. But DL is much acceptable as high scale uses need consistent erudition and these necessities to adjust to the condition of its surroundings.

2.7 Application of Energy Management in Various IoT Applications

In this segment, we will introduce proposals and signs to IoT model planners on which methods introduced in this chapter assist to accomplish the prerequisites of IoT uses in a power proficient way. The application ranges of IoT in a single space are diversified so the necessities can change in each application similar space. Such as, utilizing battery-less gadgets with BC, real time transmitting energy and data with SWIPT, as well as intermittently shifting among wakeup and sleep levels for IoT gadgets with sleep schedule and wakeup radios which may be joined with ML methods to enhance the wakeup and sleep time. These methods are exceptionally encouraging and are

Table 2.2 For and against the energy saving techniques discussed in the survey

Savvy energy saving techniques	For	Against
Machine learning	1 Performance enhances with time 2 Helps to recognize latest patterns and trends	1. Energy consumption at the time of algorithms execution 2. With the expansion of data, algorithms outmoded 3. Bad results can occur due to immaterial and vital data
Deep learning	1. Suitable for difficult task 2. Supports huge quantity of data	1. Bad results can occur due to bad quality data 2. Obscurity of neural networks 3. High data security

utilized in situations where the gadget needn't bother with to be continually working. 5G innovation is coordinated with IoT networks to help to increase communication and reaction time as well as the limit of IoT networks. Safety is likewise significant for IoT uses that utilize personal and crucial information like savvy homes, medical services and industrial IoT. IoT networks planners ought to consider safety at various levels of the network with the safety model as well as software development which not just permits to save power by changing physical safety gadgets yet additionally to have strong safety components. Green power sources additionally should be utilized for different applications to control the gadgets and guarantee the accessibility of the gadgets and continuous information.

2.7.1 Smart Home

IoT is radically changing the way in which families are controlled. Apparatus and home gadgets are furnished with communication terminals and wirelessly and automatically handled to offer an agreeable encounter for the residents. So, savvy homes facilities should be accessible whenever and with low idleness. In this way, handling information at cloud, fog and edge level permits to have a fast reaction (Mao et al. 2018). Task uploading at various stages of the model likewise upgrades the accessibility of facility and permits a fast retrieval if one of the parts quits working (Chowdhury et al. 2020). In case savvy home sensor information are communicated to cloud, fog and edge for handling and as communication expense more power than nearby computation, power proficient information communication methods need to be produced for savvy home via lightweight systems, information decrement, information collection plans (Pattamaset and Choi 2021) and social IoT (Kim et al. 2015) to utilize the connection between the gadgets to find power proficient communication paths. Savvy homes methods ought to be adaptable and contextual, and these methods may involve ML and DL methods to adjust to the client's requirements and minimize the power utilization (Jo and Yoon 2018).

2.7.2 Agriculture

IoT was launched in horticulture to enhance horticultural crop and characteristics for the rising planet populace. IoT is utilized to gather real time data about soil, temperature and humidity and send them to a stage where they can be utilized to observe the yield. IoT gadgets in SA (smart agriculture) uses should be authentic via climate conditions and accessible for a really long time without the requirement for battery substitution. Additionally green power sources and savvy wakeup/sleep methods, power proficiency in horticulture IoT gadgets may be accomplished via utilizing mobile vehicles, for example, farm trucks as energy relays and information sinks guarantee the accessibility of the application information. A blend of wakeup/sleep methods and UAV dependent mobile power collecting is done for a SA module (Jawad et al. 2019). Horticultural applications are in steady extension. Hence, sensor information management is a key test that is required to be considering via various gadgets organization methods as well as information decrease and adaptable information gathering methods.

2.7.3 Healthcare

IoT is gradually coordinated in medical services to enhance access and upgrade the nature of care. IoT medical care applications mean to accomplish great medical facilities at reasonable expenses. IoT in medical services can be utilized to observe patient actions and extricate data that will further develop prosperity, medical care and patient help. Patients' clinical data are serious, for this reason medical services application ought to safeguard information with secure transmission protocols. Medical IoT application is ought to offer continuous data about patient's wellbeing with low reaction time that may be conceivable with democratization of 5G. To guarantee proficient patient help and facilities, IoT medical care facilities ought to be accessible regardless of whether one of the parts is not working and gadgets ought to be operate through strenuous condition for an extensive stretch of time (for example wearable pulse rate observer). The most appropriate methodology to meet those necessities is social IoT. IoT gadgets in medical services are mobile generally, the association and connections between these gadgets may be utilized to guarantee the accessibility of Serious data and the module's adaptation to failure.

2.7.4 Industrial IoT (IIoT)

IoT has as of late advanced towards the business. It alludes to the connection between all the industrial resources and means to gather the highest information from these resources for accomplish ideal industrial activities. Industrial uses that are security-crucial and utilized for handle and choice making should be accessible and reputable. To accomplish IIoT necessities of real time execution, 5G techniques (Slalmi et al. 2021) that give low idleness may be joined with CR-IoT (Abbas et al. 2020) to fulfill all deadlines of jobs and improve the frequency range use and in this manner, diminish the holding up time. EC and FC may be utilized to execute power proficient computation at the edge of the module (Hsu et al. 2019) with low idleness rather than controlling them on cloud that prompts a more extended transmission time and extended power utilization. With lot of information produced in IIoT, a trade off should be made between great accessibility of IIoT and huge power utilization of information transmission. So, IoT gadgets information ought to be coordinated and decreased via information aggregation and decrease the methods introduced in this chapter to diminish the transmission information power utilization and idleness in

IIoT applications. Same as in savvy homes, IIoT applications ought to involve AI to offer an adaptable and contextual encounter and utilize cell phones to drive different gadgets via savvy cell power transfer methods (Park et al. 2020).

2.8 Summary

Along with consistent necessity of the person to get a protected, congenial and steady climate, Green Internet of Thing arose as an essential and novel innovation. This ensures abatement in power utilization of Internet of Thing. In future, the quantity of IoT gadgets will be significantly higher as large organizations (for example Google, Amazon, and so forth) are putting resources into the area. Hence, organizations and association ought to foster IoT technologies from an eco-friendly and energy effective angle. Recycling and reusability of IoT gadgets ought to be a prime focus of study of Green Internet of Thing to stay away from power producing because of the method involved with making novel savvy gadgets. Thus, to get Green and reasonable IoT, we are urgently requiring for novel IoT approaches and normalization which incorporates standards of Green IoT. We mainly emphasized on collecting an enormous scope of management of power which will assist with accomplishing Green IoT systems, in this chapter. We likewise described the latest technologies and new points of view in power management of IoT networks which was not described beforehand in literature. This is essential to introduce the different element such as QoS and security may additionally expand the power utilization of Green Internet of Thing networks. As long as, to order great exposure to the client, power absorbing arrangements which accomplish a decent QoS have frequently favored to a green explication might need a part of QoS. Another power absorbing element of Green Internet of Thing is Security. Along with great development of Internet of Thing networks and great connectivity of IoT hubs, cyber security gambles with increment in a like manner. Private and personal information will likewise course by IoT networks. Accordingly, this is essential to address the cyber security problems.

References

Abbas G, Abbas ZH, Waqas M, Hassan AK et al (2020) Spectrum utilization efficiency in the cognitive radio enabled 5G-based IoT. J Netw Comput Appl 164:102686

Abdal-Kadhim AM, Leong K (2018) Application of thermal energy harvesting in powering WSN node with event-priority-driven dissemination algorithm for IoT applications. J Eng Sci Technol 13(8):2569–2586

Al Turjman F, Kamal A, Husain Rehmani M, Radwan A, Khan Pathan AS (2019) The green Internet of Things (G-IoT)

Alharbi F, Tian YC, Tang M, Zhang WZ, Peng C, Fei M (2019) An ant colony system for energy-efficient dynamic virtual machine placement in data centers. Exp Syst Appl 120:228–238

Almasoud AM, Kamal AE (2017) Efficient data and energy transfer in IoT with a mobile cognitive base station. In: 2017 IEEE 28th annual int. symposium on personal, indoor, and mobile radio communications (PIMRC). IEEE, pp 1–5

Azar J, Makhoul A, Barhamgi M, Couturier R (2019) An energy efficient IoT data compression approach for edge machine learning. Futur Gener Comput Syst 96:168–175

Baker T, Asim M, Tawfik H, Aldawsari B, Buyya R (2017) An energy-aware service composition algorithm for multiple cloud-based IoT applications. J Netw Comput Appl 89:96–108

Bonomi F, Milito R, Zhu J, Addepalli S (2012) Fog computing and its role in the internet of things. In: Proceedings of the first edition of the MCC workshop on Mobile cloud computing, pp 13–16

Cao X, Song Z, Yang B, El Mossallamy M, Qian L, Han Z (2019) A distributed ambient backscatter MAC protocol for Internet-of-Things networks. IEEE Internet Things J

Chandnani N, Khairnar CN (2020) A comprehensive review and performance evaluation of recent trends for data aggregation and routing techniques in IoT networks. Soc Netw Comput Intell 467–484

Chowdhury C, Roy S, Ray A, Deb SK (2020) A fault-tolerant approach to alleviate failures in offloading systems. Wirel Pers Commun 110(2):1033–1055

Deng DJ, Gan M, Guo YC, Yu J, Lin YP, Lien SY, Chen KC (2019) IEEE 802.11 ba: Low-power wake-up radio for green IoT. IEEE Commun Mag 57(7):106–112

Dezfouli B, Amirtharaj I, Li CCC (2018) Emp IoT: an energy measurement platform for wireless IoT devices. J Netw Comput Appl 121:135–148

Dhall R, Agrawal H (2018) An improved energy efficient duty cycling algorithm for IoT based precision agriculture. Procedia Comput Sci 141:135–142

Feng L, Shen Y, Chen G, Xu X, Wang P, Xu C (2018) A visible light backscatter communication solution for low-cost parking space detection. In: Proceedings of the 2018 ACM int. joint conf. and 2018 int. symposium on pervasive and ubiquitous computing and wearable computers, pp 343–344

Froytlog A, Foss T, Bakker O, Jevne G, Haglund MA, Li FY, Oller J, Li GY (2019) Ultra-low power wake-up radio for 5G IoT. IEEE Commun Mag 57(3):111–117

Gharehpasha S, Masdari M, Jafarian A (2020) Virtual machine placement in cloud data centers using a hybrid multi-verse optimization algorithm. Artif Intell Rev 1–37

Hoang DT, Niyato D, Kim DI, Van Huynh N, Gong S (2020) Ambient backscatter communication networks. Cambridge University Press

Howell J (2017) Number of connected IoT devices will surge to 125 billion by 2030, HIS Markit says. IHS Markit Technology

Hsu C-W, Hsu YL, Wei HY (2019) Energy-efficient and reliable MAC offloading for heterogeneous industrial IoT networks. In: 2019 European conference on networks and communications (EuCNC). IEEE, pp 384–388

Hu Y, Li W, Xu K, Zahid T, Qin F, Li C (2018) Energy management strategy for a hybrid electric vehicle based on deep reinforcement learning. Appl Sci 8(2):187

Jawad AM, Jawad HM, Nordin R, Gharghan SK, Abdullah NF, Abu-Alshaeer MJ (2019) Wireless power transfer with magnetic resonator coupling and sleep/active strategy for a drone charging station in smart agriculture. IEEE Access 7:139839–139851

Jo H, Yoon YI (2018) Intelligent smart home energy efficiency model using artificial sensor flow engine. HCIS 8(1):1–18

John N, Jyotsna A (2018) A survey on energy efficient tree-based data aggregation techniques in wireless sensor networks. In: 2018 international conference on inventive research in computing applications (ICIRCA). IEEE, pp 461–465

Jushi A, Pegatoquet A, Le TN (2016) Wind energy harvesting for autonomous wireless sensor networks. In: 2016 Euromicro conf. on digital system design (DSD). IEEE, pp 301–308

Kaur N, Sood SK (2015) An energy-efficient architecture for the Internet of Things (IoT). IEEE Syst J 11(2):796–805

Khan JA, Qureshi HK, Iqbal A (2015) Energy management in wireless sensor networks: a survey. Comput Electr Eng 41:159–176

Kim Y, Lee Y (2015) Automatic generation of social relationships between internet of things in smart home using sdn-based home cloud. In: 2015 IEEE 29th international conference on advanced information networking and applications workshops. IEEE, pp 662–667
Ko H, Lee J, Pack S (2019) Cg-e2s2: consistency-guaranteed and energy efficient sleep scheduling algorithm with data aggregation for IoT. Futur Gener Comput Syst 92:1093–1102
Kurdi H, Al-Anazi A, Campbell C, Al Faries A (2015) A combinatorial optimization algorithm for multiple cloud service composition. Comput Electr Eng 42:107–113
Lavassani M, Forsstrom S, Jennehag U, Zhang T (2018) Combining fog computing with sensor mote machine learning for industrial IoT. Sensors 18(5):1532
Li Y, Shi R (2015) An intelligent solar energy-harvesting system for wireless sensor networks. EURASIP J Wirel Commun Netw 2015(1):179
Li S, DaXu L, Zhao S (2018) 5G Internet of Things: a survey. J Indus Inf Integr 10:1–9
Liessner R, Schroer C, Dietermann AM, Baker B (2018) Deep reinforcement learning for advanced energy management of hybrid electric vehicles. In: ICAART (2), pp 61–72
Liu X, Ansari N (2019) Toward green IoT: energy solutions and key challenges. IEEE Commun Mag 57(3):104–110
Liu Y, Yang C, Jiang L, Xie S, Zhang Y (2019) Intelligent edge computing for IoT-based energy management in smart cities. IEEE Netw 33(2):111–117
Luo J, Tang J, So DK, Chen G, Cumanan K, Chambers JA (2019) A deep learning-based approach to power minimization in multi-carrier noma with SWIPT. IEEE Access 7:17450–17460
Ma Z, Feng L, Xu F (2018) Design and analysis of a distributed and demand based backscatter MAC protocol for Internet of Things networks. IEEE Internet Things J 6(1):1246–1256
Machorro-Cano I, Alor-Hernandez G, Paredes-Valverde MA, Rodrıguez-Mazahua L, Sanchez-Cervantes JL, Olmedo-Aguirre JO (2020) Hems-IoT: a big data and machine learning-based smart home system for energy saving. Energies 13(5):1097
Magno M, Spadaro L, Singh J, Benini L (2016) Kinetic energy harvesting: toward autonomous wearable sensing for Internet of Things. In: 2016 Int. symposium on power electronics, electrical drives, automation and motion (SPEEDAM). IEEE, pp 248–254
Mao J, Lin Q, Bian J (2018) Application of learning algorithms in smart home IoT system security. Math Found Comput 1(1):63
Min M, Xiao L, Chen Y, Cheng P, Wu D, Zhuang W (2019) Learning based computation offoading for IoT devices with energy harvesting. IEEE Trans Veh Technol 68(2):1930–1941
Mounce S, Pedraza C, Jackson T, Linford P, Boxall J (2015) Cloud based machine learning approaches for leakage assessment and management in smart water networks. In: Procedia engineering, vol 119. Elsevier, pp 43–52
Nayeri ZM, Ghafarian T, Javadi B (2021) Application placement in fog computing with AI approach: taxonomy and a state of the art survey. J Netw Comput Appl 103078
O'Dwyer E, Pan I, Acha S, Shah N (2019) Smart energy systems for sustainable smart cities: current developments, trends and future directions. Appl Energy 237:581–597
Oma R, Nakamura S, Enokido T, Takizawa M (2018a) An energy-efficient model of fog and device nodes in IoT. In: 2018a 32nd int. conf. on advanced information networking and applications workshops (WAINA). IEEE, pp 301–306
Oma R, Nakamura S, Enokido T, Takizawa M (2018b) A tree-based model of energy-efficient fog computing systems in IoT. In: Conf. on complex, intelligent, and software intensive systems. Springer, pp 991–1001
Paris L, Anisi MH (2019) An energy-efficient predictive model for object tracking sensor networks. In: 2019 IEEE 5th world forum on internet of things (WF-IoT). IEEE, pp 263–268
Park S-T, Li G, Hong JC (2020a) A study on smart factory-based ambient intelligence context-aware intrusion detection system using machine learning. J Ambient Intell Humaniz Comput 11(4):1405–1412
Park JJ, Moon JH, Lee K-Y, Kim DI (2020b) Transmitter-oriented dual mode SWIPT with deep learning based adaptive mode switching for IoT sensor networks. IEEE Internet Things J

Passos D, Balbi H, Carrano R, Albuquerque C (2019) Asynchronous radio duty cycling for green IoT: state of the art and future perspectives. IEEE Commun Mag 57(9):106–111
Pattamaset S, Choi JS (2021) Efficient data aggregation for human activity detection with smart home sensor network using k-means clustering algorithm. In: Advances in computer science and ubiquitous computing. Springer, pp 9–15
Peralta G, Iglesias-Urkia M, Barcelo M, Gomez R, Moran A, Bilbao J (2017) Fog computing based efficient IoT scheme for the industry 4.0. In: 2017 IEEE Int. workshop of electronics, control, measurement, signals and their application to mechatronics (ECMSM). IEEE, pp 1–6
Perera TDP, Jayakody DNK, Sharma SK, Chatzinotas S, Li J (2017) Simultaneous wireless information and power transfer (SWIPT): recent advances and future challenges. IEEE Commun Surv Tutor 20(1):264–302
Piyare R, Murphy AL, Kiraly C, Tosato P, Brunelli D (2017) Ultra low power wake-up radios: a hardware and networking survey. IEEE Commun Surv Tutor 19(4):2117–2157
Pourghebleh B, Navimipour NJ (2017) Data aggregation mechanisms in the Internet of Things: a systematic review of the literature and recommendations for future research. J Netw Comput Appl 97:23–34
Razafimandimby C, Loscrı V, Vegni AM, Aourir D, Neri A (2018) A Bayesian approach for an efficient data reduction in IoT. In: Fortino G, Palau CE, Guerrieri A, Cuppens N, Cuppens F, Chaouchi H, Gabillon A (eds) Interoperability, safety and security in IoT. Springer International Publishing, Cham, pp 3–10
Sheikhi A, Rayati M, Ranjbar AM (2015) Energy hub optimal sizing in the smart grid; machine learning approach. In: 2015 IEEE power & energy society innovative smart grid technologies conf. (ISGT). IEEE, pp 1–5
Shin M, Joe I (2016) Energy management algorithm for solar-powered energy harvesting wireless sensor node for Internet of Things. IET Commun 10(12):1508–1521
Singh J, Kaur R, Singh D (2020) A survey and taxonomy on energy management schemes in wireless sensor networks. J Syst Arch 101782
Slalmi A, Saadane R, Chehri A, Kharraz H (2021) How will 5G transform industrial IoT: latency and reliability analysis. In: Human centred intelligent systems. Springer, pp 335–345
Song L, Chai KK, Chen Y, Schormans J, Loo J, Vinel A (2017) QoS-aware energy-efficient cooperative scheme for cluster-based IoT systems. IEEE Syst J 11(3):1447–1455
Sun X, Ansari N (2016) Edgeiot: mobile edge computing for the internet of things. IEEE Commun Mag 54(12):22–29
Thomas BL, Cook DJ (2016) Activity-aware energy-efficient automation of smart buildings. Energies 9(8):624
Wang J, Jiang C, Han Z, Ren Y, Hanzo L (2018) Internet of vehicles: sensing-aided transportation information collection and diffusion. IEEE Trans Veh Technol 67(5):3813–3825
Wang Q, Lin D, Yang P, Zhang Z (2019) An energy-efficient compressive sensing-based clustering routing protocol for WSNs. IEEE Sens J 19(10):3950–3960
Wu J, He H, Peng J, Li Y, Li Z (2018) Continuous reinforcement learning of energy management with deep Q network for a power split hybrid electric bus. Appl Energy 222:799–811
Yang G, Zhang Q, Liang Y-C (2018) Cooperative ambient backscatter communications for green Internet-of-Things. IEEE Internet Things J 5(2):1116–1130
Zhang W, Qin Y, Zhao W, Jia M, Liu Q, He R, Ai B (2019) A green paradigm for Internet of Things: ambient backscatter communications. China Commun 16(7):109–119
Zhao H, Hua Q, Chen H-B, Ye Y, Wang H, Tan SX-D, Tlelo-Cuautle E (2018) Thermal-sensor-based occupancy detection for smart buildings using machine-learning methods. ACM Trans Des Autom Electron Syst (TODAES) 23(4):1–21
Zou G, Chen Y, Yang Y, Huang R, Xu Y (2010) Ai planning and combinatorial optimization for web service composition in cloud computing. In: Proc Int. Conf. on cloud computing and virtualization, pp 1–8

Chapter 3
Integration of IoT and 5G Communication

Seema Nayak, Pankaj Jha, and Manoj Nayak

Abstract The Internet of things (IoT) is a system which describes the objects that connects to the internet. The IoT-based applications like smart cities; health care system, smart constructions, etc. have a flooding presence now. These applications need higher data rates, larger bandwidth, increased capacity; low latency, and high throughput with the fast growth of wireless communication systems technological development has allowed us to emerge new services for mobile communication such as voice, video, audio and data services. Further, it provide provision to converse faster with a high data rate between portable devices or computers within a short range. The 5G communication offers a high data rate, low profile, and compact design in a micro-strip antenna. Though micro-strip antenna suffers from low gain as well as less efficiency, but we can manage both parameters with the help of defective Ground Structure, metamaterial, Frequency selective surface, and other technologies. 5G cellular networks propose the technologies for the IOT-based organization everywhere. Many researchers, scientists, and engineers are now have many challenges in designing IoT-based systems which can well integrated with the 5G wireless communications. 5th generation wireless communication technology is now on the skyline and IoT is playing major role to ensure efficient communications with design of 5G network.

Keywords Internet of things · 5G communication · Metamaterial · Patch antenna

S. Nayak (✉) · P. Jha
Department of Electronics and Communication Engineering, IIMT College of Engineering, Greater Noida, UP, India
e-mail: seema_jessica@rediffmail.com

P. Jha
e-mail: pankaj.jha@iimtindia.net

M. Nayak
Department of Mechanical Engineering, Manav Rachna International Institute of Research and Studies Faridabad, Faridabad, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
A. Rai et al. (eds.), *Paradigms of Smart and Intelligent Communication, 5G and Beyond*, Transactions on Computer Systems and Networks,
https://doi.org/10.1007/978-981-99-0109-8_3

3.1 Introduction

With the advancement of the wireless communication system, technological enhancement is vast to emerge incipient accommodations for mobile communication such as voice, audio, video, and data accommodations. The 5G communication offers a high data rate and higher bandwidth with expeditious communication. Antenna design is the essential parameter to develop a compact system with better performance. Different amendments are done utilizing defective ground, metamaterial, frequency selective surface, and other technologies (Ghosh and Parui et al. 2010; Mutiara and Refiant 2012; Majumder 2013; Jha et al. 2020, 2021a, b, 2022; Verma et al. 2020; Dalal and Dhull 2020; Rajkumar and Kiran 2018; Prasad et al. 2017; Yang et al. 2018; Balanis 2015; Al-Khaldi 2017; Jaglan et al. 2018; Elhabchi et al. 2017; Peng et al. 2016)

The Internet of Things (IoT) is a promising technology that inclines to revolutionize and connect the whole world via various smart devices through seamless connectivity. The current scenario for machine-type communications (MTC) has resulted in a variety of communication technologies with diverse requisites to achieve the modern IoT vision. More recent cellular standards like long-term evolution (LTE) have been introduced for mobile devices but are not appropriate for low-power and low data rate contrivances such as the IoT contrivances. To address this, there is a number of emerging IoT standards. Fifth generation (5G) mobile network, in particular, aims to address the circumscriptions of anterior cellular standards and be a potential key enabler for future IoT (Akpakwu et al. 2018).

The following are the most consequential elements in the description of 5G: high performance, low latency, high reliability, more preponderant scalability and mobile communication technology of low consumption (Nayak et al. 2011) as shown in Fig 3.1.

3.1.1 The Advantages of 5G

Few of the prominent advantages of 5G are as listed below:

- It will enable identical global standards for everyone.
- Network will facilely be available for carrying out different processes.
- People will be able to utilize their laptops and kindred types of smart devices anywhere as well as any time with high-speed and good connectivity.
  - The World will become a Wi-Fi zone in an authentic sense.
  - The smart radio technology shall enable different versions of radio technologies to utilize or share the mundane spectrum with full efficacy.
  - Utilizer will be able to get the radio signal even at more preponderant altitudes.

The Internet of Things is growing at a very expeditious rate and will transmute the way things work. With enormous areas of prospective applications like health

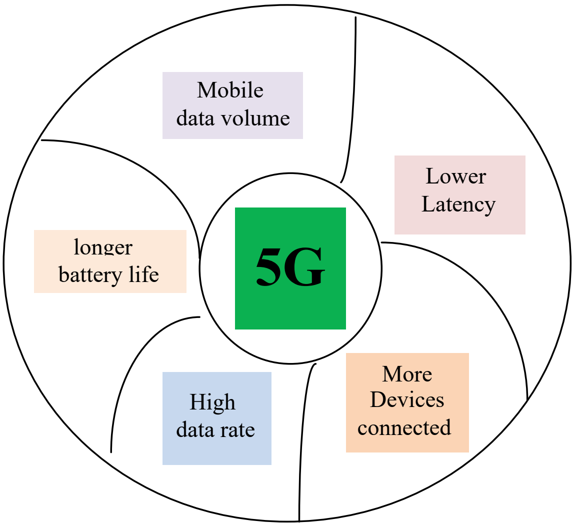


Fig. 3.1 Challenges in 5G communication

care, astute manufacturing, Keenly intellective homes, shopping, perceptive cities, wearables, etc. IoT will surely play a consequential role in our day-to-day lives. The development of 5G networks is taking place at a very expeditious rate. The advantages and features of 5G are aptly opportune to fortify IoT in an immensely colossal way. Nowadays 5G is a hot topic of interest for researchers its implementation in the era of IOT will enhance the utilization of Internet of Things technology in an astronomically immense way.

3.1.2 Enable Factors of 5G for IoT

Researchers are in the process of designing 5G to fortify 1000 times the traffic that is being handled by the subsisting networks. This gives the speed in the range of 10 Gbps. IoT will work with all types of devices in the future. Since more innovative devices will keep on integrating with the IoT network, this is one of the key requisites to fortify the magnification of IoT (Passi and Batra 2018).

Factors Of 5G For IOT

1. Higher Bandwidth: 5G support one thousand times the traffic that's being handled by the main networks. This conjointly offers a speed of the vary of ten Gbps.
2. Smaller Device Packaging: 5G operated at frequencies upto eighty gigacycle per second as compared to 4G network supporting frequencies up to half-dozen gigacycle per second thereby timid the dimensions needed for antennas up to 1–10 mm as compared to few centimeters for 4G.

3. Millisecond latency: Latency could also be outlined because the time needed to transmit one packet of information. The 4G networks offer latency as low as twenty-five ms.
4. High capable networks: Networks with high effectiveness similar to effectualness, connecting billions of physical objects at lightning speed is the key feature of 5G technology.
5. Forward compatibility: 5G technology work with all the assorted varieties of devices to be connected within the future. Since newer devices can continue adding within the IOT network, this can be one of the key needs to support the expansion of IOT (Babu and Wilson 2020).

In the nearby, 5G will integrate different networks technologies including unicast, multicast, and broadcast which will gratify all Media and Regalement requisites. The physical infrastructure will require fortifying an energy flow originating from the distributed energy resources which implicatively insinuates incipient desideratum for communication technology. 5G will play a fundamental role to achieve this goal of the energy sector Singh et al. (2017).

3.2 5G Applications in IoT

In this section, the high-speed 5G applications in the growing area of IoT are discussed for a recent communication system, and in the next section, the authors will discuss the latest technological development cognate to 5G antenna designing. The authors reviewed the technologies and different schemes that are habituated to organize IoT applications in 5G networks and summarized that cumulating mm-wave, massive MIMO, and dense diminutive networks will simplify IoT accommodations in 5G technology (Nayak et al. 2021). For future work, the author gives exploration for eliminating the challenges which are facing mm-wave, dense minute cells, and massive MIMO technology in the deployment of IoT accommodations. 5G-enabled IOT services expected growth in terms of technology and also assure to give support million of jobs across the world.

Nowadays compact wearable antennas are a core part of every wearable 5G communication system, IoT, and biomedical system. Antennas are usually used in communication systems and investigators. These possess attractive features that are vital to IoT, 5G communication, and medical systems (Jha et al. 2021b, c). These antennas are compact, lightweight, and flexible. The low-profile efficient antennas are playing a crucial role in the development of compact 5G communication system based on IoT. Due to expeditious magnification in wireless communication, IoT, and biomedical industries, there is a need to develop efficient high-gain compact antennas for 5G communication and IoT systems. The efficiency of compact printed antennas can be modified by using metamaterials and fractal structures. A novel implementation of wearable technology provides support to healthcare centers and surgical reintegration accommodations. Wireless wearable body area network is an

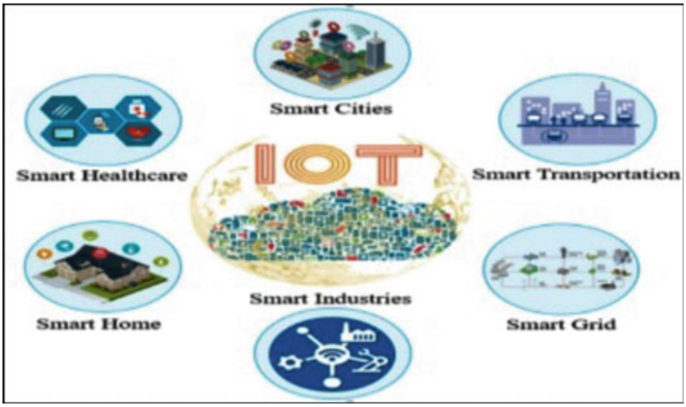


Fig. 3.2 IOT-based smart applications

emerging area as a paramount option for hospitals, medical centers, and patients. Wearable gadgets provide a service that may amend the long-term context and physiological replication of patients and healthcare customers. Wearable contrivances and emerging technology will fortify the development of personalized treatment contrivances to ameliorate patients' health with online and authentic-time feedback. Wearable medical devices with sensors can measure heartbeat, body temperature, blood pressure, and sweat rate, perform gait analysis, and quantify virtually any medical parameters of the patient who wears the medical device (Jha 2021c; Nayak et al. 2020). Figure 3.2 presents the IOT-based smart applications.

3.3 Technological Development with a 5G Antenna

Further, the study gives antenna designing and its application in 5G communication as well as integration with the internet of things.

In Wenxing et al. (2018), a low-profile compact antenna with a stable radiation pattern is proposed; with a miniaturized structure achieving a 10 dB IBW from 2.84 to 5.17 GHz, with 5 dB average gain with a stable radiation pattern, can be a potential candidate for 5G and IoT application. In Ghosh and Parui et al. (2010), a 2 × 2 micro-strip patch antenna array is designed for 5.25 GHz dedicated to WLAN application. The proposed antenna is having high gain which can be an essential requirement for transmission and reception both. In Mutiara and Refianti (2012), the oblong antenna is designed for wireless applications for lower ISM band of 2.4 GHz achieving maximum gain of 11 dB. It incorporates broad side radiation patterns and may be used as a client antenna in a laptop and is feasible for Wi-Fi IoT applications. In Majumder (2013), the patch antenna is designed and manufactured to provide an efficient and inexpensive antenna alternative to 2.4 GHz routers. This rectangular micro-strip antenna has a higher gain due to reduced losses because of the corner truncated rectangular antenna. This antenna allowed indoor communication. The discussed antenna is having linear polarization and can apply to the ISM band. In Jha et al. (2020), a coaxial feed low-profile miniaturized antenna is designed to work at 2.4GHz frequency and is suggested for future Bluetooth and many applications. In Jha et al. (2021a), multi-band antenna is designed for Sub-6 GHz applications. The 10 dB IBW of antenna is varying from 3 GHz to 5.64 GHz with a maximum gain of 3.22 dB. The designed antenna can be suitable for Wi-Fi, IoT, and Machine communication. In Verma et al. (2020), four-port MIMO is designed for 5G applications and smartphone applications. The MIMO antenna operates for dual-band of 5G. In Jha et al. (2021b), a compact CPW fed, reconfigurable antenna is designed using a PIN diode. The PIN diode is changing the electrical length of the radiating patch. The obtained results suggest that the antenna can be applicable in 5G applications. In Dalal and Dhull (2020), a flexible MIMO antenna is presented for WLAN and wearable applications. The proposed antenna maintains high isolation with good diversity performance. In Rajkumar and Kiran (2018), EBG structures are used in the antenna designing where EBG are placed close to the feed line to create band notch. Further the antenna can be used for 5G and other UWB applications. In Prasad et al. (2017), Metamaterial based gain enhancement technique is introduced in the proposed antenna that can be used in WLAN/WiMAX applications as depicted in Fig. 3.3a, b.

In Yang et al. (2018), a miniaturized antenna is designed for 1.2 GHz application. Specifically for L-band and WLAN applications. In Balanis (2015), the work exhibits a model of a self-supported BAN sensor, which comprises an electrically small triple-band antenna that covers triple operating bands including UTMS-2100, GSM-900, and TD-LTE bands with the electrical size of $0.21\lambda \times 0.2\lambda$ at 900 MHz The maximum peak gain is obtained 2.3 dB and antenna prototype is shown in Fig. 3.4.

In Al-Khaldi et al. (2017), a low-profile antenna is designed for satellite and other applications, where compact size, and less weight, make it more advantageous for IoT applications. In Jaglan et al. (2018), streamlining and creation of an exceptionally compact, improved multi-band IFA is designed The proposed antenna will work at

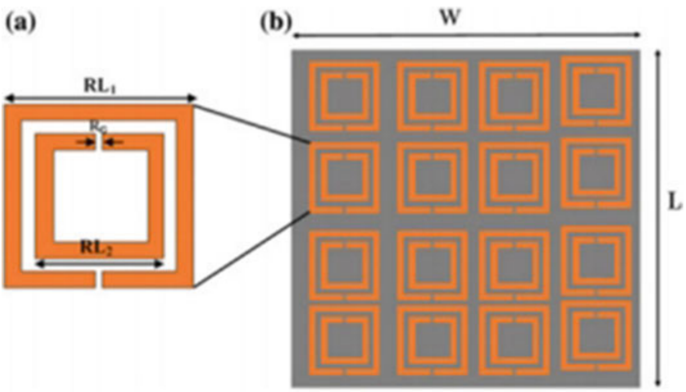


Fig. 3.3 **a** Unit cell and **b** top view of a superstrate

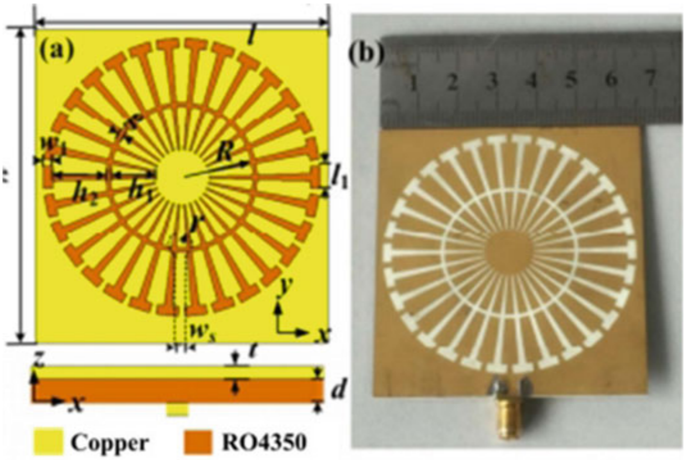


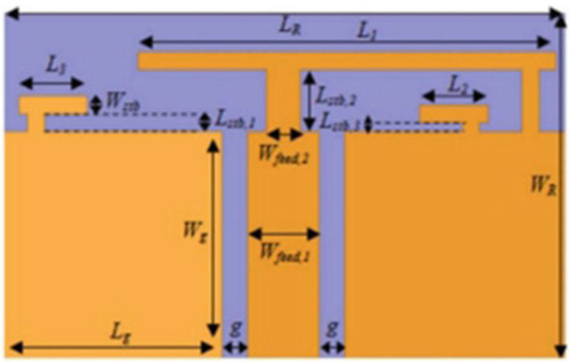
Fig. 3.4 Front view and side perspective on the antenna

2.4, 2.5, 3.5, and 5 GHz frequency for Bluetooth/WLAN, Wi-Fi, and IoT applications. The antenna layout is illustrated in Fig. 3.5.

The compactness of the proposed plans and adequacy in multi-band activity seems, by all accounts, to be promising in the chance of its insertion in convenient electronic gadgets and IoT. In Elhabchi et al. (2017), an UWB monopole antenna is designed to reject the frequency band 3.3–3.6 GHz and 5–6 GHz frequency band. The EBG and DGS technologies are used to create the band notch. The evolution steps are shown in Fig. 3.6.

In Peng et al. (2016), a novel fork-molded antenna is designed for UWB applications with the triple band-notch band. The proposed antenna, the compact antenna will be useful in 5G and other applications. In Jha et al. (2022), an antenna is designed for UWB applications with a notch band. Double mushroom-type EBG structures are used in CPW-fed antenna. In Jha et al., a compact multi-band antenna is designed

Fig. 3.5 Proposed antenna layout



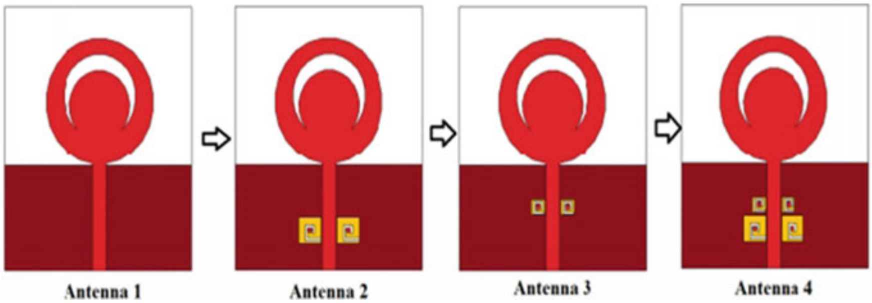


Fig. 3.6 Stepwise improvement of the proposed antenna

for different applications of 5G and WLAN applications. Strong metamaterial characteristics are displayed by the multi-band antenna with a modest structure as shown in Fig. 3.7.

In Jha et al., a modified structure of CSRR is utilized to design a MIMO antenna which can be useful in 5G mobile applications as depicted in Fig 3.8.

In Jha et al. (2021c), the flexible antenna can be used fully in 5G communication. The antenna is having simple decoupling structure and high isolation is obtained between two element antennas. The jeans material is used to design the compact antenna and the basic structure is depicted in Fig. 3.9.

In Jha et al. (2021d), a flower shaped UWB antenna is designed with CSRR and EBG technology. The SRR and EBG both are used to create a notch band to prevent the undesired band in output. The antenna will be suitable for 5G communication system. The antenna design is shown in Fig. 3.10.

A Jha et al. (2021e) triple band antenna is designed for 5G application accomplished with SRR and rectangular slots in the patch. These slots are improving the impedance matching in the antenna and SRR is improving the gain of antenna, which

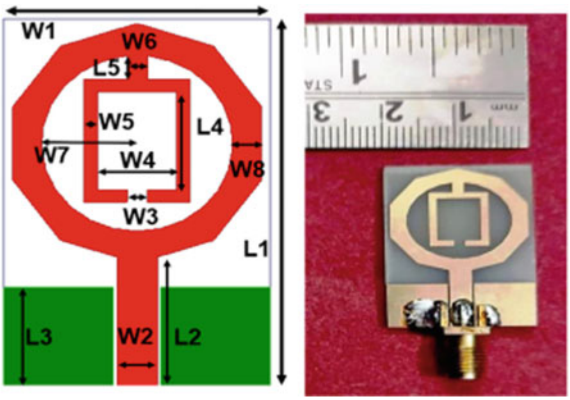


Fig. 3.7 Metamaterial-inspired antenna for 5G communication

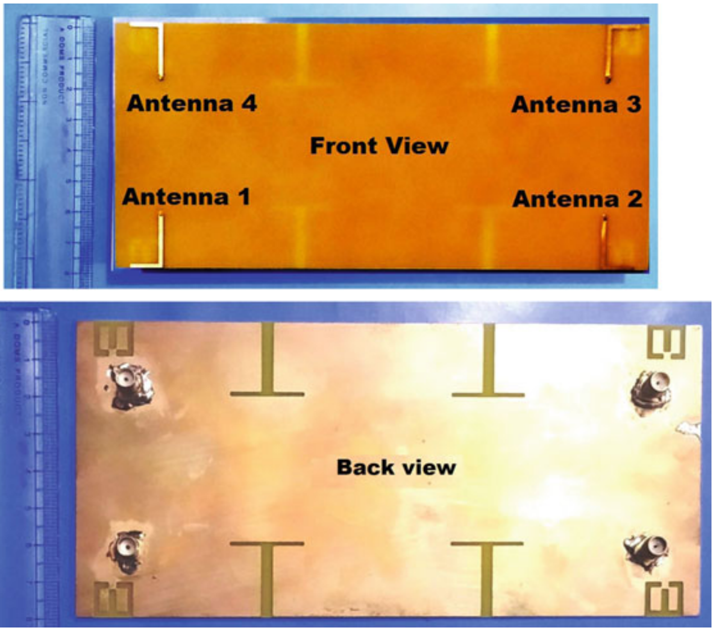


Fig. 3.8 Four port MIMO antenna for 5G mobile communication

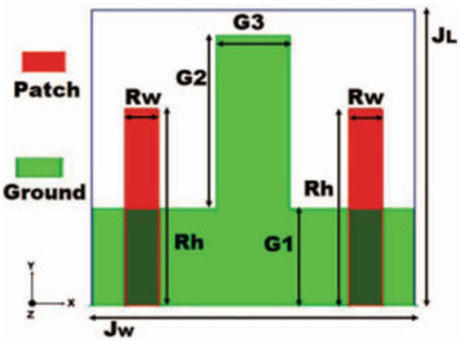


Fig. 3.9 2 port MIMO antenna designed on jeans substrate

is an essential requirement for 5G communication. The antenna layout is depicted in Fig. 3.11.

In Jha et al. (2022e), an antenna is fabricated on FR 4 substrate for 5G communication where multiple slots are created in patch to improve the impedance matching where the gain of the antenna is above 1.75 dB in the operating band. The antenna prototype is demonstrated in Fig. 3.12.

Flexible antenna is designed for WLAN application in Sharma et al. (2021), as depicted in Fig. 3.13. The compact antenna exhibits the metamaterial characteristics and the compact antenna achieves multi-band operation.

Fig. 3.10 UWB antenna accomplished with CSRR and EBG

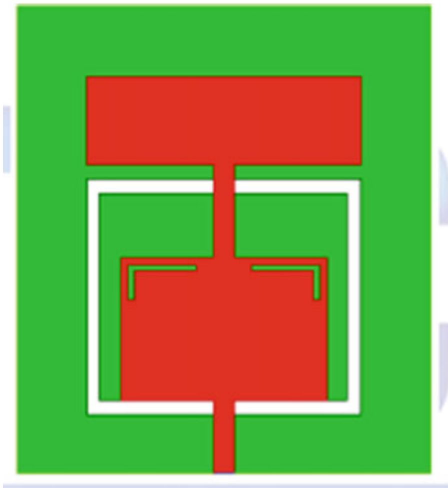


Fig. 3.11 Triple band antenna for 5G communication

Circular polarized antenna (Navneet et al. 2022; Nayak et al. 2021) which will be suitable for 5G and IoT applications as depicted in Fig. 3.14.

3.4 Summary

5G provides high-speed communication with added advantages which is desired for IoT applications. 5Gs has the potential to strengthen a massive number of static and mobile IoT devices, which have a diverse range of celerity, bandwidth, and quality of accommodation requisites. As the IoT technology grows, the flexibility of 5G

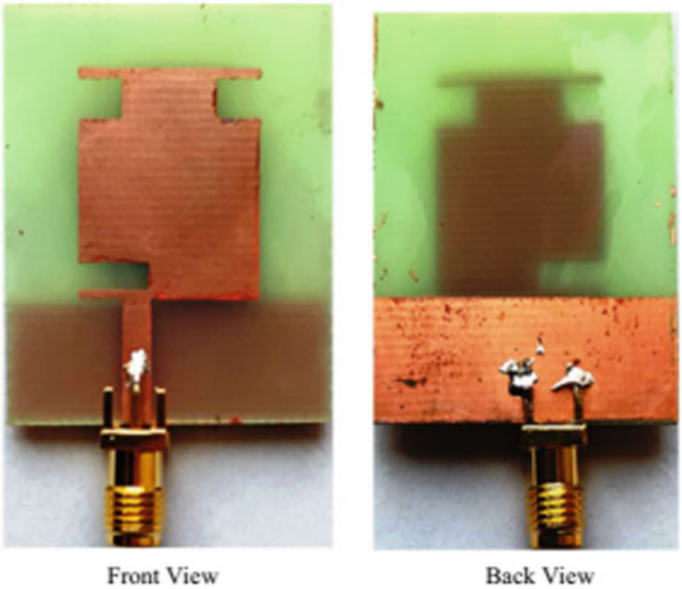


Fig. 3.12 Prototype of 5G antenna

Fig. 3.13 Prototype of metamaterial inspired antenna

will become even more paramount for enterprises. IOT is set to increase connected devices from million to billion. 5G will improve the performance and reliability of connected devices and play a key role not only in the evolution of communication but for business and society. The cyber world of things is an ecumenical network that integrates authentic-life physical objects with the virtual world through the Internet for making perspicacious decisions. In a widespread computing environment, many Smart devices have constrained of storage, battery backup, and computational capability. In such conditions, cellular networks that are evolving from 4G to 5G are set to play a key role. Various features like high bandwidth, flexibility in connectivity,

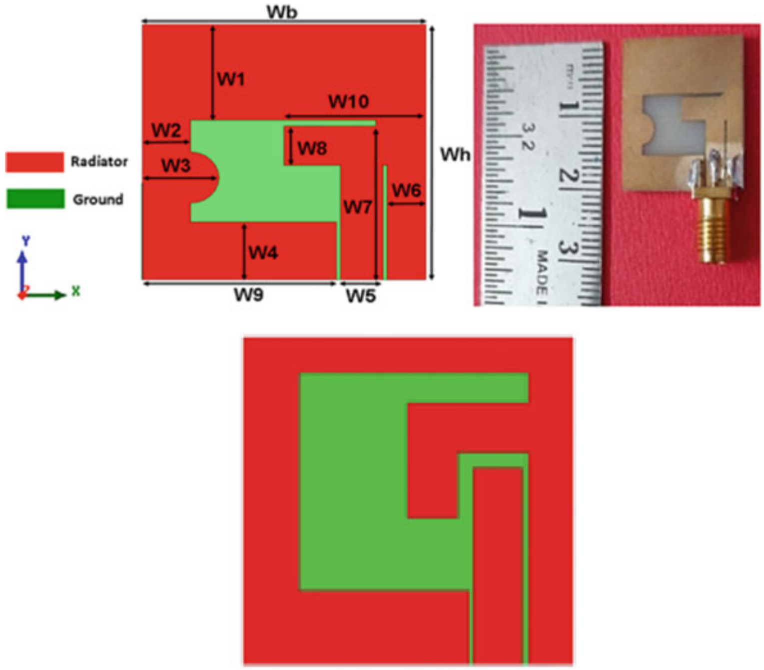


Fig. 3.14 Circular polarized antenna

wider range coverage, in-built mechanism and interface for M2M communication, etc., make 5G cellular networks an impeccable one to fortify the future IoT.

References

Al-Khaldi M (2017) A highly compact multiband antenna for Bluetooth/WLAN, WiMAX, and Wi-Fi applications. Microw Opt Technol Lett 59(1):77–80

Akpakwu GA et al (2018) A survey on 5G networks for the internet of things communication technologies and challenges. IEEE Access 6(Digital Identifier). https://doi.org/10.1109/ACCESS.2017.2779844

Balanis CA (2015) Antenna theory: analysis and design. John wiley & sons

Babu E, Wilson G (2020) IOT in 5G wireless communication. Int J Creat Res Thoughts 8(4). ISSN: 2320-2882

Dalal P, Dhull SK (2020) Upper WLAN band notched UWB monopole antenna using compact two via slot electromagnetic band gap Structure. Progress Electromagn Res 100:161–171

Elhabchi M, Srifi MN, Touahni R (2017) A tri-band-notched UWB planar monopole antenna using DGS and semi arc-shaped slot for WIMAX/WLAN/X-band rejection. Progress Electromagn Res 70:7–14

Ghosh CK, Parui SK (2010) Design, analysis, and optimization of a slotted microstrip patch antenna array at frequency 5.25 GHz for WLAN-SDMA system. Int J Electr Eng Inf 2(2):102–112

Jaglan N et al (2018) Design of band-notched antenna with DG-CEBG. Int J Electron 105(1):58–72

Jha P, Singh S, Yadava RL (2020) Wideband sub-6 GHz microstrip antenna: design and fabrication. Adv Smart Commun Imag Syst: Select Proceed MedCom 109
Jha P, Kumar A, De A, Jain RK (2021a) Modified CSRR based dual-band four-element MIMO antenna for 5G smartphone communication. Progress Electromag Res Lett 101
Jha P, Kumar A, De A, Jain RK (2021b) Flexible and textile two-port compact antenna for WLAN and wearable applications. In: 8th international conference on signal processing and integrated networks (SPIN), pp 308–311. IEEE
Jha P, Kumar A, De A, Jain RK (2021c) Flexible and textile two-port compact antenna for WLAN and wearable applications. In: 8th international conference on signal processing and integrated networks (SPIN). IEEE, pp 308–311
Jha P, Kumar A, De A, Jain RK (2021d) Design of UWB antenna based on CSRR and EBG notch for prevention of undesired band. In: 8th international conference on signal processing and integrated networks (SPIN). IEEE, pp 982–986
Jha P, Kumar A, De A, Jain RK (2021e) Bandwidth and gain improvement with SRR based defected ground in triple-band antenna for 5G application, TTIC, vol 5, pp 5–11
Jha P, Singh S, Yadava RL (2021f) Wideband sub-6 GHz micro-strip antenna: design and fabrication. In: Advances in smart communication and imaging systems, pp 109–115. Springer, Singapore
Jha P, Kumar A, De A, Jain RK (2022) CPW-fed metamaterial inspired compact multiband antenna for LTE/5G/WLAN communication. Frequenz
Mutiara AB, Refianti R (2012) Design of microstrip antenna for wireless communication at 2.4 GHz
Majumder A (2013) Design of an h-shaped microstrip patch antenna for Bluetooth applications. Int J Innov Appl Stud 3(4):987–994
Navneet AK, De A, Jain RK (2022) Compact circular polarized CPW antenna for WLAN and biomedical applications. Frequenz 76(3–4):229–237
Nayak S et al (2021) IOT in agriculture: survey on technology, challenges and future scope. In: Smart agriculture: emerging pedagogies deep learning, machine learning and internet of things, 1st edn. CRC Press Taylor & Francis Group, ISBN: 9781003227595
Nayak S et al (2020) Wearable healthcare device for elderly person. In: ICIMMI 2020 proceeding lecture notes in network & system (LNNNS). Springer Series for publication
Nayak S et al (2011) IEEE sponsored national conference on future mobile radio system 17th Sept'11, FET, MRIU, Faridabad
Peng L et al (2016) CPW fed UWB antenna by EBGs with a wide rectangular notched-band. IEEE Access 4:9545–9552
Prasad KD, Ali T, Biradar RC (2017) A compact slotted multiband antenna for L-band and WLAN applications. In: 2017 2nd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT). IEEE
Passi A, Batra D (2018) Future of internet of things (IoT) in 5G wireless networks. Int J Eng Technol 7(1.5):245–248
Patkar A (2022) 5G wireless technology. Int J Res Appl Sci Eng Technol https://doi.org/10.22214/ijraset.2022.40930. ISSN: 2321-9653
Rajkumar R, Kiran KU (2018) Gain enhancement of compact multiband antenna with metamaterial superstrate. Opt Microw Technol 91–98. Springer, Singapore
Singh N et al (2017) 5G technology and how it will boost IOT: a survey. Int J Eng Res Technol (IJERT) 5(01). ISSN: 2278-0181
Sharma N, Kumar A, De A, Jain RK (2021) Design of compact hexagonal shaped multiband antenna for wearable and tumor detection applications
Sharma N, Kumar A, De A, Jain RKJ (2021) Circularly polarized antenna for ISM (5.8 GHz), satellite communications and UWB applications. In: 2021 8th international conference on signal processing and integrated networks (SPIN), pp 303–307. IEEE
Verma RK, Kumar A, Yadava RL (2020) Compact Multiband CPW Fed Sub 6 GHz frequency reconfigurable antenna for 5G and specific UWB applications. J Commun 15:345–349

Wenxing A et al (2018) Low-profile and wideband microstrip antenna with stable gain for 5G wireless applications. IEEE Antennas Wirel Propag Lett 17(4):621–624

Yang L et al (2018) Compact multi-band wireless energy harvesting based battery-free body area networks sensor for mobile healthcare. IEEE J Electromagn RF Microwav Med Biol 2(2):109–115

Chapter 4
Role of IoT and Antenna Array in Smart Communication and 5G

Krishanu Kundu, Amrita Rai, and Ritesh Pratap Singh

Abstract IoT is nothing but an arrangement of interconnected computing devices that are assigned an unique identifiers (UIDs) with the capability of transferring data over a network in absence of a human-to-human or human-to-computer interface. The term "thing" in IoT could be either a person having a heart-beat monitoring system implanted in his body or a cow fitted with a biochip transponder in its body or any type of object having an IP address and capability of transferring data across a specified network. Broadly speaking IoT can be treated as a sensor network consisting of numerous smart devices employed for connecting people, systems as well as other applications for data collection and sharing. With the rapid gain in popularity of IoT, the need for technology supporting large amounts of data transmission over high bandwidth in a proficient manner is very obvious. Enhanced capacity with high data rates and lower latency values are the challenges to be faced by next-gen IoT devices. The deployment of new-age wireless communication, namely, 5G gears up to satisfy the complexity of IoT architectures. The current chapter focuses on the IoT protocols which are best suited for 5G. Along with IoT focus is on Designing of antenna as well as an antenna array suitable for 5G Communication.

Keywords Internet of things (IoT) · Smart communication · Antenna array · 5G

4.1 Introduction

Recently in the last era the new concept of ubiquitous network and communication has been discovered for every field of Life. This concept is that each entity in the circle of radius has been process then communication functions and connected to the network. For such type of smart communication development its significant of

K. Kundu (✉) · A. Rai
Department of Electronics & Communication Engineering, G L Bajaj Institute of Technology and Management, Greater Noida, India
e-mail: krishanukundu08@gmail.com

R. P. Singh
Haramaya Institute, Bati, Ethiopia

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
A. Rai et al. (eds.), *Paradigms of Smart and Intelligent Communication, 5G and Beyond*, Transactions on Computer Systems and Networks,
https://doi.org/10.1007/978-981-99-0109-8_4

Internet of Things (IoT) is expected a vital role in serving some useful applications like smart farming, smart health care, education, and so on recently (Lee 2017). The word IoT was firstly coined in 1985, with the concept of combining all the surroundings like people, microprocessor, and device through the communication technology in order to exchange their set of data for processing and collaborating remotely (Atzori et al. 2010). The IoT-based communication system is considered as favorable technology for few applications at the national level like disaster alerting, environmental monitoring, and so on. The basic component of IoT-based communications are sensors, heterogeneous access, information processing and networking, and lastly application and services (Chen et al. 2012). The intelligent communication requirement by industry 4.0 has several key and critical requirements of data and object connectivity, control, and processed. The industries transformation required the new and skilled personnel to cope with these challenges. Many researchers present a communication model based on Modbus and multiband communication using NB-IoT for communicating industrial equipment like PLCs, SCADA systems, etc. (Oussama et al. 2019). NB-IoT (Narrowband Internet of Things) is a part of wireless communication technology for enhancing the application area and it is also promoted by telecommunication companies for long-distance massaging and calling. This protocol is standardized by 3GPP, the standardization body that writes cellular standards, i.e., GSM, 3G, and 4G mobile technologies and their evolution (Ratasuk 2016.) Smart communication with IoT has significant usage and application in the present era, but it also required many improvements on the technical side also. This chapter gives a brief about antenna requirement and design for 5G communication. The merging of 5G communication and Internet of Things (IoT) are the subsequent accepted transfer of two emerging technologies, which assembled to create user's world convenient, easier, and more productive. Before the merger of these two technologies first, we should understand each of them. Figure 4.1 shows that in different areas of IoT-based communication system, there is a wide range of through which IoT provides the required smart solutions. These solutions lead to improved efficiencies in the smart communication between the fleet and rental companies (Rajput 2020).

The new era of communication is involving the heterogeneous smart devices which are the promising revolution of IoT technology, which connect the whole

Fig. 4.1 IoT embraces smart communication

world via a connectivity that is not broken which means seamless. The recent request intended for MTC (machine-type communications) describes data communication between two entities without the involvement of a human and has caused in a diversity of communication technologies with various service provisions to achieve the modern IoT vision. Modern cellular and wireless standards like long-term evolution (LTE) have been bringing together for wireless devices and mobile communication, but not be situated well for low-power and low-data rate communication or devices such as the IoT communication and devices. For addressing this, there is a huge number of developing IoT standards. New Fifth generation (5G) and Sixth generation (6G) mobile network, in specific, aim to solve the limitations of earlier cellular standards and be a latent key enabler for future IoT. The main reason behind 5G can deliver greater data rates than 4G technology and can be well understood from Shannon-Hartley theorem

$$\mathrm{Cap} = \mathrm{Num} * \mathrm{BW}\ \log 2(1 + \mathrm{Sg/No}) \tag{4.1}$$

where Cap represents channel capacity and its unit is bit/second, Num denotes channel numbers, the bandwidth of individual channel is denoted by BW and Sg/No equals to signal to noise ratio. In this intuitive-based theorem for achieving higher channel capacity further improvements should be incorporated. 5G, which progresses from 4G, incorporates a few well-known as well as famous practices in its planning to enhance channel capacity by the following: carrier aggregation increases bandwidth (B), MIMO architecture enhances channel allocation, allocation of new frequency bands increases bandwidth and newly adopted modulation schemes enhance the signal to noise ratio as well as bandwidth. For designing 5G compatible antennas, it's vital to recognize the challenges as well as building roadmap for addressing them. Because of stringent dimensions' constraints, recently emerged wireless devices characteristically employ active tuners for reducing antenna size. The antenna tuning basically based on altering the operating atmosphere, frequency band as well as bandwidth coverage.

Due to greater CA in 5G as well as supplementary cellular bands, it is obvious that the tuning system should be capable of supporting a greater number of tuner state as well as broader frequency bandwidth per tuner state. As per 3GPP Release 15, FR1 as well as FR2 are two elementary frequency ranges that are likely to be used for 5G:FR1 ranges from 410 MHz to 7.125 GHz whereas FR2 ranges from 24.25 to 52.6 GHz. The convergence between 5G as well as Internet of Things is the next big move in the field of these technologies constructed to make end users' lives more convenient as well as more productive. 5G compared to 4G offers speeds in the range of 7–17 Mbps for uploading and 12–36 Mbps for downloading, whereas 5G broadcast speeds can be up to 20 Gbps. In 5G latency is nearer to 10% with respect to 4G and the connection capability significantly increases that warranted the effect of convergence along with IoT.

IoT is basically a system with the feature of ever-growing difficulty; a cosmos of linked things supplying vital physical data as well as the further process of such data in the entire cloud for delivering business insights. Many establishments are forming themselves suitable for IoT, which is basically set up on components like sensors, networks, Cloud/AI as well as its applications As we combine 5G with IoT, its components will face some direct or indirect after-effects of the said combination, such as sensors will face additional bandwidth for reporting actions, similarly network will carry a large amount of information in quicker mode, in case of cloud as well as AI consideration of real-time data is reality, lastly, applications will showcase a greater number of features for covering countless options specified the widespread bandwidth supported by 5G. Advantages of utilizing 5G in the domain of IoT are basically greater transmission (up to 15–20 Gbps) as well as connection with a greater number of devices.

4.2 Basic Structure of IoT with Its Protocols

The benefit of IoT comes from permitting the components to communicate; this capability to communicate is that which helps in moving data from endpoint devices to central servers via IoT pipeline. This communication occurs by means of IoT protocols, which guarantee flawless reception of data from endpoint devices (sensors) and proper understanding of data is done by means of subsequent steps in the connected environment. Protocols are as crucial to the survival of IoT as the things themselves. Though protocols are nothing but a collective group which is vital for the smooth working of IoT, protocols aren't built in an equal fashion. Is the true fact that not all protocols function well, in each and every context, for example, some protocols are suitable for IoT implementation in buildings, whereas some protocols are suitable for IoT deployments across buildings and others perform well for national as well as global IoT use cases. There are numerous existing IoT protocols, with each one of them offers firm capabilities or blends of features that transform it preferable for specific IoT deployments. Each of the IoT protocols permits either device-to-device, device-to-gateway, or device-to-cloud/datacentric communication or groupings of those communications. Factors like geographic location, requirements of power consumption, battery-functioned options, the existence of physical barriers as well as cost determine which protocol is optimal in an IoT deployment. Technologists can opt for various communication protocols when constructing a network for serving their IoT ecosystem. Following are the four most popular IoT protocols.

4.2.1 Constrained Application Protocol (CoAP)

CoAP (Castellani et al. 2012; Thangavel et al. 2014; Ludovici et al. 2013) is basically employed for delimited gadgets. CoAP's miles are planned to use amid gadgets on the equivalent constrained community, among gadgets as well as internet nodes in addition midst gadgets over different controlled networks. In one-word CoAP is planned for IoT systems primarily built on HTTP protocols.

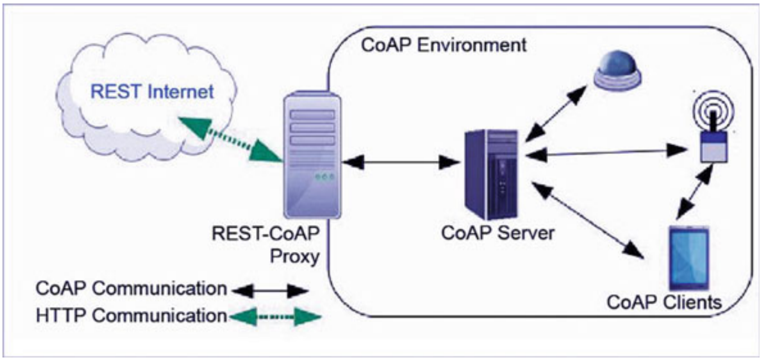


Fig. 4.2 CoAP protocol structure

As shown in above Fig. 4.2. CoAP outspreads normal HTTP clients to those clients who are having resource constraints, namely CoAP clients. It is the function of the Proxy device to cover up the gap existing between constrained environments as well as distinctive internet settings based on HTTP protocols. It is the same server that takes care of both HTTP as well as CoAP protocol messages.

CoAP protocol uses two modes for exchanging messages between the client and server. The first one is including separate response and the second one is excluding separate response. While including separate responses during message exchange it is the responsibility of the server to notify the client about the reception of the request message which will in turn increases processing time. CoAP IoT becomes unreliable due to the employment of UDP. As a result, CoAP messages ultimately reach in an unordered fashion or get lost as they reach the destination. User Datagram Protocol is utilized by CoAP for providing lightweight enactment. CoAP also utilizes RESTful architecture just like HTTP protocol. REST is a short form of representational state transfer. As per REST architecture, a Server offers admittance to resources whereas a REST client does both accessing as well as presenting the resources. CoAP makes use of DTLS for the cozy switch of statistics within the slipping layer. Out of the two layers used by DTLS, the lower layer contains record protocol and the upper one contains alert, handshake as well as application data protocol. CoAP is mostly used for machine-to-machine applications, and it also enables constrained devices to join an IoT environment, in spite of having low bandwidth, low availability, and/or low-energy devices.

4.2.2 Message Queue Telemetry Transport Protocol (MQTT)

Message Queue Telemetry Transport (Hwang et al. 2016; Soni and Makwana 2017; Dinculeana and Cheng 2019; Thangavel et al. 2014) is basically a messaging protocol, introduced by Andy Stanford-Clark et al. in 1999. MQTT is utilized for faraway tracking in IoT. The primary challenge faced by MQTT is gathering statistics from numerous gadgets as well as delivering infrastructure. It also connects gadgets as well as networks with the help of packages and middleware. MQTT protocols portray on the uppermost layer of TCP for offering informal and reliable rivulets of information (Fig. 4.3).

At the center of MQTT-based IoT structure, there exists a remote server, called **Broker**. All objects as well as services linked to it as *clients*. The broker basically forwards messages between clients. Clients can send messages as publishers and receive messages as subscribers. Published messages contain a topic that describes the message's contents. Subscribers each receive a copy of the message if they have subscribed to the topic of the published message. In MQTT, clients identify themselves using a unique ClientID. If left empty the broker will generate a random one. When a client connects to a broker, it can choose to start a new session or resume the existing one. A session contains clients' subscriptions and pending messages. Topics are used to forward published messages to subscribers. A topic is a character string indicating the message's contents. With MQTT, by default, all subscribers receive a copy of the messages published with matching topics. Upon publishing a message, a client can request that the broker keep it in memory. This is called a retained message. Any future subscription matching the retained message topic will immediately receive a copy of it. A new retained message for the same topic will discard the previous one. Existing in on the top layer of the TCP/IP network stack, MQTT proves to be a good design for providing lower bandwidth, higher latency as well as unreliable networks. MQTT's capable of sending large amount of sensor messages to analytics platforms as well as cloud solutions.

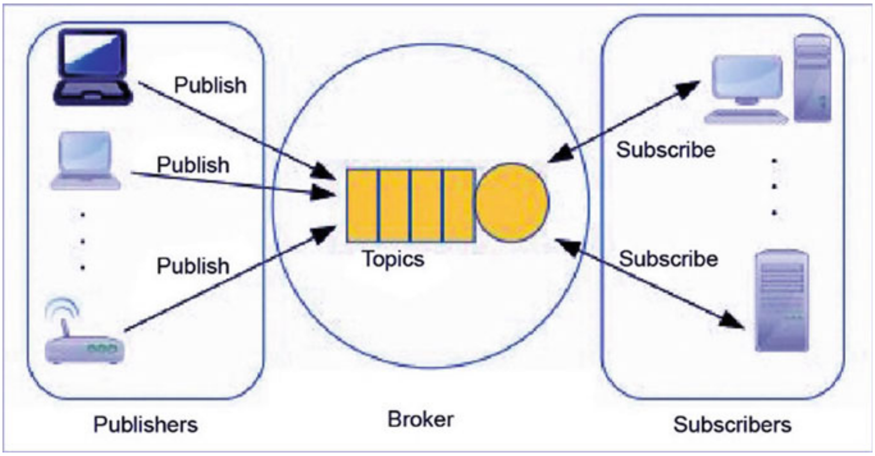


Fig. 4.3 MQTT protocol structure

4.2.3 Advanced Message Queuing Protocol (AMQP)

AMQP (Luzuriaga et al. 2015; Naik 2017) was introduced by John O'Hara. It is basically a software layer protocol applicable to message-based middleware settings. AMQP also approves dependable verbal exchange via message transport warranty primitives like at-most-once, at least once, and precisely as soon as shipping. The AMQP—IoT protocols comprised of components that route as well as save messages within a broker carrier, along with the list of rules for wiring the components altogether. This protocol also enables patron programs for talking to the dealer as well as engaging with the AMQP model. The current version consists of 3 additives as follows:

Exchange: Accepts messages from the publisher and routes them to Message Queues.

Message Queue: Keeps messages till those messages are properly processed through client software.

Binding: Establishes a secure connection between the message queue as well as the change (Fig. 4.4).

AMQP broker consists of components like exchange as well as queues. Client applications can be designated as producers. Client apps generate messages for forwarding to the broker, where the messages are being routed as well as queued. Consumers read messages from the queues and process them. Different forms of exchanges are default, direct, fanout, topic, and last but not the least header exchange.

The elementary unit of data is a frame. They are used to initiate, control, and tear down the transfer of messages between two peers. These nine AMQP frame bodies are as follows:

1. open (the connection) 2. begin (the session) 3. attach (the link) 4. transfer 5. flow 6. disposition 7. detach (the link) 8. end (the session) 9. close (the connection). Table 4.1 elaborates on a brief comparison of the above-mentioned protocols with the HTTP protocol.

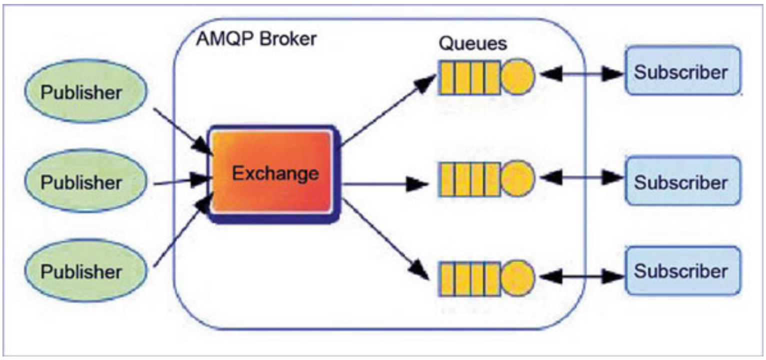


Fig. 4.4 AMQP protocol structure

Table 4.1 Comparison between AMQP, CoAP, MQTT, and HTTP protocol

	AMQP	CoAP	MQTT	HTTP
Full form	Advanced message queuing protocol	Constrained application protocol	Message query telemetry transport	Hyper text markup protocol
Developed by	It was developed by John O'Hara at JP Morgan Chase	It was developed by Internet Engineering Task Force (IETF)	It was developed by Andy Stanford-Clark and Arlen Nipper	It was developed by Tim Berners-Lee
Communication nature	It has asynchronous communication nature	This employs both asynchronous and synchronous	This employs only asynchronous	It has synchronous communication nature
Message delivery	It has guaranteed message delivery	It offers by adding labels to the messages	It has no such trait like message labeling	It has no guarantee for message delivery
Interface	It provides publish/subscribe interface	It utilizes request-response model	It employs publish-subscribe model	It provides point to point interface
Advantages	It is fast, flexible, and cost-effective protocol	It is employed in utility area networks and possess secured mechanism	It is employed in IoT applications and is secure	It is well-known, efficient and multi-purpose protocol

4.2.4 Data Distribution Service (DDS) Protocol

DDS (David et al. 2013) utilizes broker less architecture and employs multicasting for conveying high-quality QoS to applications. DDS can organize in platforms ranging from low-footprint devices to the cloud and supports green bandwidth usage. The DDS—IoT protocols showcase layers like DCPS and DLRL. DCPS plays the task of handing over the facts to subscribers, and the DLRL layer presents an interface to DCPS functionalities, permitting the sharing of distributed data among IoT-enabled objects (Fig. 4.5).

4.3 Employment of IoT and Antenna Array in 5G

As civilization runs fast into a fresh chapter of the information stage, current developments in artificial intelligence, information processing as well as cellular infrastructures have provided growth to quicker exchange of data. While 4G attains a download speed of about 2 Gbps, the expectation occurs that with 5G mobile performance will jump from 2.4 to 20 Gbps. As a result, 5G is considered to be a high-speed data networking replacement for wired fiber optic networks. As per research, it can be

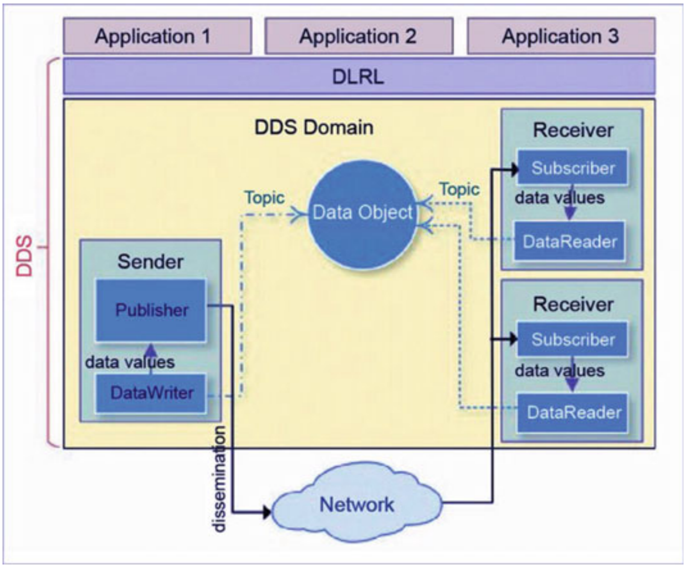


Fig. 4.5 DDS protocol structure

assumed that Internet of Things solutions will be capable of connecting more than 50 billion devices by the end of 2030. 5G comprises a typically New Radio (5G NR) framework as well as a new core network with an aim of enhancing wireless connections all over the world. At the same time 5G consist of a theory of multiple access suitable for satellites, Wi-Fi, and cellular network. Along with IoT-empowered devices, 5G connects a greater number of devices at higher speeds. As a result, 5G provides an outstanding user experience regardless of what type of application, expedient, or service we adopt. Enormous cellular IoT machineries can be categorized as a small-cost as well as small-power consumption outcome. They flourish on deep as well as extensive coverage indoors as well as outdoors. They are capable of delivering secure connectivity as well as authentication, can be easily fit on any network topology and at the same time are intended for full scope scalability and capacity upgrades.

A lot of tech-savvy regulars have already implemented a few applications of automation, for example, the automatic start of dishwashers using the mobile app or making lights on/off through remote access. A comparatively new phenomenon called the Internet of Things (IoT) can transform these automatic tasks toward advanced level by empowering objects to transfer data within interconnected devices. In the Internet of Things, mobile devices are capable of collecting and interpreting data, for example, location, known preferences, and communication with "smart objects"—it does not require any manual input. The widespread range of applications for the IoT includes media, transportation, healthcare, environment and energy conservation, and last but not least infrastructure.

With the rapid usage of smartphones, IoT has found a convenient way for collecting data and utilizing them to build strong communication between objects and smart devices. Meanwhile to attain a kickstart on creating IoT possible, it is urgent for the next peer group of mobile technology to improve smartphones.

Many researchers conclude that when 5G should address the following needs:

1. A diminished latency.
2. Improved data rates.
3. Amplified energy efficiency.

While preparing the ground for 5G, researchers are busy in designing and optimizing antennas suitable for smartphones. Side by side a lot of researchers are busy in designing compatible devices for proper interfacing of IoT and 5G. Antennas that are used in smartphones should be small size and lightweight. Planar inverted-F antennas (PIFA) as well as PIFA Arrays can be considered the right choice because of their smaller size, high power as well as better efficiency (Kundu et al. 2021a, b). As those types of antennas are able to cover several frequency bands suitable for cellular devices and WIFI, they are treated as the perfect choice for IoT-compatible devices.

4.4 Design and Simulation of Antenna and Antenna Array Suitable for 5G

With the commencement of novel IoT applications in the field of wireless communication, the problem of data traffic is growing at a very fast pace as a result of plenty of smartphone users. The main demand of present-day users is basically easy streaming as well as quick download of very high-resolution videos. Current 4G is unable to download such high-resolution videos because of their hefty data rate as well as small latency requirement. Henceforth, 5G is planned to attain the requirements of present-age smartphone users at mm-wave bands which basically tend to enhance available bandwidth. MM wave bands produce greater signal losses (Zhang et al. 2017). Due to this, in the present work, we have designed a microstrip antenna with a circular patch as well as a 2×2 Microstrip Patch Antenna Finite Array resonating at 5 GHZ. Both of these designs are able to provide higher bandwidth along with high gain as well as better efficiency. To serve against the requirements of 5G applications, main service types have been identified like media as well as virtual communication (Salleh et al. 2012).

4.4.1 Design of 5 GHz Circular Patch Antenna

A microstrip antenna with a circular patch suitable for 5G applications is being designed, simulated as well as analyzed using ANSYS software (Fig. 4.6).

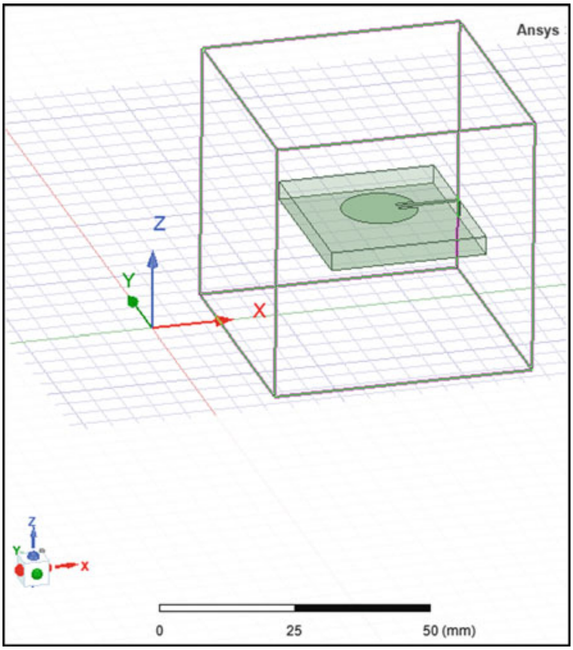


Fig. 4.6 Circular patch antenna structure for 5 GHz resonance frequency

4.4.1.1 Return Loss

Return loss or S parameter specifies the measure of electromagnetic power from the microstrip patch antenna which is reflected back and therefore it is acknowledged as the reflection coefficient. To attain an effective radiation mode, the value of return loss should be less than −10 dB. Figure 4.7 depicts the return loss versus frequency plot of the projected antenna. From the curvature, it is observed that the return loss is −21.06 dB at 5 GHz which in turn specifies the impedance match has been improved and the energy loss has been reduced (Fig. 4.8).

4.4.1.2 Gain

The parameter gain designates how effectively an antenna can transmit or receive power in a specific direction. Figure 4.9 depicts the 3D plot related to far-field gain with respect to the projected antenna and the assessed topmost value of gain is 5.9 dB at 5 GHz.

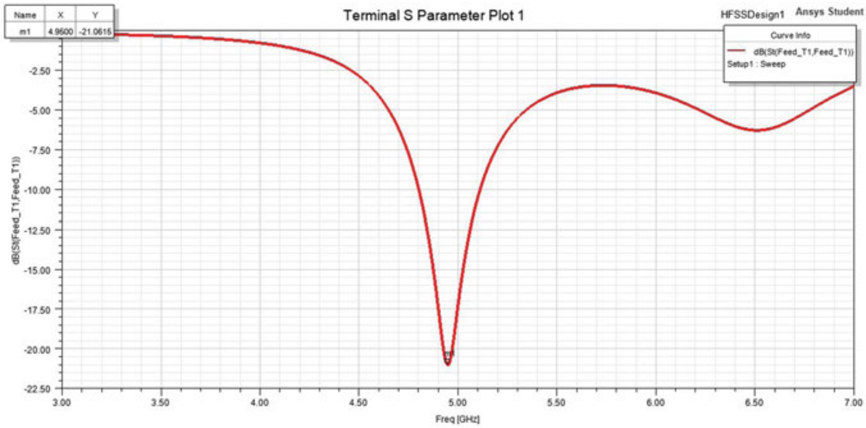


Fig. 4.7 S-parameter versus frequency curve of the projected circular patch antenna

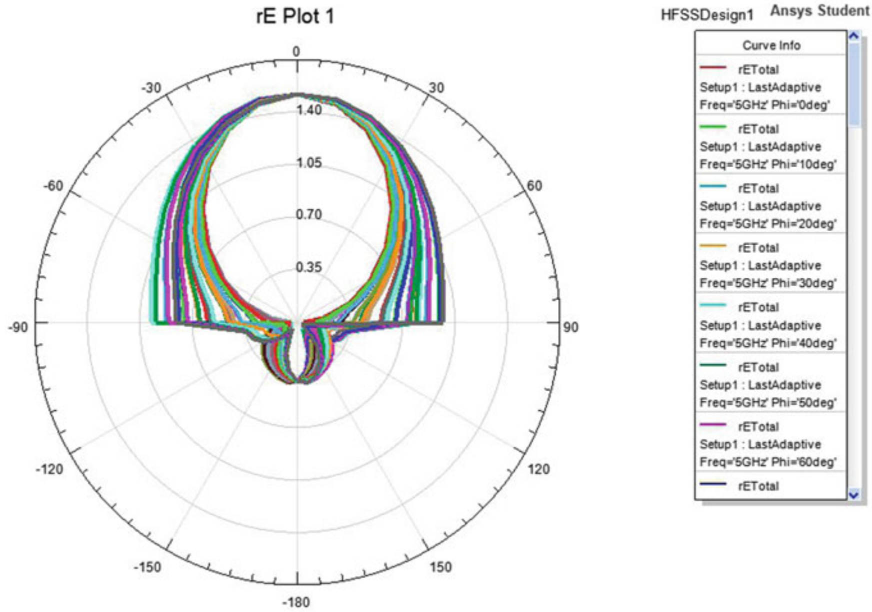


Fig. 4.8 rE plot of projected circular patch antenna

4.4.2 Design of 5 GHz 2 × 2 Microstrip Patch Antenna Finite Array

The simulation and analysis of 2 × 2 microstrip patch antenna array suitable for 5G applications are carried out by means of ANSYS software (Figs. 4.10 and 4.11).

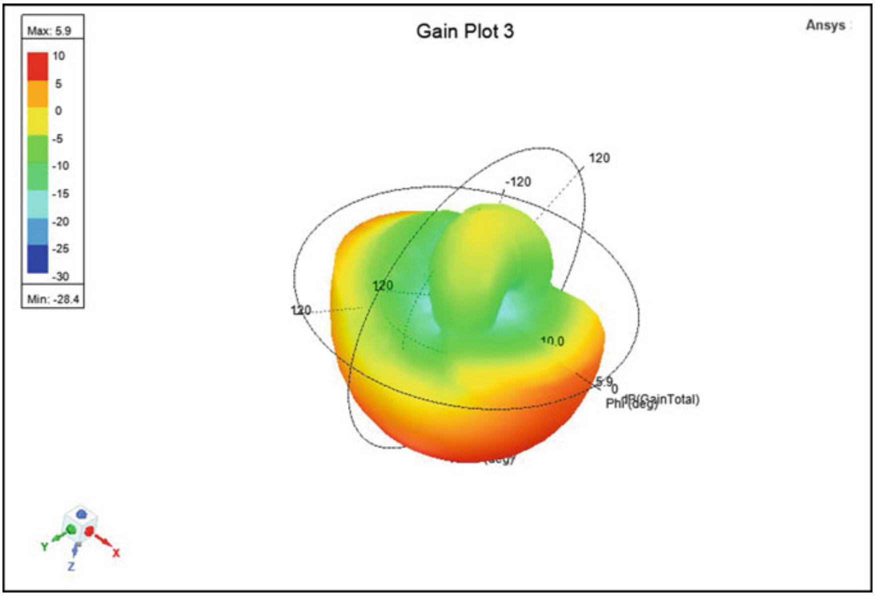


Fig. 4.9 3D plot of far-field gain at 5 GHz

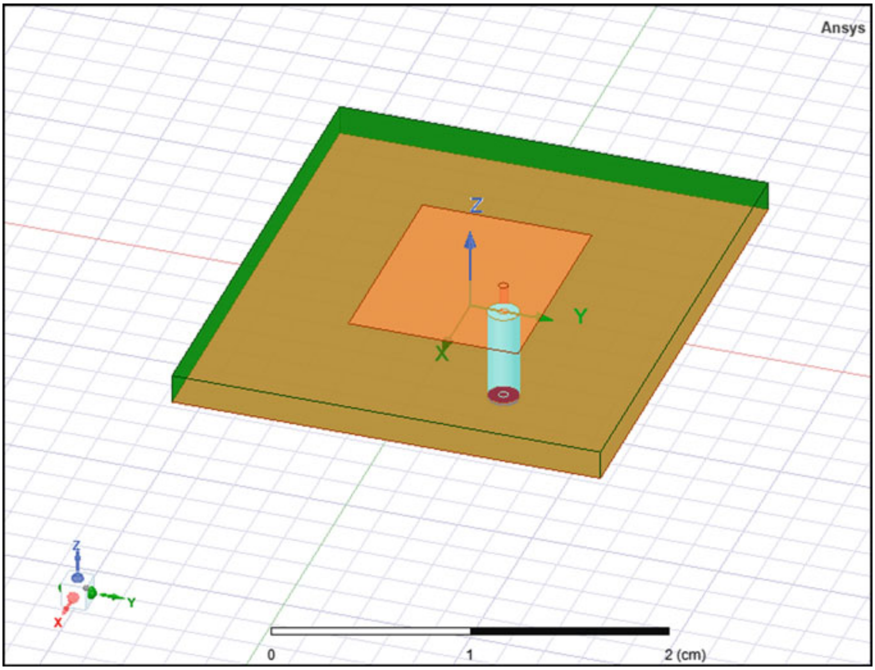


Fig. 4.10 Design of microstrip patch antenna suitable for 5 GHz frequency

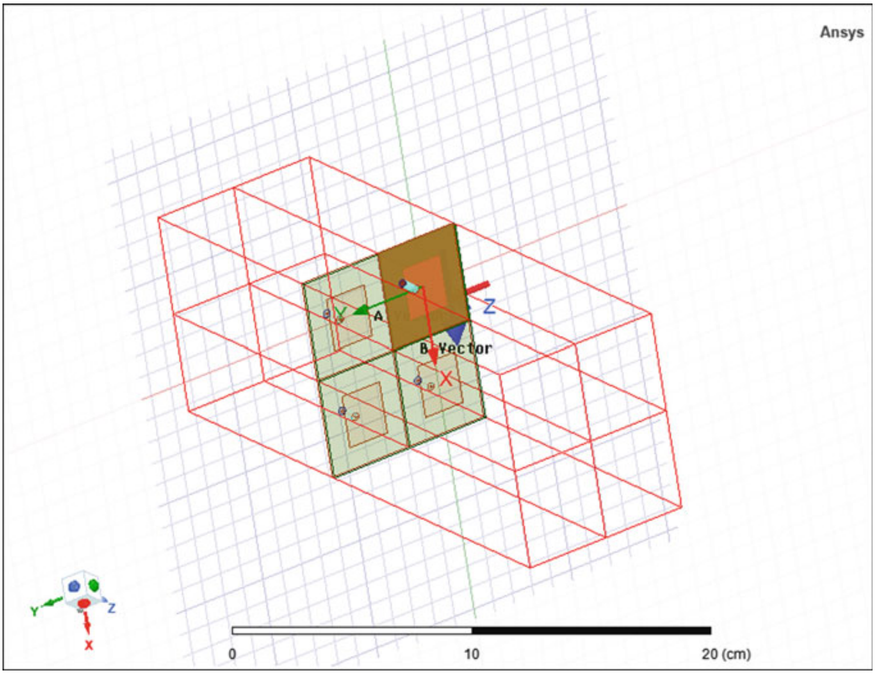


Fig. 4.11 Design of 2 × 2 microstrip patch antenna array suitable for 5 GHz frequency

4.4.2.1 Return Loss

Figure 4.12 depicts the S parameter versus frequency plot relative to the projected antenna array. It is observed that the resonance frequency relative to the anticipated antenna array is 5 GHz. From the curve, it is clear that return loss is −35.94 dB at 5 GHz which in turn concludes that the impedance match is improved as well as the energy loss is reduced. The return loss of 2 × 2 Microstrip patch antenna array is also better than a single patch antenna.

4.4.2.2 Gain

Figure 4.13 displays the 3D plot related to far-field gain with respect to the projected 2 × 2 Microstrip Patch Antenna Array. Here the assessed topmost value of gain is 14 dB at the 5 GHz resonance frequency which is much better than a single antenna.

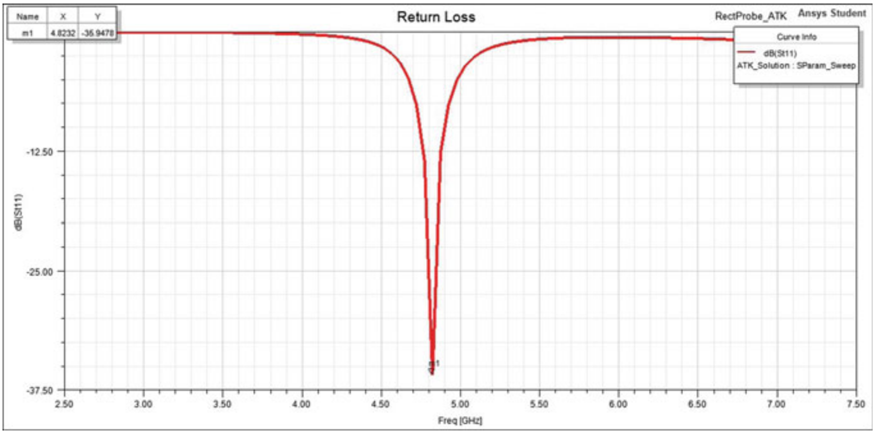


Fig. 4.12 S-Parameter versus frequency curve of the proposed 2 × 2 microstrip patch antenna array

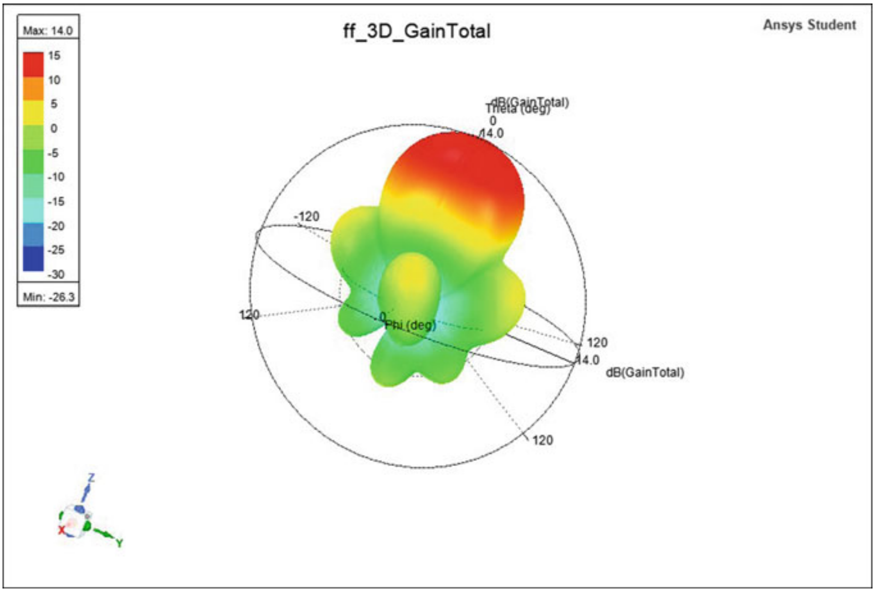


Fig. 4.13 D plot of far-field gain at 5 GHz

4.5 Applications and Examples of IoT in the Smart Communication and 5G

In view of the white papers in which motorists and merchandisers predicated their review connectivity approach against the IoT conditions with emphasis on the likely implicit risks that may arise from new connectivity results and the various reviews which have been presented on structure predicated network for M2M communication. In this chapter comprehensive study of different existing emerging LPWA IoT has been presented with different solutions for 5G mobile network and communication. The main goal of this chapter is to deliver and discuss a broad scope and application of MTC use cases development and requirements, exploring the available connectivity landscape options and promising network enablers to enable the 5G new service requirements and coming up with a context-aware congestion control (CACC) mechanism for lightweight CoAP/UDP-based IoT networks for an efficient resource utilization. Supposedly, the survey on 5G Networks for the Internet of Things: Communication Technologies and Challenges is the first audit paper to exhaustively stress a flat-out idea on 5G portable organization for the IoT (Akpakwu et al. 2018).

Figure 4.14 shows the basic model that proposes a plan of network architecture design for 5G mobile frameworks, which is all-IP based model for remote and versatile organizations interoperability. The framework comprises a client terminal (which plays a significant part in the new engineering) and various free, independent radio access advances. Inside every one of the terminals, every one of the radio access innovations is viewed as the IP connection to the external Internet world. Be that as it may, there should be a different radio connection point for each Radio Access Technology (RAT) in the versatile terminal. For a model, to approach four unique RATs, the author really wants to have four distinct access-specific interfaces in the portable terminal and to have every one of them dynamic simultaneously, with intent to have this engineering be practical (Tudzarov and Janevski 2011a).

4.5.1 Role of Smart Communication Technologies for Smart Retailing

Smart retailing is the rise of an environment of shopper-driven retailing exercises. Each action in the worth chain of retailing-including creation and plan, strategies, and warehousing, brought together acquisition, showcasing, activities, administrations, and the executives and income can be consolidated into a digitalized and astute stage. The plan of action is changing and reflected by developing working models, information dealing with, advertising approaches, and store network coordination. Arising advancements like computerized reasoning (AI), increased reality (AR), augmented reality (VR), web of things (IoT) and Big information investigation are turning out to be more modern and incorporated (Rajput 2020). The mix of these

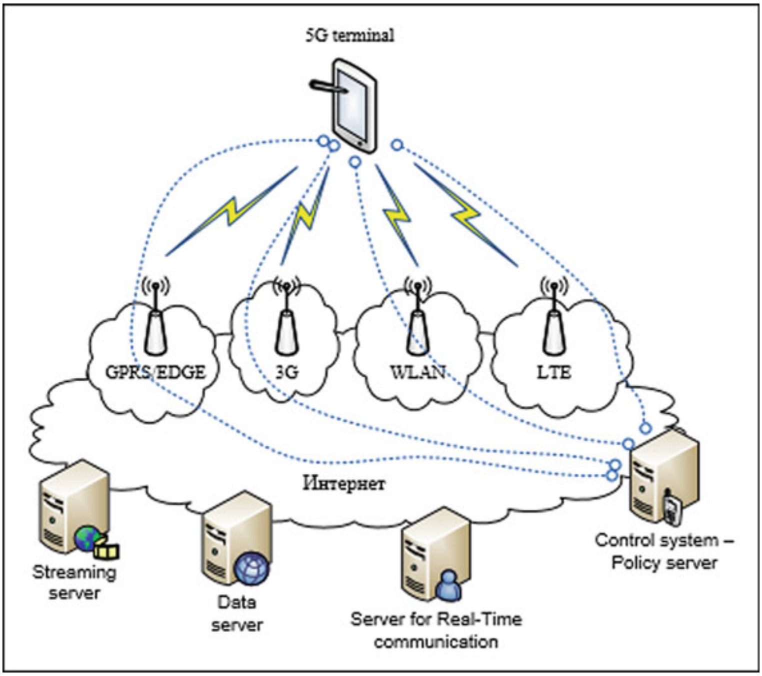


Fig. 4.14 Basic architecture of 5G mobile network

shrewd innovations with retail to make independent outlets can be named as savvy retailing. As of now, there are three fundamental business structures in the automated retail industry, including automated general stores, open retires, and savvy retail compartments (Sujata et al. 2019).

4.5.2 Impact of IoT on 5G

Over the most recent couple of centuries, technological impact of IoT in communication has transformed the unavoidable measure counting an enormous assortment of developed tenders including different sensor categories. An enormous number of trainings has been relied upon to fill in the IoT-based invention in day-to-day life, offerings before long of arranging of up to billions of gadgets with a normal of 6–7 gadgets for every individual by 2020 (Lee 2017). With a large portion of the above smart gadget and convention concerns settled throughout the most recent 10 years, the digital physical component and device-to-device collaboration assembly of sensors as well as other frameworks based on sensors are currently developing. A ton of conversation around 5G networking of the fifth age of versatile correspondence networks has been occurring as of late. As every new era shows up, they are joined by greater limits and quicker associations of IoT and 5G. The associated benefits comprise improved coverage in comparison with 4G, so that 5G cell connected towers would have enhanced capacity and due to this enhancement more cells will connect simultaneously (Atzori et al. 2010). Another collaborative advantage is the reduction in latency, due to that 5G excludes time delays in a significant manner. Quicker connection enhances in new 5G network are near about 10 times quicker than 4G connections. This is assured that the imminent 5G will be approachable, quick, and more energy-efficient (Khuntia et al. 2021; Liu et al. 2017). Though quicker moves and insufficient idleness presented through 5G will draw in cell phone clients, it is sufficient not to clarify the enormous expense Bloomberg gauges at 200 billion per year. Figure 4.15 depicted the diversity of Internet of things in any industry and the growth of that in Germany. What's more, all-out IoT interest in 2022, based on client's satisfactions and requirements.

The upcoming 5G network will chiefly add to the making of the Internet of Things as a fundamental area of the planet by establishing the groundwork to open its maximum capacity. Late portable innovation has no equivalent convenience for 20.4 billion associated devices and is setting up the trading of information without little slacks (Tudzarov and Janevski 2011b). IoT's progressively developing technologies discovered a remarkable phase of progress in five significant areas in the technology

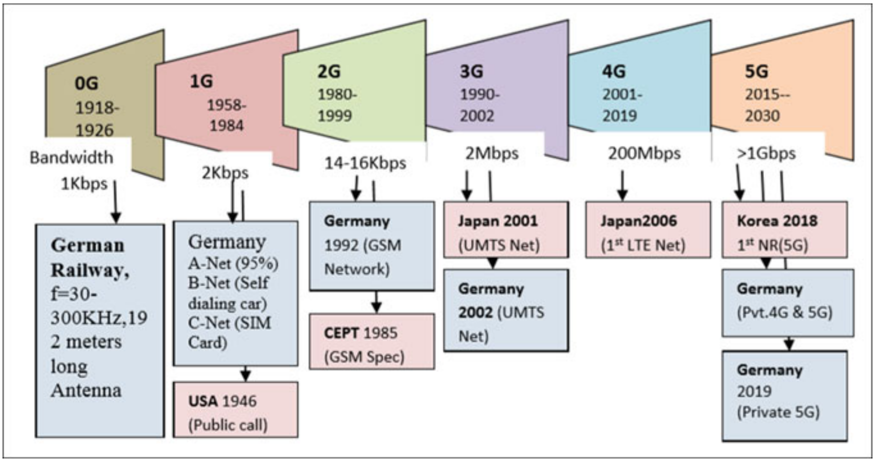


Fig. 4.15 Generation of private network technology in Germany (from 0 to 5G)

interruption first is sensing the Endpoints of IoT, next is communication as IoT communication than secure IoT, further IoT information and analytics are required with lastly acting IoT as an Artificial Intelligence (AI) (Velev 2011).

4.5.3 5G Challenges

There are many challenges in the development of 5G communication and network, the early difficulties include the following.

(i) In 5G organizations a mix of low, medium, and high channels. Every one of the versatile administrators is relied upon to carry out 5G administrations through a more modest spectrum band (Rai et al. 2021a).
(ii) All industries and administrations should be gradually shifted towards 5G from 4G, because of the basic consistent prerequisites of on-request conveyance of equipment, programming, and administrations.
(iii) Implementation of the 5G arrangements, requires significant upgrades in distributed computing organizations, in the space of virtualization and MIMO.
(iv) It is extremely critical to accomplish interoperability between client components (UE) and economically created 5G networking to approve key innovation.
(v) In ventures, 5G plans of action should be a minimal expense and elite execution executions as portrayed in Fig. 4.16. Fostering an organization foundation and application environment to assist a manageable business with displaying for 5G services (Alvi et al. 2015).

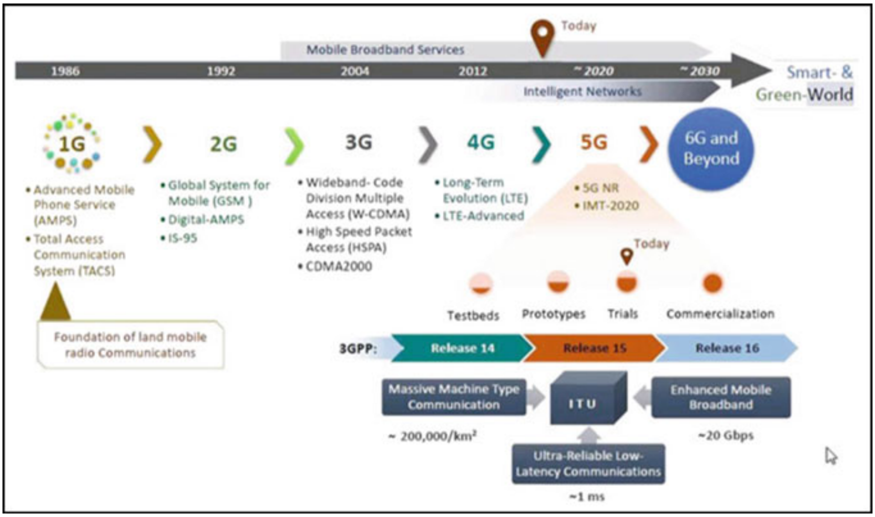


Fig. 4.16 Different phases of mobile technologies with their significance

4.6 Application of 5G Over IoT in the Different Areas

5G and IoT organized would likewise assist with putting every item on the racks to the Internet. Buyer items don't should be consistently associated with the Internet as equipment gadgets, yet they can send and get information about themselves as associated brilliant items in light of occasion-based encounters with clients and different elements through Tag Scanning, RFID readers, NFC labels, and more (Billinghurst and Kato 2002). The existing wireless framework isn't capable of managing so many organization gadgets, yet 5G will make it conceivable. Brilliant packaging and digital labels can change the manner in which retailers oversee stock and operations and give a hotbed of creative mind to involve them as a method for interfacing with purchasers in an innovative way. 4G doesn't oversee information load from the consistently expanding number of online sensors and associated gadgets, restricting what IoT can really do. The 5G is the ideal empowering agent for the Internet of Things with its high information speed, low inactivity, expanded portability, low-energy utilization, cost efficiency, and the capacity to deal with a lot bigger gadgets. 5G can assume a significant part in changing the manner we convey as well as in changing industry and society (Stergiou and Psannis 2017). There are various organizations of different fields in which 5G as well as IoT can cause interference together, few examples are as follows:

4.6.1 Automated Self-driving Cars and Other Vehicles

The sensors create huge amounts of information on self-driving vehicles, temperature estimation, traffic conditions, climate, GPS area, and so on the assortment and as re-enactment of every amount of information require a lot of energy. These vehicles additionally intensely hand off on ongoing data transmission to offer ideal types of assistance. In any case, with rapid correspondence and low inertness, this keen consideration will actually want to gather a wide range of information on a continuous premise, including time-basic information on which calculations will work freely to monitor the functioning state of the vehicle and work on future plans (Rai et al. 2021b).

4.6.2 Smart Automated Healthcare

As a wide range of clinical gadgets are fuelled by IoT, changes in their administrations will likewise be found in the clinical field. Not withstanding appropriate medical care framework, the IoT connection will extraordinarily benefit provincial regions and other comparable far-off areas. With such low dormancy, it turns into a choice to give elite medical care administrations, for example, distant medical procedures (Rai et al. 2021a).

4.6.3 Smart Logistics and Supply Chain Management

5G systems administration will further develop start to finish coordinated factors tasks with cutting edge IoT observing sensors. High rates and low inertness won't just permit information to be gotten progressively, yet additionally empower energy efficiency to produce more different data at all focuses inside a production network for quite a while. A purchaser would approach point by point data like where the fish she had recently purchased was gotten, the temperature at which it was treated during handling, and when conveyed to the vendor (Mishra and Chakraborty 2020).

4.6.4 Clean and Smart Cities and Town

The coming advanced technology 5G will permit more extensive applications from water and waste administration to brilliant metropolitan projects, traffic control to improved offices for medical care. Shrewdness urban communities should also benefited from the new-age 5G or 6G network as an ever-increasing number of gadgets arrive at the metropolitan foundation. Not exclusively can 5G handle the huge information load, but it will likewise make it a reality to join numerous brilliant frameworks that persistently collaborate with one another, bringing a genuinely associated city's fantasy nearer (Chakraborty et al. 2013).

4.6.5 Smart Marketing and Retail or Chain Store

The endeavor to figure out the client commitment and experience through mobiles or online calls, retailer IoT has a positive effect due to 5G's appearance. Further developed availability and a bigger number of organizations associated gadgets would permit new and imaginative approaches to drawing in customers to connect quickly with customers through better computerized signatures collectively. That could be achieved with expanded reality and augmented reality very well. Retailers will actually want to improve the shopping experience by executing Omni channel deals exercises all the more efficiently (Patel et al. 2021).

4.6.6 Intelligent Automotive and Smart Industries

The recent and foremost uses of 5G interfacing Devices with Augmented Virtual Reality (AVR) in all fields of daily life like medical and transportation, education and super marketing, etc. (Rath et al. 2018; Sinha et al. 2017). Upgraded vehicle normal administrations may incorporate direct vehicle-to-walker and vehicle-to-framework correspondence, along with independent driving correspondence which can be agreeable by any network. Upheld use cases would focus on vehicle solace and well-being, including continuous correspondence of direction, course arranging, coordinated driving, and local area refreshes. By consolidating 5G security into the central organization engineering, we would likewise give an incredibly solid organization to modern IoT.

4.6.7 Smart Agriculture

Farming is done in each country for ages. Agribusiness is the science and craft of developing plants. Agribusiness was the critical improvement in the ascent of stationary human development. Agribusiness is done physically from ages. As the world is moving into new innovations and executions it is a vital objective to drift up with farming too. IoT assumes a vital part in brilliant horticulture. IoT sensors are equipped for giving data about agribusiness fields. We have proposed an IoT and savvy horticulture framework utilizing robotization. This IoT-based agriculture observing framework utilizes remote sensor networks that gather information from various sensors conveyed at different hubs and send it through the remote convention. This brilliant agribusiness utilizing IoT framework is fuelled by Arduino, which comprises of temperature sensor, moisture sensor, water level sensor, DC engine, and GPRS module. Whenever the IoT-based agribusiness observing framework begins it checks the water level, dampness, and dampness level. It sends an SMS alert on the telephone about the levels. Sensors sense the degree of water assuming it goes down, it consequently begins the water siphon. Assuming the temperature goes over the level, the fan begins (Patel et al. 2021).

4.6.8 Establishment Between 5G and IoT Eco-system

The IoT drive extra opportunities to extend abilities; administrations along with dependability as greater advancement assets for 5G and their market. As per Statista survey, “The IoT (Internet of Things) based Devices introduced to almost 31 billion worldwide by 2020” (Tudzarov and Janevski 2011b) for reliable and sustainable communications, factors expected to fabricate a 5G-IoT environment are as per the preceding with some factors such as Automatic power supply, innovators, integrators, and implementers, an electronic recycling and upgradation program, with more efficient and less costly (Rath et al. 2018; Sinha et al. 2017; Priya et al. 2020).

4.7 Future Enhancement in 5G Using Antenna Array

5G cannot replace current cellular technologies all of a sudden. 5G can be treated as an alternative to the current situation, and at the same time also modified hardware is desired for taking full benefit of the supremacy of 5G, IoT's subsequent module "networks" would have additional possibilities now to deal with a widespread spectrum of frequencies as desired. The strategy is to substitute 4G communication in a stable manner with all the set-up available at present and this can be executed on numerous levels and segments such as software end, hardware implementation as well as access points. This desires a great asset for users and industries, but diversity in different geographical reasons drive dissimilar timelines to substitute 4G communication and networks and may be caused challenges in the implementation of 5G network services. This is a problem from the IoT side that behaves as the business changes. A requirement for a typical structure to accomplish communal IoT back-end responsibilities, for instance, signal processing, signal storage as well as firmware upgradation, is handling not in the proper way. This more innovative wanted structure, likely to realize diverse IoT solutions efforts with communal back-end services, will assure levels of compactness and interoperability nearly unbearable to attain with the existing group of IoT solutions. Producing such type of models is not an easy task, there are numerous hurdles as well as challenges fronting the standardization as well as the applicability of IoT solutions in addition such type of models wishes to overcome each of heralds, out of all interoperability is one of the foremost hurdles and needed more research on it.

Summary

This chapter basically focused on the different solutions for communication using technology like IoT, AI/ML, and advancement in cellular networking. Based on the literature review smart antenna can be used for smooth and fast communication over 5G networks. In this chapter, two different antennae have been studied with their parameters for 5G communication and beyond.

References

Aakpakwu GA et al (2018) A survey on 5G networks for the internet of things: communication technologies and challenges. IEEE Access 18

Alvi SA et al (2015) Internet of multimedia things: vision and challenges. Ad Hoc Netw 33:87–111

Atzori L, Iera A, Morabito G (2010) The internet of things: a survey. Comput Netw 2787–2805. Elsevier

Billinghurst M, Kato H (2002) Collaborative augmented reality. Commun ACM 45(7):64–70

Castellani AP, Fossati T, Loreto S (2012) HTTP-CoAP cross protocol proxy: an implementation viewpoint. In: 2012 IEEE 9th international conference on mobile ad-hoc and sensor systems (MASS 2012), pp 1–6. https://doi.org/10.1109/MASS.2012.6708523

Chakraborty C, Gupta B, Ghosh SK (2013) A review on telemedicine-based WBAN framework for patient monitoring. Telemed e-Health 19(8):619–626

Chen M, Wan J, Li F (2012) Machine-to-machine communications: architectures, standards and applications. KSII Trans Internet Inf Syst 6(2):480–497

David L, Vasconcelos R, Alves L et al (2013) A DDS-based middleware for scalable tracking, communication and collaboration of mobile nodes. J Internet Serv Appl 4:16. https://doi.org/10.1186/1869-0238-4-16

Dinculeana D, Cheng X (2019) Vulnerabilities and limitations of MQTT protocol used between IoT devices. MDPI Appl Sci. 27 Feb 2019

Hwang HC, Park J, Shon JG (2016) Design and implementation of a reliable message transmission system based on MQTT protocol in IoT. Wireless Pers Commun 91(4):1765–1777

Khuntia M, Singh D, Sahoo S (2021) Impact of internet of things (IoT) on 5G. Smart innovation, systems and technologies, p 153. https://doi.org/10.1007/978-981-15-6202-0_14

Kundu K, Dubey A, Dhama A, Pathak NN (2021) Planar inverted F antenna, PIFA array in 5G applications. J Phys Ser (Cite as: Krishanu Kundu et al 2021 J Phys Conf Ser 2062 012002)

Kundu K et al (2021a) Planar inverted F antenna, PIFA array in 5G applications. J Phys Conf Ser 2062:012002

Lee S (2017) Communication technology and application of internet of things (IoT) in smart home environment. Int J Control Autom 10(3):397–404. https://doi.org/10.14257/ijca.2017.10.3.32

Liu J, Wan J, Jia DY, Zeng B, Li D, Hsu C-H, Chen H (2017) High-efficiency urban-traffic management in context ware computing and 5G communication. IEEE Commun Mag 55(1):34–40

Ludovici A, Moreno P, Calveras A (2013) TinyCoAP: a novel constrained application protocol (CoAP) implementation for embedding RESTful web services in wireless sensor networks based on tinyos. J Sens Actuator Netw 2(2):288–315

Luzuriaga JE, Perez M, Boronat P, Cano JC, Calafate C, Manzoni P (2015) A comparative evaluation of AMQP and MQTT protocols over unstable and mobile networks. In: 12th annual IEEE consumer communications and networking conference, pp 931–936

Mishra KN, Chakraborty C (2020) A novel approach toward enhancing the quality of life in smart cities using cloud sand IoT-based technologies. In: Digital twin technologies and smart cities. Springer, Cham, pp 19–35

Naik N (2017) Choice of effective messaging protocols for IoT systems: MQTT, CoAP, AMQP and HTTP. In: 2017 IEEE international systems engineering symposium (ISSE). IEEE, pp 1–7

Oussama et al (2019) Smart device for multi-band industrial IoT communications. In: The international workshop on the future of internet of everything (FIoE). http://creativecommons.org/licenses/by-nc-nd/4.0/

Patel GS, Rai A, Narayan Das N, Singh RP (eds) (2021) Smart agriculture: emerging pedagogies of deep learning, machine learning and internet of things, 1st edn. CRC Press. https://doi.org/10.1201/b22627

Priya A, Rai A, Singh RP (2021) Internet of things: architecture, applications and future aspects. Advances in smart communication and imaging systems: select proceedings of MedCom 2020, vol 721, p 183

Rai A et al (2021a) Smart sensors transform healthcare system. Deep learning and IoT in healthcare systems. Apple Academic Press, pp 191–214
Rai A et al (2021b) IoT-aided robotics development and applications with AI. Emergence of cyber physical system and IoT in smart automation and robotics, vol 1
Rajput R (2020) How IoT embraces smart communication. https://readwrite.com/how-iot-embraces-smart-communication.
Ratasuk R, Vejlgaard B, Mangalvedhe N, Ghosh A (2016) NB-IoT system for M2M communication on device to device communications for 5G NETWORKS (WD5G 2016)
Rath M, Swain J, Pati B, Pattanayak BK (2018) Network security: attacks and control in MANET. In: Hand book of research on network forensics and analysis techniques. IGI Global, pp 19–37
Salleh MKM, Jaafar H, Ali MT, Subhari S, Yusof AL (2012) Improving gain performance by using air substrate at 5.8 GHz. In: International conference on computer and communication engineering, Kuala Lumpur, Malaysia
Sinha RS, Wei Y, Hwang S-H (2017) A survey on LPWA technology: LoRa and NB-IoT. ICT Express 3(1):14–21
Soni D, Makwana A (2017) A survey on MQTT: a protocol of internet of things. In: International conference on telecommunication, power analysis and computing techniques, April 2017
Stergiou C, Psannis KE (2017) Recent advances delivered by mobile cloud computing and internet of things for big data applications: a survey. Int J Netw Manag 27(3):e1930
Sujata J, Mukul P, Hasandeep K (2019) Role of smart communication technologies for smart retailing. Int J Innov Technol Explor Eng (IJITEE) 8(9S2). ISSN 2278-3075
Thangavel D, Ma X, Valera A, Tan H-X, Tan CK-Y (2014) Performance evaluation of MQTT and CoAP via a common middleware. In: 2014 IEEE ninth international conference on intelligent sensors, sensor networks and information processing. IEEE, pp 1–6
Tudzarov A, Janevski T (2011a) Functional architecture for 5G mobile networks. Int J Adv Sci Technol 32
Tudzarov A, Janevski T (2011b) Functional architecture for 5G mobile networks. Int J Adv Sci Technol 32:65–78
Velev DG (2011) Internet of things: analysis and challenges. Econ Altern 2:99–109
Zhang J, Ge X, Li Q, Guizani M, Zhang Y (2017) 5G millimeter wave antenna array: design and challenges. IEEE Wirel Commun

Chapter 5
Machine Learning and Deep Reinforcement Learning in Wireless Networks and Communication Applications

Om Prakash, Prabina Pattanayak, Amrita Rai, and Korhan Cengiz

Abstract Wireless networks and communication of the future will have to manage an ever-increasing density of mobile users using a wide range of services and apps, as well as a surge in mobile data traffic. Meanwhile, networks are becoming increasingly dense, heterogeneous, decentralized, and ad hoc, encompassing a wide range of network elements. As a result, a number of service goals, such as high throughput and low latency, must be met, as a result, resource allocation must be established and optimized. Traditional service and resource management methods that need complete and perfect system information are inefficient or inapplicable in wireless network environments due to the inherent dynamic and uncertainty. This chapter covers similar research that employs Deep reinforcement learning (DRL) to address various difficulties in 5G networks after first going through the fundamental notions of DRL. Finally, we look at some of the DRL techniques that have been presented to deal with emerging communications and networking difficulties.

Keywords Reinforcement learning (RL) · Deep reinforcement learning (DRL) · Deep-Q Networks (DQN) · Q-function · Q-values

---

O. Prakash (✉)
Sri Venkateswara College of Engineering Technology, R V S Nagar, Chittoor, Andhra Pradesh, India
e-mail: om4096@gmail.com

P. Pattanayak
National Institute Of Technology QQ5R+3WM, NIT Road, Fakiratilla, Silchar, Assam 788010, India
e-mail: prabina@ece.nits.ac.in

A. Rai
GL Bajaj Institute of Technology and Management, Plot No 2, Knowledge Park III, Greater Noida, Uttar Pradesh 201306, India

K. Cengiz
Istinye University, Istanbul, Turkey
e-mail: korhan.cengiiz@istinye.edu.tr

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

A. Rai et al. (eds.), *Paradigms of Smart and Intelligent Communication, 5G and Beyond*, Transactions on Computer Systems and Networks,
https://doi.org/10.1007/978-981-99-0109-8_5

5.1 Introduction

Machine learning (ML) is an elegant tool in designing many applications that can serve in the coming future from recommendation on e-commerce websites to content filtering on the social network by giving suggestion for making communication with each other (Qiu et al. 2016; Wang et al. 2021; Ahmed et al. 2022; Gronauer and Diepold 2022). ML techniques include conventional algorithms like Decision Tree and Regression Tree whereas Deep learning (DL) techniques include an automatic encoder and neural network. The basic block diagram of machine learning architecture is shown in Fig. 5.1. Machine learning is applied in advanced business and research for organizations today (Xiang and Foo 2021; Bellemare et al. 2017). It involves algorithms and neural network applications for improving their performance. Machine learning algorithms model mathematical simulations using referred data to take decisions.

In today's world of increasing complexity in mobile network technology, wireless access technologies have been evolving steadily. To ensure the lead for the future, Telematics has introduced 5G wireless technology to the future spectrum of mobile business and is replete with so many development opportunities that significantly improve on the previous generation and offer a whole number of new innovations. The ability to operate correctly on previously unknown inputs is the promise of applying deep learning technology in reinforcement learning. For example, image identification neural networks can recognize a bird in a photograph even if they have never seen the image or the bird. The requirement to predefine the environment is reduced because deep reinforcement learning allows raw data (e.g., pixels) as input, allowing the model to be used in a variety of situations. Deep reinforcement learning (DRL) approaches can be created to be universal because of this degree of abstraction. allowing the same model to be used for multiple tasks. Deep Reinforcement Learning blends reinforcement learning and deep learning to improve the generalization ability of policies developed with deep RL policies (DRL). It has the ability to learn end-to-end learning in robots, video games, natural language processing, computer vision, healthcare, and other domains from raw sensors or images Google Deep Mindin 2013 used Deep-Q Networks (DQN) to play Atari games, which was a watershed moment in value-based DRL.

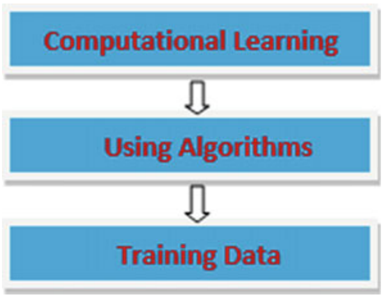


Fig. 5.1 Basics of Machine Learning and Deep Learning

Applying profound support learning (DRL) methods to improve remote enterprises' operations is becoming more and more popular. In this research, we examine three state-of-the-art DRL approaches for remote organization streamlining: Deep Deterministic Policy Gradient (DDPG), Neural Episodic Control (NEC), and Variance Based Control (VBC). We illustrate how the overall organizational improvement problem is structured as RL and provide specifics for the three remote systems administration strategies. A verifiable organization activity dataset is used for extensive analyses, and the presentation of the further developing rate and combination speed for these well-known DRL procedures is examined. While DDPG and VBC demonstrate significant promise for computerizing remote organization improvement, NEC has a significantly better union rate yet experiences the restricted activity space and doesn't perform seriously in its ongoing structure (Tanveer et al. 2022).

5.1.1 Deep Learning

A collection of inputs are transformed into a set of outputs by an artificial neural network in a process known as deep learning. Deep learning algorithms have been shown to handle more complex, high-dimensional raw input data, such photos, with less manual feature engineering than earlier systems. This has allowed for substantial advancement in fields like computer vision and natural language processing.

5.1.2 Reinforcement Learning (RL)

Reinforcement learning teaches an agent to make judgments by allowing them to make mistakes. An agent in a Markov decision process (MDP) is in a state at each time step, takes action, and receives a scalar reward before transitioning to the next stage based on the dynamics of the environment (Bellemare et al. 2017). In order to maximize profits, the agent tries to figure out a policy, or a route from observations to actions (expected sum of rewards). In reinforcement learning, the dynamics are only sampled by the algorithm (as opposed to optimal control).

5.1.3 Deep Reinforcement Learning (DRL)

The states of the MDP are high-dimensional in many real-world decision-making problems. In order to solve MDPs and build specialized algorithms that operate well in this environment, deep reinforcement learning techniques usually represent the policy or other learned functions as a neural network (Li 2019).

5.1.4 From the RL to the DRL

For each state-action pair, Q-learning necessitates the agent keeping track of and updating a set of Q-values. Wireless networks, on the other hand, are likely to be large-scale in the future, diverse, and decentralized. As a result, the number of alternative system state values grows exponentially. Furthermore, due to the tremendous variety and unpredictability of system components and surrounding parameters. Hidden system states, or even an infinite number of system states, are possible (François-Lavet et al. 2018; Nguyen Cong Luong et al. 2019). The expense of computing and storing all Q-values becomes essentially unsustainable in this case. The "curse of dimensionality" is a term used to describe this condition. DRL blends RL and deep learning methodologies to solve this problem. Deep Q-learning, which makes use of a DQN, is a common feature of DRL models.

As the data source, all input states are routed through several layers of neural networks. each with its own set of weight factors. Finally, the DQN generates a set of Q-values for each of the actions that can be taken. The purpose of learning in the DQN is to use historical data such as Q-values, actions, and state changes to train and pick the most viable weight components. Because the underlying neural network is linear in complexity, a multilayer perceptron is used to calculate the Q-values and actions for a DQN. Furthermore, the DQN's number of inputs is solely defined by the type of each state. For each input, several state values, such as distinct channel states without, can be conveyed (Packer et al. 2019; Mnih et al. 2016; Haarnoja et al. 2018; Lillicrap et al. 2016).

In order to take into account a dynamic multichannel access problem, where clients choose the channel to exchange information and several corresponding channels follow a mysterious combined Markov model, finding a method that increases the typical long-distance number of efficient transmissions is the objective. The problem is conceptualized as a somewhat recognized Markov choice cycle with enigmatic framework components. Apply the concept of support learning and implement a thorough Q-organization to primarily overcome the challenges of obscure elements and constricting calculation (DQN). Many researchers first examine the optimal approach for fixed design channel exchanging with known framework components, then demonstrate through simulations that DQN can achieve similarly optimal execution without being aware of the framework insights. Then, using both larger reproductions and actual information, compare the demonstration of DQN with a Myopic method and a Whittle Index-based heuristic, and demonstrate that DQN achieves almost optimal execution in more perplexing conditions. Finally, we suggest a flexible DQN technique that can adapt its learning in time-varying circumstances (Li and Li 2020; Zhao et al. 2022; Nair et al. 2015; Hessel et al. 2018; Zikria et al. 2020; Du et al. 2020).

The fundamental distinctions between RL and DRL when the Q-function is created using neural network are the following two features.

(1) The purpose of learning in the DQN is to use historical data such as Q-values, actions, and state changes to train and pick the most viable weight components.

Because the underlying neural network is linear in complexity, a multilayer perceptron is used to calculate the Q-values and actions for a DQN.

(2) DRL reduces the Q-values of the target and the estimated Q-values by adjusting the weight every few time steps, resulting in a more stable DRL learning process.

DRL reduces model complexity greatly when compared to RL, especially when used to handle a variety of challenges in future communication systems. DRL also excels RL when it comes to extracting system properties and predicting optimal rewards and behaviours. RL, on the other hand, considers and transfers from states to mini batches. The deep network weight factors are then trained using the historical data collected by DRL. According to this, with the support of system and environment knowledge, which provides additional information, the training process can be hastened (Table 5.1).

Table 5.1 Various DRL algorithms for different areas of application

DRL algorithms	Main techniques	
Value-based	DQN (Qiu et al. 2016)	Experience replay, target network, clipping rewards, and skipping farms
	Double DQN (Wang et al. 2021)	Double Q-Learning
	Dueling DQN (Ahmed et al. 2022)	Dueling neural network architecture
	Prioritized DQN (Rai et al. 2020)	Prioritized experience replay
	Bootstrapped DQN (Qin et al. 2018)	Deep exploration with DNNs
	Distributional DQN (Tanveer et al. 2022)	Distributional Bellman equation
	Noisy DQN (Li and Li 2020)	Parametric noise added to weights
	Rainbow DQN (Hasselt et al. 2016)	Combine 6 extensions to the DQN
	Hierarchical DQN (Wang et al. 2016)	Hierarchical value functions
	Gorila (Xiang and Foo, 2021)	Asynchronous training for multi-agents
Policy based	TRPO (Bellemare et al. 2017)	KL divergence constraint
	PPO (Yang et al. 2020)	Specialized clipping in the objective function
Actro critic	Deep DPG (Wang et al. 2018)	DNN and DPG
	TD3 (Li 2019)	Twine Delayed DDPG
	PGQ (Gronauer and Diepold 2022)	Policy gradient and Q-learning
	Soft Actro Critics (SAC) (Lazaridis, et al. 2020)	Maximum entropy RL framework
	A3C (François-Lavet et al. 2018)	Asynchronous Gradient Descent

5.1.5 Machine Learning (ML)

Machine learning is an area where a computer does not perform tasks with explicit instructions. It includes statistical models and algorithms that give patterns and inferences obtained from the data. Machine Learning algorithms use mathematics, statistics, probability, linear algebra, and calculus deeply. The next generation of wireless communication networks is becoming more and more complex due to a wide number of services, and different characteristics of applications in devices. These approaches do not support future complex wireless systems with regards to the operation and optimization, and low cost of the networks. The exponentially increasing demand in the number of people using wireless communication technologies requires innovation and research into new technologies in the coming future (Rai et al. 2020; Qin et al. 2018; Tanveer et al. 2022). Machine learning (ML) is emerging as one of the most intelligent technologies that will help researchers conceptualize and implement technologies in virtual and augmented real life as shown in Fig. 5.2.

Nowadays, the architecture of machine learning involves major industrial interaction in the automation process. This helps in the optimization of the available resources and gives results based on the data available (Li and Li 2020; Hasselt et al. 2016; Wang et al. 2016). The machine learning architecture explains different layers involved in the machine learning process. The following major steps for the transformation of data into training data for the functioning of the intelligence of a system.

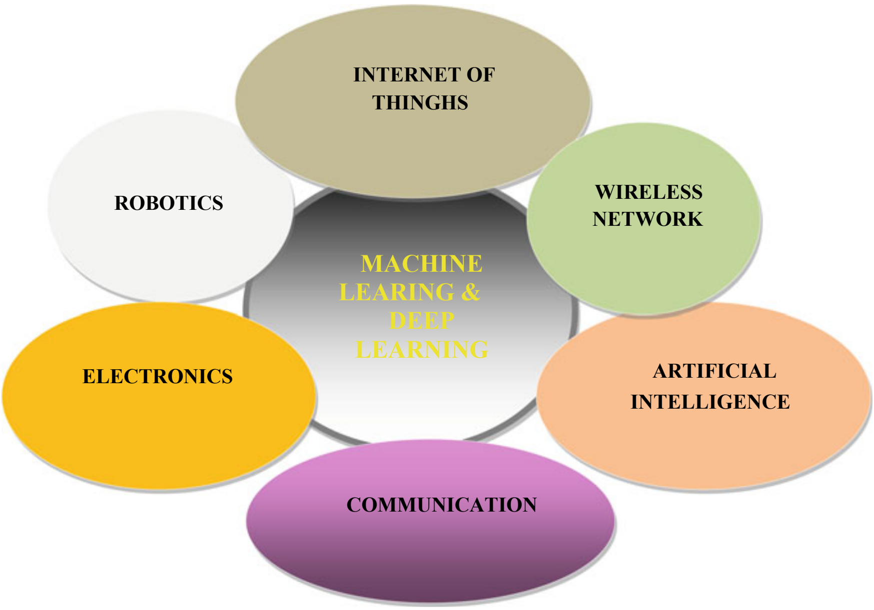


Fig. 5.2 Application area of machine learning and deep learning

5.2 Applications Deep Reinforcement Learning Techniques

Deep reinforcement learning is widely employed in a range of disciplines, as seen in Table 5.2, which is based on (Hasselt et al. 2016; Lillicrap et al. 2016; Nair et al. 2015; Du et al. 2020). This chapter will cover contemporary developments in game development, robots, Transportation, industrial applications, communication, and networking are all examples of natural language processing (NLP), and other areas. Aside from DRL, a variety of methodologies can be utilized in a programme, for example, Alpha Go is taught via supervised learning and reinforcement learning. POMDP issues are given more attention in this study, However, establishing whether an application is a firm POMDP problem is difficult. Not only do we focus on POMDP issues, but we also look at deep reinforcement learning applications in general.

5.2.1 Application in Wireless Network

Over the past 20 years, advancements in AI have been greatly impacted by reinforcement learning, one of the most popular machine learning research subjects (Yang et al. 2020). With reinforcement learning, an agent makes consistent judgments, evaluates the results, and then modifies its approach to produce the best possible policy. Although it has been demonstrated that this learning process converges, it is

Table 5.2 Tabular form of different application areas based on Various DRL

Domains	Application
Healthcare	DTRs, HER/EMR, Diagnosis
Education	Educational games, recommendation, proficiency estimation
Transportation	Traffic control
Energy	Decision control
Finance	Trading, risk management
Science, engineering, art	Math, physics, music, animation
Business management	Recommendation, customer management
Computer systems	Resource management, security
Games	Board games, card games, video games
Robotics	Sim-to-real, control
Computer vision	Recognition, detection, perception
NLP	Sequence generation, translation, dialogue

improper and inapplicable to large-scale networks since it requires to explore and learn about the entire system in order to arrive at the optimum policy. The practical application of reinforcement learning is thus constrained.

Deep learning (Yang et al. 2020; Packer et al. 2019; Haarnoja et al. 2018) has been regarded as a game-changing technique recently. It has the ability to go beyond reinforcement learning's restrictions, ushering in a new era of reinforcement learning progress known as DRL. DNNs are used in DRL to train the learning process, allowing reinforcement learning algorithms to learn quicker and perform better. As a result, DRL has been used in a variety of reinforcement learning applications, including robotics, computer vision, speech recognition, and natural language processing (see Fig. 5.3).

The DRL techniques have the following advantages in general:

- DRL is capable of resolving difficult network optimization issues. As a result, without having extensive and precise network knowledge, network controllers in current networks, base stations, for example, are capable of resolving non-convex and difficult problems including joint user association, communication, and transmission scheduling
- DRL enables network entities to better understand and learn about the communication and networking environment. Without knowing the channel model or mobility pattern, network entities, such as a mobile user, DRL can be used to discover the optimum base station, channel, handover, caching, and offloading rules.
- Independent decision-making is possible with DRL. Network entities can use DRL approaches to observe and obtain the best policy locally.
- DRL dramatically accelerates learning, particularly in circumstances with large state and action spaces. In order to manage dynamic user association, spectrum access, and transmit power for a sizable number of mobile users and IOT devices, such as IoT systems with thousands of devices, DRL offers network controllers or IOT gateways.
- Other communication and networking challenges, Games, such as the non-cooperative game, can be used to represent cyber-physical threats, interference

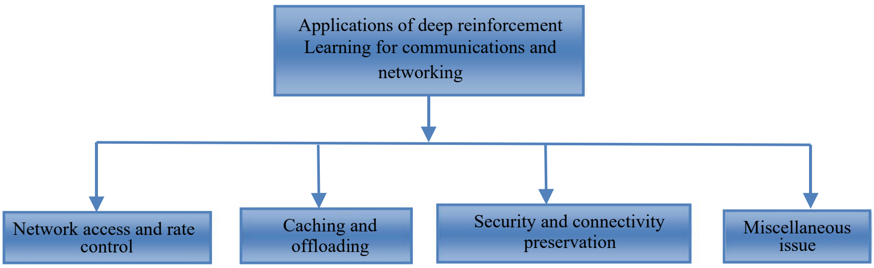


Fig. 5.3 Applications of deep reinforcement learning in communications and networking (Yang et al. 2020)

control, and data dumping. DRL has lately been used to solve games that need imperfect knowledge, such as Nash equilibrium determination.

- Wireless networks of the future are predicted to be large-scale, diverse, and decentralized (Wang et al. 2018; Xiong et al. 2019). When it comes to system parameters and states, The Q-function for the optimum actions to maximize system services and resources is calculated and updated by the agent of a network entity. The rewards earned by the agent can differ when the same behaviours are conducted with the same system states due to the uncertainty of mobile communication systems and the agent's short-sighted perspective. In this case, an appropriate updating rate, or learning rate, must be used to update the Q-function iteratively until convergence is attained. The process of updating the Q-function is referred to as "training."

5.3 DRL Applications for Future-Generation Mobile Networks

The majority of future-generation network optimization challenges have been solved using centralized optimization techniques. Such approaches are, on the other hand, based on fundamental hypotheses on the entire information requirements of network conditions. These presumptions are getting less and less useful as mobile networking scenarios become more unexpected. Unlike centralized optimization approaches, the DRL methodology makes no fundamental assumptions about the target system, making it a practical technique for dealing with a variety of difficulties in mobile 5G and beyond (François-Lavet et al. 2018; Fujimoto et al. 2018; Sharma et al. 2021; Schulman et al. 2017).

5.3.1 Power Management and Power Control

Because of the increasingly varied and dense environment in 5G networks, traditional intercell interference coordination will be inefficient, if not impossible. To reduce interference, a transmitter can lower its transmit power. The data rate, however, may deteriorate as a result. In mobile networks, interference reduction is frequently viewed as a power control optimization problem. Power management for network devices increases energy effectiveness while reducing expenses and carbon footprint. DRL enables network organizations to gain knowledge about their networks and make the best power control and management decisions possible.

5.3.1.1 Controlling Power in Cellular Networks

A DRL-based distributed dynamic power-allocation system was created by the authors of Qin et al. (2018) to take into consideration the unpredictable nature of 5G networks. Each transmitter performs the role of a system agent, and all agents are coordinated and carry out their tasks simultaneously. In order to predict the impact of current actions on future neighbour behaviour, each agent can study the consequences of its neighbours' prior activities throughout the current decision period before acting. With the aid of DRL, each agent chooses a policy that maximizes the discounted expected future benefit.

5.3.1.2 Ultra Dense Network Power Management

Ultradense networks (UDNs) consume an excessive amount of electricity due to the dense deployment of tiny BSs (SBSs). As a result, dynamically turning SBSs off is a cost-effective technique to save energy. The ON/OFF challenge in energy-harvesting UDNs has been transformed into a dynamic optimization problem. DRL is taught in order to learn a method for selecting SBS ON/OFF modes that maximize energy efficiency while accounting for the unpredictability of energy charging, channel condition information, and traffic arrivals. The ON/OFF scheduling method based on DRL consumes less energy than Q-learning.

5.3.1.3 In mm-wave Communications, Power Control is Important

Non-line-of-sight (NLOS) transmission will be a key challenge with the introduction of mm-wave communications in 5G. To increase NLOS transmission performance, the authors suggested a dynamic transmission power regulation technique. Within the restrictions of transmission power and QoS standards, In the 5G network, the overall data rate achieved by all UEs should be maximized. To forecast the Q-function, convolutional neural networks are first trained offline. The DRL is then utilized to conduct activities such as UE association and power allocation.

5.3.1.4 Computation Offloading at the Mobile Edge

A UDN with a single UE and a high number of BSs was examined. Depending on channel quality, the UE's computing responsibilities may be delegated to one of the BSs. The optimal offloading problem is tackled as a Markov decision process in order to lower the long-term anticipated cost (MDP). The aforementioned MDP challenge was addressed by developing a DRL-based online strategic computational offloading technique due to the dynamics of time-varying channel quality and the unexpected nature of task arrivals. The numerical outcomes demonstrate that the suggested strategy significantly lowers average cost. This system only has one server, as opposed to a single-server edge computing system that enables several UEs to offload computation to the server using wireless channels. The overall goal of the optimization is to reduce total delay costs and energy consumption across all UEs. The computation offloading problem is solved using a DRL-based technique to discover the best strategy and avoid the dimensionality curse.

5.3.1.5 Caching at the Edges

Edge caching can reduce transmission costs by lowering latency and easing traffic congestion on backhaul connections. Content caching, on the other hand, poses the challenge of policy control: in the face of continually changing content popularity distributions, the contents to be saved in the cache must be determined by the network edge device containing the cache. A single BS acts as a cache node for a large number of mobile users who make frequent requests for content from the BS. The authors employed DRL for the BS's cache replacement decisions to handle the requests efficiently. The objective, which is supported by evidence, is to raise the long-term cache hit rate.

5.3.1.6 Caching and Computing at the Edge

An integrated framework for coordinating computing, networking, and caching resource management in vehicular networks, in contrast to previous studies that looked at edge computing and caching concerns individually. The computational capabilities of the cache nodes' edge servers and statuses, such as the BS channel conditions, change dynamically. Furthermore, because the system state spaces are enormous, determining which resources should be allocated to which vehicle is difficult. DRL must therefore choose the best resource allocation strategy. In a similar vein, integrate computation, caching, and communication design to boost automobile networks' performance.

5.3.1.7 Transportation that is Intelligent

Vehicle networks may be able to give real-time intelligent services as 5G technology advances. Vehicular networks are anticipated to play a significant role in the development of intelligent transportation systems by enhancing the design of reliable and effective transportation systems by utilizing improved communications and data-collection methods. Based on DRL, a decentralized resource allocation system for vehicle networks. The system is used to figure out how to map incomplete observations from each car to the most efficient resource allocation. Without the necessity for precise prior knowledge, this approach can meet rigorous latency requirements for vehicle-to-vehicle links.

5.3.1.8 Slicing a Network

Network slicing is seen as the essential paradigm for satisfying varied service requirements, thanks to the greater softwarization provided in 5G networks. The necessity for slicing in mobile 5G networks is driven by the complexity of network resource management, considering the exponential increase of wireless data services, as well as various service requirements and heterogeneous wireless settings. Slicing divides, the network into a number of segments (slices), each of which is created and optimized to meet unique service demands. Right now, network slicing is a hot topic in both academia and business. An economic evaluation for distributing requests for network slices while taking into account the 5G infrastructure provider's maximum income.

5.4 Future Prediction of the Wireless Networks

Wireless networks of the future are predicted to be large-scale, diverse, and decentralized. When it comes to system parameters and states, The Q-function for the optimum actions to maximize system services and resources is calculated and updated by the agent of a network entity. The rewards earned by the agent can differ when the same behaviours are conducted with the same system states due to the uncertainty of mobile communication systems and the agent's short-sighted perspective. In this case, an appropriate updating rate, or learning rate, must be used to update the Q-function iteratively until convergence is attained. The process of updating the Q-function is referred to as "training".

For each state-action pair, Q-learning necessitates the agent keeping track of and updating a set of Q-values. Wireless networks, on the other hand, are likely to be large-scale in the future, diverse, and decentralized. As a result, the number of alternative system state values grows exponentially. In addition, due to the extreme diversity and unpredictability of system parts and environmental variables (Priya et al. 2021; Jagannath et al. 2019; Akhtar et al. 2021; Du et al. 2020; Liu and Wang 2017). There could be an infinite number of system states or even concealed system states. In this case, it becomes practically impossible to justify the expense of computing and storing all Q-values. This phenomenon is referred to as the "curse of dimensionality".

DRL combines RL and deep learning methods to overcome this issue. DRL models frequently have deep Q-learning, which makes use of a DQN.

As the data source, all input states are routed through several layers of neural networks. each with its own set of weight factors. Finally, the DQN generates a set of Q-values for each of the actions that can be taken as shown in Fig. 5.4.

The purpose of learning in the DQN is to use historical data such as Q-values, actions, and state changes to train and pick the most viable weight components. Because the underlying neural network is linear in complexity, a multilayer perceptron is used to calculate the Q-values and actions for a DQN. Furthermore, The

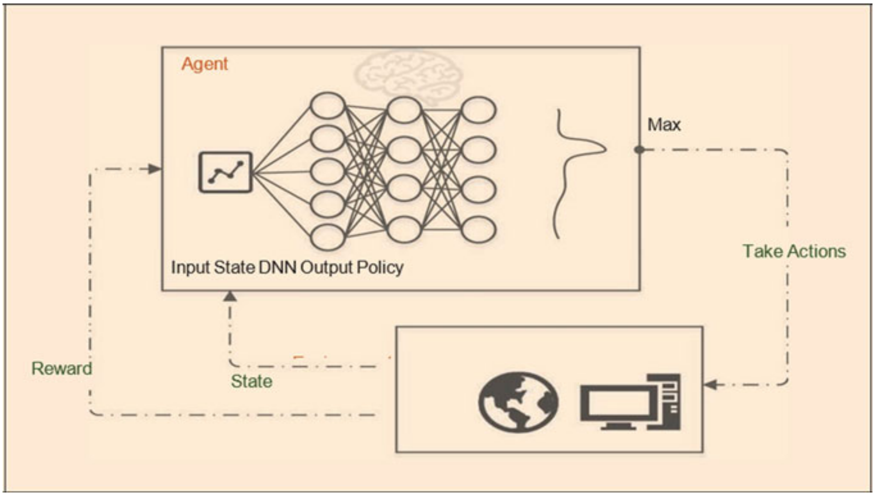


Fig. 5.4 This image depicts a DQN. The letters DNN and Max stand for deep neural network and maximum, respectively

DQN's number of inputs is solely defined by the type of each state. For each input, several state values, such as distinct channel states without, can be conveyed.

The fundamental distinctions between RL and DRL when the Q-function is created using a neural network are the following two features.

(1) The purpose of learning in the DQN is to use historical data such as Q-values, actions, and state changes to train and pick the most viable weight components. Because the underlying neural network is linear in complexity, a multilayer perceptron is used to calculate the Q-values and actions for a DQN.
(2) DRL reduces the Q-values of the target and the estimated Q-values by adjusting the weight every few time steps, resulting in a more stable DRL learning process.

DRL reduces model complexity greatly when compared to RL, especially when used to handle a variety of challenges in future communication systems. DRL also excels RL when it comes to extracting system properties and predicting optimal rewards and behaviours. RL, on the other hand, considers and transfers from states to mini batches. The deep network weight factors are then trained using the historical data collected by DRL. According to this, with the support of system and environment knowledge, which provides additional information, the training process can be hastened.

5.5 Wireless Mobile Communications and the Future of the Indian Cellular Market

Since its debut, wireless technology has advanced significantly. Over the past 10 years, there has been a remarkable increase in the number of wireless customers worldwide. Today, there is a large demand for wireless technology across a wide range of applications, including Internet and online browsing, video, and other text- and multimedia-based ones (Qiu et al. 2016). The usage of wireless technology is no longer solely restricted to wireless telephony. With more than 3 billion subscribers in over 215 countries, GSM is the most commonly used 2nd generation digital cellular standard. It gains about 1000 new users every minute. GSM, which provides wide area voice communications utilizing 200 kHz carriers and employs both TDMA and FDMA methods, was initially introduced in 1991 after being developed in the 1980s. Later, it evolved into 2.5 G standards with the introduction of packet data transmission technology (GPRS) and faster data speeds using higher order modulation methods (EDGE). For GSM/EDGE-based radio access network, they are now collectively referred to as GERAN (Wang et al. 2021). Mobile communications have quickly evolved into an essential symbol of our times since the development of the first genuinely worldwide mobile telecommunications system, GSM, in 1992. In order to protect the confidentiality of user-related information, some security mechanisms were incorporated into the original GSM. In addition to audio communication, GSM offers mobile phone communication services based on digital data exchange at up to 9.6 kbps (Ahmed et al. 2022). The majority of GSM standards were developed with operators in mind in order to minimize fraud and network abuse; operators were given the duty of integrating features relevant to user privacy (Rai et al. 2020). Even though they represent a significant evolutionary step, new modes like GPRS and EDGE will seamlessly integrate into the current GSM architecture and only slightly alter the network and base stations. In order to meet the criteria of UMTS, It's intriguing to consider how the evolutionary path can be expanded to even higher bit rates, enabling an even wider variety of services (Qin et al. 2018). There are issues with the current GSM:

(1) The end systems of GSM, a circuit-switched, connection-oriented technology, are set aside for the duration of each call. This results in inaccurate bandwidth and resource utilization.
(2) High data speeds are not supported by GSM-enabled systems (Ahmed et al. 2022).
(3) The fact that several users share the same bandwidth is GSM's biggest drawback. If there are many users, interference may occur during the broadcast. In order to get around these bandwidth restrictions, speedier technologies, such as CDMA, have been developed on networks other than GSM.
(4) According to Inc.Technology.com, another drawback of GSM is that it can interfere with some human electronics, such as pacemakers and hearing aids. Pulse-transmission technology used by GSM is what's causing this disturbance.

Therefore, it is required to turn off cell phones in many locations, such as hospitals and aircraft (Tanveer et al. 2022). Today's GSM users take for granted the ability to use their cell phones' voice roaming features to place and receive calls while travelling to other countries. The GSM mobile operators' packet-based General Packet Radio Service (GPRS) network provides a variety of services like video conferencing, multimedia apps, etc. However, this new network also makes it more difficult to accommodate the packet-based roaming scenario. Due to the projected volume of services, operators, and customers, it will be hard for operators to test all services for GPRS roaming across all partner networks. In order to ensure service continuity, consideration must be taken into account that the network capabilities of a visited GPRS network may differ from those of the home network. This takes us to a further consideration for mobile data service developers. This is especially important now that mobile operators are starting to provide 3G networks and support 3G to 2G network roaming agreements (Li and Li 2020).

5.5.1 The Growth Factor of the Telecom Sector in India

With 1.16 billion customers, India is currently the second-largest telecoms market in the world and has had impressive development in recent years. According to a report by the GSM Association (GSMA) and Boston Consulting Group, India's mobile economy is expanding quickly and will have a significant impact on the country's GDP (BCG). The US was surpassed by India in 2019 to become the second-largest market for app downloads.

Along with high consumer demand, the liberal and reformist policies of the Indian government have been essential to the sector's quick rise. The government has made it feasible for a fair and proactive regulatory structure to exist, as well as easy market access for telecom equipment, guaranteeing that customers may receive telecom services at fair prices. Due to the relaxation of Foreign Direct Investment (FDI) standards, the sector is among the five fastest growing and largest employment in the country.

5.5.2 Methodology Used in the Overall World

It is based on secondary information acquired from sources such as the Telecom Regulatory Authority of India, the Ministry of Communication, the Department of Telecommunication, government reports, and other sources. In order to assess the established targets, statistical tools such as annual growth rate, percentage, and year-by-year market share of different service providers were calculated.

5.5.3 Market Size Especially in India

In terms of overall internet users, India is the second-largest market in the world. 757.61 million people utilized the internet overall in January 2021. The number of wireless or mobile phone subscribers increased from 1,153.77 million in December 2020 to 1,163.41 million in January 2021. The second-largest telecom market in the world is found in India. As of January 2021, there were 1,183.49 million users nationwide.

In the third quarter of FY21, the telecom industry's gross revenue was Rs. 68,228 crores. Over the next five years, 500 million more Indians will have access to the internet because of rising mobile phone adoption and declining data costs, creating new business opportunities.

5.5.4 Growth Factor of Telecommunication in India

The second-largest telecoms market in the world is in India. The wireless, wireline, and internet services categories make up the telecom market. As of January 2021, there were 1,183.49 million users nationwide. Rural customers' teledensity increased from 59.05% in December 2020 to 59.50% in January 2021, indicating that there is still a significant amount of room for demand to rise in this sector. The number of wireless or mobile phone subscribers increased from 1,153.77 million in December 2020 to 1,163.41 million in January 2021.

In terms of internet users, India is the second-largest nation. One of the top data consumers worldwide is India. According to TRAI, each wireless data subscriber used an average of 11 GB of data each month in FY20. From 12.07 billion in 2017 to 19 billion in 2019, app downloads nationwide increased significantly, and it is predicted that this figure would reach 37.21 billion by 2022F. In the third quarter of FY21, India's overall wireless data usage increased 1.82% quarterly to 25,227 PB. Data consumption on mobile devices increased overall by 2.81% and 96.48%, respectively., in the third quarter of FY21 thanks to 3G and 4G data usage. the identical quarter, 2G data usage made up 0.71% of all data usage.

Gross revenue for the telecom sector in the third quarter of FY21 was Rs. 68,228 crores. The sector's development has been dependent on the government's significant policy support. The maximum amount of FDI allowed in the telecom sector has increased from 74 to 100%. FDI totaling US$37.62 billion was invested in the telecom sector between April 2000 and December 2020.

India is predicted to have the fastest-growing telecom advertising market between 2020 and 2023, with an average annual growth rate of 11%, according to a Zenith Media report. According to its National Digital Communications Policy, the Indian government anticipates investing $100 billion in the telecom sector by 2022. International manufacturers of telecom networks like Ericsson, Nokia, Samsung, and

Huawei are being enticed by the government to create all of their equipment in India using only locally available materials.

The TEPC organized India Telecom 2021 in March 2021 as a venue for the integration of technologies and business exchange (Telecom Equipment Export Promotion Council).

The Union Cabinet approved an Rs. 12,195 crores production-linked incentive (PLI) package for telecom and networking devices under the Department of Telecom. Petabytes and Forecast are abbreviations for each other.

5.5.5 *Major Market Players or Companies of Telecommunication in India*

Bharti Airtel Limited, a major international telecommunications firm with operations in 16 nations throughout Asia and Africa, is known as Airtel. Its main office is located in New Delhi, India. In terms of subscribers, the firm is one of the top three mobile service providers internationally. The company provides 2G, 3G, and 4G cellular services, mobile commerce, fixed line services, high-speed home internet, DTH, and both domestically and internationally long-distance services to carriers in India. In the remaining regions, it provides mobile shopping in addition to 2G, 3G, and 4G wireless services.

In the most recent spectrum auction held by the Department of Telecom in March 2021, Bharti Airtel paid a total of Rs. 18,699 crores for 355.45 MHz of spectrum in the Sub GHz, mid-band, and 2300 MHz bands (Government of India).

Bharti Airtel entered the advertising sector in February 2021 with the introduction of Airtel Advertisements, a productive brand interaction solution. Bharti Airtel and Qualcomm Technologies, Inc. hastened the rollout of 5G in India in February 2021. Airtel just made history in Hyderabad by testing 5G on a LIVE commercial network.

Reliance Jio's ecosystem as a whole allows Indians to completely live their digital lives. Strong broadband networks, useful applications, top-notch services, and smart devices are all components of this ecosystem, which is accessible to every Indian home. The most extensive collections of recorded and live music, sports, live and catch-up television, movies, and events are among Jio's media offerings. Jio wants to unleash the power of a young nation by developing linked intelligence for the 6 billion minds on the earth. Thanks to a triple strategy of concentrating on broadband networks, reasonably priced mobile devices, and the availability of high-quality content and applications, Jio has been successful in creating an integrated business model from the start.

Jio has the ability to provide a unique combination of media, fast bandwidth, digital commerce, telephony, and financial services. Jio Business unveiled an integrated service for micro, midsize, and medium-sized companies (MSMB) in March 2021 with the goal of providing integrated fibre connection and digital solutions. Reliance Jio announced the purchase of the spectrum in March 2021, earning the ability to utilize it across all 22 circles in India. More than 100 million feature phone customers in India were successfully upgraded to the Jio Phone network in February 2021, ushering in a new era of change for the country's feature phone consumers.

BSNL: BSNL is a technology-driven firm that offers a wide range of telecom services, including landline, mobile, and wireless local loop (WLL) telephone services, as well as broadband, leased circuits, and long-distance telecom service. With a switching network made entirely of digital technology, the firm has remained at the forefront of technology. All district, sub-divisional, tehsil, and nearly all block headquarters are covered by the nationwide telecommunications network of BSNL.

5.6 Summary

By 2020, it is anticipated that the telecom equipment industry would generate US$26.38 billion in revenue. By 2021, it's anticipated that there will be 829 million more internet members in the nation, and overall IP traffic will have increased fourfold at a CAGR of 30%.

India is predicted to have the fastest-growing telecom advertising industry between 2020 and 2023, with an average annual growth rate of 11%, according to a Zenith Media report.

IoT will be crucial in the development of the 100 smart city projects that the Indian government plans to implement. According to the National Digital Communications Policy of 2018, by 2022, the telecoms industry was expected to receive investments totaling $100 billion USD. There will be 18.11 billion app downloads in India in 2018 and 37.21 billion in 2022, according to forecasts.

References

Ahmed QW, Garg S, Rai A, Ramachandran M, Jhanjhi NZ, Masud M, Baz M (2022) AI-Based resource allocation techniques in wireless sensor internet of things networks in energy efficiency with data optimization. Electronics 11(13):2071

Akhtar T, Tselios C, Politis I (2021) Radio resource management: Approaches and implementations from 4G to 5G and beyond. Wirel. Netw. 27:693–734

Du Z, Deng Y, Guo W, Nallanathan A, Wu Q (2020) Green deep reinforcement learning for radio resource management: architecture, algorithm compression, and challenges. IEEE Veh Technol Mag 16:29–39

François-Lavet V et al (2018) An Introduction to deep reinforcement learning. Found Trends Mach Learn 11:219–354

Fujimoto S, Hoof HV, Meger D (2018) Addressing function approximation error in actor-critic methods. ArXiv, abs/1802.09477

Gronauer S, Diepold K (2022) Multi-agent deep reinforcement learning: a survey. Artif Intell Rev 55:895–943

Haarnoja T, Zhou A, Abbeel P, Levine S (2018) Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor. ICML

Hessel M, Modayil J, van Hasselt H, Schaul T, Ostrovski G, Dabney W, Horgan D, Piot B, Azar M, Silver D (2018) Rainbow: Combining improvements in deep reinforcement learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, 32(1). https://doi.org/10.1609/aaai.v32i1.11796

Jagannath J, Polosky N, Jagannath A, Restuccia F, Melodia T (2019) Machine learning for wireless communications in the Internet of Things: A comprehensive survey. Ad Hoc Netw 93:101913

Lazaridis, Aristotelis et al. (2020): Deep reinforcement learning: A state-of-the-art walkthrough. J. Artif. Intell. Res. 1421–1471

Li Y (2019). Deep reinforcement learning. ArXiv, abs/1810.06339

Li M, Li H (2020) Application of deep neural network and deep reinforcement learning in wireless communication. PLoS ONE 15(7):e0235447. https://doi.org/10.1371/journal.pone.0235447

Lillicrap TP, Hunt JJ, Pritzel A, Heess NM, Erez T, Tassa Y, Silver D, Wierstra D (2016) Continuous control with deep reinforcement learning. CoRR, abs/1509.02971

Liu Y, Wang DL (2017) Speaker-dependent multipitch tracking using deep neural networks. J Acoust Soc Am 141(2):710–721. https://doi.org/10.1121/1.4973687. PMID: 28253703

Marc G. Bellemare, Will Dabney, Rémi Munos, (2017). A distributional perspective on reinforcement Learning. In: Proceedings of the 34th International Conference on Machine Learning, PMLR 70:449–458

Mnih V, Badia AP, Mirza M, Graves A, Lillicrap TP, Harley T, Silver D, Kavukcuoglu K (2016) Asynchronous methods for deep reinforcement learning. ICML

Nair A, Srinivasan P, Blackwell S, Alcicek C, Fearon R, Maria AD, Panneershelvam V, Suleyman M, Beattie C, Petersen S, Legg S, Mnih V, Kavukcuoglu K, Silver D (2015) Massively parallel methods for deep reinforcement learning. ArXiv, abs/1507.04296

Nguyen Cong Luong Et Al (2019.), Applications Of deep reinforcement learning in communications and networking: A survey, IEEE communications surveys & tutorials, 21(4), Fourth Quarter https://doi.org/10.1109/COMST.2019.2916583

Packer, Charles; Gao, Katelyn; Kos, Jernej; Krähenbühl, Philipp; Koltun, Vladlen; Song, Dawn (2019) Assessing generalization in deep reinforcement learning 2019. In: 33rd Conference on Neural Information Processing Systems

Priya A, Rai A, Singh RP (2021) Internet of things: architecture, applications and future aspects. In Advances in Smart Communication and Imaging Systems (pp 183–190). Springer, Singapore

Qin M, Yang Q, Cheng N, Zhou H, Rao RR, Shen X (2018) Machine learning aided context-aware self-healing management for ultra-dense networks with QoS provisions. IEEE Trans Veh Technol 67:12339–12351

Qiu J, Wu Q, Ding G et al (2016) A survey of machine learning for big data processing. EURASIP J Adv Signal Process 2016:67. https://doi.org/10.1186/s13634-016-0355-xA

Rai A, Sehgal A, Singal TL, Agrawal R (2020). Spectrum sensing and allocation schemes for cognitive radio. Mach Learn Cogn Comput Mob Commun Wirel Netw 91–129

Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O (2017) Proximal policy optimization algorithms. ArXiv, abs/1707.06347

Sharma D, Singhal S, Rai A, Singh A (2021) Analysis of power consumption in standalone 5G network and enhancement in energy efficiency using a novel routing protocol. Sustain Energy, Grids Netw 26:100427

Tanveer J, Haider A, Ali R, Kim A (2022) Machine learning for physical layer in 5G and beyond wireless networks: a survey. Electronics 11:121. https://doi.org/10.3390/electronics11010121

Van Hasselt H, Guez A, Silver D (2016). Deep reinforcement learning with double Q-Learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, 30(1). https://doi.org/10.1609/aaai.v30i1.10295

Wang S, Liu H, Gomes PH, Krishnamachari B (2018) Deep reinforcement learning for dynamic multichannel access in wireless networks. IEEE Trans Cogn Commun Netw 4(2):257–265

Wang Z, Schaul T, Hessel M, can Hasselt H, Lanctot M, de Freitas N Dueling (2016). Network architectures for deep reinforcement learning. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA

Wang S, Chen M, Liu X, Yin C, Cui S, Vincent Poor H (2021). A machine learning approach for task and resource allocation in mobile-edge computing-based networks. IEEE Internet Things J 8(3), 1358–1372. [9146372]. https://doi.org/10.1109/JIOT.2020.3011286

Xiang X, Foo S(2021). Recent advances in deep reinforcement learning applications for solving partially observable markov decision processes (POMDP) Problems: Part 1—Fundamentals and applications in games, robotics and natural language processing. Mach Learn Knowl Extr, 3, 554–581. https://doi.org/10.3390/make3030029

Xiong Z, Zhang Y, Niyato D, Deng R, Wang P, Wang L-C (2019) Deep reinforcement learning for mobile 5G and beyond: Fundamentals, applications, and challenges. IEEE Veh Technol Mag 14(2):44–52. https://doi.org/10.1109/MVT.2019.2903655

Yang K, Shen C, Liu T, (2020) Deep Reinforcement Learning based Wireless Network Optimization: A Comparative Study. In: IEEE INFOCOM 2020—IEEE Conference on Computer Communications Workshops, pp 1248–1253, doi: https://doi.org/10.1109/infocomwkshps50562.2020.9162925

Zhao X, Yang R, Zhang Y, Yan M, Yue L (2022) Deep reinforcement learning for intelligent Dual-UAV reconnaissance mission planning. Electronics 11:2031. https://doi.org/10.3390/electronics11113203

Zikria YB, Afzal MK, Kim SW, Marin A, Guizani M (2020) Deep learning for intelligent IoT: Opportunities, challenges and solutions. Comput Commun 164:50–53

Chapter 6
Detection of Consumption of Alcohol Using Artificial Intelligence

Anubhav Thapliyal, Arunendra P. Singh, and Rashmi Priyadarshini

Abstract The habit of overdrinking alcohol is a major problem for the behavior of people in public places the behaviour of such people is annoying, and hence opens the door for severe security issues, especially for the surroundings. In this paper, we are proposing an accurate method to detect a drunken person without informing the person. First, the intelligent camera will detect the person by providing an alert signal, and in the second stage, the face recognition system will not allow him/her to use metros, airplanes or public transport. The system is 87% accurate and tested on over 50 people.

Keywords Artificial intelligence · Face recognition system · LWIR thermal spectrum · Alcohol detection · Public transport · Breathalyzer

6.1 Introduction

Approximately 76,446 people died between 2008 and 2017, because of road accidents nationwide, due to consumption of alcohol. According to the World Health Organization's Global Status Report based On Road Safety 2018, India was advised in terms of enforcement of drunk driving laws.

Metro has become a lifeline of the country. It makes life easy in terms of less travel time, low fares, connectivity, safety for women, cleanliness, etc., but in the metro system, there is no law prohibiting the boarding of a drunk person. We all are aware of the fact that traveling with a person who is under the influence of alcohol is annoying and insecure.

In the year 2013, Delhi Metro authorities proposed in a paper that passengers who are under the influence of alcohol when subjected to breathanalyzer tests should not be allowed to travel in the metro. The DMRC's stated that the ban on drunken

A. Thapliyal · A. P. Singh · R. Priyadarshini (✉)
EECE Department, Sharda University, Plot No. 32, 34, Knowledge Park III, Greater Noida, Uttar Pradesh 201310, India
e-mail: rashmi.priyadarshini@sharda.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

A. Rai et al. (eds.), *Paradigms of Smart and Intelligent Communication, 5G and Beyond*, Transactions on Computer Systems and Networks,
https://doi.org/10.1007/978-981-99-0109-8_6

passengers followed by a breathanalyzer test would prevent fights or quarrels in the metro and make the metro a safe travel option, but it is a manual process that is not implemented perfectly.

As India is currently actively taking steps toward drinking and driving, there are still no concrete or defined steps if a drunk person boards the metro. There are no restrictions toward this issue.

In this chapter, we will be proposing ways to overcome the issues mentioned above in a realistic sense (Bhowmik, 2011).

6.2 Ways to Detect Consumption of Alcohol

There are certain ways through which we can detect alcohol levels in the human body.

6.2.1 Breathalyzer

Breathalyzer is a device for estimating Blood Alcohol Content (BAC).

6.2.1.1 Chemistry

When the user exhales into a breathanalyzer, any ethanol present in his/her breath is oxidized to acetic acid at the anode:

$$CH_3CH_2OH(g) + H_2O(l) \rightarrow CH_3CO_2H(l) + 4H + (aq) + 4e-$$

At the cathode, atmospheric oxygen is reduced:

$$O_2(g) + 4H + (aq) + 4e- \rightarrow 2H_2O(l)$$

The overall reaction is the oxidation of ethanol to acetic acid and water:

$$CH_3CH_2OH(l) + O_2(g) \rightarrow CH_3COOH(aq) + H_2O(l)$$

An electric current is produced by the reactions, which are measured by the microcontroller. The measured current is calibrated as an indicator of the level of overall blood alcohol content (BAC) by an Alco-Sensor.

6.2.1.2 Types of Breathalyzers

There are various types of breathalyzers based on sensors: Infrared sensor, Fuel Cell gas sensor, or Semiconductor sensor.

- **Infrared sensor**

Infrared breathalyzers are more specific for the detection of ethanol. Typically, breath alcohol instruments used by police officers work on the principle of infrared spectroscopy.

- **Fuel cell sensor**

These types of sensors are based on the oxidation of ethanol to acetaldehyde on an electrode. The amount of alcohol present is directly proportional to the current produced. These types of sensors are used by traffic police because of their high stability and are calibrated after every 6 months.

- **Semiconductor sensor**

These sensors work on the principle that the conductance of a tin oxide layer increases in the presence of gas such as vaporized ethanol. The stability of these sensors is not very reliable.

6.2.1.3 Error in Measurement

Silicon oxide sensors are sometimes affected due to contamination and interference from substances other than breath alcohol. Replacement or recalibration of these substances takes place after every 6 months. Platinum fuel cell sensors are used in expensive breath alcohol testers. These sensors also require recalibration but after a long interval as compared to semiconductor devices, usually once a year.

Breathalyzer Accuracy Problems

The breathalyzer results are full of uncertainty. This uncertainty, which is also known as an error rate, ranges from 10 to 20%.

6.2.2 Identification Through Infrared Face Images

Thermoregulation is a process in which the biological organism changes the internal temperature within certain limits and is controlled by the hypothalamus. The temperature for humans in a stable state for skin is 33.5 to 37.5 °C. The process of thermoregulation can be altered through alcohol and also can generate an infused vasodilation in the skin, which can increase the loss of heat through the process of convection, resulting in a decrease in body temperature which can be related directly to the amount of alcohol consumed.

Alcohol can affect the psychic system as well as cause motor disturbances which can result in the dilation of blood vessels (Kalant and Le 1983; Buddharaju et al. 2007, 2008, 2006; Buddharaju and Pavlidis 2007), and an increase in blood pressure which can cause abnormal behavior on a biological level. In a human face, consumption of alcohol results in an increase of temperature in capillary density, in areas around the nose, forehead, and eyes. An increase in the temperature of the eyes can be useful for classification. A method is used to identify regions with high thermal variations in the face by comparing the intensity of two groups of people; one being normal, while the other is drunk. Thermal imaging is made used to gain patterns based on data received or observed during the thermoregulation process of the alcohol consumption amount by the person's thermal variation on the human face.

Experiment Data on experiment on Drunk Thermal Face database (PUCV-DTF)

Thermal Face Database: This section provides details about the achieving of Pontificia Universidad Católica de Valparaíso-Drunk Thermal Face database (PUCV-DTF).

(a) Recruitment: In the School of Electrical Engineering of the Pontificia Universidad Católica de Valparaíso a volunteer call was made to the people. People who answered the call were made aware of the research and were asked to sign a further approved consent form from the Committee of Ethics of the university.
(b) Participants: Forty-Six (46) people were selected; consisting of 40 men and 6 women. The individuals ages were ranging from 18 to 29 (17 < age of individuals < 30, Avg. age 24 years with a SD of approx. 3 years). A screening test was designed to exclude people who consume alcohol on a daily basis and to search for people who were in good health without any issues related to the consumption of alcohol
(c) Procedure: Subjects were asked to come into the robotics lab where they rested for half an hour in order to stabilize their metabolism to room temperature in the laboratory. Then they were asked to consume 355 ml of beer in a can having a temperature of 5.5°; followed by another can after 30 min. This process was repeated till four cans were consumed by the subject. Refer to Table 6.1 for details on the results of the procedure. On completion, subjects were tested with a breathalyzer test to determine their blood alcohol level. Subjects with BAC 0.8 g/L were asked to stay in a place until the BAC level was under 0.2 g/L which was again measured by the breath test.

Table 6.1 Data captured after the experiment

Class	Minimum (g/L)	Maximum (g/L)	Mean (g/L)
Sober	0	0	0
1 Beer	0	0.39	0.17
2 Beers	0.15	0.68	0.37
3 Beers	0.29	1.29	0.59
4 Beers	0.43	1.68	078

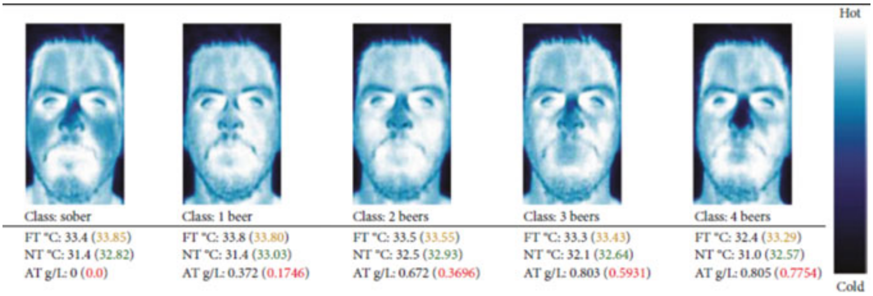


Fig. 6.1 Data of five subsets

(d) Thermal Imaging: FLIR TAU 2 (Socolinsky and Selinger 2002) was the thermal camera used which had a resolution of 640 × 480 pixels, a frame rate of 30 frames per second, thermal sensitivity of 50 mK, and a spectrum range between 7.5 and 13.5 µm. There was a total of 46 individuals (40 men and 6 women) with five subsets, so a total of 300 images were recorded including 250 images per subject and 50 per subset. Each set was classified into five subsets: "Normal—no beer consumption," "consumption of 1 beer can," "consumption of 2 beer cans," "consumption of 3 beer cans," and "consumption of 4 beer cans," corresponding to the capture of data procedure.

A reference to the end result is shown in Fig. 6.1. To highlight the patterns or information gained from the alcohol intake, the addition of color to the images captured was carried out. The figure also displays the nose temperature (NT), forehead temperature (FT), and alcohol tester (AT) readings. The average values of NT, AT, and FT are shown in parentheses. Please keep in mind that the thermal temperature of the image of the face varies due to the consumption of alcohol. However, a variation in nasal contrast can be observed after the beer intake by the subject. This variation can be the result of different actions such as the process of thermoregulation or due to the effect of breathing during the inhalation or exhalation phase.

The results achieved in this study show that the method to identify drunk people through thermal imaging of the face has a success rate of approximately 87%.

6.3 Methodology

6.3.1 Using IR Sensor Thermal Imaging Cameras

The idea is to install infrared cameras on the metro station premises, to be specific, cameras would be installed alongside metal detector gates. Using the data mentioned in the above section, we will be able to identify drunk people by observing thermal

Fig. 6.2 Thermal spectrum for drunk signs recognition purpose

images of the people. The person showing major signs of drunkenness would be spotted and stopped at the entrance itself.

Best Thermal Spectrum for Drunk signs recognition purpose

The value of emission of the spectral distribution of energy by an object is given by the product of the emissivity of the object as the wavelength function with the Planck distribution for a given temperature (Siegal and Howell 1981). The Planck distribution is at its highest in LWIR (9 μm) and around 1/6th of this maximum in MWIR during the human body temperature (37 °C) (Koukiou, 2011,2012, 2013, 2013a, 2015). The value of emissivity of human skin during the process in the LWIR is at least 0.97, and at least 0.91 in the MWIR. As a result, LWIR is favored for face recognition in the thermal infrared. Compared to MWIR, the emission of LWIR is much higher (Fig. 6.2).

Based on the data mentioned in Table 6.1, we will identify a person showing high liquor intake using LWIR thermal Spectrum.

After recognizing a said person, our next step would be to ban him/her from boarding public transport, which would prevent further disturbances to the passengers and ensure safe and sound travel.

6.3.2 Using Breathalyzers

Breathalyzers use a variety of technologies. While the majority of breathalyzers use a single technology, some analyzers use two technologies to obtain greater precision and accuracy (Table 6.2).

For the purpose of quantitative evaluation of the contents of alcohol present in the inhaled air, most of the breathalyzers are gleaned on the base principle of infrared spectroscopy. The initial models of Intoxilyzers which were created by Omicron in Palo Alto, California were 4011 A and 4011 S. The intoxilyzer was then sold to CMI, Inc. in Owensboro, Kentucky by Omicron (Farias G, 2018). The early versions were followed by the Intoxilyzer 5000, which was recently installed as a convincing BAC analyzer, and the Intoxilyzer 8000. Another example of a breathalyzer that uses IR spectroscopy technology and can detect alcohol contents by using two wavelengths

Table 6.2 Technology used in different breathalyzers

Technology	Examples of breathalyzers
Analyzers based on color metric detection due chemical reaction of reagents with alcohol	Breathalyzer 900 breathalyzer 1100
Infrared based technology	Intoxilyzer 5000, Intoxilyzeri8000, Data Master cdm
Fuel cell based technology	Alcotest models 6510, 6810, 7410, etc
	Alcosensor III and IV
Mixed technology (infrared and fuel cell)	Intox EC/IR

which are 3.37 and 3.44 m is the Data Master cdm (National Patent Analytical System, Mansfield, OH).

There are numerous different articles on breathalyzers that make use of the principle and technology of fuel cells. Some of them are Alcotest Models 6510, 6810, and 7410, Alco-Sensor III and IV (National Drager, Durango, Colorado). A porous disk covered on both sides with platinum oxide, also known as platinum black, makes up a fuel cell. Hydrogen ions being charged particles can move through the porous layer due to their saturation with acidic solutions which carry numerous electrolytes. In addition, a platinum wire is used to connect both sides of the platinum oxide-containing disk. If the exhaled air contains any alcohol content or traces, they are then converted into acetic acid, hydrogen ions, and electronsionitheiupper surface. The H ions (Hydrogen) formed are then processed to a lower surface where they integrate with oxygen from the air to form water. Now, during the process, an electric current is generated which is then passed to the microprocessor of the instrument which translates it to equivalent breath alcohol and computed blood alcohol.

In order to achieve higher sensitivity and specificity, some breathalyzers use the integrated technology of fuel cells and IR spectroscopy. Intoximeters, Inc. also makes the Intox EC/IR desktop alcohol breathalyzers, which merge reliable fuel cell analysis with infrared technology. Semiconductors are also used to make breathalyzers of low-cost which are sold to the public. It is noted that the response of the sensor can be non-specific to alcohol. For instance, semiconductor sensors will also react to particles in cigarette smoke.

A breath alcohol test is a process of inhalation and exhalation (only once) maneuver in which the subject inhales air and then exhales air in the breathalyzer equipment to the preferred volume of residue. The assumption is that the amount of alcohol in exhaled breath is the same as in alveolar air.

End-exhaled air has lower ethanol levels than alveolar air in general. A breath alcohol test requires the participant to inhale ambient air and exhale into the breathalyzer (typically 1.1–1.5 L of exhaled air is required for the test). For the accurate analysis, a smaller drinken person with a smaller lung capacity must exhale a larger fraction of air in their lungs to meet the analyzer's minimum volume requirement, and due to this reason sometimes breath test may get the level of blood alcohol more than the actual one in the case of smaller subject in contrast to a subject with bigger lung capacity.

6.4 Summary

A breath alcohol analyzer or breathalyzer, as the name implies, will assist us in detecting an intoxicated person. It will provide us with some specific information from which we can determine whether the suspected person will be allowed to ride the metro or not, as well as send a notification to the local police and their family members about their current location. If a person has had a reasonable amount of alcohol, he will be permitted to travel by metro, but if he is inebriated and unable to control himself, he will be prohibited from doing so.

Installation of Breathalyzers

Because of the presence of security guards, we can put a breathalyzer or infrared breath alcohol analyzer near the Baggage Handling Supervisor, and the suspected person will be not allowed to enter if he/she is discovered to be under the influence of alcohol.

Next step after the detection

We shall imprison the person as soon as we discover he is heavily influenced by alcohol. He will not be allowed to travel, and his last known whereabouts will be sent to the local police and family members.

References

Bhowmik MK, Saha K, Majumder S, Majumder G, Saha A, Sarma AN, Bhattacharjee D, Basu DK, Nasipuri M (2011) Thermal infrared face recognition: a biometric identification technique for robust security

Buddharaju P, Pavlidis I (2007) Multi-spectral face recognition—fusion of visual imagery with physiological information. In: Hammoud RI, Abidi BR, Abidi MA (eds) Face biometrics for personal identification. signals and communication technology. Springer, Berlin, Heidelberg, pp 91–108

Buddharaju P, Pavlidis I, Kakadiaris I (2006) Pose-invariant physiological face recognition in the thermal infrared spectrum. In: 2006 conference on computer vision and patternrecognition workshop (CVPRW'06), New York, USA, pp 53–60

Buddharaju P, Pavlidis IT, Tsiamyrtzis P, Bazakos M (2007) Physiology-based face recognition in the thermal infrared spectrum. IEEE Trans Pattern Anal Mach Intell 29(4):613–626

Buddharaju P, Pavlidis I, Manohar C (2008) Face recognitionbeyond the visible spectrum. In: Advances in biometrics. Springer, London, pp 157–180

Falk B, Burstein R, Rosenblum J, Shapiro Y, Zylber-Katz E, Bashan N (1990) Effects of caffeine ingestion on body fluid balance and thermoregulation during exercise. Can J Physiol Pharmacol 68(7):889–892

Farias G (2018) Face recognition and drunk classification using infrared face images

Foster I, Kesselman C (1999) The grid: blueprint for a new computing infrastructure. Morgan Kaufmann, San Francisco

https://www.sciencedirect.com/topics/nursing-and-health-professions/breath-alcohol-analyzer

Kalant H, Le A (1983) Effects of ethanol on thermoregulation. Pharmacol Ther 23(3):313–364

Koukiou G, Anastassopoulos V (2011) Facial blood vessels activity in drunk persons using thermal infrared. In: 4th international conference on imaging for crime detection and prevention 2011 (ICDP 2011). London, UK, pp 1–5

Koukiou G, Anastassopoulos V (2012) Drunk person identificationusing thermal infrared images. Int J Electron Secur Digit Forensics 4(4):229–243

Koukiou G, Anasassopoulos V (2013) Face locations suitable drunk persons identification. In: 2013 international workshop on biometrics and forensics (IWBF). Lisbon, Portugal, pp 1–4

Koukiou G, Anastassopoulos V (2013a) Eye temperature distribution in drunk persons using thermal imagery. In: 2013 international conference of the BIOSIG special interest group (BIOSIG). Darmstadt, Germany, pp 233–240

Koukiou G, Anastassopoulos V (2015) Neural networks for identifying drunk persons using thermal infrared imagery. Forensic Sci Int 252:69–76

Socolinsky DA, Selinger A (2002) A comparative analysis of face recognition performance with visible and thermal infrared imagery. In: Proceedings of the 16th international conference pattern recognition, vol 4, pp 217–222

Zhao S, Grigat RR (2005) An automatic face recognition system in the near infrared spectrum. In: Proceedings of the 4th international conference on machine learning and data mining in pattern recognition, pp 437–444

Chapter 7
Application of Machine Learning in Finger Vein Pattern Recognition

Rahul Dev, Krishanu Kundu, Astha Sharma, and Shiv Narain Gupta

Abstract As globalization and the standard of living have expanded, so has the use and development of consumer electronics. Convenient, easy, high-security authentication technologies are necessary to protect personal information held on mobile devices. Because of an increased focus on security and higher level of complexity with respect to currently available biometric systems, automatic individual identification employing finger vein biometrics is gaining popularity. In current work the focus is on implementing image processing in the field of biometrics for vein pattern recognition.

Keywords Finger vein pattern recognition · Machine learning · 5G · Image processing

7.1 Introduction

In the case of biometric framework, forged fingerprints as well as palm prints are very common; voice, signature as well as shapes of hands are easy to forge; similarly, biometrics, like fingerprints, recognition of iris, and face, are vulnerable to forging. So, we can say biometric identifiers are easy to duplicate and to be used for creating artifacts that may betray several biometric gadgets. The main goal is to enhance

R. Dev · K. Kundu (✉) · A. Sharma
Department of Electronics and Communication Engineering, G L Bajaj Institute of Technology and Management, Greater Noida, Uttar Pradesh 201306, India
e-mail: krishanukundu08@gmail.com

R. Dev
e-mail: rahul.dev@glbitm.ac.in

A. Sharma
Program Manager, IIITB COMET Foundation, Bangalore, India

S. N. Gupta
Department of Electronics and Communication Engineering, Greater Noida Institute of Technology, Greater Noida, Uttar Pradesh 201306, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

A. Rai et al. (eds.), *Paradigms of Smart and Intelligent Communication, 5G and Beyond*, Transactions on Computer Systems and Networks,
https://doi.org/10.1007/978-981-99-0109-8_7

recognition accuracy along with its efficiency while also becoming highly resistant to spoofing assaults. To this end, a number of experts have developed highly customized biometrics in order to improve unwavering quality while also repelling spoofs. The aim behind the current study is to offer a system that can increase the precision with which patterns of finger veins along with their characteristics can be predicted from given photos, using a publicly available dataset called "SDUMLA-HMT" (Yin et al. 2011). Image processing with MATLAB R2014a and various algorithms are included in the proposed work.

7.1.1 Literature Survey

The field of study into finger vein pattern recognition and picture enhancement is large and well-known. In the recent past, several researchers have used a variety of image processing approaches and algorithms.

Personal biometric identification using finger vein patterns by utilizing finger vein image correlation was presented (Kono et al. 2002). In this study a background-reduction filter was employed to improve vein patterns of human fingers. Specifically, a LPF was utilized to diminish background noise. The drawback was whenever background disturbances were diminished; few important background data were also filtered out automatically. Built on concepts of repeated line tracking method (Miura et al. 2004), upgraded the pursuance of finger vein detection. The highest curvature points in image profiles were utilized to extract vein pattern features, and the approach used is outlined in Miura et al. (2005). He followed a measured methodology that was computationally serious in these two systems.

Alternative approach for managing vein images via absolute angles along with distance was proposed (Wang et al. 2010). By calculating the arm's numbers beginning from a pixel, the main region was taken out from the diminished vein image. At that time, straight lines were used to connect these focuses in order to form a cross-sectional topology. For the given instance, the angle existing in between the converging points and their relative distance was determined. However, calculating relative spacing along its angle from such cross-sectional topology is a tough task.

In Rajan and Indu (2014), the proposed technique for image extraction, which used a grouping of Frangi filter, features from accelerated segment test algorithm, and fast retina key point descriptors to extract finger vein features from enhanced vein images. A biometric personal identification system based on patterns formed by finger veins (Vega et al. 2014). To obtain the vein pattern, the researchers used a Sobel detector, an enhancement filter, and a binarization method. Obtained results yielded a 27.56% equal error rate, a Genuine Acceptance Rate of 100%, and a False Acceptance Rate of 0%.

In Jain et al. (2004b), they proposed a biometric system to recognize individuals automatically founded on physiological and/or behavioral traits. It is evident from this research that any system that ensures reliable person recognition must contain a biometric component. This is not to say, however, that biometrics alone can provide a good individual recognition component. The gradient feature detector algorithm was proposed (Prathiban et al. 2014). The study of finger vein pattern recognition and picture enhancement is large and well-known. The suggested algorithm in this work extracts vein images by taking into account several characteristics like width, length of a finger vein, its junction, position as well as pixels. The proposed approach is appropriate for mobile phones as well as ATMs.

In Mulyono and Jinn (2008) employed source images acquired by a web camera, which cannot equate to a CCD camera, and attempted to improve the inferior image quality for personal identification. As a result, they attain a FAR of 0%, a FRR of 0.275%, an identification rate of 100%, and a reaction time of 0.5 s, demonstrating that biometrics are very successful for personal identification and meeting customer requirements in terms of response time.

The vein patterns of the complete hand, such as finger veins, dorsal veins, and palm veins, are fused in this work to present a novel personal recognition method (Xiangqian et al. 2010). With the use of SVM as a classifier, various vein patterns of the complete hand are recorded and demonstrated to be better than any single feature. The obtained error rate is 0.0223%.

Biometric access, control as well as management of home networks employing digital media is proposed (Corcoran et al. 2007). The authors outlined various biometric identification systems, including face recognition, fingerprint recognition, and print recognition. In this study, they explored the differences between key generation and authentication, as well as why they are currently too unreliable and computationally costly for usage in CE applications.

This study Liu et al. (2012), provides a real-time biometric recognition structure for mobile devices based on blanket dimensions as well as lacunarity implementation on a DSP platform using finger veins. After obtaining the finger vein images, ROI segmentation is performed, and then blanket dimension features along with lacunarity features are used for recognition. The EER of this approach is 0.07%, which is significantly lower than the present system. Because it uses less power and has a lower computational complexity, this system is well suited for mobile devices.

7.1.2 Problem Formulation

The more difficult part of improving finger vein detection is extracting patterns from low-contrast NIR (Near Infrared Image) images. It's still difficult to improve vein patterns and extract their features from low-contrast images.

The basic idea of this research work is to process (enhance) the finger vein images, extract their features, and find out the accuracy by comparing different algorithms and techniques. The tools and algorithms of MATLAB R2014a are used. Image enhancement is done by the use of a Local Histogram Algorithm on the pre-processed image of the database (NIR images). After this feature extraction process is done on the enhanced images with the help of a combination of Frangi filter, FAST algorithm, and FREAK discriminator.

At the end the research comes to the important phase, i.e., verification and accuracy calculation. This is being done with the help of Machine Learning's (Supervised Learning algorithm) Discriminant analysis. The accuracy is mentioned and verified after comparing with various techniques of the Supervised Learning Algorithm. After comparison with other techniques, it has been observed that Discriminant Analysis gives more accurate results than others.

7.2 Methodology

In this investigation, the techniques for image processing and feature extraction of finger vein patterns are the same as those proposed in Rajan and Indu (2014). However, a new thing is being added or precisely we can say the work has been expanded in this study by estimating the accuracy by meaningful utilization of the Discriminant Analysis algorithm and KNN Algorithm. The best techniques among those described above have been demonstrated at the end of the project by comparing the accuracy computed by these.

Certain technical procedures are followed in the proposed work, and they are listed as follows:

(A) Processing of Image

1. Pre-Processing stage
2. Binarization of the acquired image
3. Detection of edges by employing Sobel filter
4. Subdivision Of ROI
5. Enhancement of Images
6. Ultimate Image Upgradation by employing Local Histogram Algorithm.

(B) Feature Withdrawal using a Frangi filter, FAST algorithm, and FREAK descriptors.

(C) Estimation of the accuracy with respect to characteristics derived from the improved image of vein patterns. The accuracy computation is carried out using the Discriminant Analysis algorithm from Machine Learning (Supervised Learning approach) and the K-Nearest Neighbor Algorithm.

(D) Above-mentioned two Machine Learning tactics are compared to calculate accuracy.

The workflow of the proposed technique taken for the image-processing section is presented in Fig. 7.1.

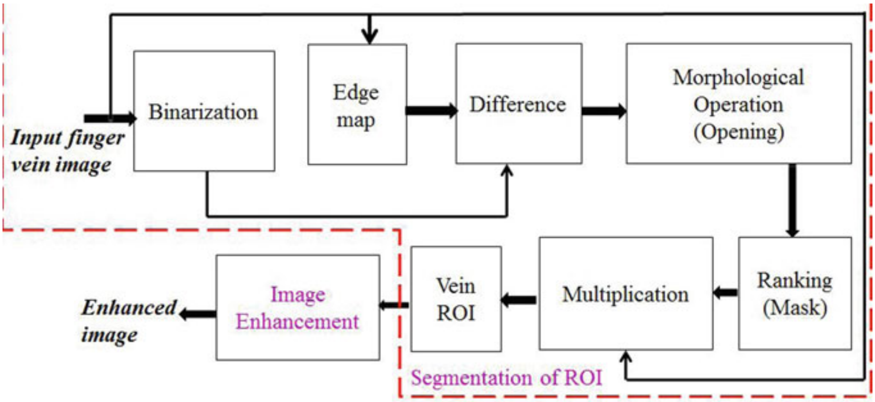


Fig. 7.1 Proposed method flow chart (image processing) (Rajan and Indu 2014)

7.2.1 Feature Withdrawal Techniques

The conceptual block diagram for obtaining vein characteristics is depicted in Fig. 7.2. The feature map is extracted using the Frangi filter, and the point of interest is determined using the FAST algorithm. Every point is centered around a FREAK descriptor (Rajan and Indu 2014).

The veins in the finger vein pattern are mostly aligned along the X coordinate and the respective diameter does not alter considerably. As a result, it is possible to consider fixing the diameter, which simplifies the problem and eliminates the necessity for multiscale analysis (Vlachos and Dermatas 2013).

7.2.1.1 Frangi Filter

Key image processing scheme is built on the evaluation of eigenvalues with respect to Hessian matrices and was first established for the detection of finger blood vessel in medical images. Alternatively, it can also be used when the line-like structure of a picture needs to be determined. The eigenvalues are employed to govern the likelihood of blood vessels existing in the region.

The Hessian eigenvalue analysis approach may distinguish bob-like, plate-like as well as tubular structure in a photograph in general.

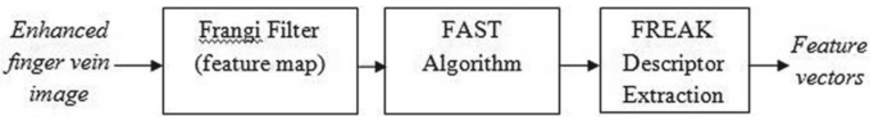


Fig. 7.2 Conceptual diagram of feature extraction (Rajan and Indu 2014)

A Hessian matrix is created from the second-order partial derivatives for the given pixel of an input image and is depicted by

$$H = \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix}. \tag{7.1}$$

In above matrix I stand for the grayscale input image and x, y for the pixel coordinates inside I. Here partial derivatives are determined as differences in pixel concentration in the pixel's instantaneous vicinity (Rajan and Indu 2014).

The programming code and processes conducted on the code are used in the proposed effort of feature extraction utilizing the Frangi filter. The FAST algorithm is then used to perform the remaining steps of corner detection.

7.2.1.2 FAST Algorithm

The FAST (Features from Accelerated Segment Test) technique is employed to detect image patch corners (points of interest). Figure 7.3 depicts the discovery of twelve-point segment test corners in a patch of the image.

Consider the parent pixel p in Fig. 7.3. Then draw a sixteen-pixel circle around the pixel p. Let t' be the chosen threshold value and S_p be the image intensity. If there are a set of l continuous pixels in the circle of sixteen pixels that are all brighter than $S_{p+t'}$ (S_p is the parent pixel's intensity) or all darker than $S_{p-t'}$, the detector classifies p as a corner point; otherwise, p cannot be a corner point. After that segment test will be applied to the remaining parent pixels after scrutinizing all pixels in the given circle. Whereas white highlights the pixels that were used for corner detection (Rajan and Indu 2014).

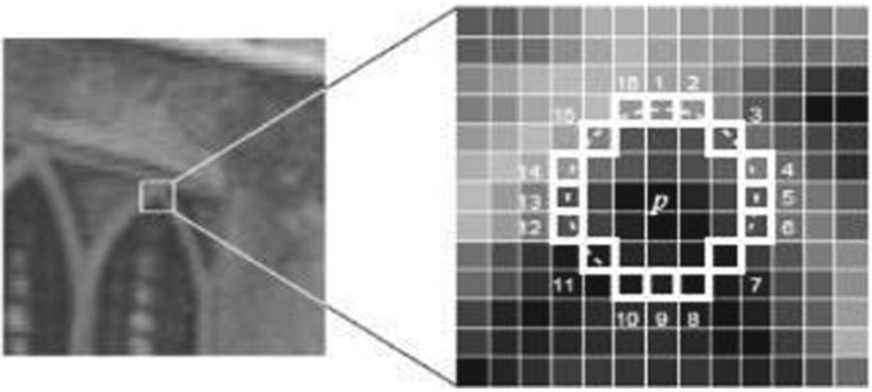


Fig. 7.3 An image patch's corner detection is tested using a 12-point segment test (Rajan and Indu 2014)

7.2.1.3 FREAK Descriptors

The human visual system, notably the human retina, inspired the Fast Retina Keypoint description. The image intensities are compared to create a binary string. As the computation of hamming distance is simpler compared to Euclidean distance, binary descriptors have the advantage of using it instead of Euclidean distance. The FREAK descriptor is designed to resemble the architecture of the retina (Rajan and Indu 2014).

7.3 Calculation and Verification of Accuracy

It is a crucial stage in ensuring the originality of the proposed work. It's also a follow-up to the findings described in Rajan and Indu (2014). Using discriminant analysis along with KNN, the accuracy of feature extraction is calculated in this proposed work.

7.3.1 Machine Learning Algorithm

A machine learning algorithm is a type of artificial intelligence that allows computers to learn without needing to be explicitly programmed. Machine learning is concerned with the development of computer programs that can self-educate in order to improve and adapt to new data. In terms of approach, machine learning is very similar to data mining. Both systems are looking for data patterns (Virtual learning environment).

However, Supervised Learning and Unsupervised Learning are the most extensively used Machine Learning Algorithms. Approximately 70% of machine learning is supervised, with the remaining 10% being unsupervised. The other two are hardly ever used.

Supervised Learning: This algorithm is trained on labeled samples in which the desired result for an input is known. Consider a piece of equipment with data points labeled "F" or "R," with "F" indicating failure and "R" indicating runs. This procedure is provided a set of proper inputs and outputs. This algorithm compares its correct output against its actual output in order to detect flaws. It then performs the appropriate model adjustments. Supervised learning is commonly employed in situations where previous data is used to predict expected future events.

Unsupervised Learning: When there are no historical labels on the data, this strategy is utilized. The system is not provided the "right response." What is being displayed must be determined by the algorithm. The purpose of this technique is to examine the data for any underlying structure. An unsupervised learning algorithm can find groups of clients with similar characteristics who can therefore be treated similarly in marketing efforts, or it can uncover the primary properties that differentiate client segments, based on transactional data.

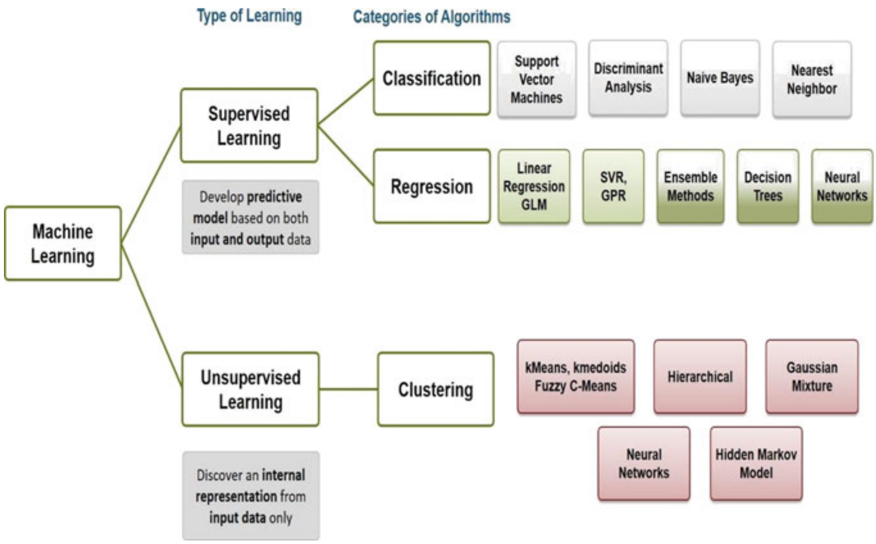


Fig. 7.4 The classification of machine learning (Wikipedia and the free Encyclopedia)

The machine learning algorithms, as well as the types and classifications of algorithm, are detailed in Fig. 7.4 (Wikipedia and the free Encyclopedia). The suggested work used Discriminant Analysis and the KNN approach, both of which are supervised learning methods. Let's have a look at some of them.

Discriminant Analysis: This is a statistical approach that predicts a clear-cut dependent variable using one or additional continuous or binary independent variables (called predictor variables) (called a grouping variable).

Discriminant analysis is employed whenever groups are known a priori (unlike cluster analysis), and each instance must have a score on one or more quantitative predictor measures, as well as a score on a group measure (Wikipedia and the free Encyclopedia).

K-Nearest Neighbor Algorithm: In pattern acknowledgement, non-parametric approaches can be employed for classification and regression. It is basically termed as lazy learning, in which the function is only approximated locally and all computations are postponed until after classification. This is one of the most fundamental machine learning approaches.

- The comparison of both methodologies, Discriminate Analysis and KNN, is the next and the proposed work's final phase.
- Finally, discriminant analysis was determined to be the most effective phase.

7.4 Results and Discussion

This part contains the results collected following the appropriate execution of the research study using the tools and algorithm previously stated. The leveled outcome is shown in this section. The paper's purpose is to find an optimum segmentation level and then use a classification technique to categorize the results. In the current situation, the complete evaluation is based on a variety of subjects and their corresponding image sets. Using a new sensor, a carefully obtained image is shown in Fig. 7.5. Image quality along the right side of the sensor image is slightly deformed, which is generally referred to as image artifacts.

The Otsu method will be used to evaluate the image in Fig. 7.6. It uses a simple cumulative probability approach based on a histogram, with the probability values displayed in the bins determined in the histogram. Users apply the Otsu method to establish an exclusive threshold, which will be used to Binarize the image, employing this probability value as well as detecting a valley in that probability in bins. The reader will note that the image quality in Binarized (Fig. 7.6) is black in the region of interest when the images 5 and 6 are compared. As a result, image quality will be judged and changes will be done.

When the images in Figs. 7.6 and 7.7 are compared, the images are used and binarized, but the most important factor is image quality; in this case, the image in Fig. 7.7 displays no artifacts in the region of interest indicated in black and white. As a result, the goal of image opening is to eliminate opening such that the region of interest is free of artifacts.

When picture noise is removed, the current image processing procedure is considered complete. Figure 7.8, which depicts such information via an image filtering technique, does a decent job of displaying it. When contrasted to Fig. 7.7, noise can be seen as little spots of white, small holes in the white zone, or undesirable artifacts in the white region, as illustrated in this image. The image quality is excellent enough to grasp the image using a simple morphing technique, as seen in the figure shown.

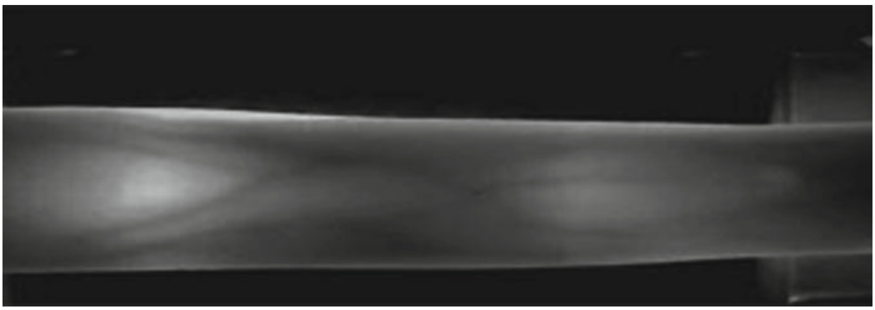


Fig. 7.5 Using an IR sensor, an image of a finger vein was recorded

Fig. 7.6 After using the Otsu method to binarize the image, the figure shows the result

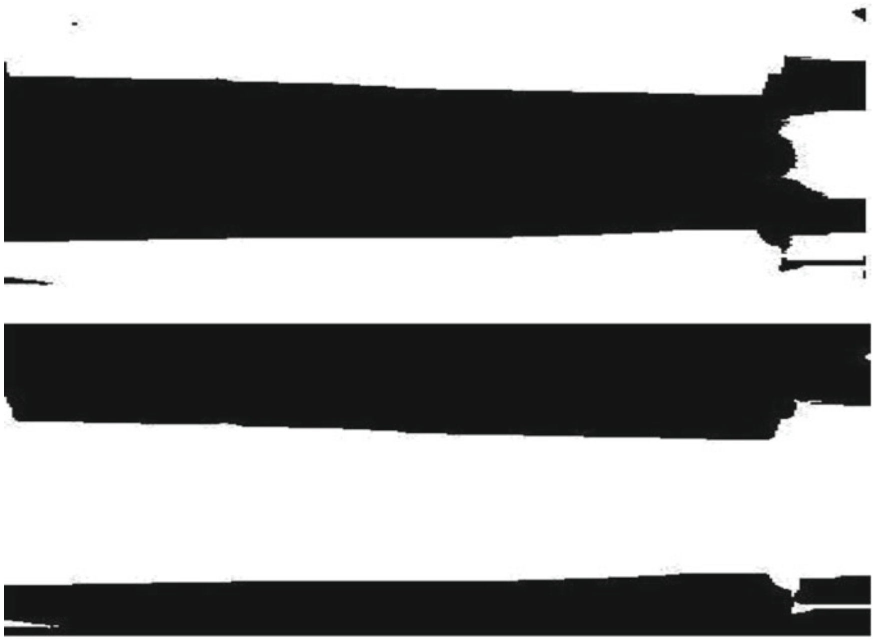


Fig. 7.7 Image in black and white obtained after morphological opening

The image quality is good in this example; however, Fig. 7.9 illustrates that the image will be improved only in the well-perceived finger region as the region of interest. This image shows the correct finger image part.

In this illustration, the image is further reduced to borders and pixels. Figure 7.10, which highlights the image's boundary region. The image will be converted to boundaries using a pixel-based operation provided in MATLAB as a function of a Sobel edge detection approach.

Fig. 7.8 After removing all noise, this image was created

Fig. 7.9 Image after the complete complementary operation

Fig. 7.10 Image after edge detection using Sobel filtering

The image quality will be examined in Figs. 7.9 and 7.10 utilizing a unique subtraction operation that employs the morphological operation and the edge image, respectively. Because an outer zone will emerge in the image, which will be identified and the border must be erased, the image quality will be obvious (Fig. 7.11).

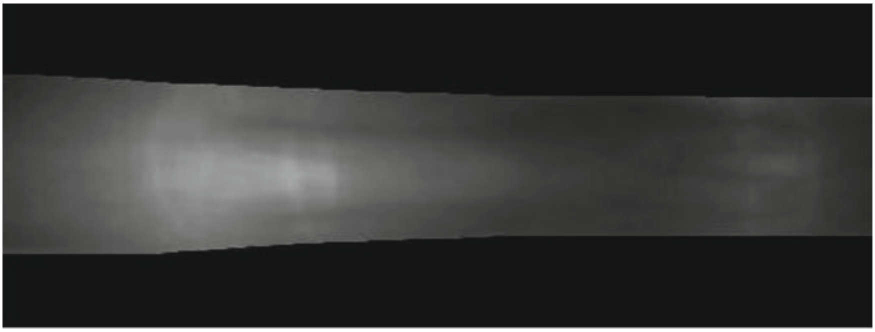


Fig. 7.11 Image evaluated using small subtraction of edge image and morphing

Figure 7.12 illustrates a high level of picturization of the region of interest. Subjects use a variety of techniques to filter and eliminate image artifacts, but the final goal is to identify that portion of the image that is completely that finger region is only visible here since it is noise-free. Technically, this image is known as a mask image, because the black part contains masks and the white portion has open space.

This component of the result shows a complete segmented picture representation of a finger divided into two sections. The prime region is a finger-shaped point of interest, with the rest of the area dark. This image was created by overlapping two images, the original image in Figs. 7.5 and 7.12.

The purpose of this image is to make the computation process easier by reducing the image's exact region by removing the background that isn't necessary for the calculation of imperative characteristics (Fig. 7.15).

Because a user is most concerned with a high-quality image, the computation performed will be very distinguishable if the image acquired in Fig. 7.14 could be brightened. In this regard, the research work suggests that if the image quality is poor after all preprocessing procedures; increase the contrast by using the local histogram idea, which is generally available as a MATLAB image processing function called

Fig. 7.12 After thorough filtering, noise removal, and subtraction, the image is shown

adapthisteq. This is the final stage of image enhancement section, here the image is enhanced.

7.4.1 Accuracy and Calculation

This section will calculate the accuracy of the findings using the algorithm used in Machine Learning's Supervised Learning. In this work, two supervised learning methodologies were used: discriminant analysis and KNN. The accuracy acquired using the KNN algorithm is 55.84%, although the precision computed using the Discriminant Analysis technique is 92.21%.

Figure 7.16 depicts the differences between the two approaches used in the study to calculate the accuracy of characteristics extracted from photographs of patients' finger vein patterns. The approaches employed are (1) discriminant analysis of a Machine Learning algorithm, and (2) KNN (K-nearest Neighbor Algorithm). The precision gained via KNN is 55.84%, however, the accuracy acquired from the discriminant analysis is significantly greater, at 92.21%.

7.5 Results Analysis

This section discusses the accuracy calculation as well as the comparison between mentioned techniques. After upgrading the captured images received from the open source (Yin et al. 2011). In this suggested research project, the correctness of the features of human finger vein patterns was determined.

The accuracy acquired using the technique of discriminant analysis is 92.21%, however, when other techniques were used to increase the accuracy, it was discovered that the accuracy gained using the current technique is the best. The other technique used for comparison is the KNN algorithm, which is a supervised learning technique. The accuracy acquired from this KNN technique is significantly less than that gained from the proposed technique, namely, 55.84%. The discussion of the result and accuracy is very well described by the above given two figures (Figs. 7.4 and 7.13).

7.6 Summary

For the accuracy or result computation in this proposed study, two algorithms, discriminant analysis along with KNN, are used. The properties of vein patterns with respect to human fingers, after being extracted from the enhanced images are utilized in the current work. These results are used to calculate accuracy. The accuracy calculated with KNN is 55.84%, while the accuracy calculated with discriminant analysis is 92.21%. By examining the aforementioned accuracy, we can compare both tactics

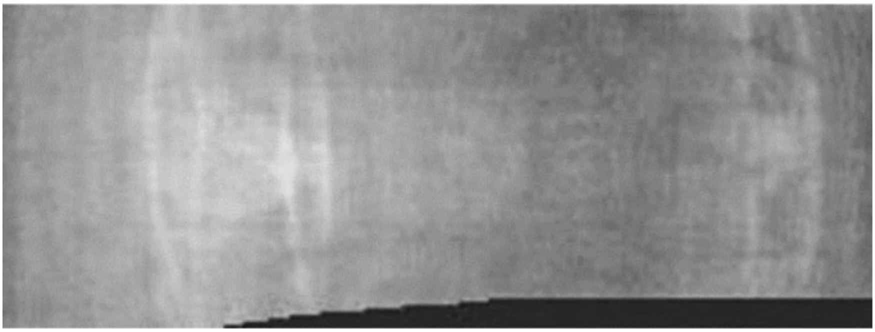


Fig. 7.13 Image after final ROI implementation

Fig. 7.14 Image shows the reduced dimension of the region of interests

Fig. 7.15 Enhanced image of the reduced dimension of finger image

and determine which is the most effective. The optimum technique for calculating the accuracy of finger vein pattern characteristics is discriminant analysis, which we can deduce.

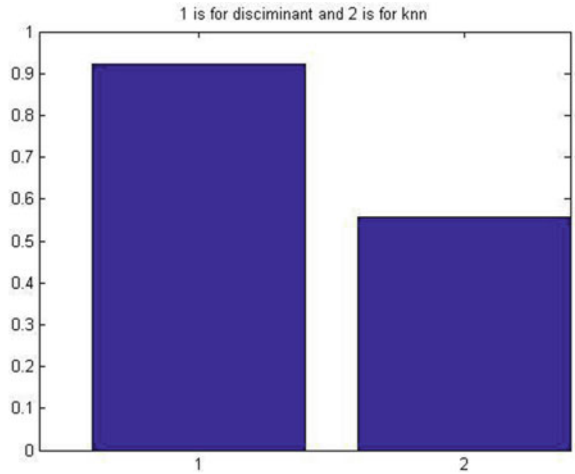


Fig. 7.16 The accuracy of two different approaches of recognition

References

Corcoran P, Iancu C, Callaly F, Cuco A (2007) Biometric access control for digital media streams in home networks

Jain AK (1989) Fundamentals of digital image processing. Prentice Hall of India, p 64

Jain AK, Ross A, Prabhakar S (2004b) An introduction to biometric recognition. IEEE Trans Circ Syst Video Technol 14(1)

Kono M, Ueki H, Umemura S (2002) Near-infrared finger vein patterns for personal identification. Appl Opt 41(35):7429–7436

Liu Z, Song S (2012) An embedded real-time finger-vein recognition system for mobile devices. IEEE Trans Consum Electron 58(2)

Miura N, Nagasaka A, Miyatake T (2004) Feature extraction of finger vein patterns based on repeated line tracking and its application to personal identification. Mach Vis Appl 15(4):194–203

Miura N, Nagasaka A, Miyatake T (2005) Extraction of finger-vein patterns using maximum curvature points in image profiles. In: Proceedings of IAPR conference machine vision applications, Tsukuba Science City, Japan, May 2005, pp 347–350

Mulyono D, Jinn HS (2008) A study of finger vein biometric for personal identification. Computer Science and Information Engineering Department, National Taiwan University of Science and Technology (NTUST), No. 43, p 63

Prathiban K, Wahi A, Sundaramurthy S, Palnisamy C (2014) Finger vein extraction and authentication based on gradient feature selection algorithm

Rajan R, Indu MG (2014) A novel finger vein feature extraction technique for authentication. In: International conference on magnetics, machines & drives (AICERA-2014 iCMMD)

The Biological Imaging Facility, Berkeley. Website http://microscopy.berkeley.edu

The Wikipedia the free Encyclopedia, Website http://en.wikipedia.org

Umbaugh SE (1998) Computer vision and image processing. Prentice Hall PTR, New Jersey

Vega AP, Travieso CM, Alonso JB (2014) Biometric personal identification system based on patterns created by finger veins. In: International work conference on bio-inspired intelligence (IWOBI)

Virtual learning environment, Website http://whatis.techtarget.com/

Vlachos M, Dermatas E (2013) Finger vein segmentation in infrared images using supervised and unsupervised clustering algorithms. Pattern Recogn Image Anal 23(2):328–334. Pleiades Publishing, Ltd.

Wang K, Ma H, Popoola OP, Li J (2010) Finger vein recognition

Wu X, Gao E, Tang Y, Wang K (2010) A novel biometric system based on hand vein. In: 2010 fifth international conference on frontier of computer science and technology

Yin Y, Liu L, Sun X (2011) SDUMLA-HMT: a multimodal biometric database" biometric recognition. In: Lecture notes in computer science, vol 7098. Springer, pp 260–268

Chapter 8
Machine Learning Techniques for Anomaly Detection Application Domains

Alka Chaudhary and Reshu Agarwal

Abstract Anomaly detection is a popular research topic these days since it influences a wide scope of uses. A variety of strategies for detecting abnormalities in various domains have been developed. The most appropriate anomaly detection technique depends on the problem characteristics and problem domain, which is why this chapter focuses on extremely common strategies for recognizing abnormalities in many application areas, namely machine learning techniques. These strategies try to make a system intelligent, allowing it to adopt new things from data sets and readily classify important information from abnormalities. This chapter also addresses the benefits and drawbacks of machine learning approaches, as well as the different types of anomaly detection applications.

Keywords Anomaly detection · Machine learning techniques · Abnormalities

8.1 Introduction

The difficulty of recognizing designs in the information that don't adjust to expected behaviour is alluded to as anomaly detection. In a variety of application fields, these non-adjusting designs are alluded to as anomalies or outliers. Anomalies and outliers are two of the most widely used phrases with regard to anomaly identification. Anomaly detection is a notion that is utilized in a wide scope of utilizations.

In a PC organization, for example, an unusual traffic example could indicate that a hacked PC is transferring delicate information to an unapproved location without permission. Typical MRI images may also suggest the existence of harmful cancer. Various investigation studies have produced an assortment of anomaly detection systems over time. A considerable lot of these systems were created specifically for

A. Chaudhary · R. Agarwal (✉)
Amity Institute of Information Technology, Amity University, Noida, India
e-mail: agarwal.reshu3@gmail.com

A. Chaudhary
e-mail: achaudhary4@amity.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
A. Rai et al. (eds.), *Paradigms of Smart and Intelligent Communication, 5G and Beyond*, Transactions on Computer Systems and Networks,
https://doi.org/10.1007/978-981-99-0109-8_8

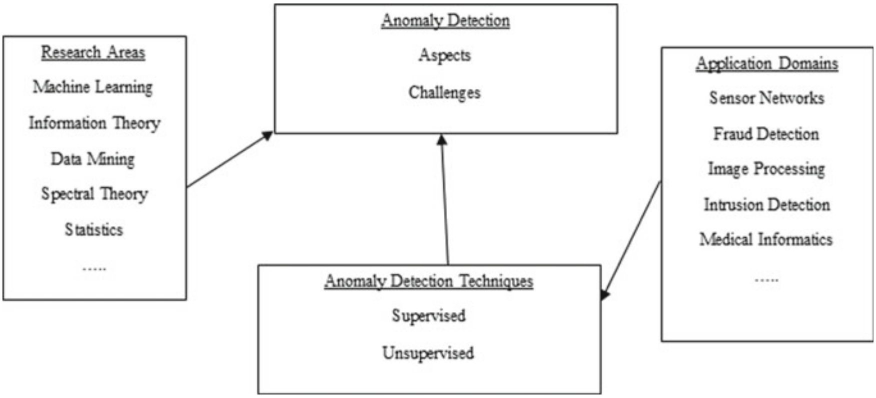


Fig. 8.1 Basic components of anomaly detection technique

specific application domains, while others are more widely applicable. The basic components of anomaly detection approaches are depicted in Fig. 8.1.

8.2 Anomaly: What Is It?

Anomalies are designs in the information that don't adjust to a well-defined definition of standard behavior. The anomalies in basic 2-layered information set are depicted in Fig. 8.2. Because the majority of observations are in these two regions, the information has two ordinary areas: N1 and N2 (Tan et al. 2005).

Anomalies in the information can be inferred from an assortment of causes, including malevolent activities. A significant and important component of anomaly detection is the real-world inference of anomalies. Anomaly detection is like, however

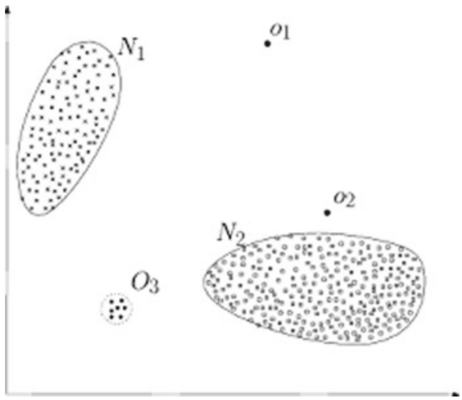


Fig. 8.2 A basic example of anomalies in a 2-layered information set

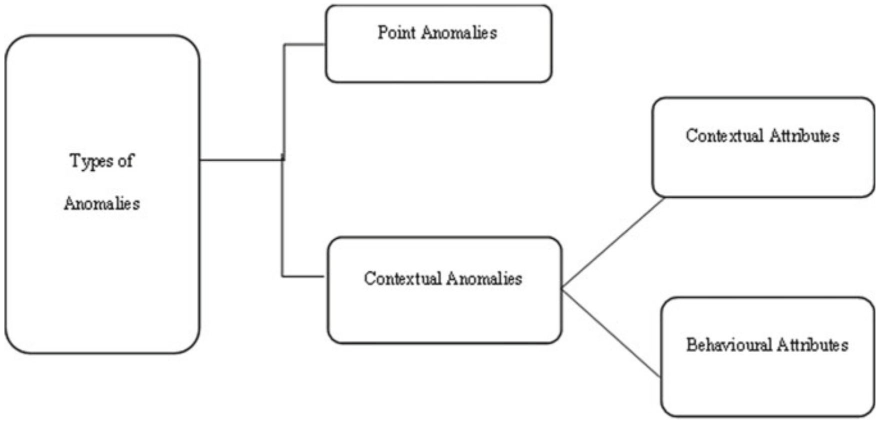


Fig. 8.3 Categorization of Anomalies

not equivalent to, noise removal and noise accommodation, which both have undesired noise in the information. Noise is characterized as an element of information that isn't significant or fascinating to the scientist however works as a hindrance to information translation.

The necessity to remove superfluous items before performing any information investigation on the specific information leads to noise removal. The categories of anomaly detection are shown in Fig. 8.3. Novelty detection (Tan et al. 2005) is another topic linked to anomaly detection that aims to find previously unnoticed (emergent, novel) patterns in data. Further, there are three sorts of anomalies.

8.2.1 Point Anomalies

A point anomaly is defined as a data instance that can be regarded as aberrant in comparison to the rest of the data. This is the most fundamental kind of anomaly, and many scientists focus on anomaly detection studies.

Let us take an instance, in the image above, the points o1 and o2, as well as the points in area O3, exist outer the regular regions' boundaries and are thus point anomalies since they differ from the typical data points. Consider the utilization of Visa misrepresentation identification for instance. It permits the informational collection to be connected to an individual's credit card purchases. Let us expect that the information is characterized by just one element, to be specific the sum spent, for lucidity. A point anomaly is a transaction wherein the sum spent surpasses the normal scope of consumption for that individual.

8.2.2 Contextual Anomalies

Contextual anomalies occur when an information instance is abnormal just in one context (but not in others) (Chawla et al. 2004). The structure of the data set infers a context's strategy, which must be specific to the problem group. The following two groups of attributes are used to describe each data instance.

8.2.2.1 Contextual Attributes

The contextual attributes are utilized to affect the occurrence's specific circumstance (or neighborhood). In spatial information sets, for example, the longitude and latitude of a certain area are relevant properties for that information set. In time series data, time is a relevant element that aids in determining an instance's position within the sequence.

8.2.2.2 Behavioural Attributes

The behavioural properties are used to describe an instance's non-contextual qualities. How much precipitation in any area, for example, is a behavioural characteristic in the geographical information set representing the normal precipitation of the whole world? The values for the behavioural traits within a certain context are used to drive the abnormal behaviour. In one situation, an information instance may be a contextual anomaly, but in another situation, it may be an identical data instance. This feature is critical for determining the contextual and behavioural attributes.

8.2.3 Collective Anomalies

A collective anomaly occurs when a group of connected data examples is abnormal in comparison to the overall information set presented.

8.3 Aspects of Anomaly Detection and Challenges

This section delves into the numerous components of anomaly detection as well as the issues that the field encounters. The formulation of a particular problem may be influenced by a variety of factors such as input data, data labelling, and application domain requirements, so various parts of the problem may change.

8.3.1 Aspects of Anomaly Detection

There are various parts of anomaly detection which are given below.

8.3.1.1 Input Data

This kind of approaching information is a critical part of any anomaly identification method. The qualities can be binary, category, or continuous, for example. Every information case might incorporate just one property (otherwise called univariate) or a few attributes (otherwise called multivariate) (likewise called as multivariate).

All attributes in multivariate information occurrences might be of a similar sort or a blend of various information types. The feasibility of anomaly detection algorithms is influenced by the type of attributes. Likewise, the type of attributes would aid in determining the distance metric to be employed in nearest neighbor-based strategies. Rather than the genuine information, a distance (or similarity) matrix may be used to represent the pairwise distance between instances. Techniques that demand unique data instances are impractical in such situations. The link between data instances can also be used to categories input data (Theiler and Cai 2003).

8.3.1.2 Labeling of Data

The labels related to an information case demonstrate whether it isn't unexpected or unusual. Obtaining accurate and representative labelled data for all kinds of actions is typically too expensive. Since labelling is frequently done physically by an expert, obtaining the labelled training data set required necessitates a significant amount of effort. Creating a labelled set of anomalous information occurrences that covers a wide range of anomalous conduct is typically more troublesome than getting labels for regular normal conduct. Furthermore, abnormal conduct is frequently dynamic. Anomalies are translated to catastrophic outcomes in some contexts, such as air traffic safety, and so are extremely rare. Anomaly detection algorithms can work in one of three modes, depending on the range of labels available.

Supervised Anomaly Detection

In supervised mode, procedures consider the accessibility of a training information collection with labelled cases for both the normal and anomalous classes. In such cases, a general method is to create a predictive model for normal vs. anomalous classes. Any beforehand concealed information example is then contrasted with the model to figure out which class it has a place with. When dealing with supervised anomaly detection, there are two primary issues to consider. First, compared to the usual instances in the training data, the anomalous values are significantly less.

Machine learning techniques (Phoha 2002) have addressed issues that develop as a result of uneven class distributions. Second, obtaining an appropriate and representative label, particularly for the anomaly class, might be difficult. Several strategies have been described for injecting artificial anomalies into a regular data collection to get a labelled training information set for ease of use (Hofmeyr et al. 1998).

UnSupervised Anomaly Detection

Strategies that work in the unsupervised mode don't need any training information; subsequently they're the most widely utilized. In the test information, the techniques in this classification make the right suspicion that normal events are essentially more normal than abnormal ones. In the event that this supposition is inaccurate, such systems will have a higher pace of deceptions. By utilizing an example of the unlabeled informational collection as training information, a few semi-administered approaches can be changed over to work in an unsupervised mode. Such variation surmises that the test information contains few anomalies and that the model picked up during preparation is impervious to these anomalies.

8.3.2 Challenges Faced in Anomaly Detection

- It's difficult to explain a typical zone that contains all conceivable normal behaviour. Furthermore, the distinction between normal and abnormal conduct is frequently blurred. As a result, an unusual observation near the border can turn out to be typical, and vice versa.
- A key and major challenge is the accessibility of labelled information for training.
- Normal behaviour continues to emerge in many fields, and a current approach to normal conduct may not be sufficient to reflect it in the future.
- When anomalies occur as a result of malevolent behaviour, the hostile opponents frequently cooperate to make the aberrant findings appear normal. As a result, defining normal behaviour becomes even more difficult.
- The precise approach to an anomaly varies depending on the application domain. A modest divergence from normal in the medical domain may be deemed an anomaly, however a comparable deviation in the financial exchange space (e.g., variations in stock value) may be considered normal but not an abnormality. Accordingly, moving a strategy created in one space to another isn't direct.

8.4 Application Domains

The application areas of anomaly detection are discussed here. Anomalies must be detected in a variety of applications in order to create and classify the correct patterns to the linked sets.

8.4.1 Medical and Public Health Anomaly Detection

Anomaly detection in this domain is often dependent on patient records and works with them. Anomalies can happen for an assortment of causes, including distorted patient circumstances, contraption slip-ups, and recording blunders. Detecting disease epidemics in a specific area has also been the focus of several strategies (Lane and Brodley 1999). As a result, anomaly detection is a major concern in this area and necessitates a great level of precision. The information is typical as records, which might incorporate an assortment of separating qualities, for example, the patient's age, blood gathering, weight, etc. The information could be both temporal and spatial in nature. In this area, the majority of the current anomaly detection approaches are aimed towards finding aberrant records (point anomalies). The collective anomaly detection algorithms are used to find anomalies in such data (Augusteijn and Folkert 2002). The most troublesome part of this current anomaly detection challenge is that the expense of classifying an anomaly as ordinary may be very high.

8.4.2 Intrusion Detection

In a PC-related framework, intrusion detection alludes to the location of threatening action, for example, break-ins, infiltrations, and different types of PC misuse (Ming 2011). From the standpoint of computer security, these harmful operations or attacks are tempting. Because an intrusion is distinct from typical system behaviour, anomaly detection techniques can be used in the intrusion detection domain. The massive amount of data in this sector poses a significant difficulty for anomaly detection. The false alarm rate is one issue that arises as a result of the big input size. Because the data consists of lots of information pieces, a small percentage of false alarms can overwhelm an analyst's analysis. Labeled data pertaining to normal behaviour is widely accessible, whereas intrusion labels are difficult to get by. As a result, in this domain, unsupervised approaches are preferable. Intrusion detection systems are partitioned into two classifications: host-based and network-based intrusion detection.

8.4.2.1 Host-Based Intrusion Detection Systems

Operating system call traces are dealt with by system call intrusion detection systems. The invasions take the form of atypical tracing sequences (collective anomalies). Malicious programmes, unlawful conduct, and policy violations are all linked to the strange sequences. While all follow contain occasions from similar letters in order, the co-event of occasions is the main perspective in distinctive ordinary from strange conduct. The information is consecutive in structure, and the letter set is comprised of individual framework calls. These calls could be made by people or by programs (Moore and Zuev 2005). The letter set is generally very immense. These system calls are executed in different orders by different programmes. The length of each program's sequence varies from one to the next.

8.4.2.2 Network Detection Systems

These frameworks are entrusted with distinguishing network information attacks. Even while certain algorithms represent the information in a sequential form and discover the abnormal subsequences, the incursions usually manifest themselves as aberrant patterns (point anomalies) (Mukkamala et al. 2005). The assaults performed by outside programmers who need illicit admittance to the organization for data robbery or to disturb the organization are the primary driver of these irregularities. An overall setting for this area is a colossal organization of PCs associated with the remainder of the world through the Internet. Intrusion detection frameworks can utilize the information at different degrees of granularity. The sort of irregularities changes over the long run as wrongdoers modify their organization assaults to evade the current interruption identification techniques. This presents a test for anomaly detection calculations in this space.

8.4.3 Industrial Damage Detection

Persistent use and ordinary mileage make modern units fall apart. To avoid future escalation and losses, such damages must be detected early. Since it is archived utilizing different sensors and gathered for investigation, the information in this area is generally alluded to as sensor information. Anomaly detection methods have been broadly used to recognize such harms in this space. Modern harm recognition can be partitioned into two classifications.

8.4.4 Fault Detection in Mechanical Units

In this domain, anomaly detection approaches screen the presentation of modern parts, for example, engines, turbines, oil stream in pipelines, and other mechanical parts to distinguish any issues that might emerge because of mileage or other unanticipated occasions. The information in this domain is regularly transient, and a few methodologies utilize time-series investigation (Peddabachigari et al. 2007). An examination in a particular climate (context-oriented peculiarities) or an unusual progression of examinations is the most well-known reasons for irregularities (aggregate oddities). Parametric Statistical Modeling, Non-parametric Statistical Modeling, etc. are a portion of the peculiarity location methods used in these areas.

8.4.5 Structural Defect Detection

Techniques for detecting structural defects and harm, like breaks in bars and stresses in airframes, are utilized to identify structural anomalies in structures. The data assembled in this area has a time part. Since they endeavor to identify assuming that there is any adjustment of the information got from a structure, anomaly detection procedures are likened to curiosity location or change point detection approaches. The average information, and consequently the learned models, is typically perpetual over the course of time. There might be a few spatial relationships in the information.

8.4.6 Fraud Detection

Fraud detection refers to the discovery of crime in business foundations, for example, banks, MasterCard firms, insurance offices, mobile phone organizations, and the stock exchange. The pernicious clients could be genuine clients of the organization or they could be acting like shoppers (otherwise called wholesale fraud). The misrepresentation happens when these clients exploit the association's assets in an illicit way. The organizations are interested in detecting such frauds as soon as possible in order to avoid any financial losses. A typical way to anomaly detection is to make a utilization profile for every client and afterward screen the profiles separately to check whether there are any varieties. The following are some examples of fraud detection applications.

8.4.6.1 Detection of Frauds in Credit Cards

Anomaly-based detection techniques are used in this domain for detecting fake credit card applications or usage. The detection of false credit card applications is comparable to the detection of fraudulent insurance applications. Data is often made up of records with many ranges, like the id of the user, the money involved, the period between using cards, etc. The fraud is usually discovered where there is a transaction involved and is associated with a large payment, purchase of things the user has never purchased before, etc. The difficulty with identifying illegal credit card use is that it necessitates detecting online fraud as soon as the fraudulent transaction occurs. To overcome this difficulty, two distinct ways are used to apply the anomaly detection techniques. The first is called by-owner, and it describes every user of the credit card on the basis of the history of his credit card usage. Whenever a new transaction takes place, a comparison of the user profile and the transactions is carried out, and if it does not match, it is flagged as an anomaly. This method is usually quite costly. The by-operation method, on the other hand, is used in detecting the anomalies from transactions occurring at a specific geographic region. The contextual type of abnormalities is detected using both by-user and by-operation strategies. Neural Networks, Rule-based Systems, and Clustering are some of the techniques of anomaly detection employed in this domain.

8.4.6.2 Mobile Phone Fraud Detection

A common activity control issue is mobile fraud detection. This detection's job is to scan a huge number of accounts, examine each one's calling behavior, and raise alarm if an account seems to be compromised. Calling activity can be demonstrated in a variety of ways, although it is most commonly expressed using the user's call records. Calls are grouped by time, such as by location or user, or call-days or call-hours, that depends on the level of granularity desired. The abnormalities are related to the high volume or calls to unusual locations.

8.4.7 Sensor Networks

Sensor networks are a hot topic of research in the current times, particularly from the standpoint of data processing, because sensor data which is generated using diverse wireless sensors has multiple distinct features. Anomalies in sensor network data indicate the malfunctioning of one or more sensors, or the detection of events (e.g., invasions) which may be of interest to analysts. Using techniques in this domain, both sensor defect detection and intrusion detection may be captured by the detection of an anomaly in a sensor. The information is obtained in a streaming fashion. The acquired data, in the area where the different sensors are located and in the communication route, is often marred with noise and missing values. The detection of anomalies in the sensor networks presents a variety of challenges. These techniques, thus, must be used only in an online environment. Due to the scarcity of resources, anomaly detection techniques must be light in weight. Furthermore, the data collection should be done in a distributed way, using distributed data mining method to analyze the data. Some of the techniques to detect anomalies in the sensor networks used in this domain include Nearest Neighbor based techniques; Rule-based systems; Spectral and Bayesian Networks.

8.4.8 Image Processing

The major focus areas of the detection algorithms that detect picture anomalies are changes in an image with time i.e., motion detection or regions which seem odd on a static image. Some examples of this domain are Satellite images, digit recognition, spectroscopy, mammographic image analysis, and video surveillance. Motion or the insertion of an external object, as well as instrumentation faults, are the sources of the anomalies. The data includes both temporal and spatial features. Every data point has a few qualities that are utilized to analyze it further, such as color, lightness, texture, and so on. Some of the techniques which were used to detect anomalies in this domain include Nearest Neighbor based Techniques, Neural networks, Regression, Support Vector Machines, a Mixture of Models, Clustering, and Bayesian Networks.

8.4.9 Text Data

The anomaly detection algorithms, in the current domain, have been used to identify new themes, news articles, news items in documents, or events. A new fascinating occurrence or an abnormal topic is to blame for the anomalies. The data in this area is very sparse, high-dimensional and temporal since the documents were gathered over time. The huge variability in the documents of a given category presents a difficulty for anomaly detection systems in this sector. The techniques utilized in this domain include Clustering, Neural Networks, Support Vector Machines, Statistical Profiling through Histograms and Mixture of Models.

8.4.10 Data Leakage Prevention

When it comes to Anomaly Detection, this area is another important application case. Sensitive data is safeguarded from harmful attackers by identifying data loss early on, ensuring that the data stays secure and integrated. In theory, it is a kind of fraud detection, however, the focus is on near-real-time analysis, so that it may be used as a preventive measure. The unusual access patterns or the anomalies are detected by logging and analyzing the access to file servers, databases and other crucial information sources.

8.5 Anomaly Detection Techniques

The application areas of the detection of anomalies have been discussed in this section. Anomalies must be detected in a variety of applications to create and classify the correct patterns to the linked sets.

8.5.1 Supervised Methods

These techniques are often known as classification methods. To build a predictive model for these methods, one needs a set that includes both normal and anomalous samples. The Supervised methods provide greater detection rates as compared to the Unsupervised or Semi-Supervised methods since the supervised methods contain more information about the sets. Nevertheless, due to some technical limitations, the approaches used are a little inaccurate, despite the fact that they should be. One of the challenges is that obtaining a collection that covers all the categories is difficult. Furthermore, getting specific labels can be difficult, and training sets frequently contain sounds, resulting in greater false alarm rates. The following are some examples of Supervised Algorithms.

8.5.1.1 K-nearest Neighbor (k-NN)

It is a well-known and moderate non-parametric way of segregating samples. First, the distance between the points on an input vector is estimated and then the points which are not labelled are assigned their nearest neighbors (k) which is an important criterion in properly constructing a k-NN range. It is to be noted that different (k) values can lead to the difference in behavior. If the value of k is larger, the neighbors who were utilized for the earlier guesses would take longer and would have an impact on the accuracy of the forecasts. A data occurrence's anomaly score is defined as its extent to its kth nearest neighbor in each data set. In the present study, the technique has been used to detect landmines from satellite pictures as well as anomalies in the DC area networks of massive synchronous turbine generators.

Three different strategies have been used by researchers to improve this. The first group of researchers tweaked the above definition to calculate a data occurrence's anomaly score. The second group of researchers manages various data kinds using various range/closeness requirements. The third group of researchers focuses on using various approaches to improve the capabilities of the basic techniques (the complication of the fundamental technique is O (N2), where N is the size of the data).

Further a developed a two-phased combinational approach for feature selection called filter and wrapper. The filter phase was used to extract the specifications with the highest information gain, and the output was then passed to the wrapper phase, which produced the final specification subset's output. K-NN classifier used the final specification subsets as inputs to classify attacks. The effectiveness of this approach was proved using a cyber-attack dataset. For feature selection and weighting, A researcher advocated a genetic approach combined with k-NN. During the training phase, all 35 primary traits were weighed, and the heaviest ones were selected for testing. To evaluate the systems, a variety of DoS assaults were used.

8.5.1.2 Bayesian Network

This is a representation that decodes probabilistic associations between variables. With the use of a variety of statistical models, this technique is commonly used to identify any breach. It provides several benefits including the capacity to encode the relationship between different variables and completed events and the ability to include past knowledge. According to recent researches, a Bayesian Network system gives a suitable mathematical basis for building simple but potentially problematic models. They claimed that Bayesian Network-based IDS should be able to identify attacks among the normal network activities by examining metrics from each and every sample of network traffic. Further a researcher employed an administered Naive Bayes classifier using 248 flow features, as well as several TCP header inherited features, to distinguish between various sorts of applications, including packet length and inter-arrival durations. To express stronger features, the correlation-based feature option was employed, and it was established that even a small part of less than 20 features can work in getting precise distribution.

The technique can be effectively used for multivariate categorical sets of the data by integrating the per-attribute posterior probability for each test and by utilizing the aggregated results to provide a class label to all test occurrences.

Several variations of the technique have been created and are being used for network intrusion detection, video surveillance novelty discovery, text data anomaly detection, and disease outbreak identification. It also presupposes that distinct parameter are independent of one another. Other versions that were proposed used Bayesian networks, one of the more complicated methods, to consider the dependencies between the differentiated parameters.

8.5.1.3 Supervised Neural Networks (NN)

The learning of this technique predicts the behavior of various users in systems. NNs have the capacity to address a variety of circumstances that rule-based techniques are unable to address if they are properly developed and executed. The fundamental advantage of NNs is their capacity to find complete findings without any prior information about the regularities of the data, as well as their resilience to imprecise data and ambiguous information. When this is combined with their ability to generalize from learning data, they have proven to be an effective way to ID. To use this method to ID, data displaying assaults and non-attacks must be supplied or introduced to the NN for it to change its network coefficients automatically throughout the duration of its training.

A rudimentary neural network-based multi-class anomaly detection system operates in two ways. To learn the different typical classes, a neural network is first qualified on the usual qualifying data. Second, each test case is fed into the neural network as an input. Network accepting the test input is considered normal; but the rejection of the test input by the network is considered an abnormality.

Several different types of neural networks have been proposed as modifications to the basic neural network technique. Only one-class anomaly detections were done with Replicator Neural Networks.

8.5.1.4 Support Vector Machines (SVM)

The next proposed model is support vector machines (SVM). The input vector is transformed into a high-dimensional feature space, then the best-separating hyperplane is found in that space. An SVM classifier was created specifically for binary classification, which involves the separation of training vectors set into two distinct classes. The training samples closest to the final conclusion judgement border. A penalty-factor is also provided by the SVM. It enables users to trade off the number of misclassified samples and the decision boundary's width. Further a model was devised to address anomaly detection challenges in the network by employing well-practiced kernels and classifier designs. They looked at the effect of type and parameter values of the kernels on the accuracy of an intrusion classification support vector machine (SVM).

In addition, this method employs a single-class learning methodology to learn a region containing the instances of the training data. The basic technique identified whether each test event fell within the learning zone for each test occurrence. A test occurrence was classified as a regular occurrence if it occurred inside the learning region; otherwise, it was classified as an anomaly.

8.5.1.5 Decision Tree

Decision Trees are common but strong tools for data classification and prediction. Each node has a unique feature parameter assigned to it, which is the most descriptive of the parameters not yet examined along the path from the root. The labelling of every leaf is done with a class or category, and a key feature value for the node's feature is the label for every arc out of a node. Beginning at the root traversing through it till the leaf node is used for the classification of data point. The classification of the data point is given by the leaf node. Quinlan's ID3 and C4.5 are only two of the many Decision Tree implementations available. The ID3 algorithm was utilised to automatically create the decision tree.

8.5.2 UnSupervised Methods

These approaches differ from supervised methods in that no training data is required. To begin, an assumption is made regarding the network that there is regular traffic very less anomaly. The second assumption is that the difference between the normal and the malicious traffic is statistical. The data is grouped according to these two assumptions and groups of the same occurrences that occur frequently are presumed to have legitimate traffic, but occurrences that occur infrequently and differ significantly from the bulk of occurrences are considered malevolent. The following are the most prevalent unsupervised algorithms.

8.5.2.1 Unsupervised Neural Network

The adaptive resonance theory and the self-organizing maps are the two most frequent unsupervised neural networks. They employ affinity or proximity to class objects as a criterion. They can detect intrusions in situations where normal behaviour is heavily fixed or concentrated around one or two centres, but anomalous behaviour and intrusions spread out beyond the regular clusters.

An unsupervised competitive learning algorithm creates the Self-organizing map (SOM). The goal of Self-organizing maps is to reduce the number of dimensions in data visualisation. The outputs are grouped into a low-dimensional grid and contain a Kohonen layer and an input layer. Every input vector is linked to a typical output in Kohonen's SOM. The network then selects the node nearest to every training case and brings the most appropriate node in the training course, which is the nearest neuron.

The adaptive resonance theory (ART) encompasses a set of neural network models for prediction, pattern recognition, and supervised and unsupervised learning. On a KDD99 dataset, The model examined the performance of ART-1 and ART-2 with binary inputs and continuous inputs respectively.

8.5.2.2 K-means

It works by separating the data into the number of clusters (k), such that the data present in every cluster is the same, and different clusters have different data. The first cluster centre was chosen at random as K data, then the rest of the data was added to that cluster that has the greatest resemblance based on how far it is from the cluster centre. This approach was repeated until none of the cluster centres changed. As a result, the data was split into K clusters.

Unfortunately, outliers are particularly sensitive to K-means clustering, and there is the possibility of the emptiness of the object collection that lies near to a centroid, preventing the updating of the centroids. For Denial of Service attacks, the K-Means algorithm was devised. To begin with, a strategy to reduce the isolated data and to reduce the noise was improved. A multimodal intrusion detection system (IDS) was recommended. The abuse model for detecting intrusion looked at the data, while the anomaly detection module looked for anything out of the ordinary. The anomaly detection module in this model was built using the unsupervised clustering algorithm that has a higher detection rate for detecting anomalies.

8.5.2.3 Fuzzy C-means (FCM)

A clustering method called fuzzy C-means is a technique that deals in allowing a single data to be contained in more than one cluster. Dunn came up with the idea in 1973 and Bezdek enhanced it in 1981. It's utilised in applications where a hard distribution of data is required but isn't very relevant or difficult to accomplish. The membership of every point is determined by a fuzzy function, and every point contributes in relocating the cluster centroid depending on the membership in the cluster. Further a method for developing a hierarchical neuro-fuzzy inference intrusion detection system model by combining numerous soft computing approaches. The component analysis neural network is employed in this proposed model to minimise feature space dimensions. To extract and handle fuzzy rules, the pre-processed data was clustered using an upgraded fuzzy C-means clustering technique.

8.5.2.4 One-Class Support Vector Machine (OCSVM)

The one-class support vector machine is a highly comprehensive example of a SVM designed to detect anomalies. The difference between the general SVM and the one-class SVM is that the quadratic optimization problem allows only a specific percentage of predetermined outliers, which makes it particularly useful for the identification of anomalies. These outliers lie between the hyperplanes and the origin. All of the data left, landed on the other side of the hyperplane, resulting in only nominal class, which gave rise to the phrase "the one class SVM". The score generated by the SVM is used to depict the distance between the data point under consideration and the ideal hyper-plane. The data values of the one class SVM output that are positive suggest normal behaviour, whereas negative data denote abnormality in behaviour.

8.5.2.5 Unsupervised Niche Clustering (UNC)

UNC is not much prone to the sub-optimal solutions as previous methods because it is based on genetic optimization. The algorithm's main strength is its ability to deal with noise and automatically calculate cluster numbers. For anomaly detection, the fuzzy set theory is coupled with the UNC and used it in detecting intrusions in the network. It was attached to every cluster as a Gaussian-shaped member function and created the radius and centre of the cluster. The normality level of a data sample will be determined by such cluster membership functions.

8.6 Pros and Cons of Supervised and Unsupervised Techniques

Multiple unsupervised anomaly detection techniques increase the effectiveness of intrusion detection systems at all levels, including clustering, feature selection, and classifications. The table below compares the most frequent methods of the distinguished anomaly detection techniques. The comparison maximises the benefits and drawbacks of each strategy. The advantages and disadvantages of supervised and unsupervised approaches are listed in Table 8.1.

8.7 Summary

The many application domains for anomaly detection using machine learning techniques are covered in this chapter. These methods are capable of detecting anomalies with a high degree of precision. Each type of technique, supervised and unsupervised, has been shown to be effective in detecting anomalies in the literature, and there is still a lot of study being done in this area. We've also highlighted the benefits and drawbacks of each category.

Table 8.1 Advantages and disadvantages of supervised and unsupervised approaches

Techniques	Pros	Cons
K-nearest neighbor	For there are fewer predictor variables, it is easier to grasp, and it is useful when developing models with non-standard data types	It necessitates a lot of storage space and is an expensive technology to use. It is highly dependent on the similarity function used to compare the instances
Neural network	It does not require reprogramming	Its operation necessitates training. Because of its unique architecture, it must be imitated. Large neural networks also necessitate a long processing time
Decision tree	It's simple to comprehend and interpret. It just necessitates a little amount of data preparation. It is based on a white box model and is quite durable. While employing decision trees, it is feasible to validate a model using statistical tests	Learners can construct extremely complicated trees that do not adequately generalise the material
Support vector machine	It finds the best hyperplane for separation. It can handle data with a lot of dimensions. When used, it usually works well	It needs a lot of CPU time and requires both positive and negative examples for the data set
Self-organizing map	It is straightforward and simple to comprehend. It has a strong ability to visualise high-dimensional data, making it particularly useful for dimensionality reduction. It can also be used with non-linear data	It takes a long time to complete
K-means	It features a simple design	It is required that the value of k be specified. It is sensitive to noise and data items that are outliers
Fuzzy-C means	It allows data to be organised into many clusters. It's a more realistic picture of gene behaviour	It is necessary to give c, or the number of clusters. It is necessary to calculate the membership cut-off value. The prior assignment of centroids has an effect on clusters

References

Augusteijn MF, Folkert B (2002) Neural network classification and novelty detection. Int J Remote Sens 23(14):2891–2902

Chawla NV, Japkowicz N, Kotcz A (2004) Editorial: special issue on learning from imbalanced data sets. ACM SIGKDD Explor Newsl 6(1):1–6

Hofmeyr SA, Forrest S, Somayaji A (1998) Intrusion detection using sequences of system calls. J Comput Secur 6(3):151–180
Lane T, Brodley CE (1999) Temporal sequence learning and data reduction for anomaly detection. ACM Trans Inf Syst Secur 2(3):295–331
Ming YS (2011) Real time anomaly detection systems for denial of service attacks by weighted k-nearest-neighbor classifiers. Expert Syst Appl 38(4):13492–13498
Moore AW, Zuev D (2005) Internet traffic classification using bayesian analysis techniques. In: Proceedings of the ACM SIGMETRICS international conference on Measurement and modeling of computer systems, New York, United States, June 2005
Mukkamala S, Sung AH, Ribeiro BM (2005) Model selection for Kernel based intrusion detection systems. In: Proceedings of the international conference in Coimbra, Portugal (2005)
Peddabachigari S, Abraham A, Grosan C, Thomas J (2007) Modeling intrusion detection system using hybrid intelligent systems. J Netw Comput ApPl 30(1):114–132
Phoha VV (2002) Internet security dictionary. Springer
Tan PN, Steinbach M, Karpatne A, Kumar V(2005) Introduction to data mining, 2nd edn. Pearson Addison-Wesley
Theiler J, Cai DM (2003) Resampling approach for anomaly detection in multispectral images. In: Proceedings of SPIE 5093, Orlando, Florida, United States 2003
Vapnik VN (1998) Statistical learning theory. Wiley, New York

Chapter 9
Application of AI & ML in 5G Communication

Lipsa Das, Biswa Mohan Sahoo, Ajay Rana, Khushi Dadhich, Smita Sharma, and Suman Avdhesh Yadav

Abstract The world is on the cusp of a new era of communication. Current networks are struggling to keep up with the demand. It is predicted that 5G will be one of the most impactful technologies in the world in the next decade. It will connect a billion people to the internet in the same way that 4G has done today, but with even faster speeds. The application of AI and machine learning in 5G communication will enable the generation of higher speeds, lower latency, and improved reliability. This will enable the next generation of internet services and greater connectivity for users. The next generation of communication networks will require a shift towards artificial intelligence and machine learning to improve the experience for users. This advancement in technology will better understand how to best utilize the limited spectrum and provide a better experience for users. The application of AI and machine learning in 5G communication will enable the generation of higher speeds, lower latency, and improved reliability. This will enable the next generation of internet services and greater connectivity for users.

Keywords Artificial intelligence (AI) · Machine learning (ML) · 5G · Wireless technique

Abbreviations

eMBB	Extreme mobile broadband
eMTC	Massive machine type communication

L. Das (✉) · A. Rana · K. Dadhich · S. Sharma · S. A. Yadav
Amity University, Greater Noida, UP, India
e-mail: lipsaentc9@gmail.com

A. Rana
e-mail: ajay_rana@amity.edu

B. M. Sahoo
Manipal University Jaipur, Jaipur, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

A. Rai et al. (eds.), *Paradigms of Smart and Intelligent Communication, 5G and Beyond*, Transactions on Computer Systems and Networks,
https://doi.org/10.1007/978-981-99-0109-8_9

URLLC	Ultra-reliable low latency communication
QoS	Quality of service
V2V	Vehicle to Vehicle
RAN	Radio access network
FDMA	Frequency-Division Multiple Access
AMPS	Advanced Mobile Phone System
MIMO	Massive multiple-input multiple-output
mmWave	Millimeter wave
SDN	Software defined network
NFV	Network functions visualization
NG-RAN	Next Generation—Radio Access Network
SBA	Service-Based Architecture
SNR	Signal-to-noise
DCNN	Deep convolutional neural network
GSM	Global System for Mobile communication
CDMA	Code-Division Multiple Access
TDMA	Time-Division Multiple Access
VR	Virtual Reality
AR	Augmented Reality
MAE	Multi-access Edge Computing

9.1 Introduction

In this current era, Wireless communication plays a very significant role in the field of various applications such as business, connecting people, entertainment, and health and safety issues. In the future wireless network, the 5G (5th Generation) technology is going to play a vital role where there is a zillion of interconnected devices, individuals, and various processes that will be required to serve simultaneously. The robust development of 5G network technology and its architecture has proven its impact on this massive Internet of Everything ecosystem.

There are various factors like virtualization, cloud technology, and software-centric network function that makes the 5G wireless technology more advanced and user-friendly, but it also brought security challenges and vulnerabilities to the system.

In communication, E2E security i.e., end to end is a very important concern, where against various threats AI & ML can be a great solution in the field of security protocol automation and design which gives higher accuracy and is very effective in various classification, identification, and automation field. Considering the pre-existing traditional security methods, it is very difficult to handle a wide range of threats, for 5G networks' where higher data rates and speed are the key points of selling. Hence AI/ML can play a significant role in the protection of virtualized components of network and data. Hence this chapter gives the idea of various implications & applications of AI/ML in 5G network security (Haider et al. 2007).

With the rapid development of communication technology, it is required to integrate various devices with different frameworks of networks. Also, the high speed and higher data rate of the 5G network and beyond enhances greatly the data communication as well as escalates the data traffic rate and brings various network security challenges. These challenges can bring various cyber-attacks financial damages, blackmailing, infrastructure & social damages, etc. to the network user. Also, various threats to democracy and threatening lives of humans are frequently reported as a news flash. Therefore to have a secure & safe communication network, the Low cost, Self-Organizing networks, various adaptive & intelligent algorithms make AI & ML can be a great solution for these security issues, as it is very difficult to deal with the pre-existing methods (You et al. 2019; Yao et al. 2019). For next-generation networks (including IMT 2020), the ITU has also established a standard Y.3172, which set the architectural model or framework and demands for various use cases of ML.

The softwareization, cloudification & dynamic virtualization of networks are the key concept of the futuristic architecture of the 5G & beyond network (Khan et al. 2019). If the safety protocols are not able to identify & prevent the data bandit's threat & attack in minimum real-time delay, then it will increase various threats to the network security and user data greatly, where the application of AI & ML can be a prevention to the above-said issues. A real-time integrated architecture of the AI/ML engine is shown in Fig. 9.1, where for detection of real-time data bandit's attack, various network parts consisting of data collection points are given to the AI/ML engine.

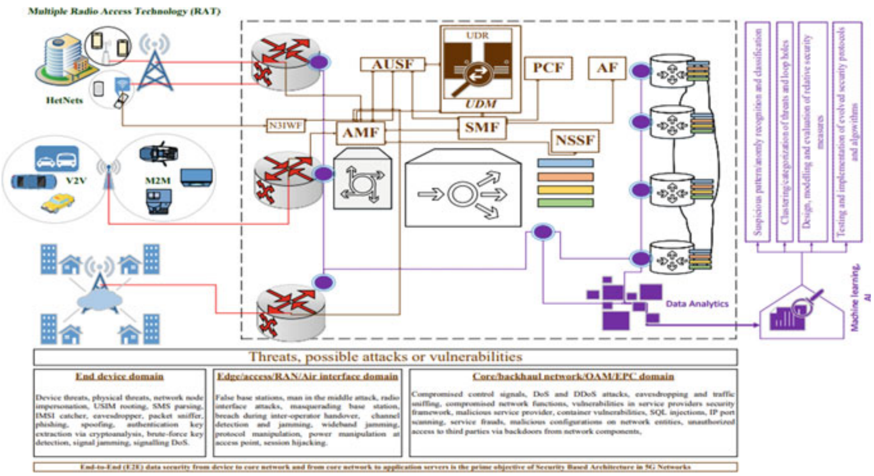


Fig. 9.1 AI & ML application in 5G Network Architecture (Haider et al. 2007)

The chapter organization is divided into six sections, where Sect. 9.1 high lights the evolution from the 1st generation to the 5th generation, 5G wireless technology & 5G network security. Section 9.2 describes the importance of AI & Machine Learning for 5G Wireless Technology. Section 9.3 presents various approaches of AI/ML for 5 g security. Section 9.4 presents a Case Study for finding types of signal modulation using DCNN. Section 9.5 describes the Challenges and Future Scopes for 5G Communications & conclusion.

9.2 Evolution from 1 to 5G

1G: The first generation of mobile phone comes into play in early 1980s and is able to do the analog voice call similar to that of the landline phone. The calling voice quality, low battery life, limited capacity and poor security are the major corner for the users. Also the data speed rate was 2.4 Kbps only. FDMA & AMPS are the key approaching techniques here whereas circuit switching is the transmission technique with no error correction mechanism (Fig. 9.2).

2G: The Second Generation enhanced the mobile service that of the first generation, by introducing a new digital technology Known as GSM, CDMA & TDMA in 1991. This has a data speed rate of approx. up to 64kbps which is enough for sending the SMS, picture messages and MMS.

2.5G: This is the advanced version of the Second Generation where the GSM technology becomes more advanced. With the enhancement of the technology, the data speed rate increased up to 64–144 kbps and with this users are able to do Web browsing.

3G: The third Generation deploys UMTS (universal mobile telecommunication system) framework in 2000. With this enhanced technology, the user is able to do constant video calls. Higher system and download speed make the technology more advanced. It is the first mobile broadband system that was able to provide the voice with some multimedia. Here both circuit switching & high-speed packet access (HSPA/HSPA+) transmission techniques are used to increase the wireless network power, as the 3G uses MIMO (Dangi et al. 2021). This 3G also comes with an error

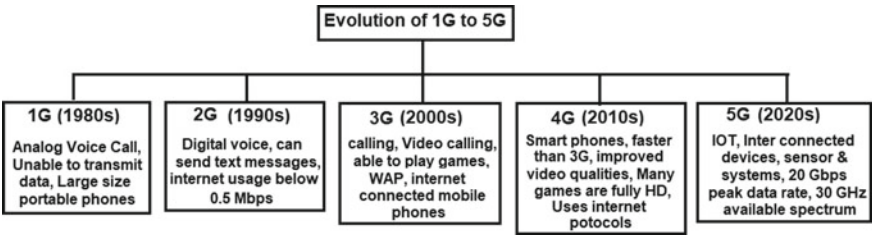


Fig. 9.2 Evolution of 1G to 5G

correction mechanism and turbo codes are used for the same. This has a data speed rate ranging from 384 Kbps to a maximum up to approx. 5Mbp.

4G: The Fourth Generation was developed by IEEE and launched in the year 2010 where LTE, OFDMA, SCFDMA, WiMAX are the access techniques and it uses packet switching as transmission techniques, with a data rate ranging from 20 to 60 Mbps (Al-Namari et al. 2017). It has also a wider bandwidth of up to 100 MHz.

4.5G: The enhanced version of 4G is the Fourth Generation LTE-A, which is faster than 4G. With this 4G's new advanced version multiple antennas can be combined with the receiver and transmitter as it uses MIMO technology, so that multiple signals & antennas can work at a time.

5G: The Fifth Generation launched in the late 2020s gives the multimedia experience as well as ultrafast speed to its users at a reduced cost. High capacity, very fast speed internet connectivity, Ultra-low latency, AR/VR media as well as the large broadcasting data in Gbps etc., makes this technology more advanced than previous ones. It is also very compatible with WWWW (wireless world wide web) (Agiwal et al. 2016). Till date it is the best internet solution for IoT devices i.e., users can connect multiple internet connections at any time anywhere as per their convenience. It is not only highly reliable & scalable but also is very energy efficient and this technology is based on the IPv6 protocol (Buzzi et al. 2016). There are three major services of 5G. Such as:

eMBB: Extreme mobile broadband is the extended service and was introduced by 4G LTE networks. It offers a high data rate across a wide coverage area.

eMTC: **This** is very cost-effective and supports the connection of an extremely high density of online devices with a low power consumption ratio. High data service rate and extended coverage with low complexity for IoT devices are the key features of this Massive machine type communication.

URLLC: This is applied to various essential applications like remote surgery, V2V, industry 4.0 or intelligent transport system, etc. where a guaranteed connection is required with low latency. In comparison with the traditional architecture of the mobile network, URLLC provides best quality of service, better reliability and ultra-low latency (Dangi et al. 2021).

5G is the next generation of wireless internet connectivity and is further divided into 6 GHz 5G and Millimeter-wave 5G.

6 GHz spectrum is a mid-frequency spectrum that works as a midpoint between capacity and coverage to offer the perfect environment for 5G connectivity. This spectrum band is primarily used for higher-capacity cellular services. In the absence of a mid-band spectrum to reduce the densification of the network, it offers a continuous channel, which makes 5G connectivity affordable at anytime, anywhere for everyone.

The mmWave in 5G wireless technology high-performance & high-speed communication network having ultra-wide bandwidth, offers a great range of services in 5G deployment planning.

9.3 5th Generation Wireless Network Technology

The cellular technology is greatly enhanced with the application of AI, ML & deep learning, which makes the 5th Generation wireless technology a more heterogeneous network and is enabled with high speed, low latency and the best quality of service. Figure 9.3 shows such a type of 5G network architecture where the 5G network is connected with all available devices. Also, there are various typical applications in the field of mixed reality, automation, autonomous vehicles and critical delay-sensitive applications, etc. (Tanveer et al. 2021; Jagannath et al. 2019). This new generation of technology is able to learn from the behavior of network traffic and come over various real-time challenges and this makes this technology a one step solution in many real-time applications like surveillance of video, spam detection, weather prediction, etc. (Zikria et al. 2020). ML-based 5G wireless communication is achievable now by using different modern networking standards; i.e., big data analysis (Qiu et al. 2016), mobile edge caching (Hou et al. 2018), mobile edge computing (Wang et al. 2020), and context-aware networking (Qin et al. 2018). The existing heuristic radio resource management (RRM) algorithms cannot oppose the essential demands of 5G due to their compound procedure (Akhtar et al. 2021). However, ML works astonishingly well from traditional approaches to compound problems that need great human interaction.

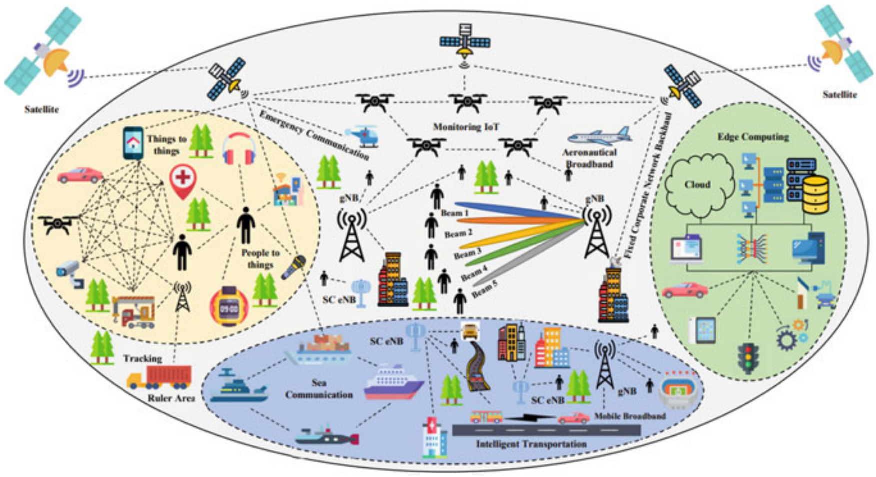


Fig. 9.3 5G network architecture (Tanveer et al. 2021)

9.4 5G Wireless Networks Security

There are various advantages of the next-generation wireless network such as low battery life, wide range coverage (100%), ultra-low latency (1 ms), a very large no of well-connected devices, 1000 × bandwidth/area, and a very low network energy usage make this 5G wireless network a very dynamic network. To fulfill the performance requirement, technologies like MIMO, mmwave, SDN, networking slicing & NFV are applied. This new enhancement of the technology also brings many challenges and threats to the privacy and security issues for the network and its user. There are number of challenges or threats to the network security of 5G wireless techniques, out of them there are 17 major key challenges/threats specified by 3rd Generation Partnership Project (3GPP) Technical Specifications Group Services & Systems Aspects (TSG SA3) in its Release 14 with their possible solutions. Various procedures architecture for security purpose to enhance the Mobile Broadband network for the 5G wireless techniques were in Release 15 (R15) in June 2019 (3GPP 2019) and the security features of forthcoming R16 and R17 are specifically focused on eMTC and URLLC to ensure the E2E security with higher-layer security protocols to support security for SBA in 5G technology. In this 5G end-to-end security architecture, specify the procedure and requirements for RAN secure connection to the end device. Hence it can be said that Network Domain Security can be defined as the combination of data protection with privacy from the access network to the core network and beyond. Figure 9.1 shows threats, possible attacks and the vulnerabilities of the 5G wireless Network framework.

The combination of AI/ML and 5G wireless technology make the future network highly efficient and more dynamic and adaptive where data of the user travel through various network layers and slices. According to the kind of challenges, there is a requirement for different security applications for different network slices & layers. It is very difficult to manage these security issues with pre-existing methods and requires security automation for the overall network. For network security, the premier industry partners are now making their plans to leverage AI/ML for 5G and beyond wireless networks (Haider et al. 2007). Figure 9.4 shows, various uses and applications of AI/ML for the next-generation 5G wireless networks.

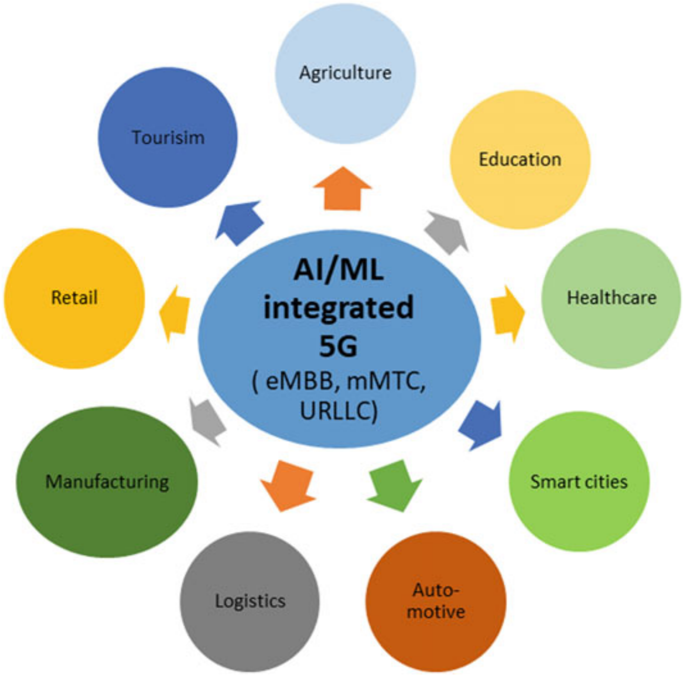


Fig. 9.4 Applications of artificial intelligence and machine learning in 5G network

9.5 Impact of AI/ML in 5G Wireless Network Technology

ML is part of AI that uses past data to make the machine learn automatically. With the help of various algorithms and available existing data to construct mathematical models, various parameters such as accuracy, reliability, latency, speed time, and power consumption of next-generation wireless technology, can be escalated greatly with the machine learning-based algorithms.

Many of the most important advances in artificial intelligence were made using what's known as traditional machine learning. This involves programming a computer to perform a specific task by feeding it vast amounts of data. Most of today's advanced AI applications are based on deep learning, which is a subset of traditional machine learning where the algorithms used are based on artificial neural networks. Unlike traditional machine learning algorithms, which require human input to improve performance, deep learning algorithms are able to learn on their own and can perform complex tasks with little or no human input. The data interpretation can be given by various algorithms used at different levels.

The applications of deep learning in security and privacy have expanded considerably in recent years. DL machines have helped identify and mitigate threats to cyber security, identify cyber-criminals, predict fraud, and provide new ways for customers to authenticate their identities.

The use of AI and ML in the field of security and privacy is not new, but they got more attention for their performance with the development of Deep Learning algorithms. Earlier all the methods were robust in nature and used to attack in a similar pattern, but after the progression of Deep Learning algorithms methods with more dynamic features such as AI and ML came into use which were able to prevent advanced threats and attacks. Because of more sophisticated methods used by attackers such as impersonation and obfuscation to attack and prevent any kind of detection. AI/ML can be used to detect threats from packet capturing to analysis to big data insights, which conventional techniques are not able to detect. The pattern-based learning at the core supported by softwareization and virtualization provides agility and robustness to timely counter threats and attacks. AI provides a positive influence on the field of Information Security. Its algorithms are acquired to address security and privacy issues. The industry of Information Security is creating more data so that AI can work as a strong antidote to prevent all the threats generated from the data. The first generation of Artificial Intelligence is developing towards detecting threats, assisting humans and scrutinizing data. Whereas the second generation of Artificial Intelligence is aiming toward forming an independent system that will solve all the issues except which require critical support and are handled by humans (Lee and Kim 2017).

From personal device to the service provider network the privacy issues and threat have increased in 5G networks because of expanded bandwidth and high data rate. Thus with the use of AI and ML techniques within the network, they are made smart enough to detect several network issues and provide solutions to these problems in real-time. To protect the personal devices that are connected to the internet, AI and ML techniques are used to provide security solutions that can handle various network issues, attacks and threats. To detect the threat and treat them with modifiable security algorithms can be included in short to medium-term plans for AI and ML whereas for the long-term plan, an automatic security mechanism that can be used to timely detect threats and attacks can be included.

5G network can support a large number of connected devices and networks as compared to the network which was used before the evolution of 5G. For example, 5G can support smart cities, smart vehicles, etc. Similarly, the introduction of IoT in 5G network provides more adaptive solutions to the security problems which occur at both the network and device sides. The security of such networks becomes much more complicated because of outside as well as local intrusion. Artificial Intelligence and Machine Learning provide the solution by classifying security links for assurance, instance, identity and authentication. The implementation of IoT in 5G provides security and privacy which will cover all the areas such as privacy, E2E protection and identity protection. For example, the key authentication framework builds an end-device to core the network and on-ward to the service provider, whereas hiding the key identifier is still a complex problem. AI and ML play an important role in the purpose of authentication to reduce attacks.

Providing security and privacy to the data with the help of different systems which has various security requirements is a difficult job. AI and ML can help to expand the mechanism of security by creating trust models, device security and data assurance to detect the problem in real-time and categorize and cluster unexpected threats for the complete IoT-5G network. This will result in less workforce skill shortage in the information security industry.

9.6 Role of AI on 5G Networks

Many internal complicacies are faced while adopting 5G networks and the best way to deal with these complicacies is by introducing Artificial Intelligence into networks. Ericsson said that they are going to combine Artificial Intelligence into 5G networks by 2020. The main aim of the combination of AI into 5G networks was to minimize building new revenue streams and optimize network performance. According to the survey, 64% of decision-makers will work on AI implementation on the network to improve its performance and 70% of responders stated that implementation of AI in the network was one of the best methods to recover the investments which are made on switching networks from 4 to 5G and 55% of responder stated that AI has improved the customer service by improving network quality and providing personalized services. Product life cycles, revenue, SLAs and networks are some other areas where AI is aimed to be implemented.

Implementation of AI into 5G networks brings up some challenges at the same time. Effectual mechanisms for collecting, structuring, and analyzing the numerous volumes of data accumulated by AI must progress. For that reason, early AI adopters who obtain solutions to these challenges will turn up as the clear frontrunners as 5G networks become connected.

9.6.1 Relevance of 5G to the Field of AI

Since the 1990s smartphones have become more and more smaller and slimmer, but the core algorithms have been the same since then, they have not advanced, which results in more power consumption and lower data rates in the 5G networks. To improve performance and reduce power consumption, conventional wireless algorithms are replaced by deep learning algorithms and AI.

Bandwidth which is used by recent cellular networks works on the radio spectrum. The electromagnetic waves (EM Waves) between the frequency range of about 10^4 and 10^{11} or 10^{12} Hz of the radio spectrum are known as radio waves. These waves are widely used in different technologies such as telecommunication. National Laws manage interference between various radio wave users. Whereas International Telecommunication Union looks after the implementation of these laws. The increasing use of wireless technologies is one of the major concerns that will overcrowd the use of airwaves which are used in devices so that they can communicate with one another. To solve this issue one method was introduced so that devices that don't broadcast on the same frequency can be developed for communication. AI algorithms would then be used to find available frequencies by enabling intelligent awareness of RF activity that was not previously feasible.

5G provides faster speed, as it is 20 times faster than 4G. 5G speeds permit developers to form applications that make full use of improved response time, including video transmission for security purposes and sporting events, due to its low latency. Also, 5G networks permit more access to real-time from various solutions. For less power consumption 5G network uses IoT sensors. This allows checking equipment conditions that change in factories and remote detection of farming irrigation levels. With the help of IoT sensors, doctors can access patient data more effectively. To apply all these applications use of AI is required.

These AI technologies have transformed the industries completely and provide the amount of value all this connectivity provides consumers by integrating big data, IoT and AI.

As 5G architecture supports AI processing, this will change the future of AI. The integration and speed of other technologies will increase because of 5G, whereas AI allows to build the systems and machines that work similarly to human intelligence. We can conclude that 5G improves the speed of the services on the cloud and AI analyses the data.

Without AI a 5G network is not completed. The combination of AI and ML in the network can be done with the use of 5G networks. It allows connection with various IoT devices which creates a large amount of data that is processed by AI and ML.

With the integration of MEC and AI/ML the wireless service provider can avail to its user:

- High automation levels from the distributed ML and AI architecture at the network edge
- Application-based traffic steering across access networks
- Dynamic network slicing to address different scenarios with varying quality of service (QoS) requirements.

9.7 5G Security: AI/ML Applications

AI and ML are data-hungry in nature which when applied data basically train the model for effective functioning. Using the application of Artificial Intelligence and Machine Learning, now it is very easy to access, maintain and analyze networks to prevent attacks and threats at a very low cost of computing and infrastructure. In the current era of 5G, it is not difficult to generate, maintain and store data because of high data growth, data sources and high computational power. Figure 9.3 given below represents all the different applications of Artificial Intelligence and Machine Learning in the field of network security. By examining the activity patterns and parameters of the network AI and ML models are able to detect suspicious activities. The model of Clustering Algorithms is used to detect various types of loopholes and threats which are generated in network security. With the help of the model of

Classification Algorithms oddities can be determined by monitoring network parameters such as error logs and throughput. Statistical Inference attacks and Generative Adversarial Networks generate a duplicate set of data to evaluate the security of the network as well as it implements and test security algorithms.

The researchers have seen extraordinary development in secure computation, privacy, federated learning and encryption while forming private AI and ML models. Hybrid models of AI and ML are formed by adopting various techniques from different fields to make faster, generic and efficient models. The differential privacy started by Google security and its privacy team is one of the most used examples (Hassan et al. 2019). Major new improvement has been seen in the field of secure computation by stocking up various protocols for fast computation (Riazi et al. 2018). Some of the most used examples for secure computation with homo-morphic encryption are TAPAS, Gazelle and Faster CryptoNets. Slalom, Chiron and Ekiden are the Artificial Intelligence and Machine learning-based models used for secure enclaves and federated learning. SecureNN is a Machine Learning based solution that is used for the purpose of comparison-based operation of neural networks for extraction and secret sharing (Buzzi et al. 2016).

To prevent unidentified attacks, the progression of robust and generic detection algorithms is one of the recent trends (Tanveer et al. 2021). Artificial Intelligence and Machine Learning-based technologies are used to deal with various applications such as fraud detection systems, spam filters, antivirus scanner systems, etc. These methods mostly work on the data which is produced by host processes, network traffic, etc. To assist humans for detecting anonymous activities, unsupervised algorithms such as neural networks and clustering are used.

A larger number of threats are generated by 5G and other networks based on third-party servers, SBA and independent decentralized networks in the form of cyber-attacks and Denial of Service. Thus to provide the best protection to these components and the whole system devoted agents from different network fields can be used.

To deal with decentralized networks various Artificial Intelligence and Machine Learning based solutions have been suggested (Jagannath et al. 2019). To deal with such attacks AI and ML-based solutions are using different Reinforcement Learning and Deep Reinforcement based techniques (Zikria et al. 2020). To deal with jamming attacks in which hackers usually seize up the Radio Frequency Signals, Deep Reinforcement Learning based solutions are provided which choose suitable frequency channels and avoid such jamming attacks using optimal policy based on previous observations. Cyber-physical attacks basically occur on autonomous systems like smart vehicles which manipulate the data to get control over the system (Qiu et al. 2016). Deep Reinforcement Learning based algorithms are used to deal with such attacks by providing the power to autonomous systems to learn from time-varying observations so that they can produce some excellent action so that the system can become more robust and dynamic. Deep Reinforcement Learning based systems have shown extraordinary progress in connectivity preservation among robots to support systematic communication.

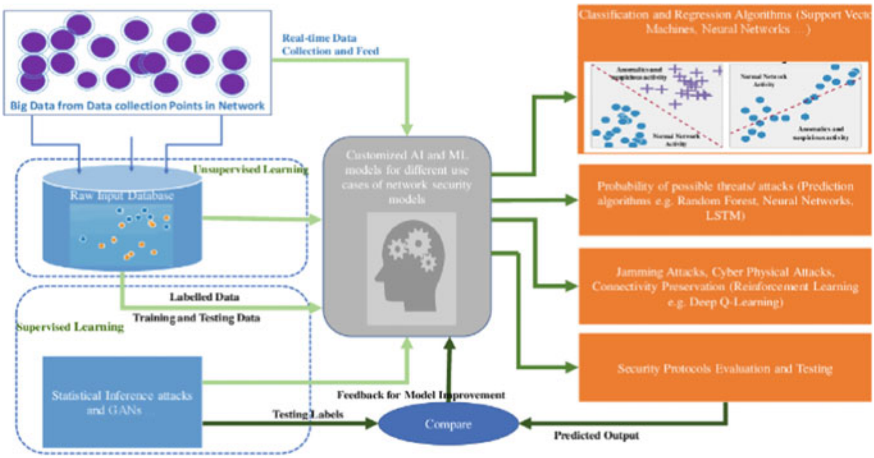


Fig. 9.5 Different application scenarios and use cases of AI/ML assisted network security (Haider et al. 2007)

Deep learning also gives some benefit to cyber security solutions as it is capable to learn patterns from previous entries automatically to prevent any kind of future intrusion and it can also identify irregular patterns. DL is established in various securities such as software-level security, infrastructure-level security and user-level security (Zhang et al. 2019).

Various developments in Deep Learning networks like auto-encoders, convolutional Neural Network and data networks are implemented in several security applications such as flooding, cyber-attacks, malware detection, DoS probing, etc. These networks are used to differentiate harmful applications of unknown traffic and spam at the software level.

9.8 Machine Learning for 5G Technology: A Case Study

9.8.1 Deep Convolutional Neural Networks Application to Detect Signal Modulation Types

The network standard for the fifth generation is described by 3GPP. One of the biggest attractions of 5G is the blazing-fast data speeds it offers, which are faster than those of past generations. 5G offers many more than that. Most notably, it offers greater connectivity, which allows users to connect to the internet anywhere, without the need for a data connection. This will enable many new uses for the internet, such as remote education and healthcare, and will help to solve many of today's biggest problems, such as slow internet speeds and data caps. The new 5G standard has been developed as an SBA and defines an NG-RAN, which is much faster than 4G.

It addresses the latency requirements of the most demanding applications, such as the feeling of being in the same room with someone, without the lag of traditional cellular networks. 5G will enable the deployment of tens of thousands of small cell sites in a single area, and millions of small cell and traditional large cell sites across a city or region, providing vastly improved coverage, capacity, and speed.

The latest technologies have been developed and are being upgraded to make three primary services such as: mMTC, eMBB and URLLC which offer faster, more efficient ways to communicate. They allow users to transmit data over longer distances, to connect to the Internet from areas that lack connectivity, and to experience faster response times when communicating.

The new technologies are being used to build better experiences for users today. One example is the use of Control and User Plane Separation (CUPS) which allows for the separation of the control plane and user plane in the network to improve performance and reliability. This technology is already used in live scenarios such as the use of mmWave which is correlated with the huge MIMO i.e., correlated with Spatial Diversity, Spatial Multiplexing and Beamforming also enabling network slicing which is used to improve network performance.

The important characteristics of the 5G wireless network system is the integration of AI/ML into the network. Considering the SBA's core network data analytic function of 5G and User device, avail the computational resources for running the ML algorithms. There is a vast possibility that has been raised in Telecom network (O-RAN specified RIC — RAN Intelligent Controller for Non-real-time and Near-real-time) with the use of machine learning and it is possible by the utilization of MEC (Multi-access Edge Computing) within 5G standardization.

Here we will discuss, how NG-RAN offers a platform for end-to-end deep learning applications to recognize and classify the modulated RF signals and accordingly modify the parameters of the channel, such that both ends of the communication accomplish optimal, effective resource usage.

9.8.2 Modulation Recognition

In telecommunications environment, signal modulation is done for transmitting a signal where differing characteristics of the carrier signal's periodic waveform is changed with a separate signal which contains the information that is to be transmitted. In this modulation process for sending the information to the desired destination, it is required to integrate the modulation signal with a high frequency carrier waveform or signal. For instance, the modulation signal might sound from a microphone, a video signal representing moving images from a camera, or a digital signal representing a sequence of binary digits.

Non-cooperative communication systems are those where the transmitter and receiver are not in the same location and have no means of directly exchanging information. The signals that are being sent between the two parties are being intercepted by other parties, and signals can be intercepted in any direction. The signals are sent using a variety of modulation types, which consist of different methods of transmitting signals. The categorization of modulation is an intermediate procedure that generally takes place between signal detection and the demodulation of signals which happen at the receiver. For the extraction of the signal which is received, the Intelligent Radio receiver must have information about the modulation type so that it can assure a successful transmission of the signals. But the major challenge is that if the intelligent radio receiver doesn't have any information about the data which is transmitted, for example, signal amplitude or frequency then the recognition of the modulation type becomes more difficult. To solve this problem AMC is used. The main purpose of AMC is to manufacture smart receivers which can find the modulation type by just analyzing the signal without providing any further details about the modulation. To recognize the modulation type, this "blind" effect is very effective, as it can reduce the information which is overloaded and improve the process of transmission.

It can be said that the classification of signal modulation has become an important topic with the continuous enhancement of 5G wireless technology. The ability to identify the modulation type automatically for the signal which is received requires a lot of civil and military techniques. Adaptive communication and cognitive radio are examples of such techniques.

Since the last two decades, various algorithms have been developed to solve the issue faced by the automatic modulation classification to complete the task. Various efforts have been done to evolve different feature-based methods to solve the purpose of differentiating modulation types. Automatic Modulation Classification basically consists of two steps: the first one is used to feature extraction whereas the second one is used for classifier training where the process or removal of features from signals is primary. For example, last recent work has proposed to mine features from the signal's frequency and phase in the time domain.

When a few of these features are formed, then the classifier is ordered to complete the job of modulation classification. For this task, any classifier can be chosen, varying from a linear to a non-linear model. Decision tree and support vector machine can be example of such models. Better the feature will be more good will be the performance of the modulation classifier for the FB method.

In the past, computer scientists and deep learning experts focused on 2 approaches to represent vision in computers: neural networks and convolutional networks. The latest developments in deep learning (deep neural networks) have shown that technology can apply as modulation classifiers which accelerate the performance of classification for better classic solutions.

Because of the sheer size of the dataset, we had to carefully think through how to scale our model. If we were to build a network at the size of the entire dataset, we'd need way too many parameters to initialize, and training would take much too long. Instead, we built a neural network in layers. The major benefit of deep learning models it can learn by themselves, almost all relevant traits for the assigned task which results to be a finer classifier. It can classify the received signal from the information received without considering the feature extraction module.

9.9 Modulation Classifier Consideration & Model Architecture

Daitan Innovation Team, November 13, 2021, created a modulation classifier to predict the modulation signals. They use PyTorch Lightning to create a deep learning model (end-to-end) and used synthetic data of GNU radio ML RML2016.10a dataset for their research work. This data set contains 3-analog and 8-digital modulation scheme with different SNRs that includes 220,000 input examples and here they contain the digital modulations such as BPSK,16QAM, 64QAM, QPSK, 8PSK,BFSK, PAM4 and CPFSK for their work and apply probability distribution for the 8-way classification problem of complex time-series representation of the received data. 5G modulated signal models are: π/2-BPSK, QPSK, 16QAM, 64QAM, and 256QAM (Referring to recommendation of 3GPP R15 protocol).

Here the architecture is based on CNN and has a sequence of the 3-Conv block (2D Conv layer) with a 1-dense block followed by ReLU non-linearity, batch normalization, and Dropout. The batch size for this model is of the input signal is (1, 2, 128).

The three convolutional blocks modify the input data is modified into their respective feature volumes with channels 64, 128, and 256 and then the representation is flattened which generates a feature vector of shape (Batch Size, 10,240, 256). Then it is passed to the dense block (with 256 neurons). Ultimately, a linear layer maps the output representation to a probability distribution atop the 8 modulation classes. The deep modulation classifier is instructed with Stochastic Gradient Descent optimization (https://daitan.com/innovation/machine-learning-for-5g-technology-a-case-study/).

9.10 Results Analysis

When the data is well-established around the SNR and around the whole classes, the categorization of the accurate score is a perfect way to analyze the performance of the classifier. To give the command to deep modulation classifier, researchers have developed a data set with 129,200 observations, where the validation set size is 6800 with 24,000 records for the test set.

Each information which is saved in the datasets has a total of 128 samples. Each dataset is distinguished in such a way so that every one of them has the same number of observation across various signal-to-noise ratios varying around −10 dB to + 20 dB. Remember that the validation dataset is the least used dataset for model hyper parameters tuning. Afterwards getting admissible score for the hyper parameters, the validation set is also comprised in the training data, which makes the size of the training data as 136,000.

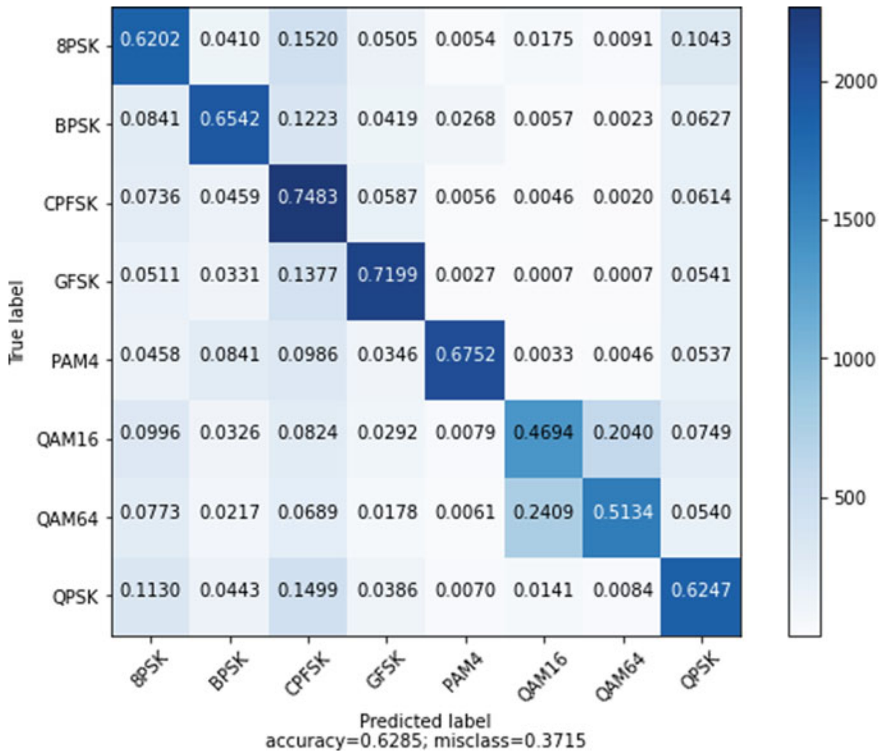


Fig. 9.6 Confusion matrix (https://daitan.com/innovation/machine-learning-for-5g-technology-a-case-study/)

Around 62.8% of accuracy can be achieved from training the models which are comparatively lighter in weight for the eight modulations. Precision in the values can be achieved aby analyzing various modulations in the confusion matrix which is mentioned below. ConvNet can be used for the completion of task of classification of eight types of modulation which comparatively has less convolutional neural network. AlexNet which is used as previously used for modulation classification can be seen to get a view of the ConvNet model, which include 62.4 million training parameters (https://daitan.com/innovation/machine-learning-for-5g-technology-a-case-study/).

It can be seen that for different SNR values, if the evaluation protocol breaks down, then with low-SNR, the above said classifier results are comparatively low (Haider et al. 2007).

Considering the comprehensive evaluation of the above classifier on the basis of various SNR values (https://daitan.com/innovation/machine-learning-for-5g-technology-a-case-study/) it is observed that the confusion tables with low-SNRs are disorganized whereas the confusion matrices with SNR $> = -4.0$ dB are very much apparently showing a higher result.

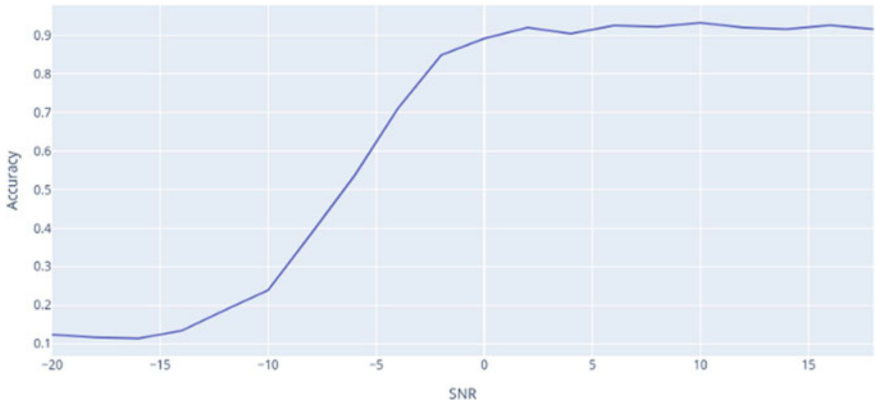


Fig. 9.7 Classifier performance versus SNR (https://daitan.com/innovation/machine-learning-for-5g-technology-a-case-study/)

Finally, it can be said that, the model based on the application of Deep learning have a robust impact for 5G wireless technology applications. Various types of modulation schemes with changing SNR ratios can be identified from the raw information or signal by using a modulation classifier based on a machine learning application with the pre-existing model. For example, we can implement a convolutional neural network architecture that has been used for the recognition of signals to detect the presence of noise in a signal. In such use case, we can solve the problem of segregating eight different modulations that are generally used in 5G wireless technology. Representations can be learned which can brief the modulations aside from relevant accuracy, even with a comparatively light model. With the high SNR, this classifier can achieve very high accuracy scores.

9.11 Challenges and Future of 5G Wireless Technology

With the increase in the advancement of 5G network the things are becoming more and more complicated these days. Nowadays, the different innovative ideas which are directly connected with the use of 5G communication are Virtual Reality, Autonomous cars, and Industrial Automation.

In other words it can be said that in the near future Machine Learning will turn into such an important aspect in building of new 5G related projects. If we talk about the new technologies, then there can be two sides of the coin i.e. Positive Impacts as well as Negative impacts of a particular thing.

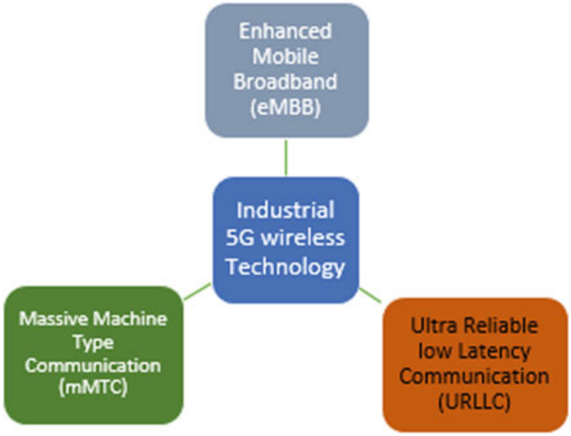


Fig. 9.8 The future industrial 5G wireless technology

9.11.1 ML Servies for 5G Wireless Communications Include

- **eMBB**: It provides a faster data transfer rate for latest application which is required for a uniform coverage area. For example ultra-high-definition video streaming and virtual reality.
- **mMTC**: An important feature of 5G communication services are the measurable connectivity requirement for increasing the actual value of wireless devices as well as transmission speed of minor capacity of data over increased coverage areas. An Application that consists of body area networks, smart homes, IoT, and drone delivery builds up this type of traffic. It should always be capable to keep the latest as well as yet unexpected uses.
- **URLLC**: Remote surgery, vehicle-to-vehicle (V2V) communications, high-speed train connectivity, mission-critical applications, connected healthcare and other industrial applications that will compute accuracy, Low latency, and flexibility over data rates.

9.11.2 Challenges for ML Application in 5G Technology

- **Data**: An Excellent quality of data is very important for Machine Learning apps as well as the category of data (unlabeled or labelled) is an essential parameter when it comes to determine the variety of learning to be implemented. Machine Learning is the one and only excellent source because of the data it gets.
- **No Free Lunch theorem**: Here in this postulate whatever data which is generated is averaged, every Machine Learning algorithm contains the same capability in case of deduced unseen data. It directly indicates one thing that the objective of Machine Learning is not only to explore the finest learning algorithm, besides

this it helps understand a particular 5G app which consists of a Machine Learning algorithm and has the finest performance in a specific data.

- **Hyperparameters selection**: Hyperparameters are the values that are fixed in the Machine Learning algorithm before training begins. These values should be chosen cautiously as they manipulate the ultimate factor which can be modified by the learning outcomes.
- **Interpretability vs accuracy**: According to the stakeholder's point of view, the complicated connection between self-sufficient variables that are quite complex to think about and perhaps it should not regularly create a business or financial sense. Thus, a tradeoff might be created between interpreting data and total precision.
- **Privacy and security**: Machine Learning algos can be a concern for legalistic attacks, such as adjustment of input case which in some-way forces the miniature to generalize into dissimilar groups by the original type.

The main objective of a completely automated cybernetic digital protection system will become an endless target, but the actual amount in merging Artificial Intelligence and Machine Learning for the present as well as for the upcoming futuristic systems will also require meticulous and ultra-careful inspection. In addition to this, they also described that nowadays cyber-hackers are taking more and more advantages of Artificial Intelligence and Machine Learning by using small algos for attacking as well as exposed leveraged.

As the studies are still going on for the usage of risk-free, intelligent and heavy-duty combination in Machine Learning as well as in Artificial Intelligence, few essential opportunities that need a very evaluative and analytic exploration.

Let's suppose, that when understanding and researching anomalies in a particular network, firstly, the evaluation and definition of normal should be clear. Network activities are rarely usual. Thus, a fully supervised or semi-supervised network would become a highly feasible network path for handling such cases. Machine Learning miniatures use big portions of data to understand and prepare a design for regression in hidden information. In conditions when modification of significant advancement of network parameters takes place or even when some differences are caught, then also the network use to get collapsed in distribution and thus, network-related retrains are needed. Mostly Machine Learning methods include DL black box in its hidden layers, as well as the observation of its statement, is restricted in nature. Data Privacy is such an important and confidential matter as Algorithms of Machine Learning and Artificial Intelligence use to get a feed of data. The usage of data inside Machine Learning increases the possibilities of attacks because the replicas used to get trained by the data which is used for data mining causes as well.

Security systems based on Machine Learning solutions every time exist at risk to new categories of practical and experienced attacks like GANs. Studies and research have suggested the testing of these risky Machine Learning entrenched security models use artificial GANs.

9.12 Summary

This chapter describes about the working of AI-assisted technological work, structures and apps for security and privacy issues of 5G and beyond wireless networks. The service oriented, component-base networks, architecture, distributed network functions and verification over 100's and 1000's of servers in 5G and Above all this, now the requirement of networks are comparatively sturdy, active and completely automated security architecture. Such architectures are made by using smart and efficient Artificial Intelligence techniques.

Artificial Intelligence is a very important aspect in the enhancement and development in safety and reliability for scattered ad-hoc setups around the globe for network infrastructure and providing complicated and diverse network-related functions.

At this period of time, a semi-automated security structure is more worthy and acceptable although with the persisting progression the AI technologies and suitable research of secure and protected usage of will determine the end game of absolute automation. Considerable studies have been found to mark problems and challenges prior to the Artificial Intelligence taking complete cover of digital automation.

References

3GPP (2019) Security architecture and procedures for 5g system. Technical specifications, 3rd Generation Partnership Project

Agiwal M, Roy A, Saxena N (2016) Next generation 5G wireless networks: a comprehensive survey. IEEE Commun Surv 18:1617–1655

Akhtar T, Tselios C, Politis I (2021) Radio resource management: approaches and implementations from 4G to 5G and beyond. Wirel Netw 27:693–734

Al-Namari MA, Mansoor AM, Idris MYI (2017) A brief survey on 5G wireless mobile network. Int J Adv Comput Sci Appl 8:52–59

Bhuyan MH, Bhattacharyya DK, Kalita JK (2013) Network anomaly detection: methods, systems and tools. IEEE Commun Surv Tutor 16(1):303–336

Buzzi S, Chih-Lin I, Klein TE, Poor HV, Yang C, Zappone A (2016) A survey of energy-efficient techniques for 5G networks and challenges ahead. IEEE J Sel Areas Commun 34:697–709

Dai Y, Xu D, Maharjan S, Chen Z, He Q, Zhang Y (2019) Blockchain and deep reinforcement learning empowered intelligent 5g beyond. IEEE Netw 33(3):10–17

Dangi R, Lalwani P, Choudhary G, You I, Pau G (2021) Study and investigation on 5G technology: a systematic review. Sensors (Basel) 22(1):26 https://doi.org/10.3390/s22010026. PMID: 35009569; PMCID: PMC8747744

Ferdowsi A, Challita U, Saad W, Mandayam NB (2018) Robust deep reinforcement learning for security and safety in autonomous vehicle systems. In: 2018 21st international conference on intelligent transportation systems (ITSC), pp 307–312, IEEE

Haider N, Baig Z, Imran M (2020) Artificial intelligence and machine learning in 5G network security: opportunities, advantages, and future research trends. arXiv:2007.04490v1

Hassan MU, Rehmani MH, Chen J (2019) Differential privacy techniques for cyber-physical systems: a survey. IEEE Commun Surv Tutor 1–1

Hou T, Feng G, Qin S, Jiang W (2018) Proactive content caching by exploiting transfer learning for mobile edge computing. Int J Commun Syst 31:e3706

https://daitan.com/innovation/machine-learning-for-5g-technology-a-case-study/

Jagannath J, Polosky N, Jagannath A, Restuccia F, Melodia T (2019) Machine learning for wireless communications in the Internet of Things: a comprehensive survey. Ad Hoc Netw 93:101913

Khan R, Kumar P, Jayakody DNK, Liyanage M (2019) A survey on security and privacy of 5g technologies: potential solutions, recent advancements and future directions. IEEE Commun Surv Tutor

Lee J-H, Kim H (2017) Security and privacy challenges in the internet of things (security and privacy matters). IEEE Consum Electron Mag 6(3):134–136

Luong NC, Hoang DT, Gong S, Niyato D, Wang P, Liang Y-C, Kim DI (2019) Applications of deep reinforcement learning in communications and networking: a survey. IEEE Commun Surv Tutor

Qin M, Yang Q, Cheng N, Zhou H, Rao RR, Shen X (2018) Machine learning aided context-aware self-healing management for ultra dense networks with QoS provisions. IEEE Trans Veh Technol 67:12339–12351

Qiu J, Wu Q, Ding G, Xu Y, Feng S (2016) A survey of machine learning for big data processing. EURASIP J Adv Signal Process 2016:1–16

Riazi MS, Weinert C, Tkachenko O, Songhori EM, Schneider T, Koushanfar F (2018) Chameleon: a hybrid secure computation framework for machine learning applications. In: Proceedings of 2018 on Asia conference on computer and communications security, ASIACCS '18, pp 707–721, ACM, New York, NY, USA

Tanuwidjaja HC, Choi R, Kim K (2019) A survey on deep learning techniques for privacy-preserving. In: International conference on machine learning for cyber security, pp 29–46, Springer

Tanveer J, Haider A, Ali R, Kim A (2021) Machine learning for physical layer in 5G and beyond wireless networks: a survey. Electronics (IF 2.397). https://doi.org/10.3390/electronics11010121

Wang S, Chen M, Liu X, Yin C, Cui S, Poor HV (2020) A machine learning approach for task and resource allocation in mobile-edge computing-based networks. IEEE Internet Things J 8:1358–1372

Yao M, Sohul M, Marojevic V, Reed JH (2019) Artificial intelligence defined 5g radio access networks. IEEE Commun Mag 57(3):14–20

You X, Zhang C, Tan X, Jin S, Wu H (2019) Ai for 5g: research directions and paradigms. Sci China Inf Sci 62(2):21301

Zhang C, Patras P, Haddadi H (2019) Deep learning in mobile and wireless networking: a survey. IEEE Commun Surv Tutor 21:2224–2287, thirdquarter 2019

Zikria YB, Afzal MK, Kim SW, Marin A, Guizani M (2020) Deep learning for intelligent IoT: opportunities, challenges and solutions. Comput Commun 164:50–53

Chapter 10
Software Defined Network-Based Management Architecture for 5G Network

Catherine Nayer Tadros, Bassem Mokhtar, and Mohamed R. M. Rizk

Abstract In today's world and surrounding cyber-physical systems, the most adopted communication devices are Internet-enabled devices and are embedded with control functions for gaining various self-capabilities. The Internet of Things (IoT) is emerged to address and manage the challenges of connecting and communicating with a massive volume of Internet-enabled devices running various services and applications. These devices can collect data from a set of sensors monitoring physical objects and they are embedded with actuators controlling the machine systems. These quick developments were the reason for the rise of the 5G mobile network to support these large data transfers. Moreover, due to the rapid evolution in the communications technology, dynamicity of IoT networks, and the emerging usage of many heterogeneous IoT devices together, the control and management architectures adopted by the traditional Internet are not able to efficiently and effectively manage networking operations and required QoS of running services and applications. Accordingly, this raises the need to have the Software Defined Network (SDN) paradigm. SDN definition is based on the dissociation of the control plane from the data plane. This separation is the reason for ensuring flexibility and simplification of network management through control plane programmability. The control layer is often comprised of a centralized unit known as an SDN controller, which is regarded as the network's brain since it holds all of the network's policies and regulations. In this chapter, the control layer is designed and studied according to the number of controllers used, their services support, and their cooperation to provide efficient

C. N. Tadros (✉) · B. Mokhtar · M. R. M. Rizk
Department of Electrical Engineering, Alexandria University, Alexandria, Egypt
e-mail: catherine@mena.vt.edu

B. Mokhtar
e-mail: bassem.mokhtar@uaeu.ac.ae

M. R. M. Rizk
e-mail: mrmrizk@ieee.org

B. Mokhtar
College of Information Technology, United Arab Emirates University, Abu Dhabi, UAE

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
A. Rai et al. (eds.), *Paradigms of Smart and Intelligent Communication, 5G and Beyond*, Transactions on Computer Systems and Networks,
https://doi.org/10.1007/978-981-99-0109-8_10

data communications in the data plane. There are different SDN control plane architectures, for managing operations and cooperation of comprised controllers in such control plane. Centralized control, distributed control, and Logically Centralized-Physically Distributed (LC-PD) control planes are the most well-known control plane architectures. The LC-PD control plane architecture is investigated in an IoT environment that uses 5G cellular network technology. The tests conducted show that implementing LC-PD control plane architecture into 5G systems increases network performance and QoS of operating network services.

Keywords Software defined network · 5G · LC-PD control plane · QoS · IoT

10.1 Introduction

Nowadays, due to technology development, everything is connected through the internet as we live now in a digital environment. So, to be able to manage computer networks smoothly, we need to enhance our traditional Internet to be more flexible and easy to troubleshoot. This can be reached by adding more programmability to the internet, which means transforming it from being hardware-based to software controlled to support dynamic changes in it. So, Software Defined Network (SDN) technology is proposed to overcome the drawbacks of traditional Internet and facilitate network management.

The Internet of Things (IoT), in which objects and systems are intelligently connected, connects everything and makes everything available from anywhere. The emergence of IoT has caused the rise of the Fifth mobile Generation (5G) due to the transfer of a huge amount of data. The 5G mobile technology is mainly characterized by its highest data transfer speed and highest Quality of Service (QoS). So, SDN is being used in this cellular network to increase network performance and improve communication efficiency and QoS of operating Internet services in the 5G mobile network. This makes SDN a good approach for solving many challenges in cellular networks. The main features of SDN are ensuring flexibility for future upgrades in the network and also the simplification of network management by means of software. SDN usage usually enhances network performance and increases network QoS also.

The SDN architecture is composed of three layers as will be explained in detail later. We are concerned mainly with the middle layer, which is called the control layer. The control layer is made up of a major unit called the SDN controller, which serves as the network's manager. Having an efficient management architecture that controls the operations and cooperation of a set of SDN controllers employed in different contexts is a challenging task. The three primary SDN control plane designs are centralized control plane, distributed control plane, and logically centralized-physically distributed (LC-PD) control plane. The usage of different control architectures depends on the application they are applied for. The distributed control plane architecture was the one widely used by many researchers to solve the drawbacks of the basic SDN control plane in different environments. This SDN control plane architecture ensures the availability of many controllers, and it mitigates the threat of a single point of failure. Also, this control plane architecture possesses usually higher reliability and scalability compared to the centralized control plane architecture. Decisions are made quicker than with a single controller since there are several dispersed controllers connected in the network. So, the distributed management architecture is used by many researchers in 5G mobile network environment. This control plane architecture solves many of the basic SDN problems, but unfortunately without reaching the required levels of QoS in the network. On the other hand, this management architecture has increased throughput in the network compared to the usage of a single SDN controller. Also, the mobility effect is considered one of the destructive impacts of this architecture as it reduces throughput and hence causes an increase in latency in the network. All of these issues in distributed control architecture motivated us to employ the LC-PD control plane design. The distributed control plane architecture as mentioned before solves many issues facing the basic SDN control plane architecture. But at the same time, there are many challenges in this SDN control plane architecture to be addressed. Some of these challenges are that all SDN controllers used possess similar information policies as their neighboring SDN controller. So only one application can be run in the network as each controller is redundant to others only, and the performance degradation of the network in case of a large network, is due to the large distance between controllers and nodes in case of the usage of the distributed control plane architecture. Hence, this will lead to increased latency and increased response time taken, and the distributed control plane architecture isn't suitable to be applied in different wireless environments. The mobility motion of nodes in a random direction in wireless environments like in cellular networks leads to a reduction in network performance.

For addressing the aforementioned challenges, we propose the usage of a modified version of the LC-PD architecture for network management. In this SDN control plane architecture, SDN controllers are physically apart but are logically centralized in control functions as if they are one controller. This on one hand has solved the problem of the long distance between communicating Internet nodes and SDN controllers and hence will reduce latency impact on the network. On the other hand, this controller management architecture has kept the benefit of a global network overview that leads to better and more efficient network decisions due to the logical centralization of SDN controllers. Our modified version of LC-PD architecture allows the usage of different data traffic at the same time contrary to the distributed management architecture. In this control plane, each controller is responsible for a certain type of application while also handling diverse types of data flow. Also, the modified LC-PD control plane architecture can face performance degradation introduced in the network due to the mobility effect.

Our main goal in this chapter is to introduce SDN-based management architectures for computer networks and their adoption for enhancing the performance of cellular networks and especially the 5G cellular network.

The rest of this chapter is organized as follows. Section 10.2 introduces SDN's main characteristics, management architecture with different controller structures, main features, and the operation of this network. Section 10.3 gives a brief on the 5G mobile network architecture, how this cellular network works, and clarifies its main features and challenges. Section 10.4 gives a deep overview of the management of the 5G network using SDN, focusing on the LC-PD control plane architecture and clarifying the controller's main operation. Section 10.5 concludes the chapter and highlights future work.

10.2 Software Defined Network

SDN is a new technology that seeks to make network administration as flexible and straightforward as controlling apps and operating systems (Foukas et al. 2015; Khan et al. 2016). The primary purpose of SDN is to improve network control by responding quickly to business changes (Naseer 2016). SDN allows a network engineer or administrator to manage network traffic from a centralized control without interacting with individual network switches (Iqbal et al. 2017). SDN is a method of employing open protocols, such as Open Flow, as a basis for developing SDN solutions (Foukas et al. 2015; Prados-Garzon et al. 2016). The Open Flow standard was the first to separate the network control and data planes (Braun and Menth 2014).

The network is currently controlled by a centralized device known as the Software Defined Network Controller, which serves as the network's single point of contact. SDN controller is a software application that controls the flow (Xu and Zhao 2016). The essential component of an SDN network is an SDN controller. It is located between forwarding devices and applications (Foukas et al. 2015). SDN Controllers are considered the brains of the network. It is the program that manages flow control to switches and routers in the SDN network and acts as a strategic control point. The SDN controller connects all of the apps to the devices. To set up network devices, an SDN controller supports a wide range of protocols. The controller selects appropriate network pathways for various application traffic (Zolanvari 2015). The SDN controller acts as the network's operating system (OS). Because network management has been removed from the devices, the controller may now manage the network automatically and facilitate the integration of business applications (Jacobsson and Orfanidis 2015). The strategic point of control in an SDN network is the SDN Controller. It communicates network information to switches/routers and applications. An SDN Controller is made up of multiple "pluggable" modules that may be used to conduct various network tasks. Among the actions the devices are on the network and their capabilities in collecting network information (Keti 2015) are determined. There are several SDN controllers available, each with its own set of specialized characteristics. Some of these SDN controllers are Open Flow, Beacon, DISCO, IRIS, Maestro, NOX, POX, Open Daylight, Flood Light, Ryu, Trema, RISE, SNAC, etc. (Khan et al. 2016; Shalimov et al. 2013).

10.2.1 SDN Architecture

SDN design divides the network into three distinct levels, which are linked via northbound and southbound APIs (Khan et al. 2016; Ali Hassan 2017; Blial et al. 2016; Kaliyamurthy et al. 2021). The application layer, control layer, and infrastructure layer are the three levels of SDN (or data plane) (Foukas et al. 2015; Keti 2015; Kaliyamurthy et al. 2021; Fatma and Qaddour 2019). The control Plane contains the controller, the brain of the network, and this layer can be designed differently according to the applied application as will be shown later (Li et al. 2018; Azizi et al. 2015). SDN was firstly introduced for wired networks but this paradigm has proved its benefits for wireless environments too. This network architecture is applied in different environments to ensure flexibility and simplify network management (Naseer 2016; Braun and Menth 2014; Jarraya et al. 2014; Bannour et al. 2018; Kobo et al. 2017; Araniti et al. 2014). Figure 10.1 provides a global view of the SDN architecture consisting of three layers which are the application layer which consists of network applications, the control layer which contains the SDN controller, and the infrastructure layer which consists of network devices.

The application layer

It generally contains the standard network applications or functionalities that businesses utilize, such as intrusion detection systems, load balancing, or firewalls (Foukas et al. 2015; Naseer 2016). These network services are used to interact with the control layer. An SDN substitutes a specialized device, such as a firewall or load balancer, in a traditional network with an application that utilizes the controller to regulate data plane behavior (Khan et al. 2016; Araniti et al. 2014). As a result, this

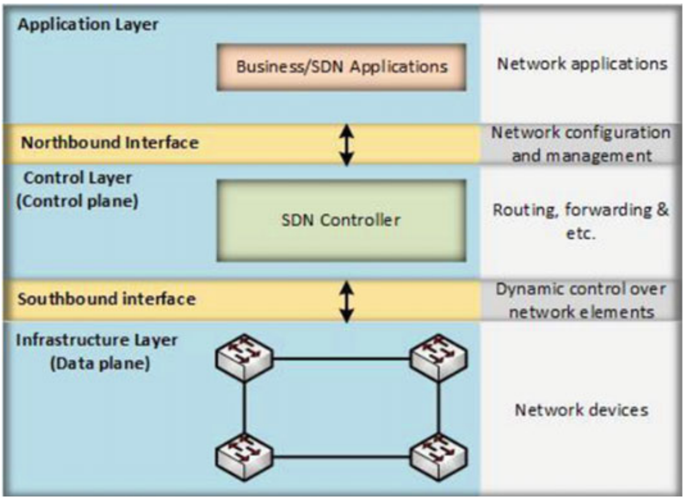


Fig. 10.1 Basic SDN architecture (Nevala 2016)

layer provides an open interface via which other levels in the design can communicate (Ali Hassan 2017; Blial et al. 2016; Jarraya et al. 2014; Bannour et al. 2018).

The control layer

This layer is comprised of a centralized SDN controller that serves as the SDN's brain by providing a global view of the whole network (Foukas et al. 2015; Araniti et al. 2014). This controller maintains network regulations as well as traffic flow (Khan et al. 2016; Naseer 2016). This layer will connect with the infrastructure layer through the Open Flow protocol (Foukas et al. 2015; Ali Hassan 2017; Blial et al. 2016; Bannour et al. 2018).

The infrastructure layer

Switches and routers are examples of physical and virtual network components in this tier. For traffic forwarding rules, all of these network devices will use the Open Flow protocol (Fatma and Qaddour 2019; Araniti et al. 2014).

Northbound and southbound application programming interfaces (APIs) are used to interact between these three levels (Blial et al. 2016). The application layer communicates with the controller via its northbound interface, whereas the controller and switches interact via southbound interfaces, such as Open Flow, but alternative protocols are available (Foukas et al. 2015; Khan et al. 2016; Naseer 2016). There is no clear standard for the controller's northbound API to match Open Flow as a broad southbound interface at the moment (Jarraya et al. 2014). The SDN design separates network control from forwarding tasks, allowing network control to be directly programmable (Prados-Garzon et al. 2016; Ali Hassan 2017; Bannour et al. 2018). Typically, the SDN architecture is:

- Directly programmable: Because network control is isolated from forwarding functions, it is directly programmable.
- Agility: By abstracting control from forwarding devices, administrators may alter network traffic flow dynamically to meet changing demands.
- Centrally managed: Network intelligence is (logically) centralized in software-based SDN controllers, which maintain a global view of the network and appear as a single, logical switch to applications and policy engines.
- Programmatically configured: SDN enables network administrators to rapidly create, administer, protect, and optimize network resources with dynamic, automated SDN programs that they can write themselves because the programs do not rely on proprietary software (Kaliyamurthy et al. 2021).
- SDN is open standards-based and vendor-neutral when implemented using open standards because instructions are delivered by SDN controllers rather than various, vendor-specific devices and protocols.

10.2.2 *SDN Management Architecture*

We will explain in detail the different SDN-based management architectures used in our studies.

Centralized Controller

A centralized controller is the most basic SDN solution, with a single controller having a global view of the network and controlling the whole network (Tadros et al. 2018). This would help for better network decisions; however, it exposes the network to the threat of a single point of failure. Furthermore, with a single central controller, there are scaling restrictions as the network size and dynamics rise, as well as dependability and congestion, which limit the widespread use of this design (Bannour et al. 2018; Kobo et al. 2017). Figure 10.2 represents the basic centralized control plane architecture used with a single SDN controller and some network devices.

Distributed Controllers

Several researchers employed numerous dispersed controllers to solve the disadvantages of the single centralized controller outlined above (a cluster-based architecture) (Syed-Yusof et al. 2020). This architecture usually consists of multiple inter-connected clusters (SDN domains), where each cluster is managed by one controller. Thus, this management architecture offers balanced centralized control and distributed operations. A network cluster with a controller is known as a domain (Oktian et al. 2017). Due to the existence of several controllers in the same network in service, the primary goal of this design is to decrease network size into SDN domains to allow control exchange across various domains (Naseer 2016; Tadros et al. 2018). Figure 10.3 shows the usage of many controllers in this case two controllers are connected to provide reliability.

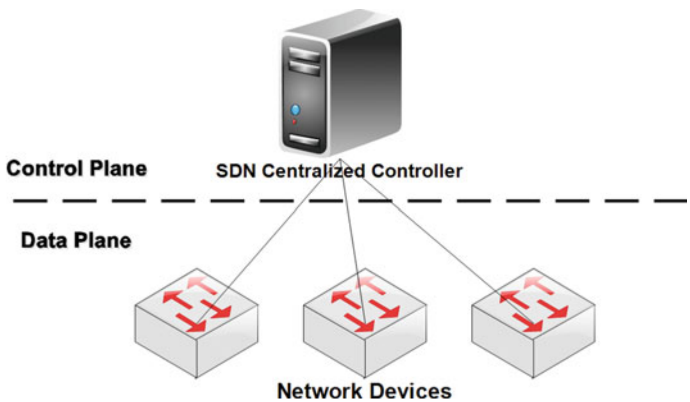


Fig. 10.2 Centralized control plane architecture

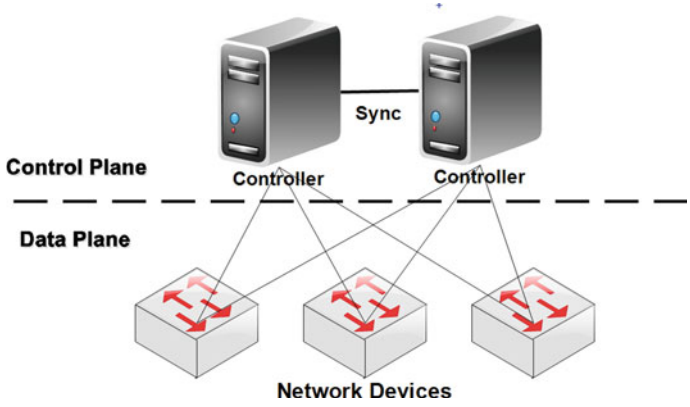


Fig. 10.3 Distributed Control plane architecture

Logically Centralized-Physically Distributed Controller

LC-PD controller operates as a central controller for the group of nodes directly connected, but in reality, many distributed controllers are physically apart from each other (Tadros et al. 2018). In this management architecture, many controllers are employed simultaneously, each of which is connected to the others, and they all share a copy of the control data and some information policies while remaining remote from one another. The number of flow requests processed per second is increased, thanks to this design, which also speeds up the processing of individual flow requests. In this architecture, each controller is responsible for a group of nodes. In the LC-PD control plane architecture, switches are linked to a nearby controller and must be reconfigured if the controller fails. Figure 10.4 represents the ordinary LC-PD control plane architecture used by many researchers in different applications with the usage of three controllers providing the same services.

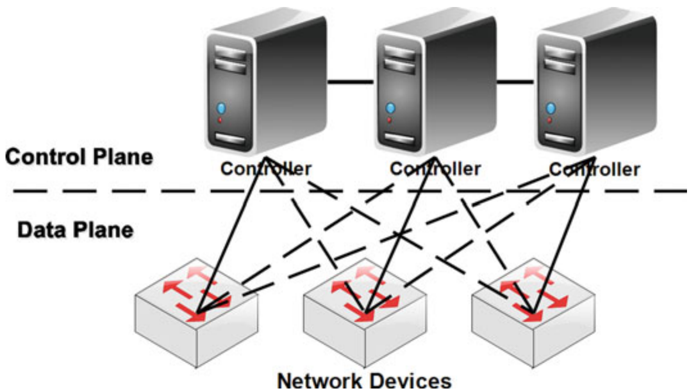


Fig. 10.4 Ordinary logically centralized-physically distributed control plane architecture

10.2.3 How SDN Works

Functional segregation, network virtualization, and automation through programmability are only a few of the technologies that SDN incorporates. The fundamental objective of SDN technology was the division of the data plane from the network management plane. The data plane typically is in charge of transporting packets from one location to another, with the control plane being the layer in charge of making decisions on the flow of packets across the network. A packet arrives at a network switch in a straightforward SDN situation, and the switch instructs where to transmit the packet following the rules created and provided to the switch by the centralized controller. The switch also gives the controller data on the traffic that it manages. Every packet is delivered by the switch along the same path and is given the same treatment.

10.2.4 Benefits of SDN

SDN has several advantages, including the flexibility to expand network resources in accordance with application and data demands, automatic load balancing, simplified physical infrastructure, and on-demand provisioning. With SDN, a network administrator has granular control and security over changing the rules of any network switch as needed. This is especially useful in a cloud computing architecture since it gives the administrator more flexibility and efficiency in managing traffic volumes. In essence, this gives the network administrator greater control over network traffic flow and the ability to employ less costly switches.

Additionally, SDN offers network control and complete visibility. Instead of setting several separate devices, the administrator works with a single centralized controller to disseminate policies to the associated switches. Since the controller can monitor traffic and put security measures in place, this feature is also helpful from a security standpoint. For instance, if the controller discovered the questionable activity, it might redirect or discard the packets. Additionally, SDN virtualizes hardware and services that were before performed by dedicated hardware, which lowers operational expenses because less hardware is required.

The feature of decoupling the data plane from the control plane in SDN usually enhances configuration, improves performance, and fosters innovation in the architecture and operations of the network. SDN control, for example, may comprise not just packet forwarding at the switching level, but also link tuning at the data link level, breaking through the layering barrier. The network programmability of SDN also provides a suitable platform for testing new techniques and encourages new network designs.

10.3 5G Mobile Network

5G is an abbreviation for fifth-generation wireless technology (Eze et al. 2018). It is the most recent generation of mobile wireless technologies. This mobile technology is more clever since it permits unlimited connections to the entire world and offers more services and advantages to the world than 4G (Llorens-Carrodeguas et al. 2018; Chaturvedi et al. 2021). This cellular technology possesses usually: greater speed of data transfer across the network up to 25 Mbps, low latency, and high throughput (Helode et al. 2017; Morgado et al. 2018; Sitan 2016). It also provides large broadcasting of data in Gigabit, and the ability to connect more devices simultaneously (Helode et al. 2017). It supports virtual private network and Wireless World Wide Web (WWWW) (Eze et al. 2018; Chaturvedi et al. 2021; Sitan 2016; Dangi et al. 2022; Sharma et al. 2016; Alexiou 2017; Tadros et al. 2020). This mobile network possesses improved spectrum efficiency, very high bandwidth that users never experienced before, better mobility support, connection to multiple wireless technologies and switching between them, and high connection density (Alexiou 2017). Multiple Input Multiple Output (MIMO) will be used in 5G to considerably boost network capacity (Eze et al. 2018; Morgado et al. 2018; Zhang et al. 2014; Ahmed et al. 2019). In comparison to the current 4G mobile standard, the 5G standard has a new wireless interface that supports greater frequencies and spectrum efficiency. This mobile technology will support IPv6 and flat IP (Dangi et al. 2022). Unlike traditional IP addresses, Flat IP Architecture allows devices to be identified using symbolic names (Helode et al. 2017; Sharma et al. 2016). 5G wireless technology employs Orthogonal Frequency Division Multiplexing (OFDM) with a frequency range of 3–300 GHz (Chaturvedi et al. 2021; Helode et al. 2017; Dangi et al. 2022; Tadros et al. 2020; Ahmed et al. 2019). Documentation, support for electronic transactions (e-Payments, e-Transactions), and other services will be provided via fifth-generation technology. Fifth-generation technology also includes features such as a camera, MP3 recording, video player, huge phone memory, audio player, and so on (Alexiou 2017).

10.3.1 5G Architecture

The 5G cellular systems model is an all-IP-based wireless and mobile network paradigm (Sharma et al. 2016; Tadros et al. 2020). Fifth-generation network design includes a user terminal and a variety of radio access technologies (RAT) (Morgado et al. 2018; Alexiou 2017). All IP-based mobile apps and services, such as mobile portals, mobile commerce, mobile health care, mobile government, mobile banking, and others, are provided via Cloud Computing Resources (CCR) in 5G Network Architecture (Alexiou 2017). Customers may use applications without installing them and access their data from any computer with an internet connection thanks to cloud computing (Zhang et al. 2014). Hence, Fig. 10.5 presents the basic 5G network architecture.

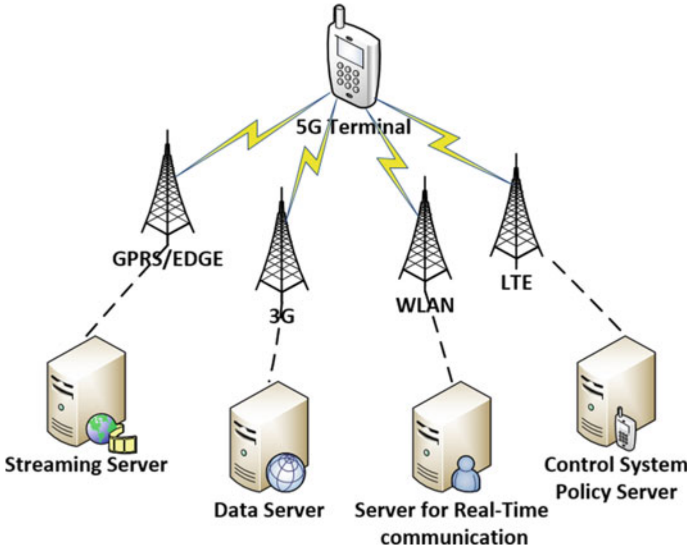


Fig. 10.5 5G architecture (Tadros et al. 2020)

10.3.2 Features of 5G Mobile Technology

This cellular network is characterized by its highest data transfer speed which is about 10 times greater than 4G, highest capacity, and lowest cost per bit (Khan et al. 2016; Eze et al. 2018; Sharma et al. 2016; Tadros et al. 2020; Ahmed et al. 2019; Lal et al. 2021). 5G technology enables global access and service portability (Alexiou 2017). Because of its great mistake tolerance, it provides high-quality services. It has a massive broadcasting capacity of up to Gigabit, which can accommodate about 65,000 connections at once (Eze et al. 2018; Sitan 2016; Sharma et al. 2016; Ahmed et al. 2019). 5G technology employs remote management to provide users with better and faster solutions (Zhang et al. 2014). 5G technology has incredibly high upload and download rates (Alexiou 2017; Lal et al. 2021). It also provides high resolution for crazed cell phone users as well as bi-directional massive bandwidth shaping (Sharma et al. 2016). It has a higher data rate near the cell's edge and a larger coverage area (Dangi et al. 2022). It also consumes less power (Morgado et al. 2018; Khan Tayyaba and Shah 2017).

10.3.3 How 5G Works

5G networks, like earlier cellular networks, are composed of cells divided into sectors. These cells use radio waves to convey data. Each of these cells is wired or wirelessly connected to the network's core. To boost network efficiency, the cell is divided into micro and pico cells. 5G integrates cutting-edge technologies such as cognitive radio, IoT, nanotechnology, and cloud computing (Eze et al. 2018; Sitan 2016; Sharma et al. 2016).

10.3.4 Challenges in 5G Network

The transition from 4 to 5G involves various issues that must be addressed (Eze et al. 2018). In addition, there are other problems associated with integrating 5G technology with various applications to deliver a wide range of services. There are also some problems faced with the new technologies used to enable this mobile technology. Because of the large number of linked devices running various services and data created, this cellular network lacks programmability, agility, security, and data management. One of the drawbacks of this mobile network is its high cost since 5G technology is incompatible with prior generations (Chaturvedi et al. 2021). This will affect the user's phone as their phone that supports 2G, 3G, or 4G will not connect to the 5G network. So, this network cannot be easily modified or upgraded. So to overcome these drawbacks, we have to make many changes in the design of cellular architecture (Eze et al. 2018).

10.4 SDN Implementation in 5G Mobile Network

As previously said, the SDN idea was established to generate greater flexibility in network administration since this new paradigm is built on the separation of the data plane and the control plane, allowing the software to handle network traffic (Foukas et al. 2015). The network is now controlled by a centralized device called SDN Controller, which serves as the network's single point of contact. SDN controllers in SDN serve as the network's "brains," as all applications connect with devices via the SDN controller. The network design, particularly the controller distribution, has an impact on the overall network QoS. As a result, the propagation latency between controllers and network devices, particularly in large networks, must be considered while developing an SDN control plane. The number of controllers and their placement in the network affects the path reliability, fault tolerance, and application requirement. So, there are different implementations of SDN controllers forming various management architectures, such as architectures comprising a centralized controller, distributed controllers, and LC-PD controller. SDN has been applied in different networking environments to solve issues of traditional networks used in conjunction. So, we will clarify in this section the management architecture and operation, the communication process, and explore the benefits of the implementation of SDN into 5G mobile network.

10.4.1 SDN Management Architecture (Proposed Approach)

We have made some modifications to the ordinary LC-PD management architecture used. In our LC-PD control plane design, we assumed that each set of nodes runs a separate application, but that each controller may also be utilized as redundancy for other controllers because they all have the same data control copy. This solution was recently employed by connecting several controllers with certain nodes running the same application. In our approach, each controller serves as a central control plane for each cluster of nodes, and each of these clusters may comprise nodes with distinct data traffic that may be supplied by other neighbor controllers that are geographically separated. Physical location and the diverse types of traffic generated (different applications) operating on nodes are unimportant in our designed architecture. Figure 10.6 presents our updated version of a logically centralized-physically distributed control plane architecture that runs several services with the usage of three controllers.

The SDN approach centralizes network control in a single controller. This increases the likelihood of network failure and availability. This will impact network performance because of the overhead in the control plane. This shows that a centralized controller puts the entire network at risk, hence it is better to use several SDN controllers simultaneously to reduce failure and overhead. Table 10.1 summarizes the main advantages and disadvantages of the different control plane architectures used in this study.

As a result, the number of controllers used in a network is governed by network characteristics as well as user needs. This makes the centralized controller suitable for small-scale networks, but for large-scale networks, numerous domains and controllers are necessary. Each SDN controller in a domain that can govern a portion of the whole network. As a result, the primary constraints must be of controller capacity, network domain count, and inter-SDN controller communication technique. Furthermore, the placement of these controllers in a network has a direct

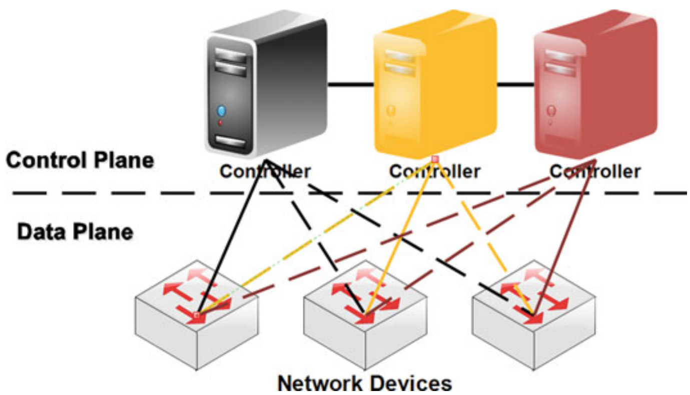


Fig. 10.6 Modified LC-PD control plane architecture

Table 10.1 Advantages and disadvantages of different SDN Control plane architecture

Control plane architecture	Advantage	Disadvantage
Centralized	– Global network view – Better and informed decisions	– The single point of failure – Unable to scale – Performance decline because of distance
Distributed	– Scalability – Quick response in decisions relative to centralized control plane architecture	– Highest costs – Unbalanced load distribution – Large amounts of data must be constantly synchronized
Ordinary LC-PD	– Scalability, Reliability, and resiliency – Quick response in decisions	– Costs – Synchronization of a lot of data continuously
Modified LC-PD	– Scalability, reliability, and resiliency – Quick response in decisions – Balanced load distribution – Efficient management for different types of applications in the same network – Less overhead – Less complexity	– Costs due to the usage of many controllers in the same network

influence on communication between switches and a controller, which impacts critical performance elements like as latency, load balancing, redundancy, connectivity, and survivability. Another consideration is controller overload, which happens when several switches are connected to the same controller; hence, we must distribute the load between controllers to avoid the controller being congested.

10.4.2 SDN Management Architecture Operation

SDN Controller Operation

All of the above control plane architecture mentioned before work like each other mostly but with small differences. The switch examines its flow table after receiving a packet. The rule action (allow or deny) will be applied to the packet if it fits a flow rule (match specific rate). But if the packet doesn't match any flow rule, the switch will either drop it or forward it to the controller. The SDN controller by turn will create a flow rule for this request. This is the operation of all control plane architecture, but the difference appears when using many controllers. When the packet mismatch happens the switch in this case will forward the packet for all controllers to check with them. Then the nearest controller to the switch will create the flow rule for this request. The flow chart in Fig. 10.7 illustrates the main operation of the LC-PD control plane architecture studied.

SDN Controller Communication

In this section, we will clarify the communication process between controllers in distributed scenarios, between Access Points and controller, and between nodes and each other.

Controller-to-Controller Communication for Distributed Architecture

In general, a controller consists of three main parts: Listener, Control logic, and Messenger. First, you need to find out the sort of case that you want the controller to listen to (e.g., Connection Up, Packet In, etc.). Then you can differentiate between distinct flows using some logic and add a proper action for a particular flow. Finally, to add the fresh rule to the Open Flow table, you send the message to the switch. So, in the case of LC-PD control plane architectures, the access point sends to the controller responsible for handling the required request not to the controller directly connected. The communication between controllers is based on the SDNi protocol as illustrated in Fig. 10.8. In a distributed control system, the East/Westbound interface of the SDNi protocol connects controllers. To provide inter-SDN domain routing, this protocol offers three different types of messages: reachability update, flow setup/tear-down/update application, and update capability.

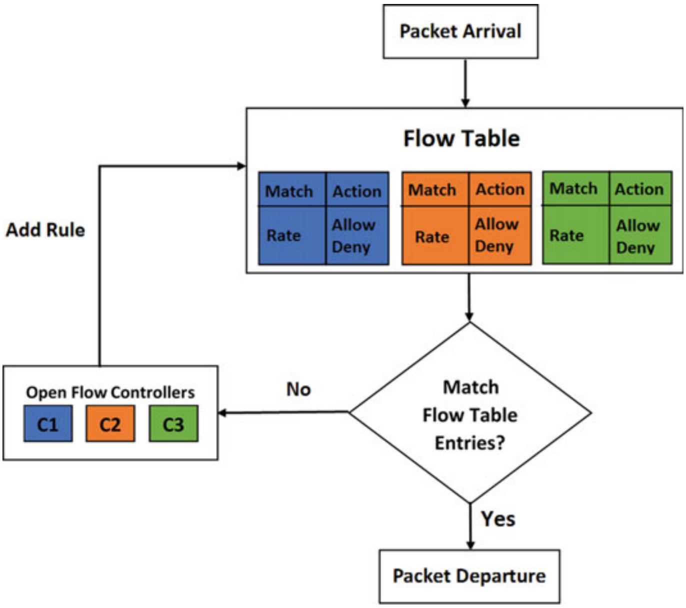


Fig. 10.7 Modified LC-PD control plane operation

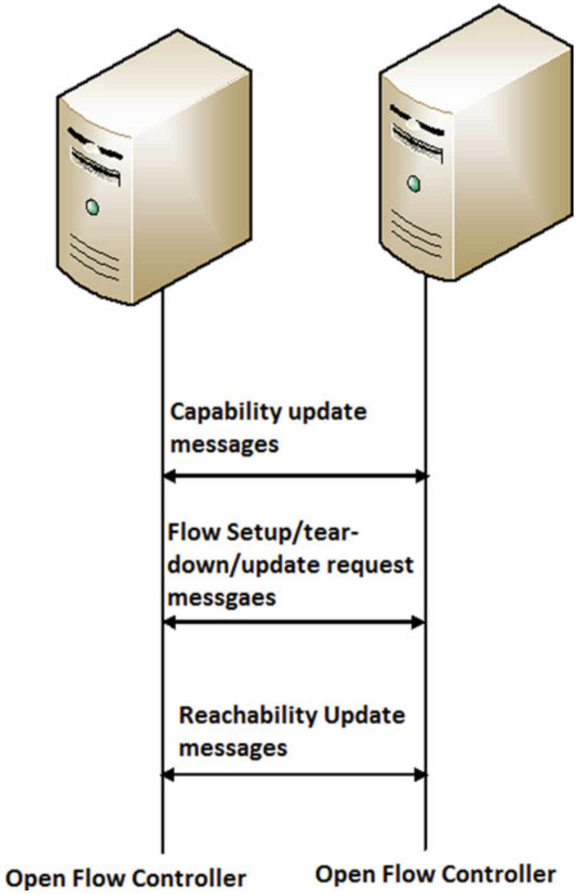


Fig. 10.8 Controller-to-controller communication using SDNi protocol

Communication Between Controller and Access Point/Switch

The communication between the controller and the access point is based mainly on messages. The connection between the controller and the access point is established after the TCP handshake. The connection between them is two-way and many messages are exchanged between them. Messages are sent between both of them to ensure that no problem in connection. Also, messages are exchanged in case of link failure, and when a flow mismatch happens. Figure 10.9 illustrates the main messages exchanged between the controller and Access Point/Switch during communication.

Communication Between Nodes

In the case of centralized control plane architecture, only one access point is available that receives all data. But in the case of distributed control plane architecture, data will be sent to the directly connected access point, which by turn forward this data to the directly connected controller to it. Similarly, in the case of LC-PD control plane

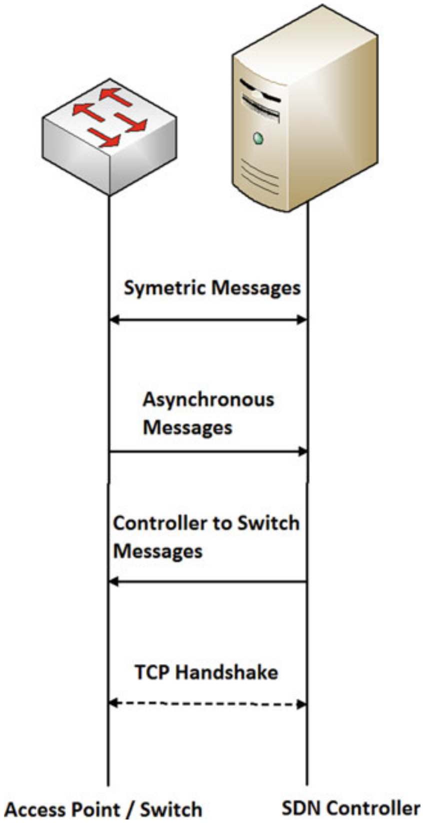


Fig. 10.9 Communication between controller and access point/switch

architecture, the difference is that the access point is connected to all controllers in the network. So, the access point forwards the data to the controller responsible for handling the required data traffic.

In the case of the 5G mobile network, nodes are called user equipment. So, to transfer data between nodes, the user will contact the access point directly connected. The access point will after that direct the data to the required user if it is registered in the flow table and if not will forward the request to the controller. The only difference appears in the case of LC-PD control plane architecture, due to the different applications running on each of the controllers. So, according to the required request, the access point will forward it to the controller that can handle it. Figure 10.10 clarifies how data is transferred between nodes according to the mentioned process.

So, to better understand how two nodes, each of them connected to an access point, which by turn is connected to a controller, are responsible for handling the different requests. We will use the LC-PD control plane architecture, supposing that we have three controllers. Controller 1 (C1) is in charge of file transmission at a rate of 3 MB/s, Controller 2 (C2) is utilized for VOIP at a rate of 5 MB/s, and Controller 3 (C3) is in charge of video streaming at a rate of 7 MB/s. Node A is connected

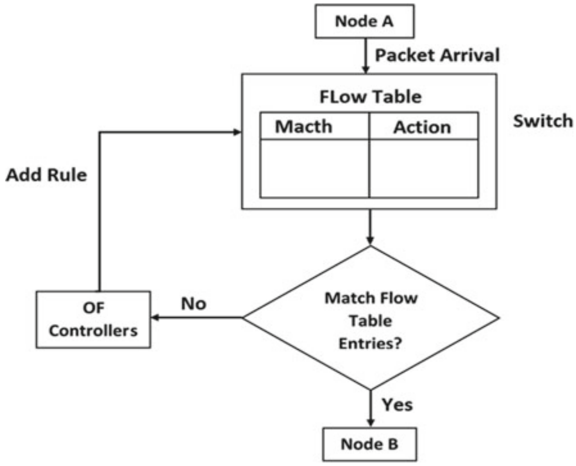


Fig. 10.10 Data transfer between nodes

to an access point, which is connected to C1. Node B is connected to an access point, which is connected to C2. Figure 10.11 illustrates the communication process between node A and Node B using the modified LC-PD control plane architecture.

So, if we suppose that Node A asks for a video from Node B, Video is handled by C3 so the request will not match the flow table record and the request will be

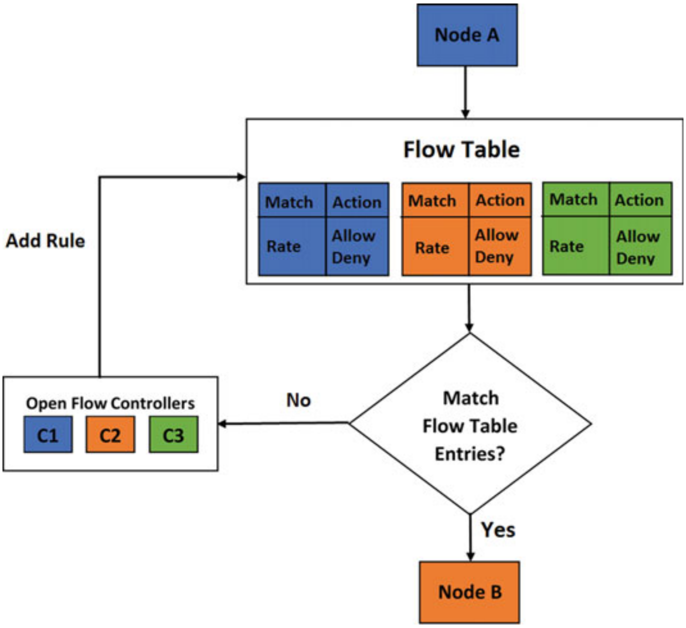


Fig. 10.11 Communication between nodes in the Modified LC-PD control plane architecture

Table 10.2 Example of communication requests between two nodes in modified LC-PD control plane architecture (Flow Table)

Request from node A	Match	Action
Video streaming (7 MB/s)	Doesn't match, forward to C3	C1 add rule
File transfer (3 MB/s)	Match	Allow
VOIP (5 MB/s)	Doesn't match, forward to C2	C1 add rule
HD video (>7 MB/s)	Doesn't match	Deny (drop request)
Video streaming (7 MB/s)	Match	Allow

transferred to all controllers through the switch. Hence, C3 will handle the request and C1 will add this rule in his flow table and so on. Table 10.2 clarifies some request examples handled between Node A and Node B.

10.4.3 SDN-Based Management for 5G Mobile Network

SDN is an intelligent network that uses less hardware. The importance of SDN in 5G mobile networks arises from its ability to support novel features like network virtualization, automation, and the development of new services on top of virtualized resources in secure and reliable networks. Furthermore, SDN principles address the Radio Resource Management issues in 5G networks in a variety of application scenarios (interference management, mobile edge computing, RAN sharing) (Haque and Abu-Ghazaleh 2016). As a result, to address all of the preceding issues in the 5G mobile network while also lowering costs, SDN was recommended as one of the main technologies that would give the flexibility necessary in the 5G cellular network design. SDN aids in simplifying network administration and setup, as well as cost reduction through the softwarization of 5G network services (Acharyya and Al-Anbuky 2016). This is accomplished by SDN virtualized network services that separate data transmission from network control. The intelligence of the entire network is maintained by the controller in SDN. While the control plane issues flow entries, the data plane is separated into several switches and routers that are in charge of flow forwarding or routing. As explained before, SDN introduced Application programmability interfaces (API) between the decoupled planes. The Open flow was identified as the first common protocol used in the southbound interface of the SDN architecture between the data plane and the control plane. The Open Flow protocol is used in SDN architecture to define all physical and logical components of a data link, specify data structures, messages, and processes, and ensure common control plane activities like packet modification, routing table maintenance, and flow management. This protocol also supports IPv6 features including VPN and IPsec tunneling, access control, and quality of service (Hicham et al. 2018). The controller will use the Open flow protocol to interface with 5G core devices, maintain network topologies, establish new flows, and gather network data to comply with the QoS requirement in 5G mobile networks. The 5G network will benefit from additional security levels brought by SDN, such as data integrity, secrecy, authentication, and access control (Llorens-Carrodeguas et al. 2018). So, Fig. 10.12 shows the deployment of SDN in a 5G network, in which the core network nodes are transformed into pure forwarding elements and control plane functionalities are exported to a centralized SDN controller node.

SDN might be used to provide the groundwork for 5G to operate over a control plane. It can also boost data flows as they pass over the 5G network. Furthermore, SDN architecture can save network bandwidth while increasing latency. SDN also allows for the management and automation of network redundancy from a centralized control plane. As a consequence, including SDN into a 5G mobile network will result in fewer modifications in network elements, centralized administration, and increased programmability in 5G, allowing numerous mobile operators to share network resources. We illustrate that most researchers use the hierarchical distributed controller management architecture to prevent the single point of failure of the basic SDN controller architecture. However, we have chosen the LC-PD control plane

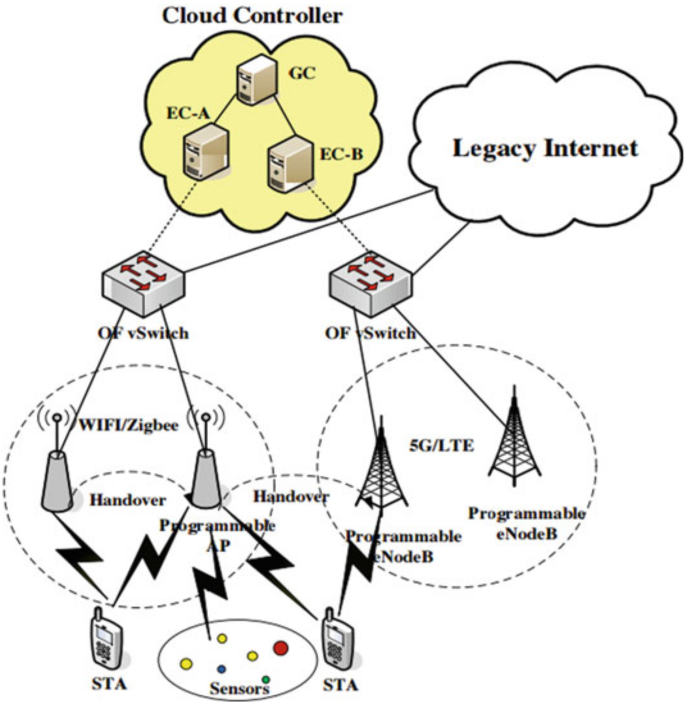


Fig. 10.12 SDN implementation in 5G network (Zolanvari 2015; Fatma and Qaddour 2019)

architecture due to its effect on network improvement. One of the important aspects affecting the network is the controller placement and the number of used controllers in the network. In our design, we supposed the usage of many controllers connected to have some caching information to manage different tasks. Each of these controllers is responsible for running different types of data traffic. The use of LC-PD architecture in combination with a 5G mobile network has decreased the network's cost. Also, it has helped in providing flexibility and simplifying network management.

10.4.4 SDN Benefits for 5G

The 5G based SDN architecture provides many features to the traditional 5G network as follows.

Directly programmable: the SDN integration with the dynamic 5G devices using a programmable control framework through controllers is the reason to make the network more intelligent (Khan Tayyaba and Shah 2017).

Open: the possibility of multi-vendor technologies in a 5G network environment will be applicable, due to the usage of SDN controllers, which are capable to operate with several Open flow devices (Hicham et al. 2018).

Agile: Network operators may dynamically design, manage, and optimize network resources, as well as change traffic flows, using dynamic and automated SDN programs(Hicham et al. 2018; Freeman and Nguyen 2017).

Abstraction: To enhance portability, SDN services are abstracted from the underlying network technologies (Hicham et al. 2018).

Automated administration: The 5G networks will be able to identify appropriate changes with the SDN concepts and how these changes can be handled automatically (Hicham et al. 2018).

Wireless resources enhancement: SDN enhances wireless resource utilization and hence increases user satisfaction.

Providing QoS requirements for 5G applications: The integration of SDN with 5G can handle the QoS parameter needs for each kind of data flow.

SDN will be the fundamental technology for QoE (Quality of Experience) management and supply in 5G mobile networks. In telecommunications terminology, QoE is a measurement used to assess how effectively a network meets the needs of its end users (Hicham et al. 2018). SDN can give flexibility and full information to aid decision-making in areas such as routing offloading, energy efficiency, and spectrum allocation (Khan Tayyaba and Shah 2017; Hicham et al. 2018; Freeman and Nguyen 2017). SDN will aid in the use of a unified, less difficult, less expansive, and flexible network behavior setup (Hicham et al. 2018). It also offers an operator-based smooth transfer at a reasonable cost (Hicham et al. 2018). The centralized flexibility of SDN can foster innovation in shared mobility, routing, and data management (Khan Tayyaba and Shah 2017; Hicham et al. 2018). Simplification of network configuration and management to reduce costs (Hicham et al. 2018). Open Flow-based SDN provides a highly secure and manageable environment: Security policy is dynamically and flexibly adjusted under SDN (Hicham et al. 2018). Interoperability of different interfaces of the other vendors (Hicham et al. 2018). To achieve the needed high capacity of 5G systems, SDN will give solutions to multi-hop wireless network restrictions and data stored on the edge network (Hicham et al. 2018).

Despite the technological advantages of SDN, especially in dynamic behavior, the division of the control plane from the data plane may result in several performance issues concerning network dependability and security, as well as the controller placement issue. The main goal of this topic is to choose the best controller location for a certain SDN architecture to get the best design. This separation notion can have an impact on both fixed and dynamic network performance. Therefore, controllers should be interconnected via an overlay network, switches should be continuously controlled, and switches should be assigned to controllers based on the shortest path. Finally, controller failures or disconnections between the control and data planes, which may cause packet loss or other network issues, should be fixed. Therefore, while formulating the controller placement problem for constructing and designing fixed and mobile networks, these SDN networks functional principles must be taken into account.

10.5 Conclusion and Future Work (Summary)

In this chapter, we are working to improve cellular network performance by implementing of SDN technology. The SDN concept adds more flexibility to the network architecture in contrary to the traditional network consisting of switches and routers used before. This flexibility in the network architecture is achieved by adding more programmability to the network thus reducing the need for hardware devices and simplifying network upgrades in the future. We have clarified everything concerning the SDN from its architecture which consists of three different layers, to its main features and how this technology works. Also, we have studied the communication protocol used between distributed controllers and the exchange of messages between switches and SDN controllers to establish connections. We have focused our study on different SDN control plane architectures. We have studied the basic SDN management architecture and different SDN-based management architectures. The distributed SDN control plane architecture is widely used in different applications. SDN was firstly designed for wired networks, but due to the emergence of IoT, SDN was applied in different wireless environments. We have studied the usage of SDN technology in 5G mobile networks. SDN adoption in 5G mobile networks often improves communication efficiency and the QoS of Internet services operating. SDN usually simplifies network management by ensuring flexibility and enhancing the quality of service.

We adapted the LC-PD architecture in managing Internet traffic adopting 5G network infrastructures. In the LC-PD architecture, we have used different types of services at the same time using multiple controllers physically apart from each other. We concluded that the LC-PD management structure is the best architecture for improving network performance used in conjunction with the various wireless environments and especially 5G mobile networks. This SDN management architecture is characterized by its quick response time and hence minimized probability of network failure. The usage of LC-PD control plane architecture has enhanced network performance in the 5G network compared to the distributed architecture used previously.

Our next set of goals is to design an intelligent SDN controller architecture adopting lightweight machine learning algorithms to be able to enhance network performance effectively, increase throughput, reduce energy consumption, losses and latency and prolong network lifetime as much as possible. Thereby providing higher QoS in case of using different data traffic. Also, we are working to present a novel SDN-based management architecture, called Software Defined Cloud of Things (SDCoT) to support running IoT services with massive user demands. This novel management architecture is based on the integration between the concept of Software Defined (SD) and Cloud Computing (CC) technologies with IoT. SDCoT also provides a resourceful interaction with large-scale IoT networks that open up new data management and security challenges.

References

Acharyya IS, Al-Anbuky A (2016) Towards wireless sensor network softwarization. In: IEEE NETSOFT 2016—2016 IEEENetSoft conference and workshops. Software-defined infrastructure networks, Clouds, IoT Serv., no June, pp 378–383. https://doi.org/10.1109/NETSOFT.2016.7502470

Ahmed AM, Hasan SA, Majeed SA (2019) 5G mobile systems, challenges and technologies: a survey. J Theor Appl Inf Technol 97(11):3214–3226. https://doi.org/10.5281/zenodo.3256485

Alexiou A (2017) 5G wireless technologies. 3(10):1–417. https://doi.org/10.1049/PBTE069E_ch

Ali Hassan M (2017) Software defined networking for wireless sensor networks: a survey. Adv Wirel Commun Netw 3(2):10. https://doi.org/10.11648/j.awcn.20170302.11

Araniti G, Cosmas J, Iera A, Molinaro A, Morabito R, Orsino A (2014) OpenFlow over wireless networks: performance analysis. In: IEEE International symposium on broadband multimedia systems and broadcasting, BMSB, no June. https://doi.org/10.1109/BMSB.2014.6873559

Azizi M, Benaini R, Ben Mamoun M (2015) Delay measurement in Openflow-enabled MPLS-TP network, vol 9, no 3, pp 90–101. https://doi.org/10.5539/mas.v9n3p90

Bannour F, Souihi S, Mellouk A (2018) Distributed SDN control: survey, taxonomy, and challenges. IEEE Commun Surv Tutor 20(1):333–354. https://doi.org/10.1109/COMST.2017.2782482

Blial O, Ben Mamoun M, Benaini R (2016) An overview on SDN architectures with multiple controllers, vol 2016

Braun W, Menth M (2014) Software-defined networking using OpenFlow: protocols, applications and architectural design choices. Futur Internet 6(2):302–336. https://doi.org/10.3390/fi6020302

Chaturvedi A et al (2021) Future of 5G wireless system, vol 06, no 01, pp 47–52

Dangi R, Lalwani P, Choudhary G, You I, Pau G (2022) Study and investigation on 5G technology: a systematic review. Sensors 22(1):1–32. https://doi.org/10.3390/s22010026

Eze KG, Sadiku MNO, Musa SM, Perry RG, View P (2018) 5G wireless technology: a primer, no July, pp 5–8

Fatma N, Qaddour J (2019) Integration of SDN and 5G and its related challenges and security issues, vol 9, no 2, pp 82–88. https://doi.org/10.5923/j.ijnc.20190902.03
Foukas X, Marina MK, Kontovasilis K (2015) Software defined networking concepts. In: Software defined mobile networks (SDMN): concepts and challenges, pp 21–44. https://doi.org/10.1002/9781118900253.ch3
Freeman B, Nguyen H (2017) Software-defined networking. In: Building the network of the future: getting smarter, faster, and more flexible with a software centric approach, pp 87–102. https://doi.org/10.1201/9781315208787
Haque IT, Abu-Ghazaleh N (2016) Wireless software defined networking: a survey and taxonomy. IEEE Commun Surv Tutor 18(4):2713–2737. https://doi.org/10.1109/COMST.2016.2571118
Helode PS, Walse KH, Karande MU (2017) An online secure social networking with friend discovery system. Int J Innov Res Comput Commun Eng 5(4):8198–8205. https://doi.org/10.15680/IJIRCCE.2017
Hicham M, Abghour N, Ouzzif M (2018) 5G mobile networks based on SDN concepts. Int J Eng Technol 7(4):2231–2235. https://doi.org/10.14419/ijet.v7i4.12194
Iqbal A et al (2017) Analytical modeling of end-to-end delay in OpenFlow based networks, pp 6859–6871
Jacobsson M, Orfanidis C (2015) Using software-defined networking principles for wireless sensor networks. In: Proceedings of the 11th Swedish national computer networking (SNCNW 2015), Karlstad, May 28–29 2015, no Sncnw, pp 1–5 [Online]. http://www.diva-portal.org/smash/record.jsf?pid=diva2:817379
Jarraya Y, Madi T, Debbabi M (2014) A survey and a layered taxonomy of software-defined networking. IEEE Commun Surv Tutor 16(4):1955–1980. https://doi.org/10.1109/COMST.2014.2320094
Kaliyamurthy NM, Taterh S, Shanmugasundaram S, Saxena A, Cheikhrouhou O, Ben Elhadj H (2021) Software-defined networking: an evolving network architecture—programmability and security perspective. Secur Commun Netw 2021. https://doi.org/10.1155/2021/9971705
Keti F (2015) Emulation of software defined networks using mininet in different simulation environments, pp 205–210. https://doi.org/10.1109/ISMS.2015.46
Khan S, Ali M, Sher N, Asim Y, Naeem W, Kamran M (2016) Software-defined networks (SDNs) and internet of things (IoTs): a qualitative prediction for 2020. Int J Adv Comput Sci Appl 7(11):385–404. https://doi.org/10.14569/ijacsa.2016.071151
Khan Tayyaba S, Shah MA (2017) 5G cellular network integration with SDN: challenges, issues and beyond. In: Proceedings of the 2017 international conference on communication, computing and digital systems, C-CODE 2017, no December 2018, pp 48–53. https://doi.org/10.1109/C-CODE.2017.7918900
Kobo HI, Abu-Mahfouz AM, Hancke GP (2017) A survey on software-defined wireless sensor networks: challenges and design requirements. IEEE Access 5:1872–1899. https://doi.org/10.1109/ACCESS.2017.2666200
Lal N, Tiwari SM, Khare D, Saxena M (2021) Prospects for handling 5G network security: challenges, recommendations and future directions. J Phys Conf Ser 1714(1). https://doi.org/10.1088/1742-6596/1714/1/012052
Li Y, Zhang D, Taheri J, Li K (2018) SDN components and OpenFlow, pp 49–68
Llorens-Carrodeguas A, Cervello-Pastor C, Leyva-Pupo I, Lopez-Soler JM, Navarro-Ortiz J, Exposito-Arenas JA (2018) An architecture for the 5G control plane based on SDN and data distribution service. In: 2018 5th International conference on software defined systems, SDS 2018, pp 105–111. https://doi.org/10.1109/SDS.2018.8370430
Morgado A, Huq KMS, Mumtaz S, Rodriguez J (2018) A survey of 5G technologies: regulatory, standardization and industrial perspectives. Digit Commun Netw 4(2):87–97. https://doi.org/10.1016/j.dcan.2017.09.010
Naseer MZ (2016) Modeling control traffic in distributed software defined networks
Nevala C (2016) Mobility management for software defined wireless sensor networks. Mälardalen Univ

Oktian YE, Lee SG, Lee HJ, Lam JH (2017) Distributed SDN controller system: a survey on design choice. Comput Netw 121:100–111. https://doi.org/10.1016/j.comnet.2017.04.038

Prados-Garzon J, Adamuz-Hinojosa O, Ameigeiras P, Ramos-Munoz JJ, Andres-Maldonado P, Lopez-Soler JM (2016) Handover implementation in a 5G SDN-based mobile network architecture. In: IEEE International symposium on personal, indoor and mobile radio communications PIMRC. https://doi.org/10.1109/PIMRC.2016.7794936

Shalimov A, Zimarina D, Pashkov V (2013) Advanced study of SDN/OpenFlow controllers

Sharma A, Kumar R, Mansotra V (2016) Proposed stemming algorithm for Hindi information retrieval. Int J Innov Res Comput Commun Eng (An ISO Certif. Organ.) 3297(6):11449–11455. https://doi.org/10.15680/IJIRCCE.2016

Sitan I (2016) Proposed technologies for solving future 5G heterogeneous networks challenges. Int J Comput Appl 142(10):1–8. https://doi.org/10.5120/ijca2016909924

Syed-Yusof SK, Numan PE, Yusof KM, Din JB, Bin Marsono MN, Onumanyi AJ (2020) Software-defined networking (SDN) and 5G network: the role of controller placement for scalable control plane. In: 2020 IEEE International RF and microwave conference (RFM), pp 1–6. https://doi.org/10.1109/RFM50841.2020.9344741

Tadros CN, Mokhtar B, Rizk MRM (2018) Software defined network based management framework for wireless sensor networks. In: 2018 IEEE 9th annual information technology, electronics and mobile communication conference, IEMCON 2018, pp 1200–1205. https://doi.org/10.1109/IEMCON.2018.8615087

Tadros CN, Rizk MRM, Mokhtar BM (2020) Software defined network-based management for enhanced 5G network services. IEEE Access 8:53997–54008. https://doi.org/10.1109/ACCESS.2020.2980392

Xu Q, Zhao J (2016) A WSN architecture based on SDN, vol 64, no Isct, pp 159–163

Zhang GA, Gu JY, Bao ZH, Xu C, Zhang SB (2014) Joint routing and channel assignment algorithms in cognitive wireless mesh networks. Trans Emerg Telecommun Technol 25(3):294–307. https://doi.org/10.1002/ett

Zolanvari M (2015) SDN for 5G mobile wireless communication, no October, pp 1–17

Chapter 11
Reversible Logic Based Single Layer Flip Flops and Shift Registers in QCA Framework for the Application of Nano-communication

Rupali Singh and Pankaj Singh

Abstract Quantum Dot Cellular Automata (QCA) and reversible logic are promising paradigms which can effectively substitute the conventional Complementary Metal Oxide Semiconductor (CMOS) based circuits in near future. Sequential circuits form indispensable part of many computational devices. It is obligatory to design low power sequential circuits so as to enhance the performance of overall system. In network communication systems flip flops and registers are widely used wherever there is a need of packet storage. QCA and reversible logic ensure the power aware and nano-size storage devices for high performance communication system. This paper focuses on the design of reversible logic based sequential circuits such as shift registers and various types of D flip flops in QCA framework. The presented circuits are designed using efficient, novel and power aware QCA layout of Fredkin gate with improved performance metrics. The proposed novel QCA layout of reversible Fredkin gate exhibits 54% better cost function and 16% lesser energy dissipation than the existing popular efficient designs. The presented QCA structure of Fredkin gate is extensively tested for associated defects and has found 66.93% fault tolerant. Considering the realistic clock distribution, the QCA layout of Fredkin gate is investigated under 2D clocking scheme. The presented gate is further utilized to realize reversible logic based sequential registers and master slave D flip flop which are unique with novel QCA architecture and not reported earlier in the literature.

Keywords QCA · Reversible logic · D latch · Shift registers · Energy dissipation · Fault tolerance

---

R. Singh (✉)
Department of Electronics and Communication Engineering, Delhi NCR Campus, SRM Institute of Science and Technology, Ghaziabad, UP, India
e-mail: rupal.rishi@gmail.com

P. Singh
School of Engineering and Technology, IIMT University, Meerut, UP, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

A. Rai et al. (eds.), *Paradigms of Smart and Intelligent Communication, 5G and Beyond*, Transactions on Computer Systems and Networks,
https://doi.org/10.1007/978-981-99-0109-8_11

11.1 Introduction

CMOS industry is continuously striving to increase the circuit density, reduce power and enhance the speed of the chips. The scientists and researchers are working to subside the limitations of CMOS technology and are trying to find its potential replacement. Moreover, it has been reported that irreversible circuits dissipate high heat energy due to loss of information (Landauer 1961). It was proved that every bit of data lost dissipates KT ln2 joules of heat. Later, it was reported that the issue can be resolved by introducing reversibility in the circuit (Bennett 1973). If the irreversible gates are replaced by their reversible counterparts then this power dissipation can be made zero ideally. A reversible gate is one which has distinctive output assignment for each input assignment. Apart from the benchmark reversible gates (Toffoli 1980; Fredkin and Toffoli 1982; Peres 1985), several other reversible gates have been presented to investigate the properties and prospects of reversible logic while designing combinational and sequential circuits (Noorallahzadeh and Mosleh 2019; Sasamal et al. 2018; Pain et al. 2019; Singh and Sharma 2020a).

The reversible circuits can be extensively implemented in various applications such as quantum computing, DNA computing, optical networks etc. However, QCA—a nanotechnology paradigm, appears to be the most appropriate platform for reversible circuits. QCA is a technology which has shown its supremacy over existing circuits in terms of size, speed and power handling capability. Moreover, the electronic blocks such as latches and shift registers are decisive components in many computing devices. Also, the power consumption is utmost concern for these devices. If the circuits are implemented using reversible logic, the power dissipation reduces to a large extent when compared with the conventional circuits. Hence, the architectural design of reversible logic based sequential circuits is targeted in this paper. This paper makes use of the blend of reversible logic and QCA to target power aware sequential circuits. Several studies have been reported showing the designs of QCA latches (Seyedi et al. 2019; Hao and Chen 2019), QCA registers (Prakash et al. 2019; Abutaleb 2018), QCA memory structures (Sadhu et al. 2019; Fam and Navimipour 2019; Majeed et al. 2019) etc. Moreover, QCA circuits employing reversible logic have been also explored in the literature for reversible sequential modules such as latches (Chabi et al. 2017; Bhoi et al. 2017; Singh and Sharma 2020b) and RAM cell (Singh and Sharma 2020a). Yet, a number of reversible circuits such as registers, memory structures and dual edge triggered flip flops are not much discussed and thus, need thorough investigation.

Fredkin gate is one of the most efficient and resilient gates in reversible logic family. The gate can be utilized effectively to implement any reversible circuit. In this article, a unique QCA layout of Fredkin gate is configured with optimized performance metrics. This QCA layout is further utilized to realize reversible circuits such as latch, master slave D flip flop, dual edge triggered (DET) D flip flop, and shift registers. The proposed reversible circuits are then investigated for different QCA parameters and energy dissipation. The computational functionality of the proposed circuits is obtained for temperature variations. This research article is divided into

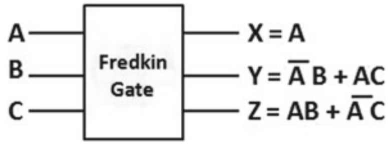


Fig. 11.1 Fredkin gate

six sections. Section 11.2 gives the basic concepts of QCA and reversible logic. The intended compact QCA layout of Fredkin gate with its cost analysis, fault characterization and power analysis is discussed in Sect. 11.3. Section 11.4 depicts the design and realization of proposed reversible circuits, Sect. 11.5 provides comprehensive assessment of the proposed circuits and the conclusion is given in Sect. 11.6.

11.2 Preliminary Overview

This section gives important insights on reversible logic and Quantum Dot Cellular Automata.

11.2.1 Reversible Logic

Reversible logic makes use of n × n function which has n distinct input and n distinct output assignments. This reversible function is used to design reversible logic gate which is designed with equal input–output vectors which are mapped bijectively. Subsequently, the output vectors can be used to recreate input vectors (Fredkin and Toffoli 1982). This can be done by two ways: either by employing logical reversibility or by attaining a physical reversibility in the circuits. Logical reversibility can be attained by using reversible gates in a reversible circuit. Fredkin gate is one of such conventional reversible gates which can be used to design reversible circuits. A conservative reversible Fredkin gate which is used in this paper to realize the proposed reversible circuits is illustrated in Fig. 11.1.

11.2.2 QCA Background

In QCA, the elementary unit is a cell. The corner positions of a cell occupy four quantum dots and two free electrons are placed diagonally (Lent and Tougaw 1997). Based on two positions of electrons, two states of polarization take place, which are encoded to logic "0" and logic "1" (Porod et al. 1999) as shown in Fig. 11.2a.

The preliminary structures used in QCA design are majority voter and an inverter. The output of majority gate, as depicted in Fig. 11.2b, follows the status of majority

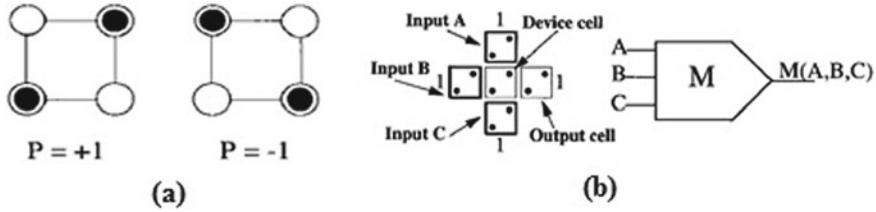


Fig. 11.2 **a** Quantum cell with polarization **b** Majority voter gate

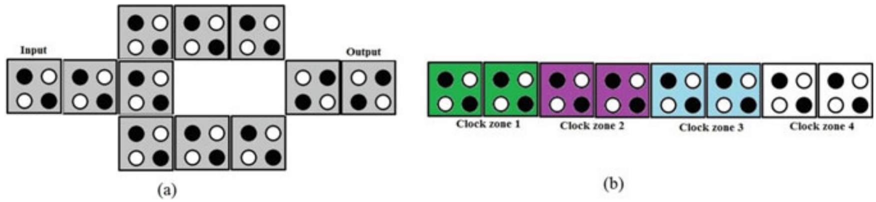


Fig. 11.3 **a** QCA inverter **b** QCA Clock zones

of input assignments and changes polarization accordingly. A majority gate can be used to realize basic logic functions such as logical AND and logical OR by making one of its inputs logic 0 or 1 respectively. As illustrated in the Fig. 11.3a if the placement of the QCA cells is at 45° then the input is complemented at the output realizing an inverter.

Further, clocking is essentially used in QCA circuits to direct the information flow. To implement clocking in QCA circuits, adiabatic switching is normally used (Lent and Tougaw 1997). This method makes use of adiabatic clock signal which has four phases. Figure 11.3b shows four clocking zones normally used to drive the QCA circuit. In each zone the clocking signal is 90º phase shifted. These clocking zones ensure the sequential computation in a circuit.

11.3 QCA Layout of Reversible Fredkin Gate—A Novel Approach

Fredkin gate is the most efficient conservative gate for design of reversible circuits. This section proposes the compact QCA structure of conservative Fredkin gate which is further used as a part of different reversible circuits discussed here. The design and simulation of QCA circuits is carried out using QCA designer 2.0.3 (Walus et al. 2004). The QCA layout of Fredkin gate is implemented as appeared in Fig. 11.4. It consists of two multiplexer functions for which the compact QCA layout is used as reported in Chabi et al. (2017). The presented QCA architecture of Fredkin gate is novel and optimized in many performance parameters. Table 11.1 describes

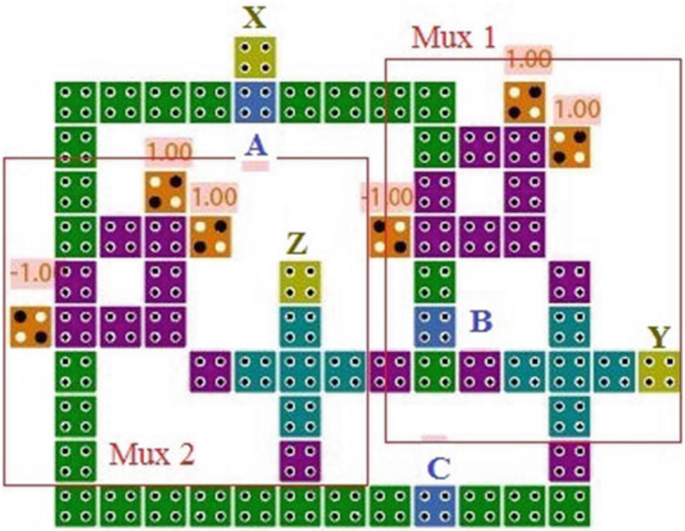


Fig. 11.4 Proposed QCA structure of Fredkin gate

the performance assessment of the proposed QCA structure which implicates that the intended QCA architecture is superior as compared to the existing layouts of reversible gates.

Table 11.1 estimates the cost function (CF), cell area (CA) and area usage (Au) of each gate by using Eqs. (11.1)–(11.3).

$$\mathrm{CF} = \mathrm{Qc} \times \mathrm{TA} \times \mathrm{D} \tag{11.1}$$

$$\mathrm{CA} = \mathrm{Nc} \times \text{ area of each cell } (18\,\mathrm{nm} \times 18\ \mathrm{nm}) \tag{11.2}$$

$$\mathrm{Au} = \mathrm{CA}/\ \mathrm{TA} \tag{11.3}$$

Here, Qc is QCA cell count, Nc is number of cells and D is delay. The delay is obtained by considering the clock zones used in the design. If a circuit uses three clock zones the delay is calculated as 0.75 clock. The suggested layout is cost efficient showing 54.56% better cost function, 22.2% lesser area and reduced delay as compared to the most efficient design reported earlier. Moreover, the reliability of QCA circuit is a measure of area usage. More the area usage more is the reliability or robustness of the circuit. Table 11.1 depicts that the area usage of the proposed QCA layout is more and hence the reliability. The designed QCA layout is compact and has revealed better performance as compared to previous Fredkin gate designs and other reversible gates in instances such as cell count, total area and cell area. circuit. Table 11.1 depicts that the area usage of the proposed QCA layout is more and hence the reliability. The designed QCA layout is compact and has revealed better performance

Table 11.1 Cost analysis of proposed Fredkin QCA design

Reversible gates	*D* (clock cycle)	Qc	CA in μm^2	TA in (μm^2)	AU	CF	% Improvement in CF of proposed layout
Thapliyal et al. (2013)	1	246	0.079	0.37	21.35	91.02	95.97
Ma et al. (2008)	0.5	147	0.047	0.16	29.3	11.76	68.7
Thapliyal and Ranganathan (2009)	0.5	117	0.037	0.11	33.63	6.435	42.81
Kumar et al. (2017)	1	177	0.057	0.24	23.27	42.48	91.33
Bhoi et al. (2017)	1	109	0.035	0.11	31.81	11.99	69.3
Chabi et al. (2017)	1	90	0.029	0.09	32.22	8.1	54.56
Mohammadi et al. (2019)	1.5	105	0.034	0.09	37.77	14.17	74.02
Singh and Sharma (2020b)	1	85	0.027	0.1	27	8.5	56.7
Proposed Fredkin design	0.75	70	0.022	0.07	31.42	3.68	–

as compared to previous Fredkin gate designs and other reversible gates in instances such as cell count, total area and cell area.

11.3.1 Fault Characterization

The fault characterization is performed for the proposed QCA layout to obtain its fault tolerance for various defects that may occur during fabrication. During the deposition phase, defects may occur in the QCA circuits which fall in three categories viz cell displacement, omission and extra cell deposition (Tahoori et al. 2004). By assuming all these defects various fault patterns have been obtained for the novel QCA architecture of Fredkin gate as illustrated in Table 11.2. Around fourteen exclusive fault patterns are attained and fault tolerance (FT%) percentage is calculated. In Table 11.2, input and expected output assignments (Vc0–Vc7) for the Fredkin gate are depicted in first two columns followed by fourteen fault patterns (F1–F14). The bold vectors illustrate faulty output and thus fault tolerance is estimated for each

Table 11.2 Fault characterization of presented QCA structure

F1–F8									
Input vctor	Output vctor	F1	F2	F3	F4	F5	F6	F7	F8
Vc0	Vc0	Vc0	Vc0	Vc0	Vc0	Vc0	Vc0	Vc0	Vc0
Vc1	Vc1	**Vc3**	**Vc3**	Vc1	Vc1	**Vc0**	Vc1	Vc1	**Vc0**
Vc2	Vc2	**Vc0**	Vc2	Vc2	Vc2	Vc2	**Vc3**	Vc2	**Vc3**
Vc3	Vc3	Vc3	Vc3	Vc3	Vc3	Vc3	Vc3	Vc3	Vc3
Vc4	Vc4	Vc4	Vc4	Vc4	Vc6	Vc4	Vc4	Vc4	Vc4
Vc5	Vc6	**Vc4**	Vc6	**Vc4**	Vc6	Vc6	**Vc7**	**Vc7**	Vc6
Vc6	Vc5	**Vc7**	**Vc7**	**Vc7**	Vc5	Vc5	Vc5	**Vc4**	Vc5
Vc7	Vc7	Vc7	Vc7	Vc7	**Vc5**	Vc7	Vc7	Vc7	Vc7
F9–F 14									
Input vctor	Output vctor	F9	F10	F11	F12	F13	F14	%FT	
Vc0	Vc0	Vc0	Vc0	Vc0	Vc0	Vc0	Vc0	100%	
Vc1	Vc1	**Vc0**	**Vc3**	Vc1	**Vc3**	**Vc3**	Vc1	42.80%	
Vc2	Vc2	**Vc3**	**Vc1**	**Vc3**	Vc2	**Vc0**	**Vc3**	42.80%	
Vc3	Vc3	Vc3	Vc3	Vc3	Vc3	Vc3	Vc3	100%	
Vc4	Vc4	Vc4	Vc4	Vc4	Vc4	**Vc7**	Vc4	85.70%	
Vc5	Vc6	**Vc7**	**Vc4**	Vc6	Vc6	**Vc7**	**Vc7**	42.80%	
Vc6	Vc5	**Vc4**	**Vc7**	**Vc4**	**Vc7**	**Vc4**	Vc5	35.70%	
Vc7	Vc7	Vc7	Vc7	Vc7	Vc7	**Vc4**	Vc7	85.70%	

input–output vector set. The average fault tolerance of the suggested Fredkin QCA layout is 66.93% which is better than many earlier reported fault tolerant gates (Bhoi et al. 2017; Sen et al. 2014a, b).

11.3.2 Energy Dissipation Analysis of the Presented QCA Structure

Estimation of power dissipation across the QCA circuits is critical aspect to be discussed. Timler and Lent came up with the explicit model of power dissipation in QCA (Timler and Lent 2002). The total power of a QCA cell can be obtained as given by using Eq. (11.4).

$$Pt = \frac{d}{dt}E = \frac{\hbar}{2}\left(\frac{d}{dt}\Gamma\right)\lambda + \frac{\hbar}{2}\Gamma\left(\frac{d}{dt}\lambda\right) = P1 + P2 \tag{11.4}$$

Here, $P1$ suggests the power gain and $P2$ is the power consumed by the cell. The tool for power estimation, QCA Pro, was introduced in Ref. (Srivastava et al. 2011) using

Table 11.3 Power dissipation of the proposed QCA structure

Reversible gates	Average switching energy (meV)			Average leakage energy (meV)			Total energy dissipation (meV)			% Improvement in proposed QCA layout
	0.5Ek	1.0Ek	1.5Ek	0.5Ek	1.0Ek	1.5Ek	0.5Ek	1.0Ek	1.5Ek	1.0Ek (for total energy dissipation) (%)
Fredkin and Toffoli (1982)	213	175	143	101	283	483	314	458	625	68.90
Kumar et al. (2017)	194	227	191	105	293	506	299	520	697	72.10
Roohi et al. (2018)	193	166	141	58	168	294	251	334	435	57.30
Bhoi et al. (2017)	109	93	78	52	154	272	161	247	350	42.50
Singh and Sharma (2020b)	93.7	80	67.6	32	89.2	153.2	125	169	220	16
Proposed QCA layout of Fredkin Gate	90.59	78.5	66.8	21	64	114	111	142	180.4	–

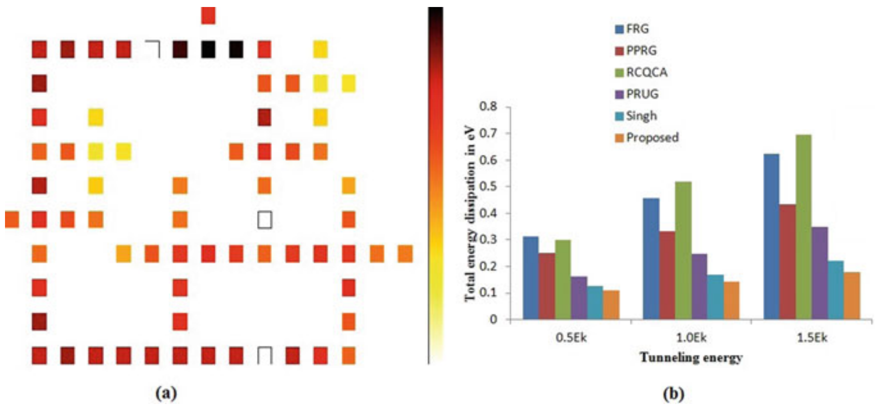


Fig. 11.5 **a** Energy map of proposed QCA layout of Fredkin gate at 1.0Ek **b** Total energy dissipation for various reversible gate layouts

upper bound model. The proposed QCA layout is simulated on the QCA Pro tool and energy dissipation results have been computed for standard tunneling energies.

Figure 11.5a shows the energy dissipation plot of presented QCA gate at 1.0EK. Table 11.3 shows that average switching energy, average leakage energy and their sum as total energy. The total energy of designed gate is lower than the earlier reported reversible QCA layouts. The percentage improvement in whole energy dissipation of the presented design at 1.0Ek with respect to existing reversible QCA layouts is estimated in Table 11.3. The energy dissipation of presented QCA structure is explicitly less as compared to the earlier designs. Figure 11.5b indicates the graph showing total energy dissipation across the proposed layout and referred gates at three different tunneling energies.

11.3.3 QCA Layout of Fredkin Gate with 2D Clocking Scheme

Apart from the proposed novel QCA layout, the realistic clock realization of Fredkin gate is discussed in this section. Hennessy et al. presented the realistic approach for clock distribution in QCA circuits. The arrangement of QCA circuit and its underlying clocking mechanism is depicted in Hennessy and Lent (2001). Few researchers have shown diverse clocking schemes to implement QCA layouts with realistic clock distribution. These schemes include Landaur clocking (1961), Bennett clocking (Lent et al. 2006), Niemier trapezoidal clocking (2001), 2D clocking (Vankamamidi et al. 2008), USE clocking (Campos et al. 2016), clocking scheme for regular cells (Nath et al. 2017) etc. Here, the QCA layout of Fredkin gate is redrawn using 2 D clocking scheme to investigate its performance in realistic clock distribution. Figure 11.5a shows the two dimensional (2D) clocking scheme. It depicts the square shaped

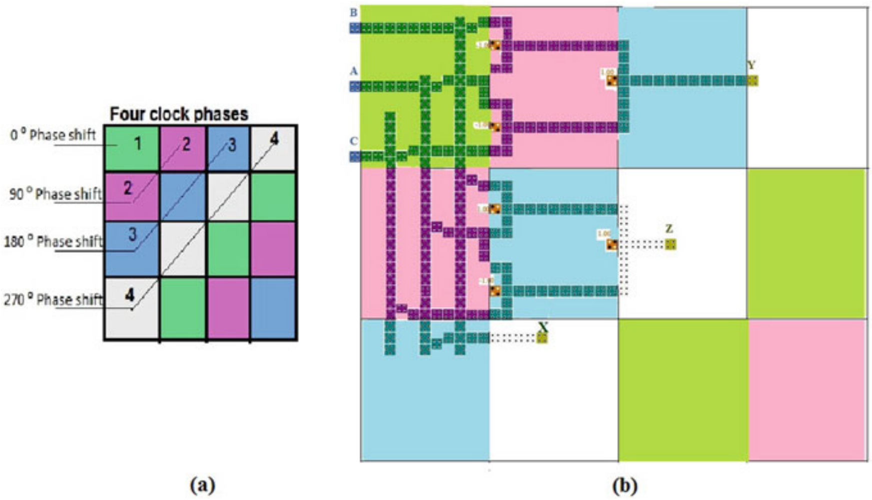


Fig. 11.6 **a** Mechanism in 2D clocking scheme **b** QCA Realization of Fredkin gate in 2D clocking scheme

clocking zones organized in an array such that similar clock zones lie on diagonals. It indicates that the underlying clocking wires run diagonally to provide necessary phase shifts in the clocking zones. Figure 11.6b depicts the Fredkin gate layout in 2D clocking scheme. It occupies 0.41 μm^2 total area, 0.074 μm^2 cell area, and 230 QCA cells.

11.4 Proposed Reversible QCA Circuits

This section presents the design and realization of QCA based, reversible sequential circuits. Initially, D latch circuit is designed using the proposed cost efficient, power aware and fault tolerant Fredkin based QCA layout. Further, the reversible QCA circuits of master slave D flip flop, DET flip flop and shift registers are designed using the proposed reversible D latch.

11.4.1 QCA Based Reversible D Latch

This section portrays the configuration of reversible D latch which is one of the rudimentary components of reversible computing circuits. Many sequential elements are based on D latch circuit. Hence it is important to design optimized, high performance D latch. The standard expression to represent input–output relation of D latch is represented by QN = DE + E’Q where D is input, E is Enable, Q is previous state output and QN is present state output. The schematic of Fredkin gate based D latch and QCA architecture is depicted in Fig. 11.7a, b respectively. The circuit simulation is shown in Fig. 11.7c to validate correctness of the circuit. In Fig. 11.7c, it can be seen that when E = 0 then QN = D and when E = 1 then QN = Q i.e. the output follows the previous state. Thus, QCA circuit behaves as a D latch with the latency of 0.75 clocks (3 zones). The structural examination of the suggested D latch is given in Table 11.4. The cost function (CF) is estimated by using Eq. (11.5).

$$\text{CF} = \text{number of gates} \times \text{area} \times \text{delay} \tag{11.5}$$

The presented design of latch is improvised by 20% in cost function, 20% in area as compared to the most cost efficient design reported earlier.

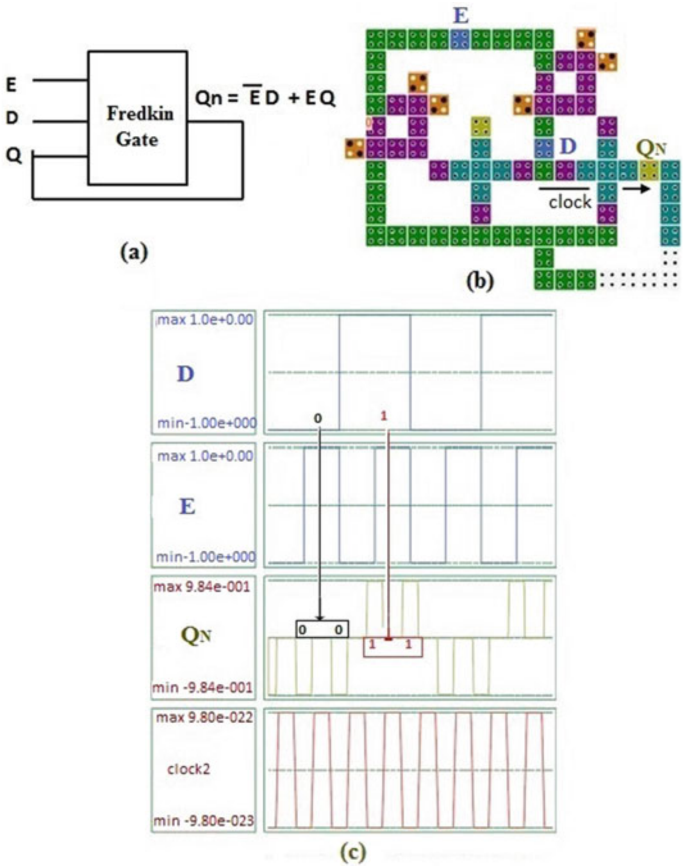


Fig. 11.7 **a** Schematic of Fredkin based D latch **b** proposed QCA layout **c** simulation waveform

Table 11.4 Cost assessment of proposed reversible D latch

Reversible D Latch	# Of gates	Latency (clock)	TA (μm)^2	CF	% Improvement in CF of proposed QCA structure (%)
Thapliyal et al. (2013)	1	1	0.37	0.37	83
Chabi et al. (2017)	1	1	0.11	0.11	45
Kumar (2017)	1	2.25	0.24	0.54	88
Bhoi (2017)	1	1	0.14	0.14	57
Mohammadi (2019)	1	2	0.14	0.28	78.57
Singh and Sharma (2020a, b)	1	0.75	0.1	0.075	20
Proposed D latch layout	1	0.75	0.08	0.06	–

11.4.2 QCA Based Reversible Master Slave D Flip Flop

Further, the structure of reversible master slave flip flop is taken into consideration using Fredkin based latch. It uses two latches, one as a master and the other as a slave as depicted in Fig. 11.8. The input D and clock signal CLK is applied to master flip flop while the slave flip flop is connected with the output of master flip flop as well as inverted clock. The simulation waveforms are presented in Fig. 11.9. The final output is obtained at Qn+1. The latency of the circuit is 1.75 clock as marked by the arrow. The output Qn+1 follows input D after 1.75 clocks as depicted in Fig. 11.8.

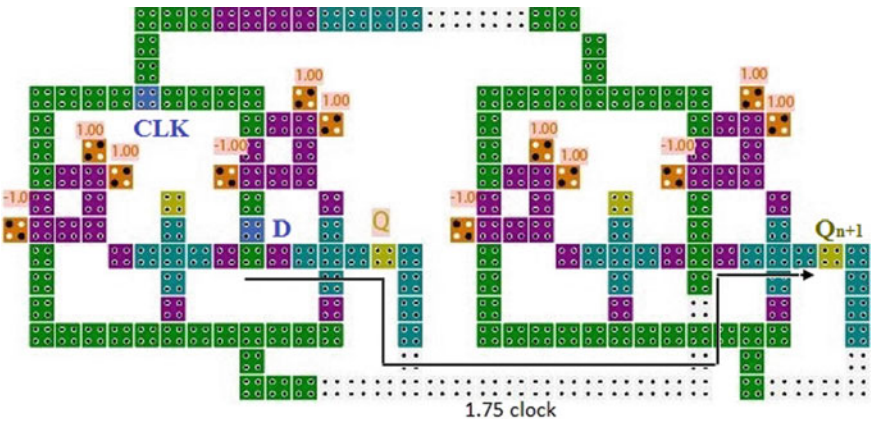


Fig. 11.8 Proposed QCA structure of reversible master slave flip flop

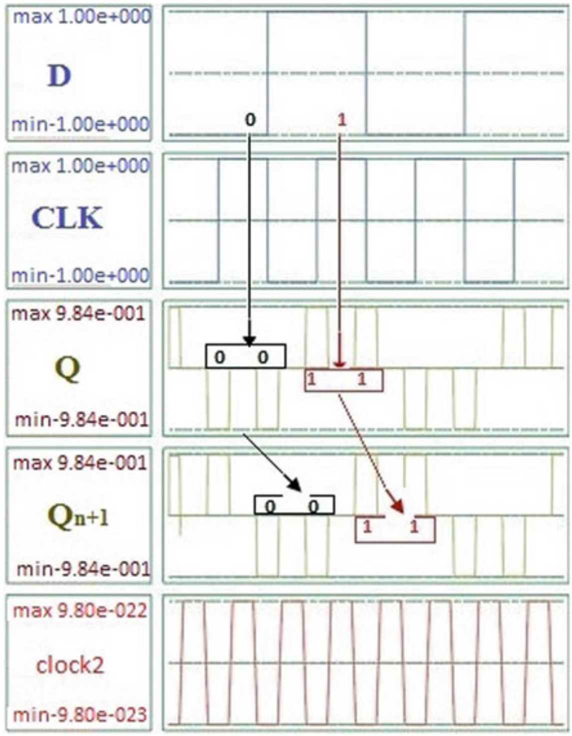


Fig. 11.9 Simulation waveform of proposed reversible master slave flip

11.4.3 QCA Based Reversible DET Flip Flop

DET flip flop consists of two reversible D latches and one multiplexer formed using the proposed Fredkin layout as shown in Fig. 11.10. The negative edge triggered latch produces output Q1 = D when CLK = 0 and positive edge triggered latch produces output Q2 = D when CLK = 1. Both the outputs are then given to multiplexer which selects either Q1 or Q2 depending on the clock edge. Thus final output QDET follows input D at positive as well as the negative edge of the clock. It gives high throughput as compared to conventional SET flip flop. The simulation waveform is depicted in Fig. 11.11 which verifies the accuracy of the proposed QCA circuit. The delay taken up by the proposed reversible DET is 1.75 Clocks.

11.4.4 Design of Proposed Reversible Shift Registers

Many computing machines make use of a shift register as one of its vital components. Thus, it is essential to construct power aware, area efficient and fault resistant reversible shift register. Few QCA based shift registers have been reported in

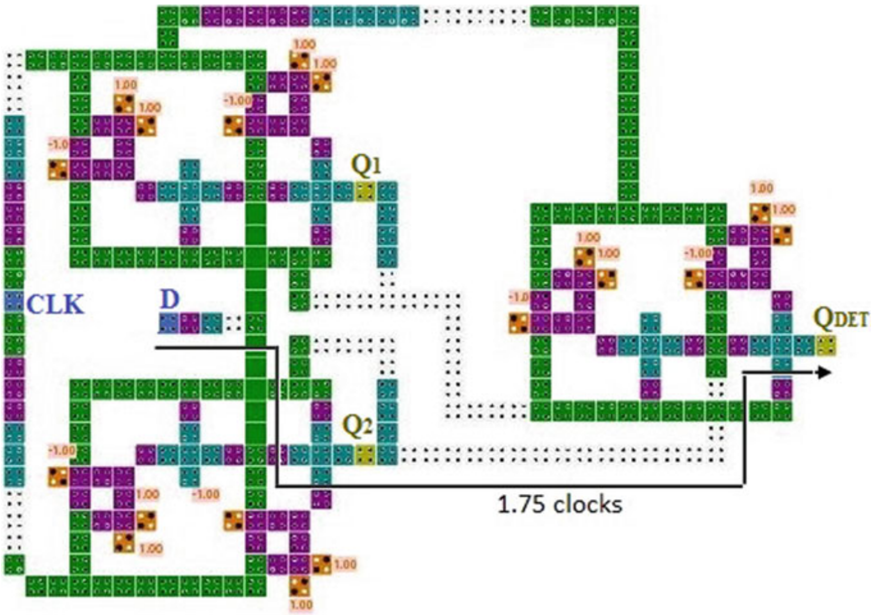


Fig. 11.10 Proposed QCA structure of reversible DET flip flop

literature (Prakash et al. 2019; Divshali et al. 2018; Das and De 2017). But, the design of reversible logic based shift register is not yet reported in the literature. This section describes the structures of QCA based reversible, 4 bit shift registers using the proposed layout of Fredkin based D latch. Parallel in parallel out (PIPO), serial in serial out (SISO) and serial in parallel out (SIPO) shit registers are exhibited here.

Firstly, the QCA based PIPO register is presented in Fig. 11.12. The circuit consists of assembly of 4 D latches. Four inputs are applied to four latches as D1, D2, D3, D4 and the outputs are from all the latches separately as Q1, Q2, Q3, Q4. The common clock is applied to all the latches. The latency of each output is 0.75 clock (3 zones).

The simulation waveforms are illustrated in Fig. 11.13 which validates the correctness of the circuit. The waveform verifies that the output from each latch follows its corresponding input. Secondly, the QCA based SISO shift register is presented in Fig. 11.14. The circuit depicts four D latches cascaded so that the output of first D latch is attached to the input of next D latch and so on. The input provided to the first D latch is D and the output taken from the last latch is Qn. As the input D is travelled through four latches, each with a clock latency of 0.75 clock, the final output Qn is produced after 3.75 clock (15 zones) as marked in Fig. 11.14. The simulations are exhibited in Fig. 11.15. The simulation of reversible SISO register explains that the output Qn is identical to input D but delayed or shifted by 3.75 clock cycles.

Lastly, the design of QCA based reversible SIPO shift register is presented as illustrated in Fig. 11.16. This QCA circuit also consists of 4 units of proposed D latches. As the circuit has serial input, only the first D latch is assigned with the input

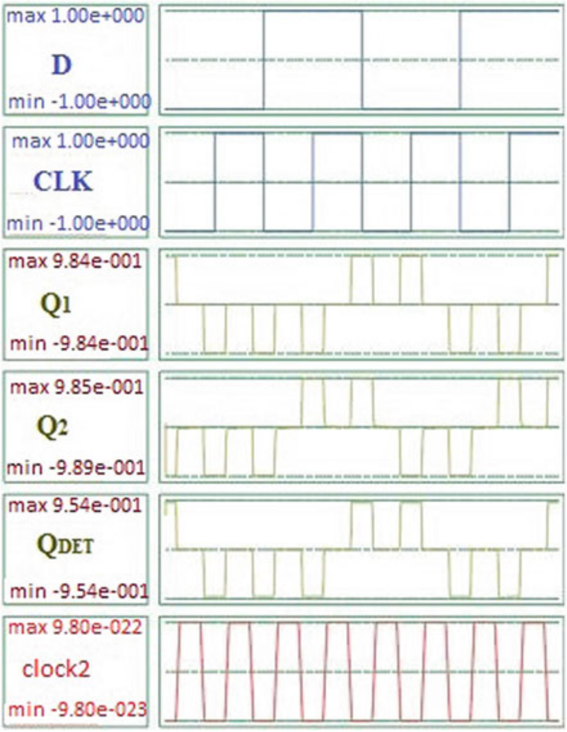


Fig. 11.11 Simulation waveform of presented reversible DET flip flop

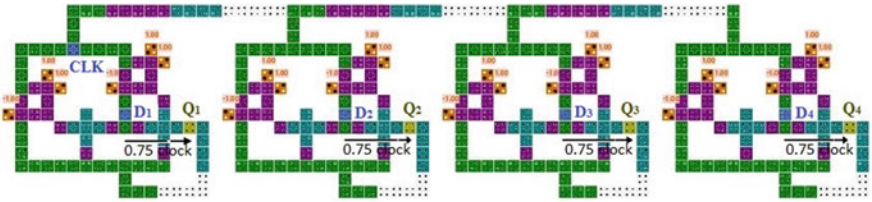


Fig. 11.12 QCA layout of proposed reversible PIPO shift register

D and the output of each latch is absorbed as input for the subsequent latch. Further, to have a parallel output, the output is taken from each latch as Q_1, Q_2, Q_3 and Q_4. As the input is travelled from first latch to forth latch, the latency of each output is different. The latency of the output Q_1 is 0.75 clock (3 zones) which is then fed to the second latch as input.

Thus, the second output Q2 has latency of 1.75 clock (7 zones). Similarly, the outputs Q3 and Q4 have latencies of 2.75 clocks (11 zones) and 3.75 clock (15 zones) respectively as shown by arrows (C1, C2, C3, C4) in Fig. 11.16. The output

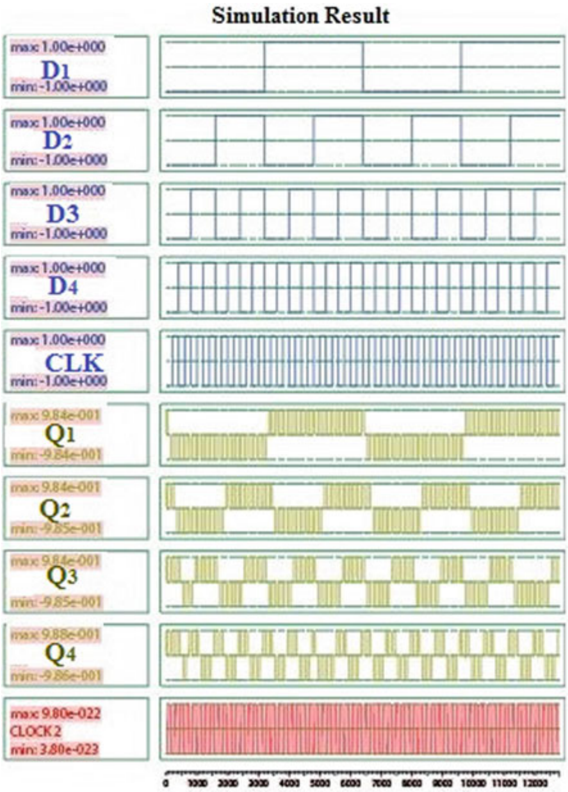


Fig. 11.13 Simulation waveform of proposed reversible PIPO shift register

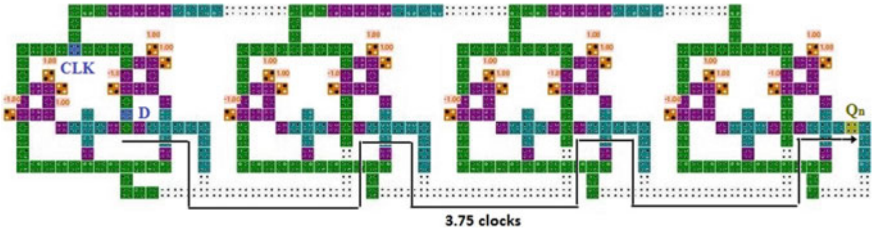


Fig. 11.14 Proposed reversible SISO shift register in QCA framework

waveform is validated from Fig. 11.17. It shows that each output follows the input D with specified delay.

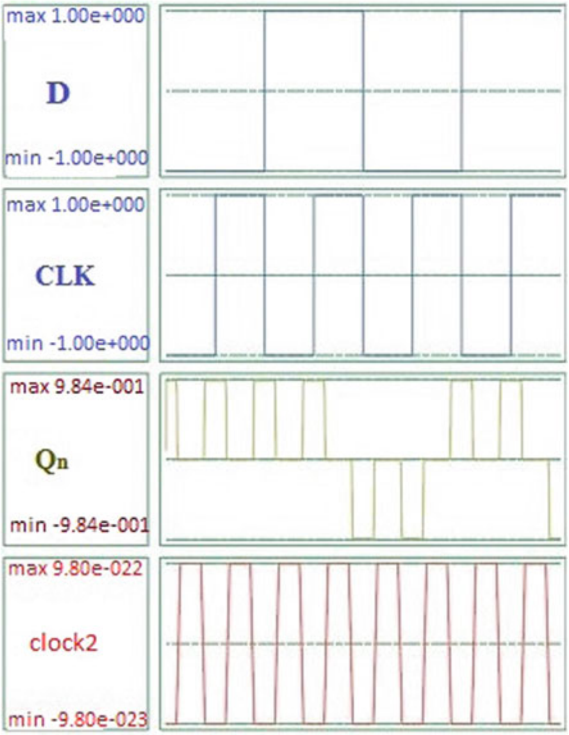


Fig. 11.15 Simulation waveform of proposed reversible SISO shift register

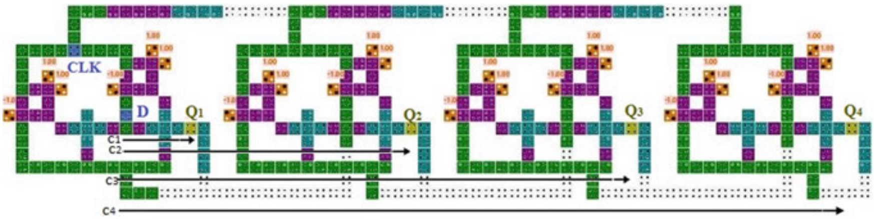


Fig. 11.16 QCA layout of proposed reversible SIPO shift register

11.5 Performance Analysis of Proposed Reversible QCA Circuits

The proposed circuits of shift registers are functioning in the desired manner. Now, the performance of these circuits can be tested on the basis of the decisive metrics such as cost function, energy dissipation and temperature dependencies. The QCA parameters of all the designed circuits are depicted in Table 11.5. In Table 11.5, critical path indicates the longest path between input and output terminals. The cost

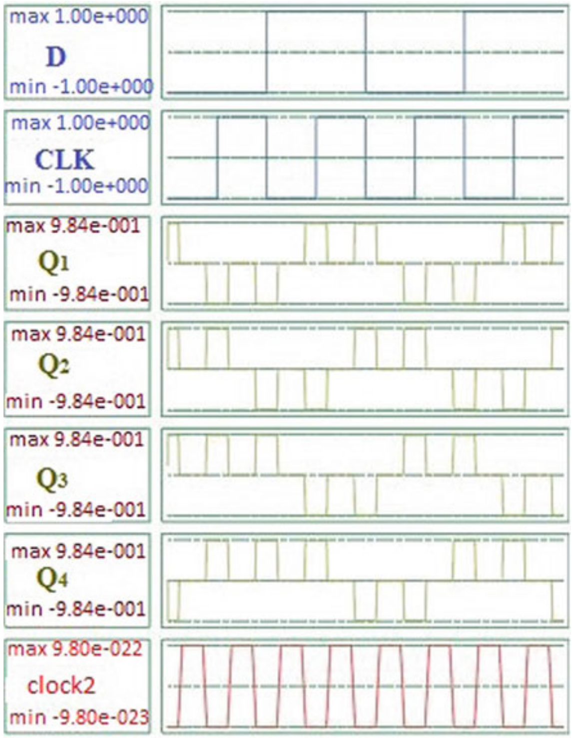


Fig. 11.17 Simulation waveform of proposed reversible SIPO shift register

function is estimated by taking a product of latency and an effective area of QCA layout. It is observed that the proposed QCA layout of reversible master slave D flip flop attains faster output response. The latency observed is 1.75 clocks which is very less as compared to the conventional QCA structures utilizing multiple reversible gates.

Also, the estimated cost function is 0.06 which is an optimum value. Furthermore, the proposed QCA structure of DET flip flop is optimally designed with 337 QCA cells, 0.43 μm^2 area and latency of 1.75 clocks. Thus, an improvement of 25.8% in area and 42% in latency is achieved for the proposed DET as compared to the layout in Singh and Pandey (2018). Also, 4 bit reversible shift registers such as SISO, PIPO, SIPO need 4 reversible gates, latency of 3.75 clocks and area of 0.4 μm^2. Thus, the cost function of 1.5, 0.3, and 1.5 is obtained for SISO, PIPO and SIPO respectively which are minimal values when compared with the cost function of the conventional structures.

The proposed reversible circuits are further tested for energy dissipation on QCA Pro. The total energy dissipation values for every proposed circuit are depicted in Table 11.6 at three different tunneling energies of 0.5Ek, 1.0Ek and 1.5Ek. The energy maps for D latch, master slave flip flop and PIPO register are shown in Fig. 11.18. The energy dissipation values for SIPO and SISO registers are found approximately same

Table 11.5 Performance assessment of presented reversible QCA circuits

Reversible gate	Category	# Of reversible gates used	# Of gates in critical path	Latency (clock cycles)	Cell count	Effective area (μm^2)	Cost function
Proposed reversible circuits	D Latch	1	1	0.75	82	0.08	0.06
	Master Slave D Flip flop	2	1	1.75	197	0.2	0.35
	DET D Flip flop	3	2	1.75	337	0.43	0.7525
	4 bit SIPO Register	4	4	3.75 (maximum)	420	0.4	1.5
	4 bit PIPO Register	4	1	0.75 (for each output)	384	0.4	0.3
	4 bit SISO Register	4	4	3.75	420	0.4	1.5

which are illustrated in Table 11.6. Figure 11.19 shows the graphical illustration of the total energy dissipation for each reversible circuit at different tunneling energies.

Further, all the intended reversible QCA circuits are tested for temperature dependencies. Mean output polarization of the output of each circuit at different temperatures is obtained on QCA designer using coherence vector engine and plotted as shown in Fig. 11.20. It can be seen from the graph that after 10° k the value of mean polarization for each output drops and get distorted. Hence, the proposed QCA circuits can be operated exclusively within 1–10° k. All the shift registers have shown the similar response for temperature variations and hence a single line is drawn to represent all the three shift registers.

11.6 Summary

This paper explores the design and optimization of reversible sequential circuits such as shift registers, master slave D flip flop, DET flip flop and D latch mapped in QCA. For designing these circuits, firstly, a novel and compact QCA structure of Fredkin gate is proposed. The intended QCA structure is investigated for various performance parameters and it is found that it has 54% improved cost function, 66.93% fault tolerance and 16% less energy dissipation than the available efficient reversible gates reported in the literature. Secondly, the novel architecture of Fredkin gate is used to design D latch which in turn is further used to design, master slave

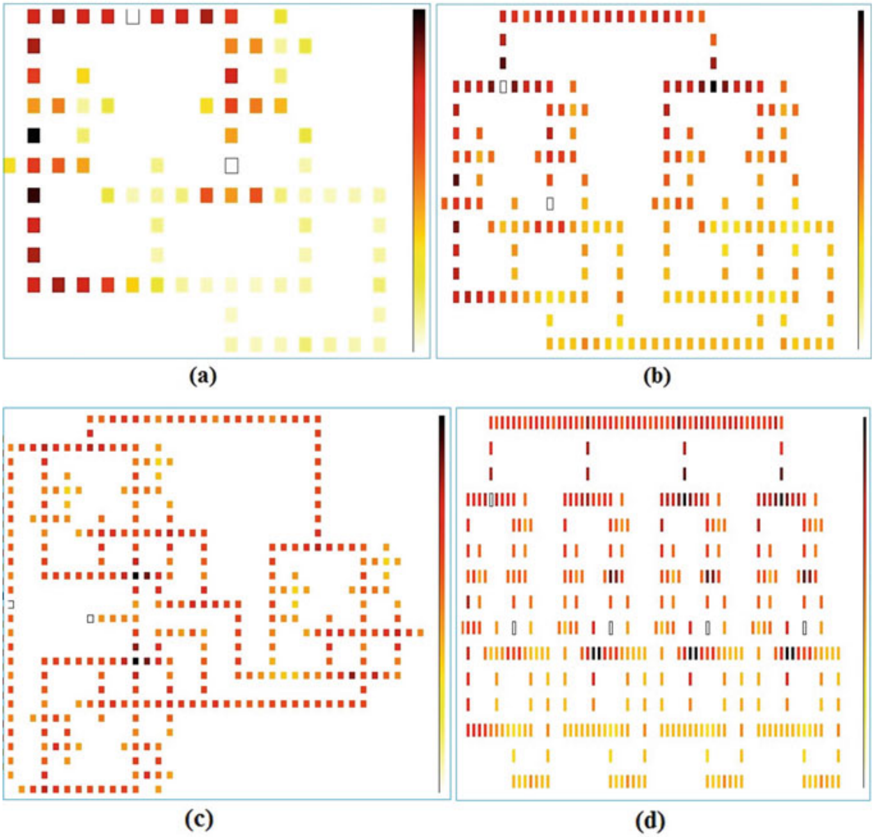


Fig. 11.18 Energy map of proposed **a** reversible D latch **b** reversible master slave flip flop **c** reversible DET flip flop **d** reversible PIPO register

flip flop and shift registers. Furthermore, comprehensive assessment of the proposed QCA circuits is exhibited for various QCA parameters and energy dissipation. It is found that the cost function of reversible master slave D flip flop, PIPO, SIPO and SISO shift registers are 0.35, 0.3, 1.5, 1.5 respectively and the total energy dissipation values are 0.4193, 0.8736, 0.8861, 0.8861 respectively at 1.5Ek tunneling energy. Moreover, the circuits are analyzed for temperature perturbations to study temperature dependencies. The investigation of reversible logic based shift registers and master slave D flip flop using QCA is novel and reported for the first time in the manuscript. All the circuits are optimized in respect of cell count, area, cost function and energy dissipation and hence can contribute extensively to the development of reversible computers. The circuits are found to be suitable for store and forward storage devices in nano-network communication.

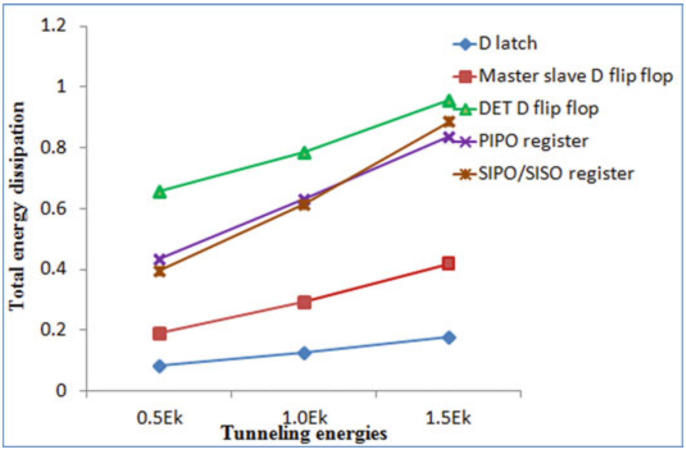


Fig. 11.19 Energy dissipation of proposed circuits at various tunneling

Table 11.6 Energy dissipation values for proposed circuits

Proposed reversible circuits	Average switching energy (meV)			Average leakage energy (meV)			Total energy dissipation (meV)		
	0.5Ek	1.0Ek	1.5Ek	0.5Ek	1.0Ek	1.5Ek	0.5Ek	1.0Ek	1.5Ek
D latch	59.3	52.1	44.7	24.7	74.1	132.6	84	126.1	177.2
Master slave D flip flop	130.6	114.5	98.3	58.3	177.7	321.1	188.8	292.2	419.3
DET D flip flop	560.4	491.4	425.5	96.1	29.5	53.2	656.5	786.3	957.8
PIPO register	315.9	275.7	235.8	117	355.1	637.7	432.9	630.8	873.6
SIPO/SISO register	273	240	206.2	120.8	373.7	679.9	393.8	613.7	886.1

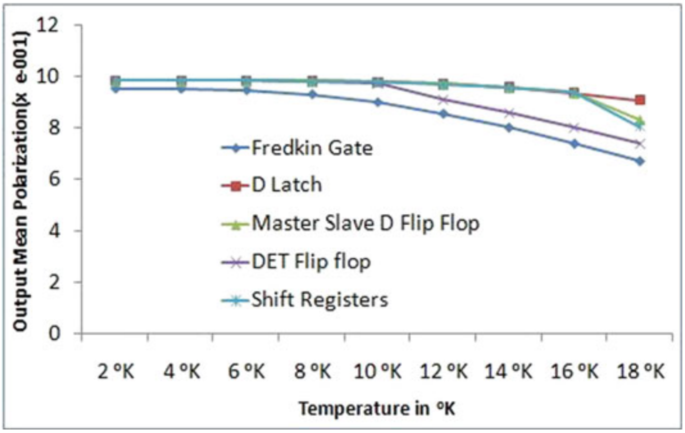


Fig. 11.20 Temperature dependencies of proposed QCA circuits

References

Abutaleb MM (2018) A Novel configurable flip flop design using inherent capabilities of quantum-dot cellular automata. Microprocess Microsyst 56:101–112

Bennett CH (1973) Logical reversibility of computation. IBM J Res Dev 17:525–532

Bhoi B, Misra NK, Pradhan M (2017) Design and evaluation of an efficient parity-preserving reversible qca gate with online testability. Cogent Eng 4:1–18

Campos CAT, Marciano AL, Neto OPV, Torres FS (2016) USE: a universal, scalable, and efficient clocking scheme for QCA. IEEE Trans Comput Des Integr Circuits Syst 35:513–517

Chabi AM, Roohi A, Khademolhosseini H, Sheikhfaal S, Angizi S, Navi K, Demara RF (2017) Towards ultra-efficient qca reversible circuits. Microprocess Microsyst 49:127–138

Das JC, De D (2017) Operational efficiency of novel siso shift register under thermal randomness in quantum dot cellular automata design. Microsyst Technol 23:4155–4168

Divshali MN, Rezai A, Karimi A (2018) Towards multilayer qca siso shift rgister based on efficient d-ff circuits. Int J Theor Phys 57:3326–3339

Fam SR, Navimipour NJ (2019) Design of a loop-based random access memory based on the nanoscale quantum dot cellular automata. Photonic Netw Commun 37:120–130

Fredkin E, Toffoli T (1982) Conservative logic. Int J Theor Phys 21:219–253

Hao P, Chen S (2019) Comparison of D-Flip-Flops and D-Latches: influence on set susceptibility of the clock distribution network. Nucl Sci Tech 30:1–10

Hennessy K, Lent CS (2001) Clocking of molecular quantum-dot cellular automata. J Vac Sci Technol B Microelectron Nanom Struct 19:1752–1755

Kumar N, Wairya S, Sen B (2017) Design of conservative reversible sequential logic for cost efficient emerging nano circuits with enhanced testability. Ain Shams Eng J 9:2027–2037

Landauer R (1961) Irreversibility and heat generation in the computing process. IBM J Res Dev 5:183–191

Lent CS, Tougaw PD (1997) A device architecture for computing with quantum dots. Proc IEEE 85:541–557

Lent CS, Liu M, Lu Y (2006) Bennett clocking of quantum-dot cellular automata and the limits to binary logic scaling. Nanotechnology 17:4240–4251

Ma X, Huang J, Metra C, Lombardi F (2008) Reversible gates and testability of one dimensional arrays of molecular QCA. J Electron Test 24:297–311

Majeed AH, Alkaldy E, Zainal MS, Navi K, Nor D (2019) Optimal design of ram cell using novel 2:1 multiplexer in QCA technology. Circuit World 46:147–158

Mohammadi Z, Navi K, Sabbaghi-Nadooshan R (2019) Design of testable reversible latches by using a novel efficient implementation of fredkin gate. Int J Electron 107:859–878

Nath T, Ferrão P, Fournier J, Lacarrière B, Corre OL (2019) A new clocking scheme for quantum-dot cellular automata based designs with single or regular cells. Energy Procedia 117:466–473

Niemier MT, Kogge PM (2001) Problems in designing with QCAs: layout = timing. Int J Circuit Theory Appl 29:49–62

Noorallahzadeh M, Mosleh M (2019) Efficient designs of reversible latches with low quantum cost. IET Circ Dev Syst 13:806–815

Pain P, Das K, Sadhu A, Kanjilal MR, De D (2019) Power analysis attack resistable hardware cryptographical circuit design using reversible logic gate in quantum cellular automata. Microsyst Technol 2:1–13

Peres A (1985) Reversible logic and quantum computers. Phys Rev A 32:3266–3276

Porod W, Lent CS, Bernstein G, Orlov AO, Amlani I, Snider GL, Merz JL (1999) Quantum-dot cellular automata: computing with coupled quantum dots. Int J Electron 86:549–590

Prakash G, Darbandi M, Gafar N, Jabarullah NH, Jalali MR (2019) A new design of 2-bit universal shift register using rotated majority gate based on quantum-dot cellular automata technology. Int J Theor Phys 58:3006–3024

Roohi A, Zand R, Angizi S, Demara RF (2018) A parity-preserving reversible QCA gate with self-checking cascadable resiliency. IEEE Trans Emerg Top Comput 6:450–459

Sadhu A, Das K, De D, Kanjilal MR (2019) Area-delay-energy aware SRAM memory cell and M × N parallel read/write memory array design for quantum dot cellular automata. Microprocess Microsyst 102944
Sasamal TN, Mohan A, Singh AK (2018) Efficient design of reversible logic alu using coplanar quantum-dot cellular automata. J Circ Syst Comput 27:185–204
Sen B, Dutta M, Goswami M, Sikdar BK (2014a) Modular design of testable reversible alu by QCA multiplexer with increase in programmability. Microelectronics J 45:1522–1532
Sen B, Dutta M, Sikdar BK (2014b) Efficient design of parity preserving logic in quantum-dot cellular automata targeting enhanced scalability in testing. Microelectronics J 45:239–248
Seyedi S, Darbandi M, Navimipour NJ (2019) Designing an efficient fault tolerance d-latch based on quantum-dot cellular automata nanotechnology. Opt Int J Light Electron Opt 185:827–837
Singh R, Pandey MK (2018) Analysis and implementation of reversible dual edge triggered flip flop using quantum dot cellular automata. Int J Innov Comput Inf Control 14:147–159
Singh R, Sharma DK (2020a) QCA-based ram design using a resilient reversible gate with improved performance. J Circ Syst Comput 29:1–26
Singh R, Sharma DK (2020b) Fault tolerant reversible gate based sequential quantum dot cellular automata circuits: design and contemplation. J Nanoelectron Optoelectron 15:331–344
Srivastava S, Asthana A, Bhanja S, Sarkar S (2011) QCAPro—an error-power estimation tool for qca circuit design. In: IEEE international symposium on circuits and systems, pp 2377–2380. IEEE Press, Brazil
Tahoori MB, Huang J, Momenzadeh M, Lombardi F (2004) Testing of quantum cellular automata. IEEE Trans Nanotechnol 3:432–442
Thapliyal H, Ranganathan N, Kotiyal S (2013) Design of testable reversible sequential circuits. IEEE Trans Very Large Scale Integr Syst 21:1201–1209
Thapliyal H, Ranganathan N (2009) Conservative QCA gate (CQCA) for designing concurrently testable molecular QCA circuits. In: 22nd international conference on VLSI design—held jointly with 7th international conference on embedded systems, pp 511–516, New Delhi, India
Timler J, Lent CS (2002) Power gain and dissipation in quantum-dot cellular automata. J Appl Phys 91:823–831
Toffoli T (1980) Reversible computing. In: International colloquium on automata, languages, and programming. Lect Notes Comput Sci 85:632–644
Vankamamidi V, Member S, Ottavi M (2008) Two-dimensional schemes for clocking/timing of QCA circuits. IEEE Trans Comput Aided Des Integr Circuits Syst 27:34–44
Walus K, Dysart TJ, Jullien GA, Budiman RA (2004) QCA designer: a rapid design and simulation tool for quantum-dot cellular automata. IEEE Trans Nanotechnol 3:26–31

Chapter 12
Machine Learning Technique for Few-Mode Fiber Design with Inverse Modelling for 5G and Beyond

Bhagyalaxmi Behera, Abhishek Das, and Mihir Narayan Mohanty

Abstract The next generation of communication, 5G and beyond works to connect people and things via intelligent networks. It is difficult to handle this massive data traffic with the current network architectures. The vision of the 5G network is to form a truly smart city or a connected society. But before 5G converts to the truth we need to reconstruct the network architecture to handle Trillions of Megabits and Billions of connected devices. To handle this huge data traffic the spatial domain of the fiber is highly useful. In this work, we have reviewed the requirements of 5G networks and how these can be handled by spatial multiplexing and mode multiplexing through a few-mode optical fiber. The conventional design of few-mode fiber with a complex structure is time-consuming and fixed for a given fiber structure. This article demonstrates the machine learning-based inverse modeling of the triangular-ring-core few-mode fiber profile with weak coupling optimization.

Keywords Board-learning-system · 5G communication · Few-mode fiber · Machine learning · Mode division multiplexing · Spatial division multiplexing

12.1 Introduction

12.1.1 Optical Fiber in 5G and Beyond

The new era of communication networks focuses on technical innovations to support high data rate, low latency, high security and high degree of computing, providing the connection between things and people. The 5G telecommunication networks promise to fulfil future requirements of the fastest and most securely connected society. The true meaning of the smart city and Internet of Things (IoT) is possible with the advancement of the 5G network. 5G will hugely influence the world economy like

B. Behera · A. Das · M. N. Mohanty (✉)
ITER, Siksha 'O' Anusandhan (Deemed to Be University), Bhubaneswar 751030, India
e-mail: mihir.n.mohanty@gmail.com; mihirmohanty@soa.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

A. Rai et al. (eds.), *Paradigms of Smart and Intelligent Communication, 5G and Beyond*, Transactions on Computer Systems and Networks,
https://doi.org/10.1007/978-981-99-0109-8_12

that of IoT. It is the booming sector of telecommunication networks and motivates the researcher to establish the network infrastructure that will connect Billions of devices and transfer Trillions of megabits before making 5G real. The 5G network is not only restricted to the telecommunication sector but also applicable in the development of smart communities/cities/buildings, vehicle automation, virtual medicines for preventive care, etc. (5G networks impact on fiber-optic cabling requirements 2022). The deployment of cellular communication started from 14.4Kbits/sec (1G) and reached to 10Gbits/sec (5G) to facilitate IoT connectivity with multiple Gbits/sec (5G networks impact on fiber-optic cabling requirements 2022) as shown in Fig. 12.1. This is possible only due to the development of high-speed and secure optical networks. Optical networks are used to serve a variety of communication services, like residential, business, and mobile services. Liu in Liu (2019), gives a brief review of the deployment of optical networking and communication for the 5G paradigm. The author gives a pictorial demonstration of 5G optical end-to-end network architecture with access, metro, and core networks as shown in Fig. 12.2. The requirements to support 5G communication are low latency, high bandwidth, security, network slicing, and synchronization as shown in Fig. 12.2. The three main requirements as low latency, highly secure, and high bandwidth makes the optical fibers as the potential transmission medium for the 5G network infrastructure. Fibers are the best medium for connecting 5G small cells "fronthaul" connections as well as wireless "backhaul" networks due to their high bandwidth, low attenuation, low electromagnetic interference, security, and small size (5G networks impact on fiber-optic cabling requirements 2022). This makes optical communication the heart of every access network to accommodate the exponential growth in data traffic every year. The author in Winzer (2010) has given a detailed study regarding "beyond 100G Ethernet". In that article, he has demonstrated the evolution of the capacity of WDM system and serial bit rate over the years since 1986 to till date which is exponentially grown from 10Gbits/sec to 100Tbits/sec. A major concern for the research society is how to accommodate future communication, including video streaming and IoT. The solutions are finding the way to design new types of fibers to support 5G and beyond.

Abbreviations used in are as follows:

ODN optical distribution network

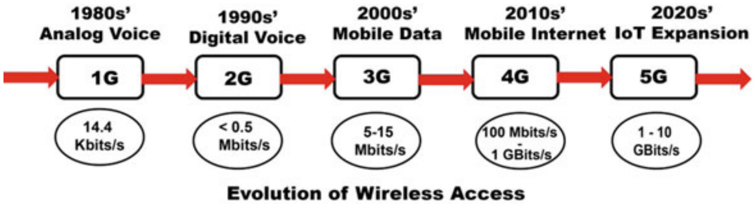


Fig. 12.1 Deployment of cellular access networks (5G networks impact on fiber-optic cabling requirements 2022)

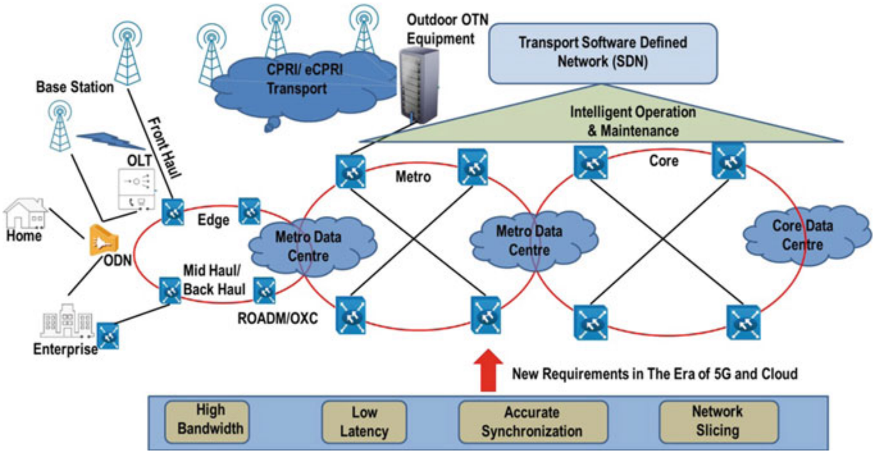


Fig. 12.2 Demonstration of optical end-to-end (access, metro, core) network for 5G and beyond to provide diverse applications (Liu 2019)

OLT	optical line terminal
OTN	optical transport network
ROADM	reconfigurable optical add/drop multiplexer
OXC	optical cross-connects
CPRI	common public radio interface
eCPRI	evolved CPRI

12.1.2 Types of Fiber Used in 5G Networks

This section of the introduction emphasizes the types of fibers used for 5G networks. Optical fibers are the key requirement for the 5G networks. The choice of optical cables for 5G infrastructure must satisfy the current requirements as well as the future. The 5G networks are designed to provide reliable services to clients with strong connectivity, high-speed, and low latency. This needs a huge number of 5G base stations due to limited network coverage and a higher frequency band. There is an estimation that by 2025 there will be more than 6.5 million 5G global base stations are required to establish a truly smart society (Community).

This motivates the researcher to design optical fibers to meet future requirements. There are five types of optical fibers are required for 5G network infrastructure, such as bend-insensitive fibers, optical multi-mode (OM) fibers, micron-diameter optical fibers, and ultra-low-loss with large effective-mode-area fibers, fiber cables for both indoor and outdoor connections. The main challenge in the 5G infrastructure is to connect the indoor micro-base stations with the macro base station and for this, the bend-insensitive fibers are most appreciable. Table 12.1 gives the list of commercially available bend-insensitive fibers used for 5G communication with the ITU-T

Table 12.1 List of commercially available bend-insensitive fibers for 5G networks (Community)

Company	Product name	ITU standards	Bend radius (1 turn around a mandrel)	Induced attenuation (dB)
Corning	Clear Curve LBL fiber	G.652.D, G.657.A2/B2	7.5 mm	$\leq$0.4
YOFC	Easy Band® Ultra BIF	G.652.D, G.657.B3	5 mm	$\leq$0.15
Prysmian Group	Bend Bright XS fiber	G.652.D, G.657.A2/B2	7.5 mm	$\leq$0.5

standard and minimum acceptable bending radius. Different OM fibers are available commercially, but the OM5 fibers are best suited for 5G core networks to connect the data centres with a maximum speed of 400G but at an 850 nm optical window. To accommodate many optical fibers in a limited space to enhance the spectral density micron-diameter fibers are required for 5G networks. The current wavelength division (WDM), and dense-WDM (DWDM) optical networks with minimum channel spacing of 50 GHz are reached the capacity limits due to the onset of intensity-dependent fiber nonlinearities (Essiambre et al. 2010). With the increasing demand for high data rate and low latency, among millions of simultaneous connectivity, the optical power level increases through the fiber core for long-haul communication. This causes the capacity crunch in recent optical networks due to the nonlinear nature of light. To go along with the capacity demand and increasing power level in 5G networks another important requirement is large effective-mode-area fibers with ultra-low-loss features (Desurvire 2006). Lastly, the fibers are used for 5G networks must be suitable for both indoor and outdoor connections. The important features of these fiber structures are adaptable in both dry and wet conditions, these fibers are small in size, easy to install, and with low splicing loss.

12.1.3 Role of Few-Mode Fiber in 5G Networks

The need for fibers for next-generation communication 5G and beyond must handle large power with low nonlinear coefficients, large effective mode-area, bend-insensitive, small in size, low splicing loss, easy to install, and can be used for both indoor and outdoor connections. To accommodate millions of users simultaneously multiplexing is one of the important points of discussion for 5G networks. All the dimensions of multiplexing like polarization, wavelength, and time are exhausted and reach the limit. So, the spatial dimension for multiplexing must be explored for 5G communication (Lagkas et al. 2020). Space division multiplexing (SDM) is currently being researched as it appears to be the next significant leap in fiber transmission capacity required to address the impending capacity crisis mentioned previously (Richardson et al. 2013). The usage of SDM, in which various spatial routes in a single fiber are employed to permit concurrent data transmission of independent data channels, is the hopeful approach for increasing capacity limits (Su et al. 2021). The SDM is jointly known as mode division multiplexing (MDM) (Memon and Chen 2021) through vector modes or linearly polarized (LP) modes of a few-mode fiber (FMF) (Kitayama and Diamantopoulos 2017), and core-multiplexing using multi-core fibers (MCFs) (Richardson et al. 2013). The other direction of SDM is also possible by orbital angular momentum (OAM) fibers (Rjeb et al. 2020). The FMFs (Wang et al. 2021)and MCFs (Gasulla and Capmany 2019) both are gaining attention towards the establishment of 5G networks. But FMFs are dominating over MCFs due to the following, size is comparable with single-mode fibers (SMF), splicing and amplification are better than MCFs, effective mode-area is more than MCFs, and fabrication is easy at the cost of multi-input and multi-output (MIMO) signal processing at the receiver end (Kitayama and Diamantopoulos 2017). The MDM transmissions are classified as weakly-coupled and strongly-coupled depending on the few-mode fiber structure. The strongly-coupled long-haul MDM links are greatly affected by the signal degradation due to large differential-mode-delay (DMD) and need coherent detection and MIMO processing at the receiver side (Rademacher et al. 2019). However, for 5G networks, we have to establish a high-speed connection between nearby base stations and for that direct-detection (MIMO less) is more advisable than that of coherent detection to reduce the receiver complexity (Wang et al. 2021). MIMO less MDM transmission with direct direction is achieved with low modal crosstalk or large separation between the guided modes, whereas the signal degradation due to large DMD always requires MIMO processing (Liu et al. 2018). This motivates the researcher to design weakly-coupled FMFs for short distance direct-detection MDM links (Behera et al. 2021).

To date, different types of weakly-coupled FMFs are demonstrated by various authors by varying the structure and material. These are a special type of multi-mode fibers (MMF) designed to guide a few tens of LP modes or vector modes through single-core, where each guided mode is treated as a unique data channel for spatial multiplexing. For weak coupling between the side by modes, the effective index difference ($\Delta \mathrm{n}_{\mathrm{eff}}$) between the adjacent LP modes should be greater than 1×10^{-3} whereas between vector modes is greater than 1×10^{-4} (Chen et al. 2020). The single-core weakly-coupled FMFs are available with step-index (SI) (Soma et al. 2017), graded-index (GI) (Sillard et al. 2014), and ring-core profiles (RC) (Liu et al. 2018). Each of the structures is having some advantages and disadvantages. SI-FMFs have a limited number of design parameters so easy to control the number of guided modes, and easy to fabricate but demonstrate weak coupling between the adjacent modes with small V (normalized frequency parameter) values. This allows for guiding a smaller number of modes through the conventional SI-FMFs. Hence, need MIMO processing for SI-FMF-MDM links (Liu et al. 2018). Whereas the GI-FMF the profile allows guiding of LP mode-groups (MG) with degeneracy. The inter-MG separation (Δn_{eff}) is very large as compared to intra-MG separation and this requires less MIMO equalization for intra-MG spatial channels with low DMD between intra-MGs for MDM transmission.

Secondly, the GI-FMF profile equalizes the DMD parameters between the guided modes (Sillard et al. 2014). In SI-RC-FMF the radial mode index is limited to 1. This fixes the number of degenerate modes in a higher-order MG to 4, i.e. increases the azimuthal mode order. As a result, the crosstalk and DMD between the side by modes decrease and allow MIMO less MDM transmission using higher-order MGs of RC-FMF. RC-FMF-based MDM solutions are quite appealing because of these properties (Jin et al. 2016).

12.1.4 State-Of-Art in the Design of Weakly-Coupled FMFs

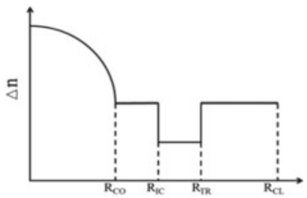
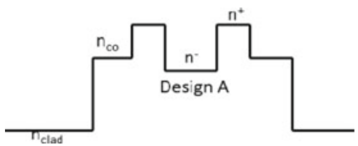
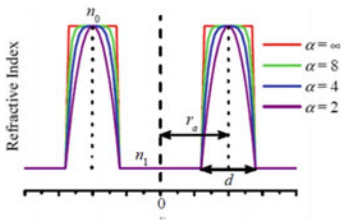
The above works of literature give an impact on the requirements of 5G and beyond that can be achieved jointly through weakly-coupled SDM and MDM. To establish a weakly-coupled SDM or MDM transmission link we need weakly-coupled FMFs to guide a few tens of LP modes with a large effective-mode-area (A_{eff}), low bending loss, low splicing loss, low DMD, low dispersion, and low nonlinear coefficients. Table 12.2 gives a comparative analysis of the designed FMFs that are claimed to be the best candidate for next-generation communication using SDM or MDM transmission across the C-Band.

The above works of literature emphasize the conventional design of FMFs through the parametric sweep. This conventional method of FMF design is very tedious for a complex structure with many design parameters. This process is also not reusable when the structural parameters are changed. This motivates the researcher to demonstrate and conduct an additional study on fiber design using machine learning (ML) models.

12.1.5 Machine-Learning in FMF Design

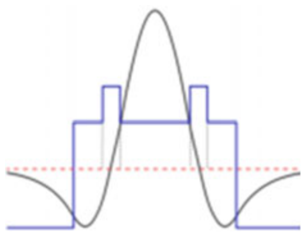
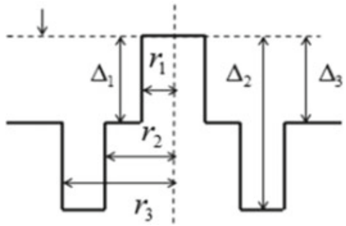
ML models and artificial neural networks (ANN) are new technologies that has emerged in recent years. This can be applied in prediction, forecasting, optimization, and classification problems of different fields. This can also be used in the fiber industry for both forward and inverse design processes. The ML models are either used to predict the accurate modal parameters of a fiber structure over an optical window through forward design. The authors in Chugh et al. (2019), demonstrate the prediction of modal parameters of a photonic-crystal fiber using forward-ANN with 3 hidden layers, and 50 neurons per layer. They have achieved mean square error (MSE) in the range of 0.00134–0.0065 with 5000 epochs. Another development for the fiber business is inverse modeling via machine learning. This can be used for any sophisticated FMF designs to improve the weak coupling optimization. To enhance the capacity crunch in recent years and to establish backhaul wireless networks and frontend optical networks for high-speed 5G and 6G infrastructure the FMFs are the alternative solutions. The design of FMFs through inverse modeling is fast and more

Table 12.2 Comparison analysis of different weakly-coupled FMFs proposed by various authors for SDM and NDM applications

References	Profile	LP modes	min Δn_{eff}	minA_{eff} [μm^2]	Min DMD [ps/m] and Optical loss [dB/km]
Zhang et al. (2019)		12	~1×10^{-3}	150	<0.0413, 10^{-4} @7 mm bendig radius
Ge et al. (2018)		6	1.49×10^{-3}	100	–
Han et al. (2018)		3	>1×10^{-3}	100	– 10 dB/turn@ 10 mm bending radius

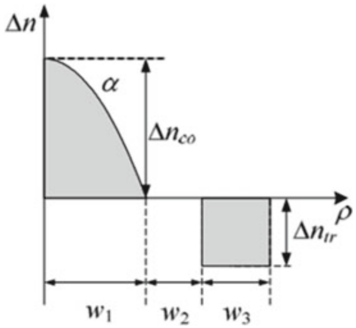
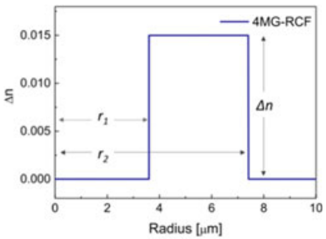
(continued)

Table 12.2 (continued)

References	Profile	LP modes	min Δn_{eff}	minA_{eff} [μm^2]	Min DMD [ps/m] and Optical loss [dB/km]
Jiang et al. (2018)		7	1.8×10^{-3}	–	– 0.23 dB/km
Shen et al. (2018)		4	–	–	– 0.168 dB/km

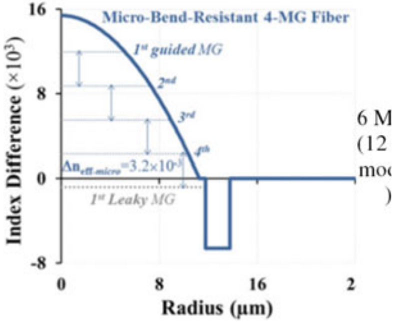
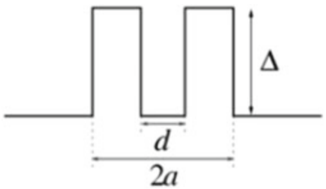
(continued)

Table 12.2 (continued)

References	Profile	LP modes	min Δn_{eff}	minA_{eff} [μm^2]	Min DMD [ps/m] and Optical loss [dB/km]
Ferreira et al. (2017)		12	–	–	<12 ps/km, –
Jung et al. (2017)		4 MG (7 spatial)	Large	–	– 0.32 dB/km

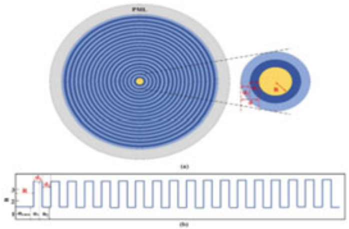
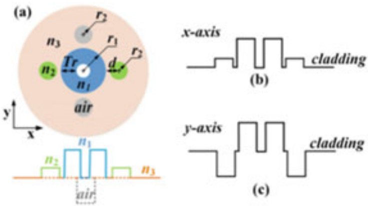
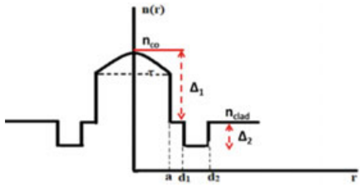
(continued)

Table 12.2 (continued)

References	Profile	LP modes	min Δn_{eff}	minA_{eff} [μm^2]	Min DMD [ps/m] and Optical loss [dB/km]
Sillard et al. (2017)		6 MGs (12 LP modes)	Large	75	<100 ps/km, 0.218 dB/km
Kasahara et al. (2014)		3	>2.5 × 10^{-3}	80	<10 ns/km, –

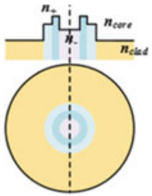
(continued)

Table 12.2 (continued)

References	Profile	LP modes	min Δn_{eff}	minA_{eff} [μm^2]	Min DMD [ps/m] and Optical loss [dB/km]
Su et al. (2020)		7	2×10^{-3}	64.28 μm^2	8.4 ns/km, 10 dB/Km at 10 mm
Chen et al. (2020)		8-spatial	1.96 × 10 − 4	114.28	3.00 × 10–6 dB/km@5 cm bending radius
Behera et al. (2021)		10	>1 × 10^{-3}	150	0.3 ps/m

(continued)

Table 12.2 (continued)

References	Profile	LP modes	min Δn_{eff}	minA_{eff} [μm^2]	Min DMD [ps/m] and Optical loss [dB/km]
Ge et al. (2019)		4	1.49×10^{-3}	109	5.4 ps/m 2.46×10^{-10} dB/km @10 mm

reusable than the forward design. Authors in He et al. (2020) have used an NN-based ML process for the inverse design of FMF. The authors exhibited an inverse design for weak coupling optimization in step-index few-mode fiber with 4-rings to guide 4, 6, and 10 modes, as well as optimizing the profile parameters for 6-ring to achieve 20 modes with weak coupling, in their study. The ML-based inverse modeling method of few-mode fiber design is very precise, quick, and reusable. The ML models are accurately mapping the inputs and outputs, which makes the design process fast. This method is universally applicable and can be extended further to optimize other parameters like loss, dispersion, DMD, and effective area.

This motivates us to keep working on inverse modeling to predict the profile parameters of a triangular-core FMF (T-FMF) for a set of LP modes while minimizing the mode coupling $\left(\min \Delta n_{\mathrm{eff}} \geq 1 \times 10^{-3}\right)$ between neighbouring modes.

The organization of the article is as follows: Sect. 12.2 emphasizes the design methodology of the proposed T-FMF, and a brief discussion about the board-learning-system (BLS) regression model. Section 12.3 is describing the results of the inverse machine learning process, and the modal characteristics of T-FMF. The article is concluded in Sect. 12.4.

12.2 Proposed FMF Structure

This work presents triangular-core index profiles for the few-mode operation. Figure 12.3 describes the radially varying core index profile of T-FMF. In this work, we predict the profile parameters of the proposed fiber using an ML-based inverse process. The range of profile parameters for generating the data set for the training process is selected by varying the dispersion parameter of the proposed T-FMF. This type of fiber profile is used as dispersion-shifted fiber (DSF) with nearly zero dispersion slope at 1550 nm LP_{01} (Agrawal 2012). Hence, we are modifying the conventional triangular DSF structure in Behera and Mohanty (2019) for few-mode operation with low mode coupling. The fiber structure and the core material are modified to achieve a few-mode operation with zero dispersion at 1550 nm. Then these profile parameters are varying within a specific range to guide 5 to 15 LP modes with 50% of the guided modes satisfying the weak coupling criteria to create the data set for training.

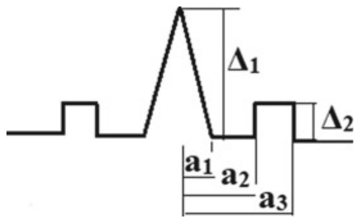


Fig. 12.3 Radially varying refractive index profile of T-FMF

12.2.1 T-FMF Structure

The proposed T-FMF is a special type of graded-index core fiber with a high-index ring in the cladding. This type of fiber profile is the generic of W-type fiber structures. The special usages of this type of structure are to make the waveguide dispersion parameter $D_{wg}(\lambda)$ more negative at 1550 nm and to compensate for the material dispersion parameter $D_{\text{mat}}(\lambda)$. This in return provides zero dispersion at that particular wavelength. The index profile of T-FMF concerning radial distance is shown in Fig. 12.3. Here, the core resembles a graded-index profile with the index exponent (α) fixed at 1 to achieve a triangular profile. The profile of T-FMF is given in Eq. (12.1). The cladding is treated as the host material and presumed to be pure silica with refractive index n_2. The index difference between the core and cladding is Δ_1 whereas between the high-index ring and cladding is Δ_2. The change in refractive indices is achieved by changing the molar concentration of the dopant. The highest index in the centre of the core is n_1. The profile parameters are listed in Eq. (12.1) as core radius a_1, profile exponent α, and r is the radial distance. The refractive index of the high-index ring in the cladding is noted as n_3. The empirical relationship between the operating wavelength and refractive index is found in this proposed work through the well-known Sellmeier equation as given in Eq. (12.2) (Behera et al. 2021). In Eq. (12.2), the operating wavelength is λ in the micrometre, and A_i, C_i is the Sellmeier coefficients. The material dispersion of the host and dopant materials is studied with their dispersive nature. And then the mole percentage of the dopants is adjusted to obtain the desired number of modes in a range of 5–15 with zero dispersion for LP_{01} mode. These mode solutions (effective index of guided modes n_{eff}) along with the profile parameters $[a_1, a_2, a_3, \Delta_1, \Delta_2]$ are used for the training of the ML model. The proposed T-FMF is designed by using a finite-difference method-based solver.

$$n(r) = \begin{cases} n_1\left[1 - 2\Delta_1\left(\frac{r}{a}\right)^{\alpha}\right]^{1/2} & r \le a_1 \\ n_1[1 - 2\Delta_1]^{1/2} \approx n_2 & a_1 \le r \le a_2 \text{ and } r \ge a_3 \\ n_2/[1 - 2\Delta_2]^{1/2} \approx n_3 & a_2 \le r \le a_3 \end{cases} \tag{12.1}$$

$$n^2(\lambda) = 1 + \frac{A_1\lambda^2}{\lambda^2 - C_1} + \frac{A\lambda^2}{\lambda^2 - C_2} + \frac{A_3\lambda^2}{\lambda^2 - C_3} \tag{12.2}$$

12.2.2 Design Methodology

The entire design process of the proposed FMF is divided into two steps: forward design and inverse design. To create a data set for the ML models, the forward design method is employed. Figure 12.4 depicts the full design technique. The ML models are used to accurately map the target outputs (primary parameters n_{eff}) with the input profile parameters $[a_1, a_2, a_3, \Delta_1, \Delta_2]$ to obtain a large separation (Δn_{eff}) between adjacent modes. The objective is to minimize the Δn_{eff}, (the secondary parameters), but as the primary parameters are directly related to profile parameters hence the model is trained by the n_{eff} rather than Δn_{eff} to predict the accurate inputs for lower Δn_{eff}. Inverse modeling is sped up and improved with machine learning.

The steps involved in the design and validation of the proposed T-FMF are as follows:

- The profile is designed and the modal parameters are obtained by using the finite-difference method solver to select the range of profile parameters. For this, the commercially available OptiFiber tool is used to create the training data set. This is known as forward design.
- The training data set is arranged in a 2000 × 16 matrix and undergoes min–max normalization before the training process. The training data set must be created with a precise number of elements to reduce computational time and achieve a high degree of accuracy.
- Over a large range of n_{eff}, the design parameters are modified to obtain a maximum of 15 modes, and more than 50% of the mode solutions satisfy the criterion of weak mode coupling. The range of design parameters listed in Table 12.3 are considered to guide 5–15 modes through T-FMF with zero dispersion slope for the LP_{01} mode.

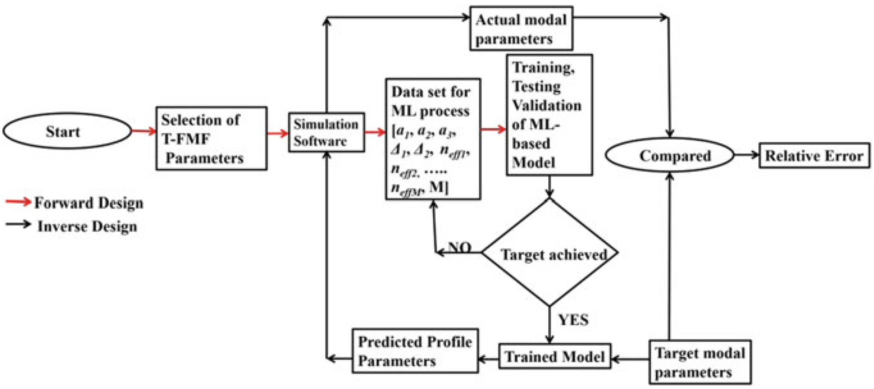


Fig. 12.4 The flow diagram for the ML-based inverse design of T-FMF

Table 12.3 Range of design parameters for T-FMF to guide 5–15 modes

Parameters	Minimum	Maximum
a_1 [μm]	7	9.5
a_2 [μm]	8.4	10.8
a_3 [μm]	10	16.5
Δ_1 [%]	0.024	0.034
Δ_2 [%]	0.003	0.008

Table 12.4 Predicted parameters of the proposed T-FMF through inverse modeling to guide 15 modes

Parameters	Predicted value
a_1 [μm]	9.5
a_2 [μm]	10.82
a_3 [μm]	15.32
Δ_1 [%]	0.034
Δ_2 [%]	0.008

- We have achieved the best-predicted results by splitting 75% of the data for training and 25% for testing.
- The trained model is validated by predicting the profile parameters for a target number of guided modes within 5–15 with minimum coupling between the side by modes. The proposed T-FMF is designed with the predicted parameters by accurately mapping the outputs of M-modes [neff_1, $\text{neff}_2 \ldots \text{neff}_M$] with the input parameters [a_1, a_2, a_3, Δ_1, Δ_2] through the ML process. This process of design is treated as inverse modeling.
- To guide fifteen LP modes, the profile parameters of the designed T-FMF are predicted using the trained model with large mode separation among the side by modes. The performance parameters of the model are listed in Table 12.4.
- The actual modal parameters of the proposed T-FMF with the predicted design parameters are obtained through inverse design and compared with the target parameters to evaluate the relative error.

12.2.3 Machine Learning Model

This is the first time BLS models have been used to predict several design parameters of proposed T-FMF structures. This work can be expanded to include various FMF architectures and will demonstrate potential uses in the fiber modeling sector. The next section briefly describes the BLS regression model used for prediction.

A broad learning system-based machine learning model (Chen and Liu 2017) is utilized for an inverse design that predicts the required parameters to design the proposed T-FMF. The BLS model is formed by enhancing the traditional random vector functional-link neural network (RVFLNN). The RVFLNN models are directly trained with the input data and passed through enhancement nodes in the next layer. In BLS, the input data is first utilized to develop the feature maps and these features are used for training along with the enhancement nodes. The detailed structure of the proposed BLS model is shown in Fig. 12.5.

Figure 12.5, X represents the input matrix, whereas Y represents the predicted values. Input data X is mapped to n feature maps and n groups resulting in the prediction of Y as the expression given in Eq. (12.3) and (12.4).

$$Y = \left[Z_1, \xi\left(Z_1 W_{h_1} + \beta_{h_1}\right) | \ldots Z_n, \xi\left(Z_n W_{h_n} + \beta_{h_n}\right)\right] W^n \tag{12.3}$$

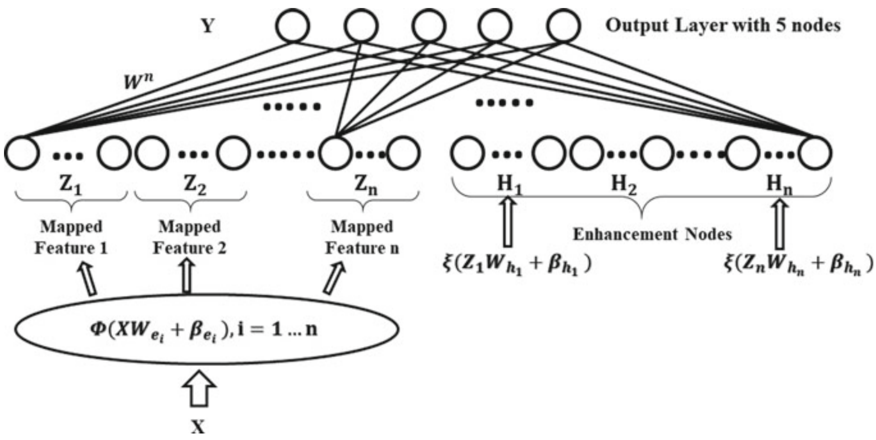


Fig. 12.5 Architecture of the proposed BLS model

$$\cong [Z_1 \ldots Z_n|\xi\left(Z_1 W_{h_1} + \beta_{h_1}\right) \ldots, \xi\left(Z_n W_{h_n} + \beta_{h_n}\right)]W^n \quad (12.4)$$

where $Z_i, i = 1 \ldots n$, represents the feature nodes given by Eq. (12.5), W_{e_i} and β_{e_i} are randomly generated.

$$Z_i = \Phi(X W_{e_i} + \beta_{e_i}) \quad (12.5)$$

The mth groups of enhancement nodes H_m are given by Eq. (12.6).

$$H_m = \xi\left(Z^n W_{h_m} + \beta_{h_m}\right) \quad (12.6)$$

Deep learning models have largely hidden layers and large parameters to be evaluated during the training process and it consumes lots of time. BLS learning relies on the concept of incremental learning. The new enhancement nodes are added to the corresponding layer until it reaches the optimum solution.

The new nodes are formed means new columns are added to the input data matrix X. The new data matrix X^{m+1} is expressed as in Eq. (12.7)

$$X^{m+1} = [X^m|\xi\left(Z^n W_{h_{m+1}} + \beta_{h_{m+1}}\right) \quad (12.7)$$

where $W_{h_{m+1}}$ and $\beta_{h_{m+1}}$ are mapped to the p additional enhancement nodes and are randomly formed.

The new weights can be represented as

$$W^{m+1} = \begin{bmatrix} W^m - DB^T Y \\ B^T Y \end{bmatrix} \quad (12.8)$$

where $D = (X^m)^+ \xi\left(Z^n W_{h_{m+1}} + \beta_{h_{m+1}}\right)$

$$B^T = \begin{cases} (C)^+, if C \neq 0 \\ \left(1 + D^T D\right)^{-1} B^T (X^m)^+, if C = 0 \end{cases} \text{ and}$$

$$C = \xi\left(Z^n W_{h_{m+1}} + \beta_{h_{m+1}}\right) - X^m D \tag{12.9}$$

In this work, the input data X is a matrix of size 2000 × 16 representing 2000 sets each containing 16 parameters of optical fiber as discussed in Sect. 2.2.

12.3 Discussion of Proposed model with RMSE and MSE

The robustness of the model is evaluated by determining different error functions like mean square error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and correlation coefficients using Eq. (12.10–12.12). Here, y_i is the ith element of actual data and $\widehat{y_i}$ is the predicted value of y_i. The performance of the model is depicted in Table 12.5. Figure 12.6a–e shows the actual and predicted outputs over the data index for all the five design parameters (a_1, a_2, a_3, Δ_1, and Δ_2) to validate the accuracy of the trained model. The design parameters are then predicted by the trained model and assigned to OptiFiber to yield the first-order parameters n_{eff}. The second-order parameters [$n_{\text{eff-i}} = n_{\text{eff-i}} - n_{\text{eff-(i+1)}}$] are calculated from the first-order parameters and compared to the target data set to determine the relative error, as illustrated in Fig. 12.7. The relative error in Fig. 12.7 demonstrates the BLS regression is one of the best models to predict the profile parameters of T-FMF. The robustness of the model is determined by computing MSE, RMSE, MAE, and accuracy of the model and is listed in Table 12.5. The predicted parameters to guide 15 modes through the T-FMF with maximum separation among the adjacent modes are given in Table 12.4. The proposed fiber is designed with the parameters in Table 12.4 using Optifiber. The modal properties of T-FMF with inverse design are shown in Table 12.6 for 15 LP modes at 1550 nm. The range of Δn_{eff} is found to be 1×10^{-3} to 3×10^{-3} with the predicted parameters. Whereas the minimum A_{eff} is 80μm^2, the minimum chromatic dispersion parameter is 0 [ps/nm km] for LP_{01} mode, and the range of bending loss is found to be less than 0.4 dB/km at a bending radius of 7.5 mm to satisfy ITU standard as discussed in Table 12.1. The modal properties shown in Table 12.6 summarize the proposed T-FMF as a promising candidate for 5G and beyond. The proposed FMF satisfies the characteristics of fibers needed for 5G networks as discussed in Sect. 1.2. With large mode separation, large effective-mode-area, low dispersion, and low bending loss the proposed T-FMF claims to be useful for MDM transmission for 5G networks.

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \widehat{y_i}\right)^2 \tag{12.10}$$

Table 12.5 Performance of BLS model for multi-output regression

	Parameters of BLS regression model			
	MSE	MAE	RMSE	Accuracy [%]
a_1	4.62×10^{-5}	3.16×10^{-7}	1.25×10^{-6}	99.18
a_2	2.63×10^{-7}	4.36×10^{-8}	6.44×10^{-5}	99.06
a_3	3.69×10^{-2}	1.93×10^{-2}	2.73×10^{-3}	97.92
Δ_1	4.68×10^{-1}	8.42×10^{-2}	4.43×10^{-2}	98.96
Δ_2	1.39×10^{-7}	3.12×10^{-7}	3.49×10^{-7}	99.15

$$\text{MAE} = \sum_{i=1}^{N} \left| y_i - \widehat{y_i} \right| \tag{12.11}$$

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(y_i - \widehat{y_i} \right)^2} \tag{12.12}$$

12.4 Summary

This article gives an idea about the usage of fiber in 5G and beyond. The characteristics of fibers used for 5G networks are briefly discussed. The requirement of 5G networks and their implementation using spatial multiplexing is described in detail. Brief literature about the weakly-coupled FMFs in the application of MIMO-less MDM/SDM links for next-generation communication is discussed. At the end, we demonstrate the inverse modeling of T-FMF using BLS regression models for the first time to our knowledge. The BLS model for multi-output regression exhibits high accuracy in the inverse design process, with an accuracy of 99.1% and a low relative error of 10^{-3} between the target and actual parameters. The desired T-FMF profile parameters for guiding fifteen modes with large mode separation are illustrated. The proposed FMF satisfies the characteristics of fibers needed for 5G networks as discussed in Sect. 12.1. With large mode separation, large effective-mode-area, low dispersion the proposed T-FMF claims to be useful for MDM transmission for 5G networks.

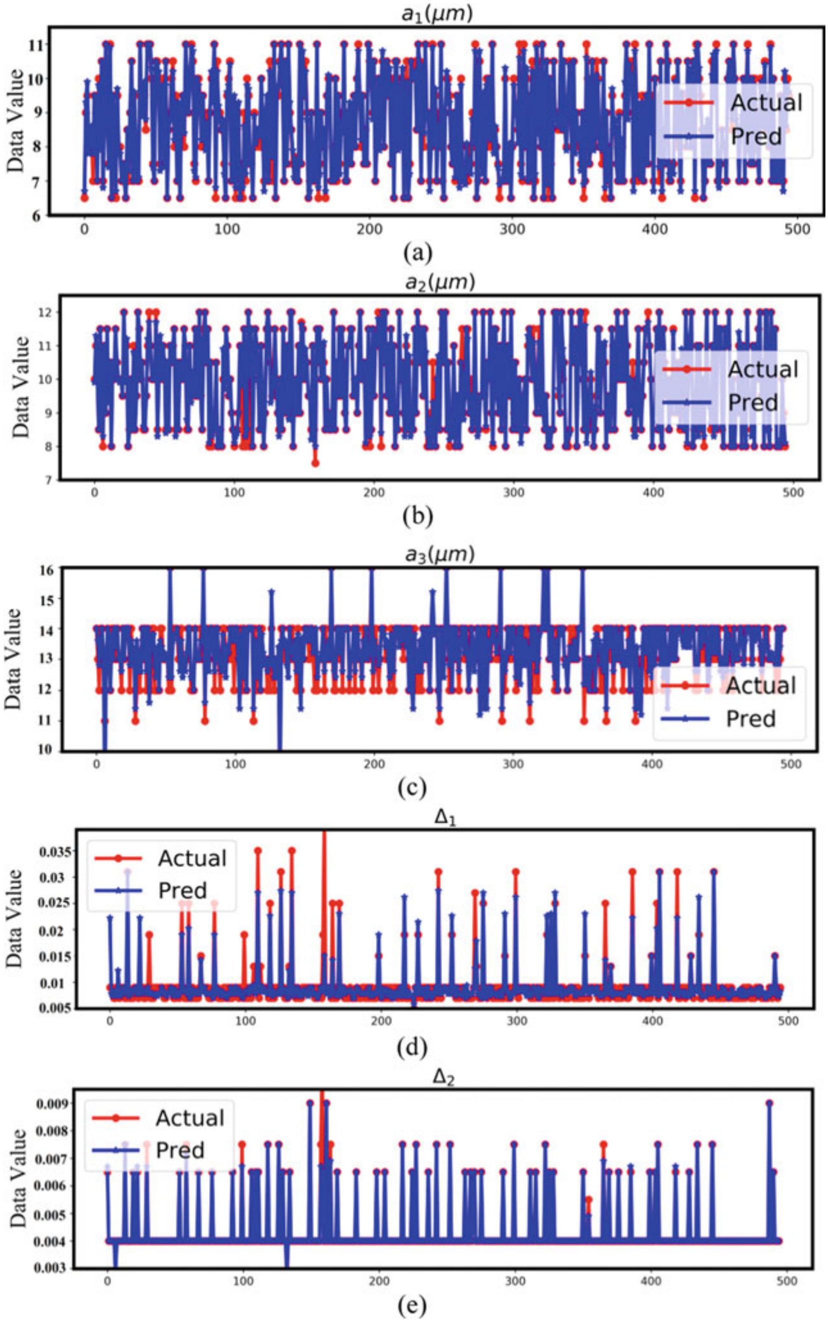


Fig. 12.6 **a–e** Mapping between actual and predicted structural parameters (a_1, a_2, a_3, Δ_1, and Δ_2) using BLS regression

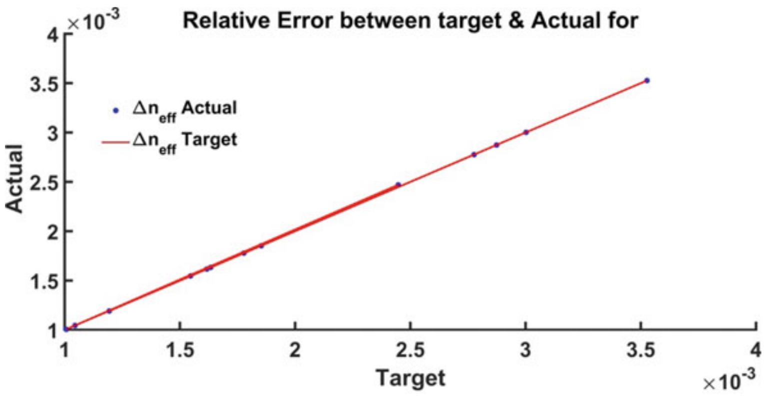


Fig. 12.7 Relative error between the target and actual Δn_{eff} for T-FMF designed through the inverse process

Table 12.6 Modal properties of T-FMF designed with the predicted parameters

LP modes	LP01	LP11	LP02	LP21	LP12	LP03	LP31	LP22
n_{eff}	1.4702	1.4677	1.4647	1.4627	1.4616	1.4604	1.4584	1.4574
$\Delta n_{\mathrm{eff}}\ \times 10^{-3}$	0.0025	0.003	0.002	0.0011	0.0012	0.002	0.001	0.001
$A_{\mathrm{eff}}\ \mu\mathrm{m}^2$	80	187	100	195	198	300	210	235
LP modes	**LP13**	**LP32**	**LP41**	**LP51**	**LP04**	**LP61**	**LP23**	
n_{eff}	1.4564	1.4549	1.4524	1.4494	1.4483	1.4471	1.4461	
$\Delta n_{\mathrm{eff}}\ \times 10^{-3}$	0.0015	0.0025	0.003	0.0011	0.0012	0.001	–	
$A_{\mathrm{eff}}\ \mu\mathrm{m}^2$	250	210	284	189	220	340	265	

References

5G networks impact on fiber-optic cabling requirements (2022)

Agrawal GP (2012) Fiber-optic communication systems. John Wiley & Sons

Behera B, Varshney SK, Mohanty MN (2021) Structure for fast photonic medium on application of SDM communication using SiO_2 doped with GeO_2, and F materials. IET Nanodielectrics 4:107–120

Behera B, Mohanty MN (2019) Design of bend-limited large-mode area dispersion shifted few-mode fiber for fast communication. In: 2019 international conference on applied machine learning (ICAML), pp 277–281

Chen CP, Liu Z (2017) Broad learning system: an effective and efficient incremental learning system without the need for deep architecture. IEEE Trans Neural Netw Learn Syst 29:10–24

Chen S, Tong Y, Tian H (2020) Eight-mode ring-core few-mode fiber using cross-arranged different-material-filling side holes. Appl Opt 59:4634–4641

Chugh S, Gulistan A, Ghosh S, Rahman B (2019) Machine learning approach for computing optical properties of a photonic crystal fiber. Opt Express 27:36414–36425

Community F. 5 types of optical fibers for 5G networks

Desurvire EB (2006) Capacity demand and technology challenges for lightwave systems in the next two decades. J Lightwave Technol 24:4697–4710

Essiambre R-J, Kramer G, Winzer PJ, Foschini GJ, Goebel B (2010) Capacity limits of optical fiber networks. J Lightwave Technol 28:662–701
Ferreira FM, Costa CS, Sygletos S, Ellis AD (2017) On the feasibility of mode-division multiplexed transmission over few-mode fibres. In: 2017 SBMO/IEEE MTT-S international microwave and optoelectronics conference (IMOC), pp 1–5
Gasulla I, Capmany J (2019) Multicore fibres for 5G fronthaul evolution. Optical and Wireless Convergence for 5G Networks, pp 79–100
Ge D, Gao Y, Yang Y, Shen L, Li Z, Chen Z, He Y, Li J (2019) A 6-LP-mode ultralow-modal-crosstalk double-ring-core FMF for weakly-coupled MDM transmission. Optics Commun 451:97–103
Ge D, Li J, Zhu J, Shen L, Gao Y, Yu J, Wu Z, Li Z, Chen Z, He Y (2018) Design of a weakly-coupled ring-core FMF and demonstration of 6-mode 10-km IM/DD transmission. In: 2018 optical fiber communications conference and exposition (OFC), pp 1–3
Han J, Li Y, Zhang J (2018) Design of an improved radially single-mode and azimuthally multi-mode ring-core fiber for mode-division multiplexing systems. In: 2018 Asia communications and photonics conference (ACP), pp 1–3
He Z, Du J, Chen X, Shen W, Huang Y, Wang C, Xu K, He Z (2020) Machine learning aided inverse design for few-mode fiber weak-coupling optimization. Opt Express 28:21668–21681
Jiang S, Ma L, Zhang Z, Xu X, Wang S, Du J, Yang C, Tong W, He Z (2018) Design and characterization of ring-assisted few-mode fibers for weakly coupled mode-division multiplexing transmission. J Lightwave Technol 36:5547–5555
Jin X, Gomez A, Shi K, Thomsen BC, Feng F, Gordon GS, Wilkinson TD, Jung Y, Kang Q, Barua P (2016) Mode coupling effects in ring-core fibers for space-division multiplexing systems. J Lightwave Technol 34:3365–3372
Jung Y, Kang Q, Zhou H, Zhang R, Chen S, Wang H, Yang Y, Jin X, Payne FP, Alam S, Richardson DJ (2017) Low-loss 25.3 km few-mode ring-core fiber for mode-division multiplexed transmission. J Lightwave Technol 35:1363–1368
Kasahara M, Saitoh K, Sakamoto T, Hanzawa N, Matsui T, Tsujikawa K, Yamamoto F (2014) Design of three-spatial-mode ring-core fiber. J Lightwave Technol 32:1337–1343
Kitayama K-I, Diamantopoulos N-P (2017) Few-mode optical fibers: original motivation and recent progress. IEEE Commun Mag 55:163–169
Lagkas T, Klonidis D, Sarigiannidis P, Tomkos I (2020) 5G/NGPON evolution and convergence: developing on spatial multiplexing of optical fiber links for 5G infrastructures. Fiber Integr Opt 39:4–23
Liu X (2019) Evolution of fiber-optic transmission and networking toward the 5G era. Iscience 22:489–506
Liu J, Zhu G, Zhang J, Wen Y, Wu X, Zhang Y, Chen Y, Cai X, Li Z, Hu Z (2018) Mode division multiplexing based on ring core optical fibers. IEEE J Quantum Electron 54:1–18
Memon A, Chen K (2021) Recent advances in mode converters for a mode division multiplex transmission system. Opto-Electron Rev 29
Rademacher G, Luís RS, Puttnam BJ, Eriksson TA, Ryf R, Agrell E, Maruyama R, Aikawa K, Awaji Y, Furukawa H, Wada N (2019) High capacity transmission with few-mode fibers. J Lightwave Technol 37:425–432
Richardson D, Fini J, Nelson LE (2013) Space-division multiplexing in optical fibres. Nat Photonics 7:354
Rjeb A, Seleem H, Fathallah H, Machhout M (2020) Design of 12 OAM-Graded index few mode fibers for next generation short haul interconnect transmission. Opt Fiber Technol 55:102148
Shen L, Chen S, Sun X, Liu Y, Zhang L, Hu T, Li J (2018) Design, fabrication, measurement and MDM transmission of a novel weakly-coupled ultra low loss FMF. In: 2018 optical fiber communications conference and exposition (OFC), pp 1–3
Sillard P, Bigot-Astruc M, Molin D (2014) Few-mode fibers for mode-division-multiplexed systems. J Lightwave Technol 32:2824–2829

Sillard P, Molin D, Bigot-Astruc M, Jongh KD, Achten F, Antonio-López JE, Amezcua-Correa R (2017) Micro-bend-resistant low-differential-mode-group-delay few-mode fibers. J Lightwave Technol 35:734–740

Soma D, Wakayama Y, Igarashi K, Tsuritani T (2017) Partial MIMO-based 10-mode-multiplexed transmission over 81km weakly-coupled few-mode fiber. In: Optical fiber communication conference, vol 4. Optical Society of America, pp M2D

Su Y, Zhangsun T, Ren F, Zhang Y, Yin J, Chen W, Fan X, Wang J (2020) Design of solid-core Bragg few-mode fiber for short-reach MDM networks in O+ C+ L band. Optics Commun 461:125245

Su Y, He Y, Chen H, Li X, Li G (2021) Perspective on mode-division multiplexing. Appl Phys Lett 118:200502

Wang G, Zhang J, Zhang H, Wang F, Yan X, Zhang X, Li S, Cheng T (2021) A low crosstalk multi-core few-mode fiber with composite refractive index profile and air-hole embedded trench assistance. Optics Commun 499:127258

Winzer PJ (2010) Beyond 100G ethernet. IEEE Commun Mag 48:26–30

Zhang H, Zhao J, Yang Z, Peng G, Di Z (2019) Low-DMGD, large-effective-area and low-bending-loss 12-LP-mode fiber for mode-division-multiplexing. IEEE Photonics J 11:1–8

Chapter 13
IoT for Landslides: Applications, Technologies and Challenges

Amrita Mohan, Ramji Dwivedi, and Basant Kumar

Abstract The most common existing human-induced and natural geohazard causing damage to infrastructure and property loss are landslides occurring in hilly regions of India. Landslide involves excessive surface movements, including rock failure, slope failure, debris flow, etc. Landslide prevention is based on the principles of slope stability engineering and developing techniques to reduce its effects on natural resources, river ecosystems, and infrastructure loss. Early warning and monitoring of landslides are the most critical detection/prevention strategies. With advancements in the evolution of the Internet of Things (IoT) technique, the recent innovative concept involves the utilization of IoT to strengthen particularly capability and accuracy of early warning and monitoring frameworks to prevent landslides. This chapter presents a brief overview of related research and technological advancements in IoT techniques used for landslide studies. First, it studies the importance and application of IoT for landslide monitoring and early warning system design, then investigates the fundamental layers in IoT along with the application and key technologies involved in landslide prevention. This chapter also highlights challenges with IoT-based landslide early warning and monitoring systems.

Keywords Internet of Things (IoT) · Landslide Detection · Wireless Sensor Network

A. Mohan (✉) · R. Dwivedi
GIS Cell, MNNIT Allahabad, Prayagraj, UP, India
e-mail: er.amritacs@gmail.com

R. Dwivedi
e-mail: ramjid@mnnit.ac.in

B. Kumar
ECED, MNNIT Allahabad, Prayagraj, UP, India
e-mail: singhbasant@mnnit.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
A. Rai et al. (eds.), *Paradigms of Smart and Intelligent Communication, 5G and Beyond*, Transactions on Computer Systems and Networks,
https://doi.org/10.1007/978-981-99-0109-8_13

13.1 Introduction

Landslide disasters are induced by natural and anthropogenic activities like abnormal changes or geodynamic movements due to the influence of internal, external (or human) actions on the earth's surface. The conditions that trigger landslides are slope oversteeping either by undercutting or by adding material on the upper portion of the hilly area, increase in the weight of slope material due to saturation of rock material and soil with water, vibrations due to blasting or earthquakes. In saturation, water adequately diminishes rock and soil materials by lowering friction and cohesion between particles.

The reason for landslide occurrence involves a disturbance in the natural stability of the slope in hilly areas accompanying intense rainfall, earthquake, volcanic eruption etc. Types of landslides include mudflows, debris flow, rockfall, etc.

Mudflow generally starts on steep slopes and is activated due to the accumulation of water on the ground surface, resulting in water-saturated rock and debris flow. The landslide occurrence was based on the change in slope from stable to unstable conditions. Many factors are causing unstable slopes. The natural causes of unstable slopes include; groundwater pressure to destabilize the slope, barely stable slopes are affected by earthquakes, and river or ocean waves causing erosion at the bottom. The anthropogenic factors of landslide occurrence include unnecessary construction destabilizing fragile slopes and vibrations due to machinery or traffic. Hilly regions with highly steep slopes are prone to landslides. Besides human lives, it also destroys agricultural lands with vast tracks, roads, buildings etc. Therefore, prevention and early warning frameworks for landslides are critical. Research on different case studies for landslide detection, early warning and monitoring system design exists. Although, the complex geographical conditions, variations and unpredictability of influential factors depict challenges in the initial stage of landslide prevention.

Landslide detection is based on methods like visual inspection (Mohan et al. 2021) based on digital aerial photographs (Rosin and Hervás 2005), and statistical techniques. The data-driven approach for landslide detection includes utilizing wireless sensor networks (WSN) (Mei et al. 2020). The importance of landslide detection is to limit the spontaneous collapse by using the concept of early detection of its motion, reducing the loss of lives caused by landslides. With the advancements in IoT technology (Cruz et al. 2018), IoT-based equipment, techniques and applications have made tremendous progress. The primary purpose of applying IoT techniques is to strengthen both the accuracy and efficiency of the landslide early warning and monitoring system. Traditionally, in the risk assessment framework, only limited consideration of soil behavior based on some approximations and experimental approaches (Oguz et al. 2020), with GIS, aerial photographs, and landslide inventory data play an essential role in determining landslide risk in a specific area.

The conventional method and technique for landslide detection are small in terms of data frequency, quality, and handling terms, i.e., deployment difficulty, high estimated cost, and restricted access to the area of interest. It is essential to monitor landslides and alert the public on a prior basis. Several methodologies have been proposed for landslide detection, but they are not feasible compared to IoT-based systems.

This chapter presents a survey on related research and technological advancements in the IoT used in landslide studies. It first surveys the background concepts of IoT along with its components and importance. Also, the application of IoT in landslide detection and recent technologies is discussed in subsequent sections and summarizes the challenges in IoT-based landslide detection systems followed by conclusion and future directions.

13.2 Related Concepts

This section describes the basics of IoT, and its application areas, followed by the description of essential layers involved in the architectural implementation of the IoT framework for landslide detection.

13.2.1 Internet of Things

The word 'Internet of Things (IoT)' (Kramp et al. 2013) was initially introduced by Kevin Ashton in 1998 during his invited talk for the Auto-ID center at the Massachusetts Institute of Technology (MIT). However, it was formally introduced by the International Telecommunication Union (ITU) in the internet report in 2005 (Kramp et al. 2013). IoT (Kramp et al. 2013) is a collection of devices connected with different sensors or electronic devices via a network allowing the devices to assemble required data and relocate the autonomous flow of data using modern techniques. The IoT-based approach is also used to perform analysis, monitoring, detection and rescue operations (Sinha et al. 2019). The primary application areas of IoT include:

1 Prevention and minimization of disaster risk by the geographic information system (GIS) and spaceborne communication techniques monitor disaster-prone areas and implement early warning systems (Sinha et al. 2019).
2 Providing real-time communication for appropriate support and response measures during an emergency (Sinha et al. 2019).
3 Providing disaster recovery such as finding a missing person online and fund management systems (Sinha et al. 2019). The IoT network smartly identify, locate, track, monitor and manage the devices using GPS (Global Positioning System), infrared sensors, RFID (radio frequency identification), and other types of sensing devices-based o defined arrangement (Mei et al. 2020; Nord et al. 2019)

The fundamental architecture of IoT is composed of three layers, shown in Fig. 13.1:

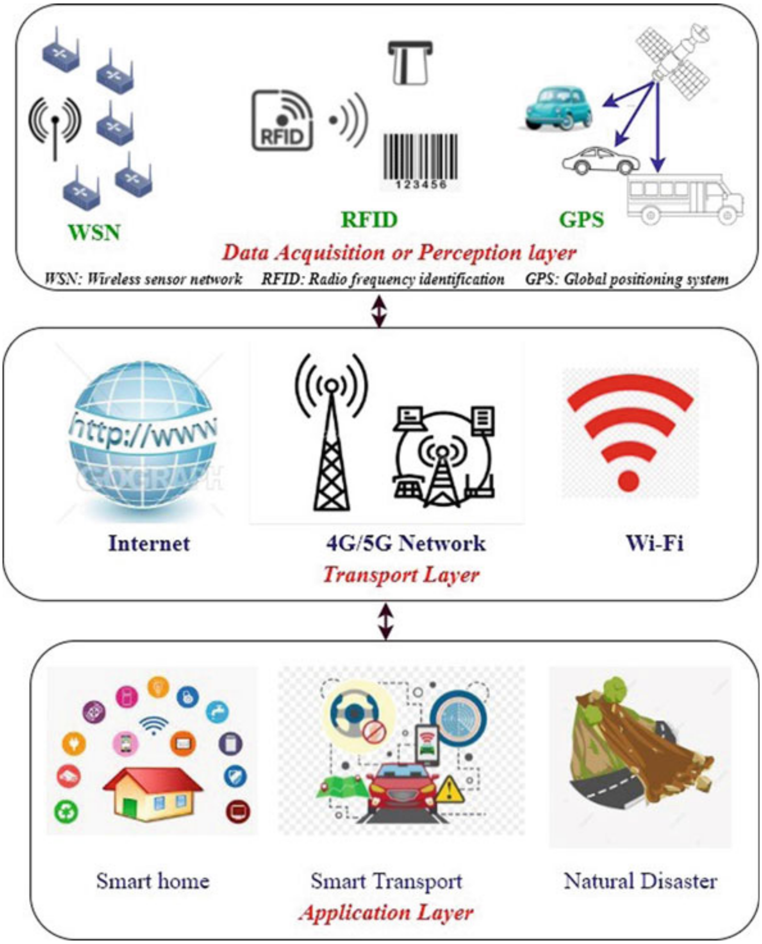


Fig. 13.1 Layers in IoT architecture

(a) ***Perception layer***: As the name implies, the perception layer means to perceive. It is also termed a sensor or sensing layer. To resolve the major concern of data collection amidst the real world and exhaustive data, the perception layer collects data using cameras, sensors and other devices (Mei et al. 2020). The fundamental techniques of the perception layer include sensor networks, RFID etc. (Ray 2018).

(b) ***Transport layer***: This layer is responsible for transmitting the collected data via an existing communication network, internet, television or radio network, special networks etc. (Montori et al. 2018). The fundamental technique used in the transport layer includes long-range wireless and wired communication scheme to process mass information, intelligent processing technology, and network fusion technology (Mei et al. 2020).

(c) ***Application layer*:** It is also an abstraction layer. This layer incorporates the IoT technique with specialized industrial applications to achieve an extensive range of intelligent solutions like intelligent transportation (Lu et al. 2014) and smart city (Jin et al. 2014). The key technique in the application layer includes machine learning, cloud computing, edge computing, etc.

13.2.2 IoT Application for Landslide Prevention

A landslide phenomenon involves the downward sliding of the dry mass of earth and rock along the slope under the influence of human and natural activities (Aleotti and Chowdhury 1999). Some researchers have applied IoT to overcome the shortcoming of traditional landslide monitoring methodology.

For ex. A WSN was developed by Sofwan et al. (2017) to measure the physical parameters that activate landslides. Wang et al. (2013) proposed an early warning and emergency monitoring framework using IoT to monitor unattended and real-time landslides.

To monitor landslide hazards in remote areas, a WSN-based monitoring system is introduced by Khoa and Takayam (Khoa and Takayama 2016). The early signs of catastrophic and rapid movement and to detect the beginning of slope movement, a monitoring technique defined by Moulat et al. (2018) Bian soongnern et al. (2016) implemented a system of early warning for landslides by utilizing a microcontroller-based sensing node. Several Ideas and solutions for designing a landslide early warning framework have been given (Intrieri et al. 2012). The Geocube Wireless Network (Benoit et al. 2014) was designed to monitor landslide displacement. The web-based interface for the continuous and automatic monitoring of landslips has been demonstrated in Frigerio et al. (2014). The challenges faced during the field deployment of WSN were summarized by Ramesh (2014). The early warning and monitoring system for analyzing slope stability has been given in Fan et al. (2016), Yang et al. (2017). A landslide phenomenon involves the downward sliding of the dry mass of earth and rock along the slope under the influence of human and natural activities (Aleotti and Chowdhury 1999). Some researchers have applied IoT to overcome the shortcoming of traditional landslide monitoring methodology. For ex. A WSN was developed by Softwan (Sofwan et al. 2017) to measure the physical parameters that activate landslides. Wang et al. (2009) proposed an early warning and emergency monitoring framework using IoT to monitor unattended and real-time landslides. Landslide monitoring in remote areas, a WSN-based monitoring system, is introduced by Khoa and Takayam (2016).

The early signs of catastrophic and rapid movement and to detect the beginning of slope movement, a monitoring technique defined by Moulat et al. (2018). Bian soongnern et al. (2016) implemented a system of early warning for landslides by utilizing a microcontroller-based sensing node. Several ideas for designing a landslide early warning framework have been given (Intrieri et al. 2012). The Geocube

Wireless Network (Benoit et al. 2014) was designed to monitor landslide displacement. The web-based interface for the continuous and automatic monitoring of landslips has been demonstrated by Frigero et al. (2014). The challenges faced during the field deployment of WSN were summarized by Ramesh (2014). The early warning and monitoring system for analyzing slope stability has been given in (Fan et al. 2016; Yang et al. 2017).

13.3 IoT Technology for Landslide Studies

13.3.1 Overview

The online investigation and early warning system for landslides has achieved initial results providing acceptable probability. Several technologies are involved in applying IoT to conduct landslide research. The IoT technologies include fiber optic sensing techniques, cloud computing and big data analysis, implementation of sensor networks and landslide early warning systems.

13.3.2 Sensor Network

The fundamental need of this study is to describe the importance of sensor networks required to monitor and detect landslides. The selection of appropriate geophysical sensors requires thorough knowledge along with a deep understanding of landslide mechanisms, triggering parameters, and geomorphology of the deployment area is needed. There are many reasons for landslide occurrence, including heavy rainfall, deforestation, volcanic activity, change in water level etc. Rain infiltration on the slope causes vulnerability, reduced safety factors, differences in the height of the water table, reduced shear strength holding soil or rock, increment in soil weight and reduced angle of repose (Ramesh 2014). The sensor network is composed of the base station and the sensor node required for real-time field information acquisition. The displacement gauge and rain gauge are the two main components of sensor nodes. Their primary function is to collect information related to mountainous displacement, rainfall and different other types of data required for real-time monitoring of natural disasters (Mei et al. 2020; Ramesh 2014).

The networking of sensor nodes involves both wired and wireless engineering. The monitoring information about nodes present in the network is spontaneously transferred to the base station via wired or wireless networks (Mei et al. 2020). Further, it is connected with a geohazard information transmission network at a remote location via the base station, providing real-time transmission and monitoring of data (Mei et al. 2020). It is difficult to monitor geohazard in harsh and complex environments due to the failure of the sensor node.

The reason for failure may include physical damage or environmental interference. If the sensing device fails, it is significant to sustain the sensor network's task without any suspension and with high fault tolerance (Mei et al. 2020; Hoblos et al. 2000). Furthermore, the sensor network demands to expand the amount of low-cost sensor devices to monitor various scales of geohazards (Mei et al. 2020; Cerpa et al. 2001) to improve network scalability (Cerpa et al. 2001; Rabaey et al. 2000). During geohazard monitoring for an extended period when devices are unattended in a severe atmosphere, the sensor node must be effective and have long battery life under available power resources (Lee et al. 2017). The intelligent disaster prediction system was developed by Tu et al. (2014) ensuring that node works properly and sleeps regularly. By using lower power consumption, sensor nodes are employed in sleep mode and variable working frequency, which is used by the whole network, to respond to modifications in the environment. The significant issues ensuring the reliability of the sensor network include: minimizing the casualty rate of message transfer among adjacent nodes (Xu and Zhang 2008), the position of WSN nodes (Hongwen et al. 2009), network requirement, networking of large-scale sensor nodes and flexibility of communication protocols (Integrated and Sensors 2020).

As presented in Fig. 13.2 landslide early warning and monitoring scheme consists of three components i.e., wireless sensor network layer (WSN) i.e., perception layer, transmission network and data collection layer. To acquire the real time landslide monitoring data, WSN layer uses a displacement sensor and different other types of sensors. The network layer is capable to transmit the collected data via internet, spaceborne networks, television and radio networks. The data layer analyzes, store, and calculate the collected data. The analysis result obtained, helps professionals in early warning investigation, and judgement along with frequent dispatch of contingency resources providing detailed data assistance for exhaustive research as shown in Fig. 13.3.

13.3.3 Fibre Optic Sensing Technology

The optical fiber sensing technology (Mei et al. 2020) is an internal monitoring technique which is instantly utilized to monitor the deformed rock and earth masses. Implementing both monitoring and sensing networks involves the application of optical fiber internally and on landslide surfaces. By using a monitoring system, relative displacement size across the internal part of the landslide might be recognized, and the status of the whole deformed condition of the landslip may be collected. At the same time, deformity of the mudslide sliding layer and spatial information about landslide occurrence can be captured precisely (Mei et al. 2020). Presently the most ideal distributed optical fiber technology corresponds to quasi-distributed optical fiber sensing technology called FBG (Lau et al. 2001), Brillouin-based visual time domain reflectometer (BOTDR) (Mei et al. 2020), the thoroughly divided optical fiber sensing technique called Brillouin optical time domain analysis (BOTDA) (Mei et al. 2020). An FBG overcomes the conflict of the light intensity fluctuation, optical

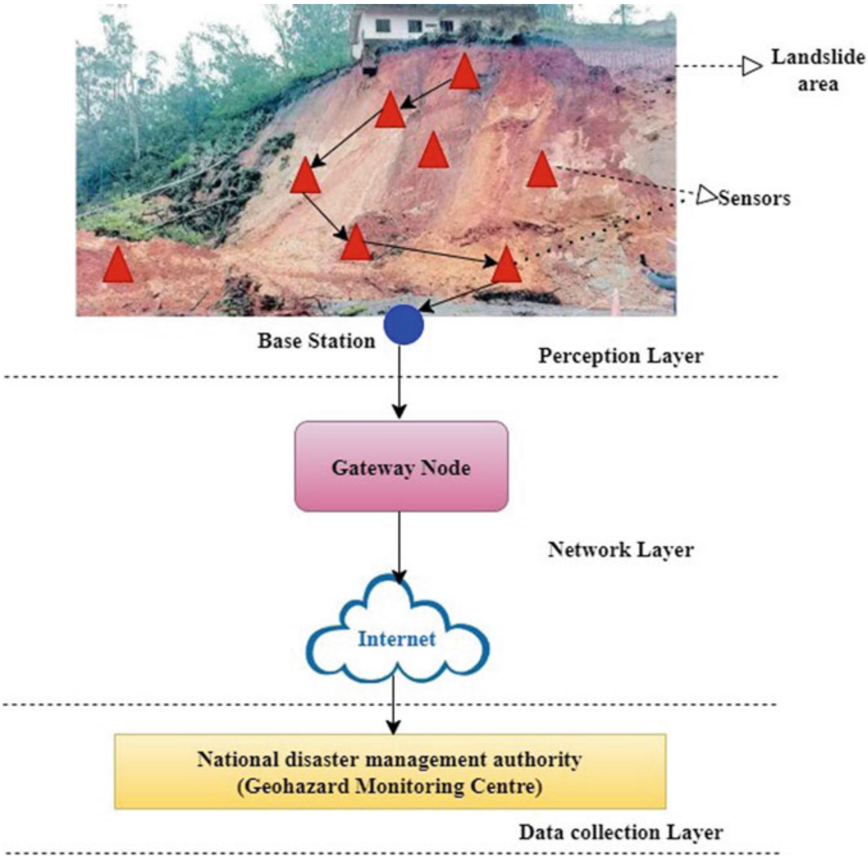


Fig. 13.2 WSN for landslide monitoring system

fiber bending loss and traditional measuring instruments. Due to the high sensitivity of sensors, landslides at various points across the entire region can be simultaneously measured. However, information about landslide deformation is distributed in the space in a non-continuous manner and could be proficient simply by using a quasi-distributed measurement technique for distinct points.

The two terminal path system with a wide measurement range, strong monitoring signal and high spatial resolution is followed by BOTDA (Soto et al. 2011), but breakpoints are not measured. Furthermore, the system structure is complex with high costs (Mei et al. 2020). As compared to BOTDA, the BOTDR has been used (Zhang et al. 2014) for monitoring because of its single-end access property. The measurement range, along with the accuracy, is finite by the signal intensity. BOTDR (Zhang et al. 2014), the monitoring technology, is based on a simple structure and single-ended facts, and further, it is utilized for measuring breakpoints. In contrast, to the conventional monitoring approach like Global Navigation Satellite System

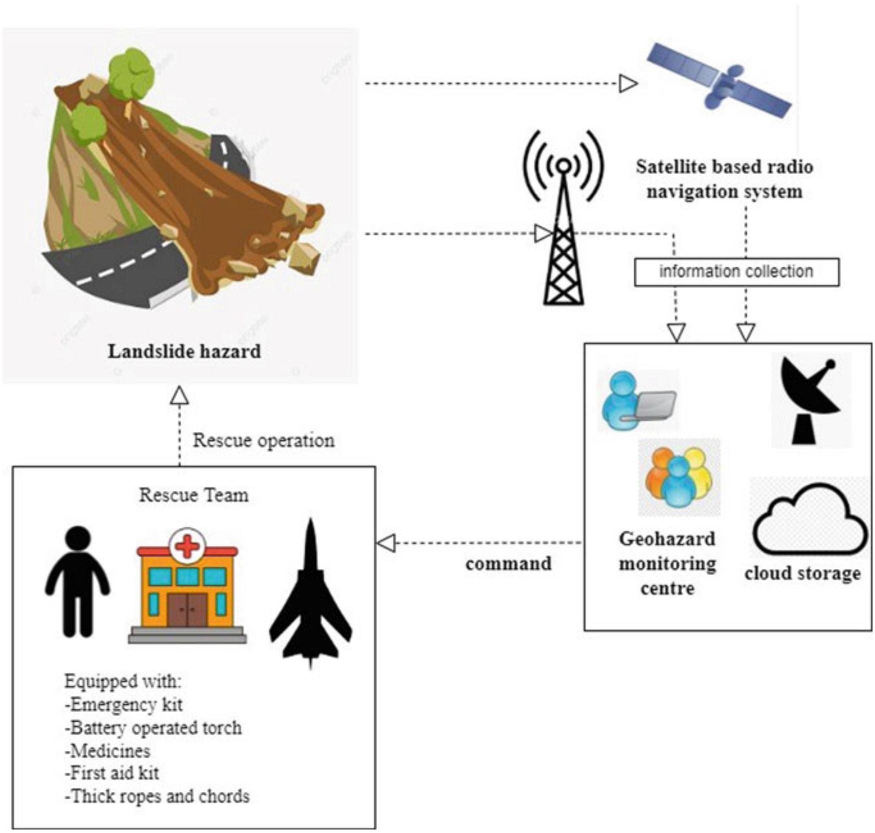


Fig. 13.3 Monitoring and rescue operation during landslides

(GNSS) and geodetic techniques, the fiber optic sensing technique to monitor natural hazard has several advantages like high sensitivity, anti-electromagnetic interference, fast response, corrosion resistance, variable shape, small size, huge transmission channel, and a reusable distributed measurement.

The landslide deformation data is quickly and accurately monitored. The slope position and sliding amount could be calculated in small duration. The surveillance data collected from the unknown risk points of geohazards are stored on the cloud timely and the information about the early warning might be issued based on the monitoring condition. The investigation, data mining and cloud platform helps to use the monitoring information to explore the reason for incident and settlement behaviors of natural disasters. The prediction of landslides becomes easy in terms of time and scale of occurrence (Mei et al. 2020). In conclusion, optical fiber sensing technique is continuous in nature while monitoring geohazard and it performs well in landslide monitoring (Mei et al. 2020).

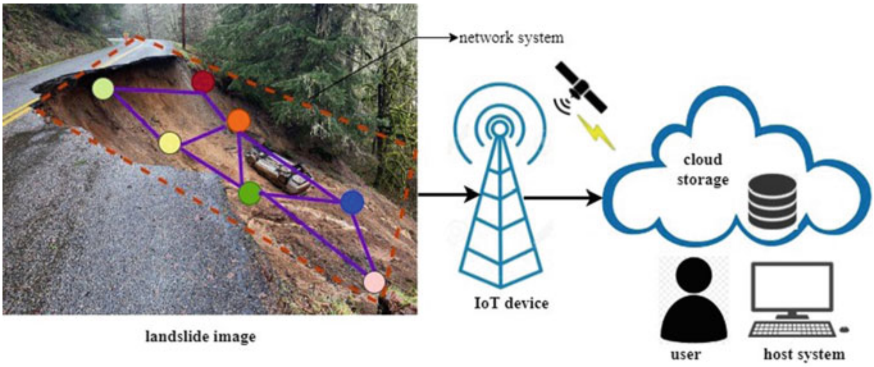


Fig. 13.4 Application of cloud computing network in a disaster area

13.3.4 Cloud Computing Platform

The increased distribution of web-based monitoring equipment and natural disaster monitoring information is also growing. The historical and real-time monitoring data plays an essential role in geohazard prohibition. It is necessary to employ massive data investigation technology to retrieve effective data from these excessive monitoring data collection (Mohammadi et al. 2018). Especially the monitoring information should be precise, updated and examined to accomplish better as well an authentic early warning system for disaster prevention. Based on the recent trends, the cloud monitoring concept in the context of huge data is a better option for integrating massive geo-information along with processing and inefficient analysis (Coppolino et al. 2019). The cloud computing approach, as shown in Fig. 13.4, allows fast and efficient computation on traditional hosts/servers, diminishing the software and hardware demands of local computing equipment (Mei et al. 2020). The distribution of a predictable early warning solution means ensuring the detection abilities of both initial signs of massive failure and landslide triggering factors like heavy rainfall, floods etc. (Pascal et al. 2014).

13.4 Challenges with IoT-Based Monitoring System

The reliability of the monitoring system (Wang et al. 2014) based on IoT is the ability of the procedure to satisfy the serving requirements under simultaneous destructive causes frequently. There are many factors involved in making IoT monitoring systems unreliable. Natural disaster monitoring systems based on IoT require comprehensive, extensible and high geared service to support real-time supervision in complicated and rigid geographic environments. There are several challenges encountered while achieving this reliability (Chernyshev et al. 2018).

(a) Sensor networks for landslide early warning and monitoring system

Implementing a landslide monitoring system involves the usage of sensor networks because long-run monitoring is required and generally caused due to rainfall events. The purpose of a monitoring system based on sensor networks demands less power (because battery replacement is complicated) and serves distantly. The essential components of a landslide monitoring system are nodes equipped with an accelerometer to measure and detect vibrations caused by a landslide (Hart and Martinez 2020). The critical consideration while using the sensor network is energy efficiency. Since the hazardous condition in the environment is complex and harsh and it requires an extended monitoring period. The battery replacement of the sensor node is difficult in hilly areas. While nodes communicate together within the landslide monitoring framework, severe circumstances like rainfall will immensely affect the communication channel quality. In such cases, if the destination nodes desire accurate sensor details, the source node must start the Zigbee stack transmit scheme (Mei et al. 2020) conveying so that the node expends supplementary energy. To preserve useful energy, the sensors should be capable of identifying individual reports and event data timely.

Thus, it must be essential to establish a sensing network, ensuring communication quality in complicated surroundings and concurrently employing the minimum quantity of power. Geological risks such as mudflow and rock falls are damaging and generally must be limited period. Shocks generated in the midst of geohazard will conduct fast damage to different transmission capabilities, making it complicated to gather, store and sense data during a disaster. Thus, it is required to develop a sensor that can deal with rapid geohazard at both software and hardware levels (Mei et al. 2020).

(b) Unifying technical standards of IoT

The IoT-based geohazard early warning and monitoring framework design has gained huge progress, the production and selection of sensors familiar to the field conditions, interfaces, lack in technical standards of communication network and gateway manufacturing, online early warning and monitoring of geohazards also needs to be established. The series of standards contributes a guarantee for massive application and promotion. To establish a standard service interface and to expedite the implementation, formulation, and common standards utilization verification, simulation and testing of several standards are required (Trappey et al. 2017).

(c) Security of IoT devices

There are more significant security issues among emerging technologies compared to other areas. In the past ages, numerous research has been conducted to figure out the numerous security problems associated with the IoT (Ferrag et al. 2019), such as key management issues (Roy et al. 2018) and policy implementation (Sicari et al. 2018). To address the merging threats and implement robust security standards for IoT devices and systems (Yang et al. 2017), the IoT manufacturing company must work strictly with standardization organizations and regulators.

(d) Artificial intelligence and big data

For early warning and assessment of landslides, spatial and temporal data play an essential role. The critical indicator for early warning systems is the real-time warning of geohazards. Therefore, it is necessary to achieve better accuracy, the intelligence of data cleaning and real-time capability, conversion and large-scale geohazard monitoring must be achieved. Furthermore, under the big data environment, the extraction and identification of underlying disaster triggering agents are the two major issues to be resolved to prevent a natural disaster. Within artificial intelligence, deep learning offers an effective means to process and analyze big data (O'Leary 2013); concurrently, big data highlights the necessity for the evolution of deep learning techniques (O'Leary 2013) with huge sample data.

By utilizing the experimental outcome of big data founded based on enormous instances of the temporal and spatial information about geological hazards depending upon the structure of deep learning mechanism using the non-linear nature and high precision fitting of the difficult natural disasters, the additional ideal theoretical scheme and a more precise hazard evaluation and prediction framework for natural disaster have been developed (Mei et al. 2020).

(e) IoT and 3S

The '3S technology' comprises a Geographic information system (GIS), Global positioning system (GPS) (Mei et al. 2020) and Remote sensing technology (RS) (Ding et al. 2014). The important trends and challenges for further research include: storing, managing and analyzing the function of geographical data via GIS (Hossain et al. 2009), the rapid data acquisition of spatial location using GPS (Wang et al. 2013) and utilizing high-resolution disaster information and satellite images given by RS technique (Pohl and Genderen 1998).

13.5 Summary

The review of application areas, techniques and difficulties of the IoT while monitoring and preventing landslides has been demonstrated in this chapter. The literature reviews shows that the IoT is extensively used in different types of landslide analysis like debris flow, rockfall, etc. The important technologies involved in implemention of IoT-based early warning and monitoring systems have been discussed in this chapter. The advantage of IoT based framework for landslide monitoring and early warning system design includes fast, accurate, safe and smart performance. However, there are still various challenges to be discussed. The sensor networks play a crucial role in landslide assessment and monitoring, but the sensor reliability and power consumption should be improved. In future work, emerging techniques like integration of 5G (fifth generation) with IoT-based monitoring system design, deep learning algorithm for landslide monitoring and early warning system design must be encouraged.

References

Aleotti P, Chowdhury R (1999) Landslide hazard assessment: summary review and new perspectives. Bull Eng Geol Environ 58(1):21–44. https://doi.org/10.1007/s100640050066

Benoit L, Briole P, Martin O, Thom C (2014) Real-time deformation monitoring by a wireless network of low-cost GPS. J Appl Geod 8(2):119–128. https://doi.org/10.1515/jag-2013-0023

Biansoongnern S, Plungkang B, Susuk S (2016) Development of low cost vibration sensor network for early warning system of landslides. Energy Procedia 89:417–420. https://doi.org/10.1016/j.egypro.2016.05.055

Cerpa A, Elson J, Hamilton M, Zhao J, Estrin D, Girod L (2001) Habitat monitoring: Application driver for wireless communications technology. SIGCOMM LA 2001—Work. Data Commun. Lat. Am. Caribb., no. February, pp. 20–41. https://doi.org/10.1145/371626.371720

Chernyshev M, Baig Z, Bello O, Zeadally S (2018) Internet of things (IoT): research, simulators, and testbeds. IEEE Internet Things J 5(3):1637–1647. https://doi.org/10.1109/JIOT.2017.2786639

Coppolino L, D'Antonio S, Mazzeo G, Romano L (2019) A comprehensive survey of hardware-assisted security: from the edge to the cloud. Internet of Things (Netherlands) 6. https://doi.org/10.1016/j.iot.2019.100055

Da Cruz MAA, Rodrigues JJPC, Al-Muhtadi J, Korotaev VV, De Albuquerque VHC (2018) A reference model for internet of things middleware. IEEE Internet Things J 5(2):871–883. https://doi.org/10.1109/JIOT.2018.2796561

Ding H, Sun LS, Wang JL, Wang X (2014) Design and realization of 3S technology-based geological disaster database system in Liaoning Province. Appl Mech Mater 580–583:2708–2712. https://doi.org/10.4028/www.scientific.net/AMM.580-583.2708

El Moulat M, Debauche O, Mahmoudi S, Brahim LA, Manneback P, Lebeau F (2018) Monitoring system using internet of things for potential landslides. Procedia Comput Sci 134:26–34. https://doi.org/10.1016/j.procs.2018.07.140

Fan Z, Li F, Zheng B, Cui F (2016) The mudslide monitoring and early warning system based on WebGIS in Nujiang. Int Conf Geoinformatics 2016. https://doi.org/10.1109/GEOINFORMATICS.2015.7378633

Ferrag MA, Derdour M, Mukherjee M, Derhab A, Maglaras L, Janicke H (2019) Blockchain technologies for the internet of things: research issues and challenges. IEEE Internet Things J 6(2):2188–2204. https://doi.org/10.1109/JIOT.2018.2882794

Frigerio S et al (2014) A web-based platform for automatic and continuous landslide monitoring: The Rotolon (Eastern Italian Alps) case study. Comput Geosci 63:96–105. https://doi.org/10.1016/j.cageo.2013.10.015

Hart JK, Martinez K (2020) Sensor networks and geohazards, 2nd edn. Elsevier Inc. https://doi.org/10.1016/b978-0-12-818234-5.00037-7

Hoblos G, Staroswiecki M, Aitouche A (2000) Optimal design of fault tolerant sensor networks. IEEE Conf Control Appl Proc 1:467–472. https://doi.org/10.1109/cca.2000.897468

Hongwen X, Weiming Y, Yuxiang S, Qizhi Z (2009) Genetic algorithm for nodes location in wireless sensor network to monitor abrupt geological hazards. In: 2009 international workshop intelligent systems and applications, ISA, pp 1–4. https://doi.org/10.1109/IWISA.2009.5072887

Hossain MS, Chowdhury SR, Das NG, Sharifuzzaman SM, Sultana A (2009) Integration of GIS and multicriteria decision analysis for urban aquaculture development in Bangladesh. Landsc Urban Plan 90(3–4):119–133. https://doi.org/10.1016/j.landurbplan.2008.10.020

Wang H, Guo Tuo X, Yu Zhang G, Ling Peng F (2009) Panzhihua airport landslide (Oct. 3rd 2009) and an emergency monitoring and warning system based on the internet of things. J Mt Sci 10(5):873–884. https://doi.org/10.1007/s11629-013-2368-3

Integrated W, Sensors N (2020) Wireless sensors networks. Encycl Wirel Netw 43(5):1505–1505. https://doi.org/10.1007/978-3-319-78262-1_300743

Intrieri E, Gigli G, Mugnai F, Fanti R, Casagli N (2012) Design and implementation of a landslide early warning system. Eng Geol 147–148:124–136. https://doi.org/10.1016/j.enggeo.2012.07.017

Jin J, Gubbi J, Marusic S, Palaniswami M (2014) An information framework for creating a smart city through internet of things. IEEE Internet Things J 1(2):112–121. https://doi.org/10.1109/JIOT.2013.2296516

Van Khoa V, Takayama S (2018) Wireless sensor network in landslide monitoring system with remote data management. Meas J Int Meas Conf 118:214–229. https://doi.org/10.1016/j.measurement.2018.01.002

Kramp T, van Kranenburg R, Lange S (2013) Introduction to the internet of things. https://doi.org/10.1007/978-3-642-40403-0_1

Lau KT, Yuan L, Zhou LM, Wu J, Woo CH (2001) Strain monitoring in FRP laminates and concrete beams using FBG sensors. Compos Struct 51(1):9–20. https://doi.org/10.1016/S0263-8223(00)00094-5

Lee HC, Ke KH, Fang YM, Lee BJ, Chan TC (2017) Open-source wireless sensor system for long-term monitoring of slope movement. IEEE Trans Instrum Meas 66(4):767–776. https://doi.org/10.1109/TIM.2017.2657838

Lu N, Cheng N, Zhang N, Shen X, Mark JW (2014) Connected vehicles: solutions and challenges. IEEE Internet Things J 1(4):289–299. https://doi.org/10.1109/JIOT.2014.2327587

Mei G, Xu N, Qin J, Wang B, Qi P (2020) A survey of internet of things (IoT) for geohazard prevention: applications, technologies, and challenges. IEEE Internet Things J 7(5):4371–4386. https://doi.org/10.1109/JIOT.2019.2952593

Mohammadi M, Al-Fuqaha A, Guizani M, Oh JS (2018) Semisupervised deep reinforcement learning in support of IoT and smart city services. IEEE Internet Things J 5(2):624–635. https://doi.org/10.1109/JIOT.2017.2712560

Mohan A, Singh AK, Kumar B, Dwivedi R (2021) Review on remote sensing methods for landslide detection using machine and deep learning. Trans Emerg Telecommun Technol 32(7). https://doi.org/10.1002/ett.3998

Montori F, Bedogni L, Di Felice M, Bononi L (2018) Machine-to-machine wireless communication technologies for the Internet of Things: taxonomy, comparison and open issues. Pervasive Mob Comput 50:56–81. https://doi.org/10.1016/j.pmcj.2018.08.002

Nord JH, Koohang A, Paliszkiewicz J (2019) The internet of things: review and theoretical framework. Expert Syst Appl 133:97–108. https://doi.org/10.1016/j.eswa.2019.05.014

O'Leary DE (2013) Artificial intelligence and big data. IEEE Intell Syst 28(2):96–99. https://doi.org/10.1109/MIS.2013.39

Oguz EA, Robinson K, Depina I, Thakur V (2020) IoT-based strategies for risk management of rainfall-induced landslides: a review. 733–738. https://doi.org/10.3850/978-981-11-2725-0-is13-2-cd

Pascal B, Thierry V, Emmanuelle K, Yves G (2014) Cloud monitoring: an innovative approach for the prevention of landslide risks. Landslide Sci Pract Early Warn Instrum Monit 2:665–670. https://doi.org/10.1007/978-3-642-31445-2_87

Pohl C, Van Genderen JL (1998) Review article multisensor image fusion in remote sensing: concepts, methods and applications. 19(5). https://doi.org/10.1080/014311698215748

Rabaey JM, Ammer MJ, Da Silva JL, Patel D, Roundy S (2000) PicoRadio supports ad hoc ultra-low power wireless networking. Computer (Long. Beach. Calif) 33(7):42–48. https://doi.org/10.1109/2.869369

Ramesh MV (2014) Design, development, and deployment of a wireless sensor network for detection of landslides. Ad Hoc Netw 13:2–18. https://doi.org/10.1016/j.adhoc.2012.09.002

Ray PP (2018) A survey on Internet of Things architectures. J King Saud Univ Comput Inf Sci 30(3):291–319. https://doi.org/10.1016/j.jksuci.2016.10.003

Rosin PL, Hervás J (2005) Remote sensing image thresholding methods for determining landslide activity. Int J Remote Sens 26(6):1075–1092. https://doi.org/10.1080/01431160512331330481

Roy S, Chatterjee S, Mahapatra G (2018) An efficient biometric based remote user authentication scheme for secure internet of things environment. J Intell Fuzzy Syst 34(3):1403–1410. https://doi.org/10.3233/JIFS-169435

Sicari S, Rizzardi A, Miorandi D, Cappiello C, Coen-Porisini A (2016) Security policy enforcement for networked smart objects. Comput Netw 108:133–147. https://doi.org/10.1016/j.comnet.2016.08.014

Sinha A, Kumar P, Rana NP, Islam R, Dwivedi YK (2019) Impact of internet of things (IoT) in disaster management: a task-technology fit perspective. Ann Oper Res 283(1–2):759–794. https://doi.org/10.1007/s10479-017-2658-1

Sofwan A, Sumardi, Ridho M, Goni A, Najib (2017) Wireless sensor network design for landslide warning system in IoT architecture. In: Proceedings of the 2017 4th international conference on information technology, computer, and electrical engineering (ICITACEE), vol. 2018, pp 280–283. https://doi.org/10.1109/ICITACEE.2017.8257718

Soto MA, Bolognini G, Di Pasquale F (2011) Long-range simplex-coded BOTDA sensor over 120km distance employing optical preamplification. Opt Lett 36(2):232. https://doi.org/10.1364/ol.36.000232

Trappey AJC, Trappey CV, Hareesh Govindarajan U, Chuang AC, Sun JJ (2017) A review of essential standards and patent landscapes for the Internet of Things: a key enabler for Industry 4.0. Adv Eng Informatics 33:208–229. https://doi.org/10.1016/j.aei.2016.11.007

Tu JF, Yang Y, Li CH, He AP, Li L (2014) A context-adaptive and energy-efficient wireless sensor network for debris flow monitoring. In: Proceedings of the 2014 international conference on wireless communication sensor network. WCSN, pp 157–162. https://doi.org/10.1109/WCSN.2014.39

Wang H, Liu M, Hong S, Zhuang Y (2013) A historical review and bibliometric analysis of GPS research from 1991–2010. Scientometrics 95(1):35–44. https://doi.org/10.1007/s11192-012-0853-z

Wang C, Xing L, Vokkarane VM, Sun Y (2014) Reliability and lifetime modeling of wireless sensor nodes. Microelectron Reliab 54(1):160–166. https://doi.org/10.1016/j.microrel.2013.08.001

Xu H, Zhang Q (2008) Wireless sensors network design for real-time abrupt geological hazards monitoring. In: Proceedings of the interenational conference on computer science and information technology (ICCSIT 2008), pp 959–962. https://doi.org/10.1109/ICCSIT.2008.155

Yang Y, Wu L, Yin G, Li L, Zhao H (2017) A survey on security and privacy issues in internet-of-things. IEEE Internet Things J 4(5):1250–1258. https://doi.org/10.1109/JIOT.2017.2694844

Yang Z et al (2017) A multi-source earlywarning system of MEMS based wirelessmonitoring for rainfall-induced landslides. Appl Sci 7(12). https://doi.org/10.3390/app7121234

Zhang Y, Wu X, Ying Z, Zhang X (2014) Performance improvement for long-range BOTDR sensing system based on high extinction ratio modulator. Electron Lett 50(14):1014–1016. https://doi.org/10.1049/el.2014.0909

Chapter 14
A Review: Dust Cleaning Approach of Solar Photovoltaic System Using IOT & ML

Kunwar Babar Ali, Zaiba Ishrat, and Seema Nayak

Abstract Governments have been pushed to boost the installation of Non-Conventional Energy Sources as a outcome of rising energy need and trouble about global warming. Their development is captious in light of the 2030 program's sustainable development goals, which include boosting the part of Non-Conventional energy sources in global power mix. Solar power is a viable energy source that has received widespread concern due to its availability and lack of fuel expenses, resulting in the development of a number of uses, including photo-voltaic (PV) panels. It is fairly obvious that dust formation is one of the most fundamental factors influence the performance of Photovoltaic panels. Atmospheric factors such as atmospheric temperature, dirt formation, partial shade, and so on influence the efficiency of photovoltaic panels. This article gives detailed overview of cleaning approaches for solar photovoltaic panels that are currently available and being utilized by researchers. The study's main goal was to examine the literature on solar photovoltaic module cleaning approaches based on IOT and Machine Learning, in order to determine research gaps in the field of solar panel cleaning.

Keywords Renewable energy · Photo-voltaic panels · IOT · Machine learning

14.1 Introduction

Now on a daily basis, power-related angles are getting very much important. They imply, as an example, a logical use of resources, the atmospheric collision associated with the impurity emission and therefore the utilization of unconventional resources. For these causes, there's a consistent worldwide absorption in defendable energy manufacturing and energy reduction (Renewables 2020 global status report). Now

K. B. Ali
Department of Computer Science & Engineering, NIU, Greater Noida, India

Z. Ishrat (✉) · S. Nayak
ECE, IIMT COE, Greater Noida, India
e-mail: hellozaiba@gmail.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

A. Rai et al. (eds.), *Paradigms of Smart and Intelligent Communication, 5G and Beyond*, Transactions on Computer Systems and Networks,
https://doi.org/10.1007/978-981-99-0109-8_14

a day's solar energy received a wide attraction of the government due to its lavish availability and less cost as compared to fossil fuel (Al Shehri et al. 2016, 2017; Zorrilla-Casanova et al. 2011; Solar panel cleaning—automatic cleaning system).

The main problem now a day on earth is global warming. Due to global warming the temperature of the earth rises 2 °C which is responsible for many issues such as shortage of groundwater, flooding, disappearing of many animal species, and increasing the heat during summer days. To overcome this issue, the prime solution is to less use of fossil fuels, otherwise it will be very harmful to human acts (Fig. 14.1).

Due to decrement in renewable energy resources solar energy plays a very important role. It is extremely set to become inexpensive in future time and rising as better technology in terms of both cost and applications. The total incident sun power on earth is 1366w approx on earth which is totally free of cost. The main property of solar power is that the sunlight directly converted into solar power by using a small PV cell. Researchers are doing research in the field of manufacturing PV panel to convert sunlight efficiently into solar energy. This solar energy also required low manpower as compared to other conventional energy sources like fossil fuel. The solar PV sector has come a long way, with significant developments in recent years in terms of installations (including off-grid), costs, and technology, as well as the creation of important associations for solar energy (Fig. 14.2). Undoubtedly, solar energy will remain a crucial renewable energy source in the ensuing decades.

Photovoltaic (PV) may be a procedure of generating wattage by changing radiation into power using semiconductors that revel the PV effect. The amount of solar energy incident on a PV cell is expressed as watts per square meter, which is also known as solar irradiation. Solar irradiation depends on the location it will vary from one location to another and sometimes at one location it also changes according to time. The ability of power generation from a PV element turn on the quantity of solar photospheric that reaches the solar panel (Adinoyi and Said 2013).

Photons incident on solar panel can produce both voltage and current. The prime requirement of this process is the material those have the property of absorption of light by which electron will move from a lower energy start to a higher energy level. Than these electrons will move from a higher energy level to externally connected

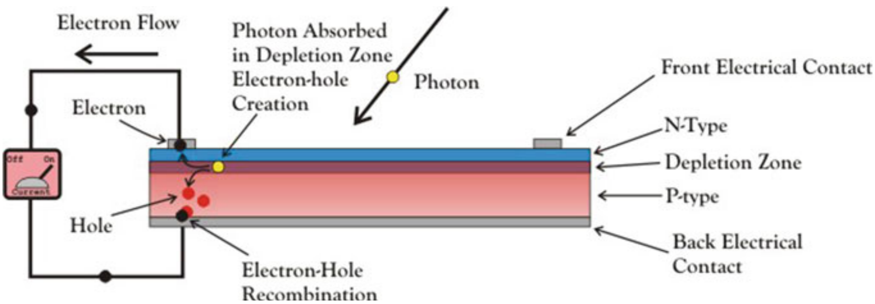


Fig. 14.1 Reaction of solar energy

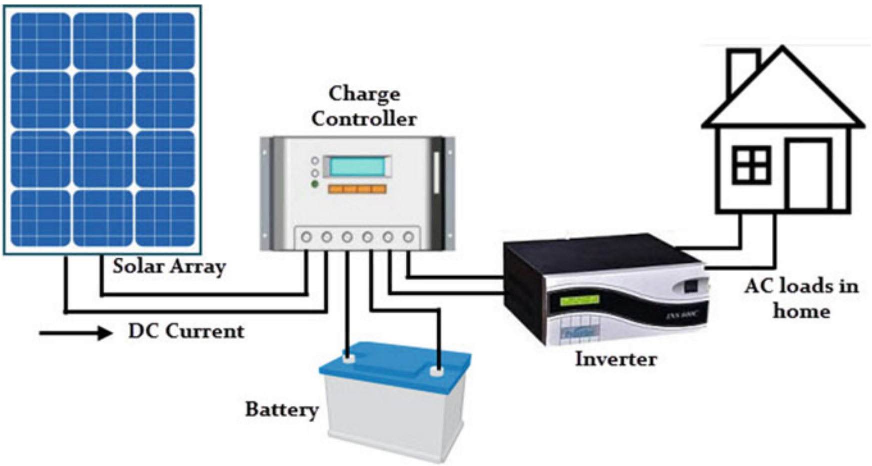


Fig. 14.2 Generation of solar energy

circuit. Different materials can be used for the photon to electricity conversion but mostly used material is a semiconductor device.

Mono-crystalline Solar Panels: in this type of PV panel used material for manufacturing is pure silicon. This is a highly effective and expensive material so this type of panel can give an efficiency of upto 20%.

Polycrystalline Solar Panels: The difference in the manufacturing of mono-crystalline and polycrystalline solar panels is that melted trashs of the silicon is used to Polycrystalline Solar Panels. This method offers less construction cost, but this method offers less efficiency by upto 15–17% efficiency. Although this efficiency can be improved by using novel technologies.

Thin-Film Solar Panels: The thin-film solar panel uses different material instead of silicon that are cadmium telluride (CdTe), amorphous silicon (a-Si), and Copper Indium Gallium Selenide (CIGS) or with any mixture of a different material. It offers minimum efficiency as compared to other PV panels which is around 11%.

These are the following steps for the conversion of photon energy into electricity: Figs. 14.1 and 14.2.

- Due to the incident of light electrons will be generated
- The compilation of the light-generated electrons to generate a current
- The generated current is responsible to generate the voltage across the cell
- The generated power will dissipate across the external electronic circuit.

Thus the output of the module is affected by many factors that the factors that may be controlled a number of these factors include: the sort of PV material, radiation intensity received, cell temperature, parasitic resistances, clouding, inverter efficiency, soil, width orientation, earthly location, cable thickness, dust accumulation. It is a remark that depending on the location the daily sun energy loss will be up to 1% and the monthly decrement in efficiency is up to 80%which is quite countable.

The specific mass and size of the object determines the extent of efficiency degradation. Deposition of dust particles on the PV module surface. As the amount of deposited dust grows, As the size gets smaller, the module's efficiency and power output decline. As more radiation is blocked by tiny particles on the PV module surface, power production declines.

The deposition of various pollutants may take the form of red soil, ash, sand, calcium carbonate, silica, etc. The energy yield of PV panels may significantly decline in the presence of air pollution; even immediately following a brief time of the panels' outside exposure (for instance, two months) without cleaning cause the generation of energy to drop by about 6.5% (Sarver et al. 2013).

The current paper point to produce an assessment of dirt issue and scour approaches based on emerging technologies. Due to the effect of soiling two problems arise named as hard shading and soft shading effect.

Due to hard shading issue on the PV Panel tiny patch of dirt accumulate therefore Voc of the PV panel will decrease.

Due to the soft shading effect dust will distribute on solar panel uniformly therefore it will reduce the Isc.

There are many tactics like robotic cleaning, automatic water dust removal, nano film implementation and self-cleaning which are in current practice for dust removal. The decision of cleaning strategy depends on panel size, location and also the decision of optimal time for cleaning is also mandatory otherwise in an unnecessary way economic wastage will take place. The application of data science and machine learning for cleaning operation decision-making can be a significant step forward for solar PV energy systems. Such an enhanced Making a cleaning selection would undoubtedly aid in lowering the LCoE (levelised cost of energy). The machine learning technique decides a schematic way for a cleaning interval scheduler that could serve as a basis for future research. Another significant challenge with effective solar PV is soiling operation. The buildup of outside particles on the panel surface doesn't just prevent solar energy from entering the panel radiation, but it also raises cell temperature, producing a dual impact.

Section 14.1 discuss the natural cleaning method, Sect. 14.2 manual cleaning method, Sect. 14.3 Sprinkle based cleaning, Sect. 14.4 cleaning strategies supported by emerging technologies, Sect. 14.5 ECD based strategy, Piezoelectric based and self-cleaning approach, Sect. 14.6 discussed a comparative analysis of different approaches and in last we conclude the paper.

14.2 Natural Cleaning System

The native powers are working to get rid of the dirt, like breeze generation, planetary motion and also the wipe of the aqua. The importance of cleaning so as to keep up planning and therefore the remarkable of uncontaminated cleaning by cloudburst haven't been extensively studied in several weather. A one-only native cloudburst event was adequate to scrub the console to grade that refreshment output to within 1%

of the dynamically cleaned panel. Given the success of native cloudburst in washing the console, appropriate group of rule for maintaining superlative regulation will be set on for various weather conditions. The main problem of this method atmospheric condition because we can't predict about the rain. Breeze movement, mainly when it's an average rate and doesn't cause dirt, can naturally wipe the Photovoltaic plane facing its motion. In spite of the fact that forecasting hustle times and quantity isn't possible with high precision, rain can rise the planning of the Photovoltaic plane by fend off or lessen the buildup of impurities (Gaier et al. 1990; Chen et al. 2020).

Gaier et al. (1990) announce that they'd deliberate the rationality of this wind procedure. It is feasible that the solar cell arrangement will be go round to erect or oblique location to get rid of the dirt easily when first flush of morning, late at, night and a time period. However, the turning of the big electric cell arrangement is extremely strenuous.

14.3 Manual Cleaning System

Manual cleaning is one in every of the foremost used ways of wiping of evidentiary solar console. Standing on hierarchy or ascend on roofs, this is arise employment for knowledgeable because it is hazardous and compulsory the proper material. Even so, this procedure may need little or no affect on the annually yield but it gets done frequently. Particularly if the task is completed by a fabricator this may be very bad contribution there the labour costs will exceed the profitable return (Moreno et al. 2006). Today the foremost common thanks to clean solar panels is to use dishwashing soap and/or commercial cleaning detergents. Running soap and/or water from the rooftop can stain the edges of a house if there's not an overhang.

However, there's the likelihood that direct contact with the surface of the PV may cause slurry because the movement of the brushes and their pressure on the surface of the painting cannot be certain to be in equal pressure along the PV area. This method is pricey compared to the above methods because it requires skilled labour (Solar panel cleaning—automatic cleaning system; Sai Lalitha Chowdary et al. 2017).

Al Shehri et al. (Solar panel cleaning—automatic cleaning system; Sai Lalitha Chowdary et al. 2017) conducted an empirical study to see the efficient use of both nylon, cloth, and synthetic rubber foam brushes as shown in Fig. 14.3.

Zorrilla-Casanova et al. (Jadhav 2019) found that using water to scrub PV is that the optimal method. This method is also very expensive in PV stations with extended space, and therefore the cost increases when employed in areas affected by water scarcity Fig. 14.4.

Fig. 14.3 Manual cleaning by rubber and cloths

Fig. 14.4 Manual cleaning by running water and soap

14.4 Mechanical Cleaning Techniques

The mechanical tactic employs brushes, blowing, vibration and ultrasonic driving to remove the dirt from the solar panel. The brushing method is the same as wind screen wiper, it is just like the booming process. This technique can clean the solar panel in a very less time and offers high efficiency. The main disadvantage of this method is that this tactics does not offer high efficiency due to the small and sultry property of dirt particles. Also these types of machines have wear and tear loses and due to the large size of the plant this method offers high cost.

14.5 Sprinkle System

To enhance the PV efficiency Sprinkler method is regularly employed in dusty regions to stay PV clean. It's the identical cleaning outcome as rainwater and can clear panels at a comparatively low price. A similar system is available on the web site of Heliotex (Vaghani et al. 2019), exist out of a water filtration system and soap dispensing system. Even though this is often a comparatively superior working arrangement,

Fig. 14.5 Sprinkle based cleaning

there is a pair of drawbacks to that. This technique is suitable for very dusty regions where dust stacks up quickly but will not be cleaned by brush only Fig. 14.5.

14.6 Cleaning Approach Based on IOT

There are many dusts cleaning strategies based on IoT the main aim of using IoT In solar energy systems is to provide the automation for cleaning systems. To clean solar panels, the researcher created an integrated microprocessor and voltage sensing system. The system has a wiper and the ability to add other features such as water and soap. Because the hardware can operate in low vision, there will be no impediment to electricity production during the sunlight. And it was tested on a 2.2 KW plant. A cost–benefit analysis is carried out, as well as a payback period calculation. Deploying it results in a lot of extra power generation, as well as being cost-effective and dependable. Instead of employing LDRs and dust sensors, they use a voltage-sensing circuit that detects the panel voltage directly. Because sensors must be cleaned or reset on a regular basis. As a result, this solution solves the problem Figs. 14.6 and 14.7.

Researcher developed the narrative structural design of a dust cleaning system for Solar panels using IoT, Each PV panel is connected to a dust sensor and a cleaner system in order to integrate into the reference level in the environment. The (Raspberry Pi) work as a communication gateway to connect these smart devices. The dust sensor data is collected by the communication gateway and sent to the knowledge base and server system. Data is sent to the knowledge pre-processing in the database/server system, where it is transformed into a clear format before being sent to the dust cleaner decision algorithm. This program compares the dust brink value to the real data. The system sends a notification to the user/admin if the dust data

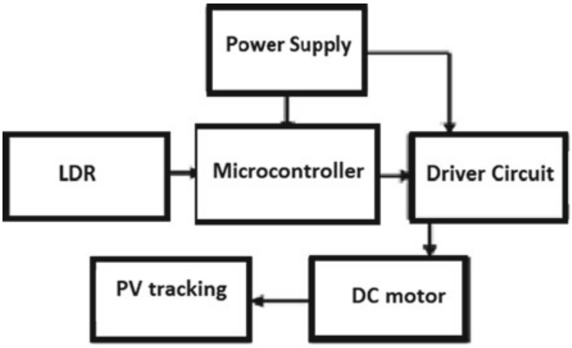


Fig. 14.6 Microcontroller based cleaning robot

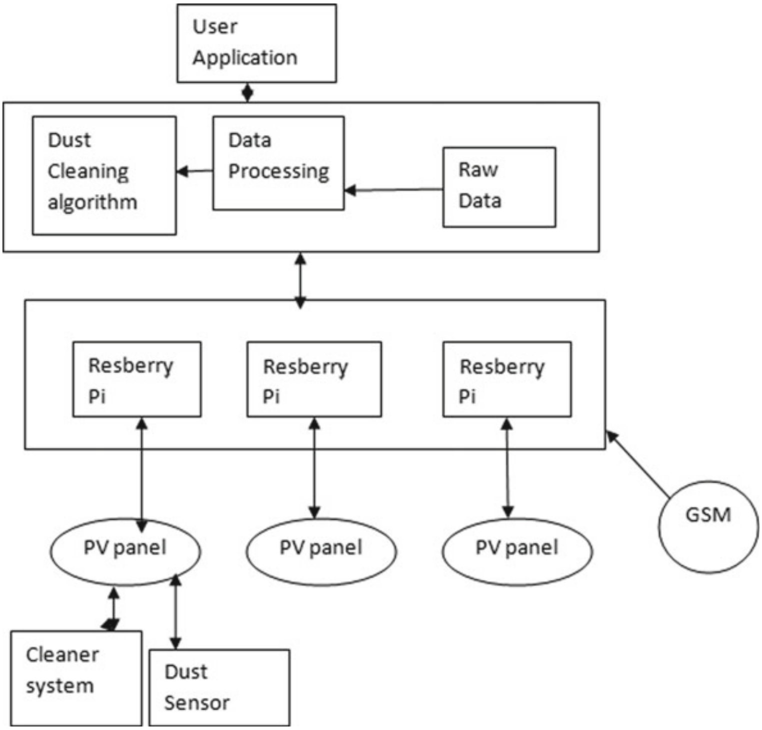


Fig. 14.7 Dust cleaner system based on IOT

value exceeds the threshold value. Then user gives the command to robot to clean the PV panel and continuously also check the dust threshold value if this value becomes less than the threshold than automatically dust cleaner off Fig. 14.7.

An electrical device cleaning system operated by mobile supported IOT is introduced by the researcher. The solar battery cleaning structure receives power from a

Fig. 14.8 A wiper based dust cleaning system supported by IOT

rechargeable battery (12 V), which is activated by a switch using the simple mobile app. For the cleaning purpose a simple wiper is used which is connected by a gear motor and the gear motor is connected with the rack and pinion tool. The pinion is responsible for the entire wiper movement mechanism. The rack guides the pinion, and the rack is directed by the gear motor. During this cleaning process, the water from the cleaning tool pump is sprayed on top of the panel. Simultaneously wiper wipes electrical device and thus the dusty water flows away from the PV panel and finally by an off signal from the mobile app cleaning process will stop in Fig. 14.8.

A researcher offered a prototype which is used to rub the panel face. This prototype consists of a cleaning robot and interface with cloud: A moving cleaning robot is used and which can move back and forth, which has the rotary brush. A cloud interface is used to maintain the human and machine interface. Moreover, to identify the act of a far-positioned solar farm, a detecting unit consisting of a detector was added to the present structure. Furthermore, a month's worth of data from completely clean and dusty panels was analysed using statistical procedures, and the resulting regression model was encoded into the sensing unit. Because it forecasts the best time for cleaning action, the sensing unit combined with the regression model is referred to as an autonomous unit.

According to a system evaluation performed on an example PV module, the designed system can clean dry dust that has accumulated on the panel's surface. Furthermore, by mounting the metal train tracks on a long-term solar battery, the technology appears to be scalable to a large-scale solar farm. Figure 14.9 (Khadka et al. 2020). It is observed that the most effective strategy is automatic cleaning for a huge PV farms those are placed in semi-arid regions, where cleaning frequency is elevated due to regular sand accumulation.

To examine voltage, calculate the current value and power production from the PV panel, a system was designed using a C programme and compiled with Arduino IDE. Proteus 8 Professional was used to create the circuit. The system was capable to detect power loss owing to dust accumulation on the panel face, and as a consequence, based on simulation result the motor drive for the cleaning mechanism responded appropriately. Two DC motors were connected to the Arduino UNO via the control circuit of the L293D motor driver.

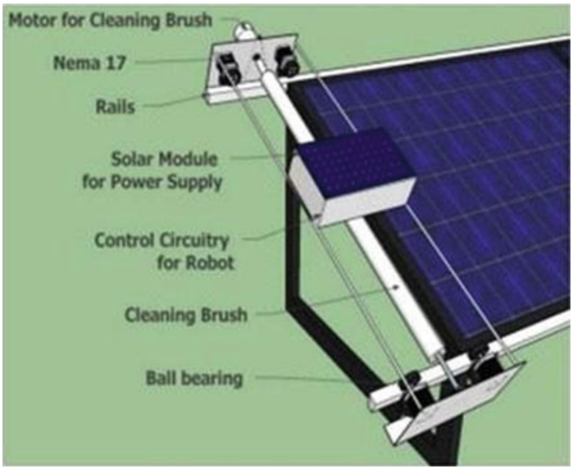


Fig. 14.9 Automated dust cleaning robot by using two DC-motor

The optical sensor and the monitoring circuit were also linked to the Arduino UNO board. The detector was used to sense the existence of optical signal, and if it is a clear day or the output is "HIGH," the microcontroller will respond by measuring the output voltage/current from the panel. Only when the efficiency or output power is low will the cleaning process be activated (20% below average value). The cleaning arms move in rotation until the output power efficiency is more than 20% (Tsamaase 2017).

Swanand S. Wable et al. (2017) researcher proposed the cleaning system for the PV Panels Farms positioned in dusty areas specially in steamy countries. An autonomous cleaning robot is being developed to remove the accumulated dust from the panels at regular intervals. The cleaning procedure depends on a control panel, a DC motor, and micro-fibre (bristles). This research gives you an idea of how the robot will work and how it will affect solar farm energy production. It will also assist in comprehending the issue that arises from the failure to clean solar cells (Wable and Ganiger 2017).

Author Mani et al. discovered in 2010 that solar panel cleaning needs continuous cleaning during dry days single time in a week, and this cleaning is raised to once a day during the significant amount of dust collection. As indicated in Fig. 14.8, therefore a mechanical cleaning automation is utilized. The mechanism is controlled by a particular controller with the help of sensors. In areas where water cleaning is not possible, this procedure can be quite useful. In this scheme, the power expenditure is high as compared to other techniques, and due to the use of mechanical equipment system the problem of wear and tear so needs regular maintenance. Therefore this method does not show full effectiveness as compared to other strategies based on IOTs Fig. 14.10 (Mani and Pillai 2010).

The author Anderson et al. in 2010 introduced a mechanical device to clean PV panel by utilizing the water which is directly injected into the PV surface. The study's

Fig. 14.10 Mechanical cleaning automation

Fig. 14.11 Inclined dust cleaning method

findings revealed a 15% increase in the output of the solar system under consideration Fig. 14.11 (Anderson et al. 2010).

Moreno et al. (2006) offered a less weight mechanical dust removal mop and observed that the efficiency of the solar panel after cleaning is enhanced by 7%. The main cons of this type of system are high cost and require composite mechanical and control structure, as shown in Fig. 14.12 (Moreno et al. 2006).

An Arduino-based solar panel cleaning system is fabricated to clean the dust from solar panels. The projected solar panel cleaning system is waterless, cost-effective, and self-contained. This system uses a two-step mechanism that includes an exhaust fan that also works as an air blower and a wiper that removes dust from the system. It is projected that in dry places, the system's efficiency can be lowered by around 50% and about 15% power losses can occur (Ashtaputre and Bhoi 2019). As a result, it is vital to maintain the solar panel's surface as clear as possible. The proposed solution does not use water and relies on a two-step cleaning process. As a result, there is no waste of water with this technology. The exhaust fan works as an air blower, first clearing the dust from the solar panel's surface. The residual dust on the surface is then wiped away with a wiper. The main parts of this tactics are PV panel,

Fig. 14.12 A Mop based dust cleaning strategy

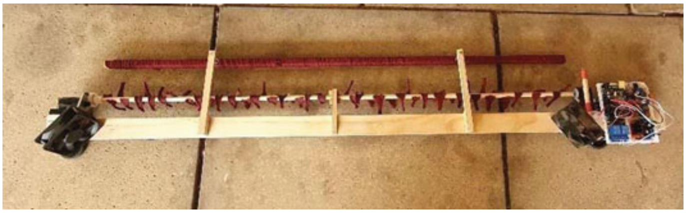


Fig. 14.13 An Arduino based solar panel cleaning system

microcontroller (Arduino Uno), dc gear motor, buck boost converter, and motor drive module Fig. 14.13 (Rawshan Habib et al. 2021).

14.7 Cleaning Approach Based on Machine Learning

PV panels must be cleaned and monitored on a regular basis to maintain efficiency; artificial intelligence is employed to support the cleaning idea. There are numerous methods. To identify dust accumulation and enhance solar panel power efficiency, a researcher used artificial intelligence and computer vision. A drone with a camera and an end effector is used to monitor the panels and perform the cleaning task. Machine Learning algorithms are used to analyze the collected pictures and then determine the category of dust particles, prompting the drone to clean the contaminated particles. Each solar panel's efficiency is continuously calculated based on its output power and supplied into a centralized structure. The drone is controlled by the ground subsystem which is placed on the surface to perform its functionality if the efficiency falls below a specific threshold.

A Human Machine Interface (HMI) screen on a ground subsystem displays abnormalities, battery signs, and allows the controller to control operations if necessary and in an urgent situation. The picture of the inefficient cell is captured by the camera

Fig. 14.14 Dust cleaning robot

mounted to the drone, and analysis is carried out using several vision techniques. These approaches detect the locations where irregularities are located, and calculating tactics are done over these utilizing the drone's concluding effectors (Prasad and Rithika 2020).

A Smart Cleaning will be carried out by robot activities utilizing a dust analysis method, as well as an HMI that will transmit all information to the user about the present status of the machine and allow the user to operate it remotely. The dust analysis algorithm is used to make the cleaning smart. The AI-based Solar Panel Cleaning Robot is capable of determining the power production by gathering the numbers for individual panel current and voltage. The project not only focuses on basic cleaning but also goes above and beyond, resulting in smart cleaning. In addition, this technique is incredibly cost effective. This enables users to make efficient use of renewable energy Fig. 14.14 (Ashtaputre and Bhoi 2019).

Kadam Bhagwat et al. built an automation system that cleans the PV panel in a less amount of time and at a low cost. As a result, the project background is to automate the cleaning of the solar plate. The IR sensor is used to detect dust accumulation on the surface. If the sensor sends a single signal to the microcontroller, this indicates that no dust has gathered or that its density has no effect on the solar panel's performance. When it sends a 0 to the controller, it implies that dust must be removed using the cleaning mechanism. The microcontroller performs the actions that have been programmed into it. It operates the drive mechanism under the supervision of limit sensors and completes one cleaning cycle. After that, check the IR module for dust on the panel. If it is clean, wait for the dust to accumulate as the cycle continues. (Bhagwat et al. 2017).

Piezoelectric based system: The piezoelectric-based cleaning system works on the principle of pressure force which is working in between the wiper and PV panel surface. By using the vibration dust particles moves away from the PV panel and then the wiper wipes all the dust from the surface effectively.

Electro Static Dirt removal Strategy: Electricity is used for the electrostatic dust removal technique. There are two methods for charging the electrons one is to an incident of UV light on a PV panel and the second is the tribo electric charging.

Due to the high voltage at solar cell an electrostatic force will be developed by which the panel will apply attraction force on dirt particles. But dirt particles face repulsion forces in between of each other so these particles will try to move far away from the panel. Rain plays a limited role in this tactic (Syafiq et al. 2018).

Electric curtain system: A behavior discussed by the machine of the appropriate electric field to the dust lying on the PV panel may be proscribed in places with dry dust. The charging of dust phenomenon is based on electrostatic theory in which charged dust particles make electric curtain those have standing waves (Mazumder et al. 2014), which have the amplitude and direction at every location. This approach is distinguished by its quickness in eliminating dust as compared to other strategies and as its low power expenditure and lack of a complicated system because the structure is prohibited by an easy regulator coupled to sensors. The probability of screen degradation owing to ultraviolet illumination is a possible downside of this device (Mazumder et al. 2014).

A very high voltage is required to generate the electric field, and due to this generating efficiency decreases by 15%. According to research, this approach is ineffective in eliminating damp dust particles or those originating from cement, and its effectiveness is restricted to micro and microscopic particles. In dry locations, this technology can extend the PV's lifetime and lessen the damage caused by ultraviolet irradiation by replacing screens with polymer or weather-resistant glass Fig. 14.15 (Mazumder et al. 2014; Al-Waeli et al. 2017c).

Self-cleaning mechanism: In this method to avoid the collection of dust on the solar cell a transparent nano film is placed all over the pv surface. This nano film made of Super wet ability or super-hydrophobicn. To remove the dust from the solar panel super hydrophilicity technique is used which spread the water on the whole surface of the PV panel (Park et al. 2011). For the nano film coating mostly used material is TiO_2 that clears the PV panel in the following ways:

Photocatalytic process is the first step in which UV light incident on the panel, and this UV signal reacts with TiO_2, in the response of this reaction dirt particles splited.

In the second step with the help of rain water all the dirt particles are cleaned off, although the main cons of this method are the requirement of rain fall which is not human dependent and it is hard to get rain in dry seasons and sandy regions. The second issue is that solar panel contact angle should be more than 150°, therefore

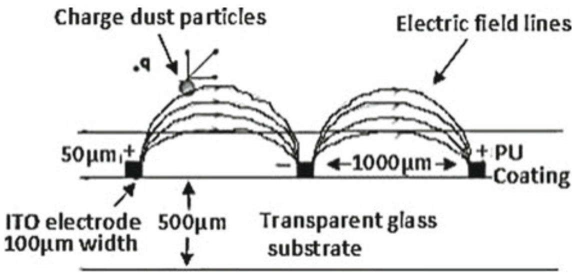


Fig. 14.15 ECD based dust cleaning approach

when rain drops incident on the panel than from the surface dirt would be drained off quickly with water. It is observed that with coating film PV Panel have daily wastage of energy due to the soiling effect is around 2.5% and if without coating this loses increases up to 3.3%. In the observation period of one year it is also observed that coated PV module does not face any deprivation level in the coating of film and PV module.

14.8 Summary

In this paper author reviewed, discussed an assortment of cleaning methodology for solar panel which enhances the efficiency of solar panel. The decision of the right way of cleaning depends on many parameters like the size of the solar panel, location, availability of water and designing methodology. Moreover, many of researchers recommended that techniques of cleaning depend on frequency. Although there is no standard related to the position of panel and strategies to be used it totally depends on the location, climate condition and availability of recourses so according to the availability of resource optimum solution should be used among all the existing strategies.

References

Adinoyi MJ, Said SA (2013) Effect of dust accumulation on the power outputs of solar photovoltaic modules. Renew Energy 60:633–636

Al Shehri A, Parrott B, Carrasco P, Al Saiari H, Taie I (2017) Accelerated test bed for studying the wear, optical and electrical characteristics of dry cleaned PV solar panels. Sol Energy 146:8–19. https://doi.org/10.1016/j.solener.2017.02.014

Al Shehri A, Parrott B, Carrasco P, Al Saiari H, Taie I (2016) Impact of dust deposition and brush-based dry cleaning on glass transmittance for PV modules applications. Sol Energy 135:317–324. https://doi.org/10.1016/j.solener.2016.06.005

Al-Waeli AHA, Sopian K, Chaichan MT, Kazem HA, Ibrahim A, Mat S, Ruslan MH (2017c) Evaluation of the nanofluid and nano-PCM based photovoltaic thermal (PVT) system: an experimental study. Energy Convers Manag 151. https://doi.org/10.1016/j.enconman.2017.09.032

Anderson M, Grandy A, Hastie J, Sweezey A, Ranky R, Mavroidis C, Markopoulos YP (2010) Robotic device for cleaning photovoltaic panel arrays. Mob Robot Solut Challenges 367–377. https://doi.org/10.1142/9789814291279_0047

Ashtaputre G, Bhoi A (2019) Artificial intelligence based solar panel cleaning robot. IOSR J Electron Commun Eng (IOSR-JECE) 14(3):12–15. Ser II. e-ISSN: 2278-2834, p-ISSN: 2278-8735

Bhagwat K, Gajanan M, Nilesh R, Satish S, Sagar Y, Nemade KV (2017) Artificial intelligence solar panel cleaning mechanism. IJARIIE 3(3). ISSN(O) 2395-4396

Chen J, Pan G, Ouyang J, Ma J, Fu L, Zhang L (2020) Study on impacts of dust accumulation and rainfall on PV power reduction in East China. Energy 194. https://doi.org/10.1016/j.energy.2020.116915

Gaier J, Davis P, Marabito M (1990) Aeolian removal of dust types from photovoltaic surfaces on mars. In: 16th AIAA/NASA/ASTM/IES space simulation conference. Albuquerque, NM

He G, Zhou C, Li Z (2011) Review of self-cleaning method for solar cell array. Procedia Eng 16:640–645

Jadhav D (2019) Dust cleaner system for PV panel using IoT. Int J Comput Appl 178(10). 0975-8887

Khadka N, Bista A, Adhikari B, Shrestha A, Bista D (2020) Smart solar photovoltaic panel cleaning system. In: International conference on sustainable energy and green technology. 2019 IOP conference series: earth and environmental science, vol 463. IOP Publishing, p 012121. https://doi.org/10.1088/1755-1315/463/1/012121

Mani M, Pillai R (2010) Impact of dust on solar photovoltaic (PV) performance: research status, challenges and recommendations. Renew Sustain Energy Rev 14:3124–3131. https://doi.org/10.1016/j.rser.2010.07.065

Mazumder M, Horenstein M, Stark J, Hudelson JN, Sayyah A, Heiling C, Yellowhair J (2014) Electrodynamic removal of dust from solar mirrors and its applications in concentrated solar power (CSP) plants. In: 2014 IEEE industry applications society annual meeting IAS, pp 4 1–7. https://doi.org/10.1109/IAS.2014.6978347

Moreno L, Cabás R, Fernández D (2006) Low mass dust wiper technology for MSL rover. 1–7

Park YB, Im H, Im M, Choi YK (2011) Self-cleaning effect of highly water-repellent microshell structures for solar cell applications. J Mater Chem 21:633–636. https://doi.org/10.1039/c0jm02463e

Prasad A, Rithika, Rathika PD (2020) Illustration of automatic panel cleaning system. Int J Eng Res Technol (IJERT) 9(08). http://www.ijert.org. ISSN: 2278-0181

Rawshan Habib Md, Shahnewaz Tanvir Md, Suhan AY, Vadher A, Alam S, Shawmee TT, Ahmed K, Alrashed A (2021) P5 IEEE 2021 automatic solar panel cleaning system based on Arduino for dust removal

Renewables 2020 global status report. www.ren21.net

Sai Lalitha Chowdary K, Prameela G, Jayaharsha K, Ravi Kishore D (2017) Influence of dust removal mechanism on solar power plant and its performance. Int J Mod Trends Sci Technol 03(04). ISSN 2455-3778. http://www.ijmtst.com

Solar panel cleaning—automatic cleaning system: Heliotex. solarpanelcleaningsystems.com

Syafiq A, Pandey AK, Adzman NN, Rahim NA (2018) Advances in approaches and methods for self-cleaning of solar photovoltaic panels. Sol Energy 162:597–619

Tsamaase K (2017) Automated dust detection and cleaning system of PV module. IOSR J Electr Electron Eng (IOSR-JEEE) 12(6):93–98

Vaghani M, Magtarpara J, Vahani K, Maniya J, Gurjwar RK (2019) Automated solar panel cleaning system using IOT. Int Res J Eng Technol (IRJET) 06(04). p-ISSN: 2395-0072. e-ISSN: 2395-0056. www.irjet.net

Wable SS, Ganiger S (2017) Design & manufacturing of solar panels cleaning system. Int J Res Appl Sci Eng Technol (IJRASET) 5(VII)

Zorrilla-Casanova J, Piliougine M, Carretero J, Bernaola P, Carpena P, Mora-Lopez L, Sidrach-de-Cardona M (2011) Analysis of dust losses in photovoltaic modules. 2985–2992. https://doi.org/10.3384/ecp110572985.

Chapter 15
Prediction of Heart Disease Using Hybrid Machine Learning Technique

Nagaraj M. Lutimath, Chandra Mouli, B. K. Byre Gowda, and K. Sunitha

Abstract Researchers have paid close attention to the field of medicine. Several factors have been blamed for human early mortality by a sizable number of researchers. The relevant research has established that diseases are brought on by a variety of factors, one of which is heart-related illnesses. Numerous scholars suggested unconventional ways to prolong human life and aid medical professionals in the diagnosis, treatment, and management of the cardiac disease. Some practical techniques help the expert make a conclusion, yet every effective plan has limitations of its own. In data mining, support vector machines (SVMs) are an important classification technique. It is a method of supervised classification. It locates a hyperplane to classify the intended classes. A variety of heart-related illnesses make up heart disease. Vascular problems such as arrhythmia, weak myocardium, congenital heart disease, cardiovascular disease, and coronary artery disease are included in this category. A common form of heart disease is coronary artery disease. It causes a heart attack by decreasing the blood supply to the heart. Support vector machines are used in this study to assess the data set from the UCI machine learning repository made up of heart disease patients. Patients with cardiac disease are accurately classified, as expected. Python is used as the programming language for implementation.

Keywords Prediction · Hybrid · Heart disease · SVM · MAE · MSE · RMSE

N. M. Lutimath (✉)
Department of Computer Science and Engineering, Dayananda Sagar Academy of Technology and Management, Opp. Art of Living, Udayapura Road, Kanakapura, Bengaluru, Karnataka, India
e-mail: nagarajlutimath@gmail.com

C. Mouli
Department of Computer Science and Engineering, East Point College of Engineering and Technology, Jnana Prabha, Virgo Post, Bengaluru, Karnataka, India

B. K. B. Gowda
Department of Information Science and Engineering, Sir M. Visvesvaraya Institute of Technology, International Airport Road, Hunasamaranahalli, Yelahanka, Bengaluru, Karnataka, India

K. Sunitha
Department of Electronics and Communication, Government Women's Polytechnic, Holenarasipura, Karnataka, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
A. Rai et al. (eds.), *Paradigms of Smart and Intelligent Communication, 5G and Beyond*, Transactions on Computer Systems and Networks,
https://doi.org/10.1007/978-981-99-0109-8_15

15.1 Introduction

Heart disease has emerged as a serious health concern for many individuals due to its high mortality rate throughout the world. Detecting cardiovascular disorders including heart attacks, coronary artery diseases, etc. by routine clinical data analysis is a critical task; early detection of heart disease may save many lives. Making informed decisions and precise predictions is made possible by machine learning (ML). The application of machine learning techniques in the medical sector has advanced significantly. By more effectively identifying diseases at an earlier stage, ML can assist in lowering the rate of readmissions to hospitals and clinics. The ability to find and create novel treatments that have a great chance of aiding patients with complex illnesses has also advanced thanks to technology. Machine learning is used in various fields around the world. Healthcare is no exception. Machine learning is very important in determining if movement disorders, heart disease, and other medical conditions are present. With enough information up front, physicians can gain valuable insight and customize diagnosis and treatment strategies for each patient.

People can be impacted by diseases both physically and mentally since acquiring and coping with sickness can change a person's outlook on life. A condition that damages an organism's components but is unrelated to any recent exterior injuries. It is common knowledge that diseases are medical illnesses linked to particular symptoms and indicators. Arteria coronaria disease (blood flow obstruction), cerebrovascular illness, and lower respiratory infections are the three most lethal conditions that affect people. Heart diseases are the most unforeseen and unpredictable. Using machine learning techniques, we can anticipate cardiac disease. Since cardiac illness has a complex character, it requires cautious management. Failure to do so could harm the heart or result in premature death. To identify different types of metabolic syndromes, data mining and the perspective of medical research are employed. Heart disease prediction and data analysis both greatly benefit from data mining with classification.

The main cause of death in the world is heart disease. Heart disease is often recognized by symptoms such as shortness of breath, physical limitations, and swollen feet (HD). Sometimes there are no clinical experts available to treat the coronary condition, and exams take a long period. A specialist typically diagnoses HD after looking at the patient's clinical history and compiling a report on their physical examination. However, the results are frequently wrong.

The medical profession is still far from being able to treat patients suffering from various diseases. The abnormality of the heart, which cannot be seen with the naked eye and manifests itself right away when it reaches its limits, is one of the most deadly. No hospital can afford to have a patient die as a result of poor medical decisions. A suitable and affordable computer-based therapy and support system can be created to help people make wise decisions. The major goal of this study was to develop a model that could analyse historical data from the database of heart illness and deduce unknown information (patterns and correlations) connected to heart disease. It can answer difficult heart disease detection questions and assist medical professionals in making wise clinical judgments that differ from the norm.

It is challenging to pinpoint cardiac disease due to the multiple risk factors that contribute to it, including diabetes, high blood pressure, excessive cholesterol, an irregular pulse rate, and many others. Finding out the severity of heart disease in humans has been tasked with a number of data mining and neural network methods. Several methods, including the K-Nearest Neighbor Algorithm, Decision Trees, Logistic Regression, and Support Vector Machine, are used to categorise the severity of the illness. Because the health of heart disorders is complex, the disease must be handled with care. Failure to do so could harm the heart or result in premature death. Medical science and data mining perspectives are applied to identify various types of metabolic illnesses.

Non-communicable diseases (NCDs), commonly known as chronic diseases, are diseases that cannot be transmitted from person to person. They often survive long and move slowly. They are the product of a confluence of behavioral, environmental, physiological, and genetic factors. Diabetes, malignancies, chronic respiratory disease, and cardiovascular disease are the four main categories of non-communicable diseases. Noncommunicable diseases (NCDs) are currently a public health concern. In fact, they are responsible for over 70% of deaths on Earth. According to the World Health Organization (WHO)'s first report on the status of noncommunicable diseases in 2010, of the 57 million deaths worldwide in 2008, 36 million (nearly 60%) were mainly caused by non-communicable diseases. Cardiovascular disease, cancer, diabetes, chronic pneumonia.

From 36 to 41 million people died in 2016, accounting for 70% of the global mortality rate. The leading cause of NCD mortality (17.9 million per year) is cardiovascular disease, which is then followed by cancer (9 million), respiratory conditions (3.9 million), and diabetes (1.6 million). The primary risk factors affecting non-communicable illnesses are a sedentary lifestyle, an imbalanced diet, cigarette use, and excessive alcohol intake. The 2020 Sustainable Development Program faces a significant challenge from non-communicable diseases. Reducing the risk factors for these diseases is a crucial control strategy. Preventive medicine holds that risk factors can help make an early diagnosis of patients or offer health advice.

This includes the static analysis and machine learning techniques such as K-Nearest Neighbors, Decision Trees, Logistic Regression, Neural Networks, and Support Vector Machines. These predictive analytics techniques are used to detect fraud, reduce risk, improve operations, simplify marketing, and more. Data mining (decision trees, neural networks, regression, classification, clustering, etc.), machine learning (supervised and unsupervised learning, etc.), and deep learning are used to predict NCDs, especially cardiovascular diseases. Various predictions are made.. Techniques (such as Autoencoder and Softmax) have been created.

The WHO estimates that 12 million people worldwide die prematurely from heart disease each year. Cardiovascular disease is responsible for over 50% of her deaths in the United States and elsewhere. In many countries it is one of the leading causes of death. It is considered the leading cause of death in adults. Heart disease can be classified as either coronary artery disease or cardiovascular disease. The term "cardiovascular disease" refers to many conditions that affect the heart, blood vessels, and the body's circulatory and pumping systems. A variety of illnesses, disabilities, and deaths are caused by cardiovascular disease. One of the most important and demanding tasks in medicine is the diagnosis of diseases.

Cardiovascular disease is one of the most common causes of death worldwide and in developed countries. Hypertension, obesity, stress, diabetes, alcohol, high cholesterol, and smoking are other risk factors for cardiovascular disease that can be prevented and managed through healthy behavioral adaptations. However, other risk variables such as age, gender, and family history may not be controlled. Early detection of cardiovascular disease can reduce mortality because lack of awareness prevents people from knowing the cause of cardiovascular disease early. is taking place. The disease is usually discovered only in the late stages or after death.

In order to properly diagnose cardiac disease, machine learning is essential. Machine learning techniques include decision trees, neural networks, Naive Bayes classification, genetic algorithms, regression, and support vector machines, to name a few. The decision tree method is used to identify patterns that can be utilized to forecast cardiac disease. With the aid of the Cleveland data set, the C5.0 decision tree technique was completed. When compared to the other algorithms, it has an accuracy value of 85.33% (Ngom et al. 2020) (Maru et al. 2021). It was discovered to be superior to other data mining techniques. Data on the patient was entered via a graphical user interface, and a Weighted Association rule-based Classifier was used to determine whether or not the patient had heart disease. Results indicated that, in comparison to other Associative Classifiers presently in use, the Weighted Associative Classifier offered better accuracy. A classifier that uses probability is called Nave Bayes (Atallah and Al-Mousa 2019). To predict heart disease, medical factors like blood pressure, age, and sex were used. The implementation was done using MATLAB. By lowering the size of the tree, a prediction model that combines pre- and post-pruning of decision tree learning increased classification accuracy (Motarwar et al. 2020).

By utilising the most recent technologies, the health sector can be upgraded, extending the average population's life expectancy. Leading causes of death worldwide include cancer and heart disease. Cardiovascular disease-related deaths are rising alarmingly quickly each year. According to a World Health Organization report for 2016, heart attacks and strokes accounted for 85% of all cardiovascular disease-related deaths globally, accounting for 31% of all fatalities. In both industrialised and developing nations, the rising use of alcohol and tobacco directly raises the risk of heart disease. In industrialised nations like the United States, England, Canada, and New Zealand, the prevalence of obesity is rising, which increases the chance of developing heart-related issues.

When you consider how cardiovascular diseases affect the world's population, a machine-learning model for early identification is quite helpful. To address this growing, enormous problem, ongoing efforts are conducted employing a variety of technological advances. In order to address the ever-expanding health issues, various bioengineering approaches have been created in recent years. The improvement reduction rate is benefiting from the ongoing study in the field.

Regular Neural Networks (NNs) and Convolutional Neural Networks (CNNs) both have a wide range of applications and differ in terms of architecture. The heart disease diagnosis issues in this article were solved using both of the two machine learning models. We put the algorithms into practise, adjusted the settings, and ran a number of tests. We compare the two models' predictive abilities with various parameter values. For the diagnosis of cardiac illness, we used the Cleveland database, which was taken from the UCI learning dataset repository. According to the experimental findings, NNs typically outperform CNNs in terms of prediction accuracy.

Prediction analysis can also make use of additional data mining techniques like regression, neural networks, support vector machines, and genetic algorithms. Support vector machines with linear and sigmoid kernel functions are contrasted in this article. The UCI machine learning data set repository made the dataset accessible. The structure of this study is as follows:

The relevant research is discussed in Sect. 15.2, followed by a discussion of methodology in Sect. 15.3, examples of feature engineering in Sect. 15.4, examples of prediction analysis in Sect. 15.5 and a conclusion in Sect. 15.6.

15.2 Related Work

In order to anticipate the people who will have this disease, experts have focused their attention on heart disease. Usually, hidden patterns in the data set are extracted using a knowledge discovery technique. Data mining is an information extraction process used in knowledge discovery. It is crucial in the process of diagnosing the illness. Data mining techniques are used to categorize the data collection, including decision trees, neural networks, Naive Bayes classification, support vector machines, and genetic algorithms (Kohli and Arora 2018). Using appropriate medical data, decision tree C4.5 and Fast Decision trees were investigated (Basha et al. 2019). The UCI repository was utilized for medical data sets. Decision trees were accurate to a degree of 69.5%, and fast decision trees were accurate to a degree of 78.54%.

Analysis and forecasts of coronary artery heart disease were carried out using a data collection of 335 records representing the various 26 variables (Lin et al. 2020). The data set was pre-processed using the correlation concept. The characteristics were located and extracted using Particle Swarm Optimization (PSO). Fuzzy, decision tree, fuzzy regression, and neural network models were among the models. The data set was used with the neural network model. The accuracy percentage was found to be 77%. It was also used for regression modeling. The accuracy as a result was 83.5%. There were no significant changes in the other fuzzy and decision tree models.

The data set was then optimized using the pre-processing technique. We used K-means clustering, feature extraction and selection with PSO, and correlation. One of the methods, or a combination of them, was used to categorize the data set. The accuracy of the regression model's results was 88.4%. The data set was also subjected to a hybrid model. The accuracy of classification techniques increased from 8.3 to 11.4%.

With the use of the Gini Index and support vector machines, a second study on the prediction of heart disease was completed (Gavhane et al. 2018). The classification of the data was then finished using the proper classification techniques. Sequential Minimal Optimization (SMO) and Naive Bayes probability classification were the techniques employed for classification. Artificial neural network models and SMO with bagging were also incorporated for analysis. SMO with bagging had an accuracy of 93.4%. Naive Bayes probability classification has a 75.51% accuracy rate. SMO accuracy was 94.08, whereas neural network models had an accuracy of 88.11. The 10-cross-fold validation procedure was used to complete the results' verification.

Using an appropriate medical data set, a Transaction Reduction Method (TRM) Apriori approach was used to diagnose heart illness (https://www.alivecor.com/how-it-works; https://doi.org/10.1007/s10462-01). The outcomes were contrasted with some of the traditional methods. The algorithm produced an accuracy of 93.75%. 92.09% accuracy was attained when SMO was used. 89.11% accuracy was attained when SVM was applied. The accuracy of the C4.5 decision tree was 83.85%, and the accuracy of the Naive Bayes probability classification was 80.15%.

There is no shortage of documentation about the symptoms experienced by patients having heart attacks. However, they are not being used to their full potential to help us predict comparable possibilities in otherwise healthy adults. For illustration: According to the Indian Heart Association, 25% of all heart strokes in Indians happen before the age of 40, and 50% of heart attacks happen before the age of 50. The risk of heart attacks is three times higher in urban areas than in rural ones (Gavhane et al. 2018).

To maintain heart health, many experts advise a balanced diet and moderate exercise. The following are the factors that were taken into account when creating the system for the study and have a high-risk percentage for heart disease:

Age, sex, blood pressure, heart rate, diabetes, and high cholesterol are the first six factors.

AliveKor It is available as a bracelet or a touchpad that connects to your cell phone over a wireless network. Through Bluetooth, the touchpad simulates the patient's ECG on his mobile device. As a result, all the relevant parameters, including blood pressure and heart rate, are readily available. On the wristband, however, the pulse function is shown on the dial through finger contact. Additionally, it may signal atrial fibrillation (https://www.alivecor.com/how-it-works). MyHeart, B. This system uses a variety of on-body sensors to wirelessly transmit physiological data to a PDA. The data is processed, and the analysis is used to provide the user with health suggestions (https://doi.org/10.1007/s10462-01). C. HealthGear HealthGear is a programme for keeping track of the most popular indicators, including physical and lab measurements (Boudi 2016). Fields consist of: Blood pressure, haemoglobin, WBC, RBC, and platelets are among the physical indicators, along with height, weight, and BMI—[Lipids]: Triglycerides, HDL, LDL, and VLDL—[Sugar] Fasting Glucose, HbA1C, and After Meals (Boudi 2016). Fitbit, D This sensor is used to monitor one's health and has functions for detecting heart rate, blood pressure, and calories burned. After conducting this analysis, we came to the conclusion that Fitbit would be the most convenient and cost-effective way to gather data, while HealthGear would be used for all other metrics.

All of the methods discussed above deal with predictive analysis utilizing conventional techniques. When employing appropriate medical data sets, classification methods such as decision trees, Naive Bayes, support vector machines, or neural networks should be taken into account.

Purusothaman et al. (2015) introduced the common data mining classification approaches such as ANN, fuzzy logic, Neural Networks, Decision trees, data mining genetic Algorithm, and Nearest Neighbor method. In the paper, hybrid applied data mining techniques were suggested. The importance of big data analytics was emphasised in the work by Cheryl Ann Alexander et al. for diagnosing, treating, and predicting chronic diseases (Sharma and Rizvi 2017). The study put forth the concept of IoT and cloud computing technologies in the field of medicine.

Johnson-Coyle et al. (2012) developed an efficient data analysis model for the prediction of severe heart syndromes by utilising a variety of classification techniques. It is common for datasets to have noise features, which can suddenly damage good data. As a result, they aimed to reduce the noise by cleaning and pre-processing the dataset as well as by reducing its dimensionality. They discovered that neural networks may produce results with good accuracy.

Peripartum cardiomyopathy (PPCM) is a form of enlarged cardiomyopathy with an unknown cause, according to Leah Johnson-Coyle et al. Although the rate is low—less than 0.1% of pregnancies—disappointment and death rates are significant, ranging from 5 to 32%, and they occur in already healthy women in the final month of pregnancy and up to 5 months following delivery. While for some women, PPCM progresses to cardiovascular disappointment and even abrupt cardiac passing, for others, the clinical and echocardiography status improves and may return to normal. Clearly linked to the restoration of cardiac function is the guess of PPCM. Failure of the heart to return to its normal size is linked to increased mortality and gloom. As described in Sebastian et al. (2012), dilated cardiomyopathy is characterised by left ventricular enlargement that is associated with systolic brokenness. Right ventricular function can become impaired and diastolic dysfunction can occur. Affected individuals have the risk of both left and right ventricular failure. A fundamentally hereditary or explosive foundation underlies a sizable portion of DCM cases. Although distinct aspects of heart remodelling educate forecast and provide restorative advice, estimation of LV size and launch percentage remains essential to conclusion, risk classification, and treatment. Evaluation of myocardial fibrosis forecasts both the risk of sudden cardiovascular death and the likelihood of LV functional recovery, and may influence the patient's decision about the insertion of a cardioverter-defibrillator. Finding preclinical DCM could greatly save suffering and fatalities. Preclinical DCM detection could significantly save suffering and fatalities by enabling early evaluation of cardio-protective analyses.

Heart disease prediction as a method of medical diagnosis has been the subject of numerous investigations. First, a study using a neural network to analyse data from a self-applied questionnaire (SAQ) in order to create a system for predicting cardiac disease has been proposed by R. W. Jones, M. Clarke, Z. Shen, and T. Alberti. The study highlights not just typical illness risk variables but also additional SAQ data. By comparing the results of the neural network with the "Dundee Rank Factor Score," which statistically correlates three risk factors (blood pressure, smoking, and blood cholesterol) with sex and age to assess the risk of developing heart disease, the work's validity was confirmed. A multi-layered feedforward neural network that was trained using the backpropagation algorithm was employed in the study. The neural network they employed included three layers: input, hidden, and output. By expanding the neural network's input quantities, the performance was enhanced to a Relative Operating Characteristic (ROC) area of 98%. The best categorization strategy for the intended system, according to Ankita Dewan and Meghna Sharma, who explored numerous techniques for constructing a heart disease prediction system. Additionally, they suggested employing genetic algorithms to overcome the limitation of local minima in backpropagation algorithms. The suggested methodology was designed to be used in the future with accuracy close to 100% or with few mistakes.

S. Y. Huang, A. H. Chen, C. H. Cheng, P. S. Hong, and E. J. Lin have planned and carried out a further investigation on the prediction of heart disease. The learning Vector Quantization Algorithm, one of the Artificial Neural Network learning techniques, was used to train the classification and prediction. Their approach consisted of three steps. The first step was to choose three of the 13 clinical features—age, cholesterol, kind of chest pain, exercise-induced angina, maximum heart rate, fasting blood sugar, number of vessels coloured, old peak, resting ECG, sex, slope, thal, and trestbps—that are more significant than the others. The second used a classification system based on artificial neural networks. Finally, a technique for predicting heart disease was created. The study's results showed a prediction accuracy rate of close to 80% (Motarwar et al. 2020). D. R. Patil and Jayshril S. Sonawane have developed a new Artificial Neural Network technique for heart disease prediction. The Vector Quantization Algorithm uses random order incremental training to train the employed network. The network in use has three levels: input, hidden, and output layers. In the input layer, there were 13 neurons, which is equivalent to the number of clinical data in a database of cardiac diseases. To achieve fewer mistakes and greater accuracy, the neurons of the buried layer could be modified. In the layer of output that indicates whether or not there is cardiac disease, there was only one neuron. Training with a variety of neurons and training epochs increased the system's performance. The outcome demonstrates that, when compared to other researchers, they had the best accuracy (85.55%), as claimed in the report (Kohli and Arora 2018). Another work by Majid Ghonji Feshki and Omid Sojoodi Shijani uses a feature selection and classification approach using a particular dataset to predict cardiac disease. There were three steps in the suggested method. The procedure of splitting the dataset into two subsets as sick and healthy people was the initial step. The complete characteristics were divided into 8192 subsets in the second step. The optimal subset with the highest accuracy was identified in the third stage by combining the PSO algorithm with the Feed Forward Backpropagation Algorithm, a classifier algorithm. Four classifier methods were employed in the methodology: C4.5, Multilayer Perceptron, Sequential Minimal Optimization, and Feed Forward Backpropagation. The most effective approach was identified as a neural network with the PSO algorithm using feature selection and backpropagation. The study's results showed a 94.94% accuracy rate. The goal of a study by R. R. Manza, Shaikh Abdul Hannan, R. J. Ramteke, and A. V. Mane is to predict the diagnosis of cardiac disease using an artificial neural network as a classifier. There were 5 steps in the suggested process. Step 1 involved gathering information regarding prescription medications and heart disease patients. Step 2 involved converting heart disease symptoms and medications into binary form (0 or 1), where 1 denotes the presence of a symptom or medication. In phase 3, the Radial Basis Function was trained. Step 4's performance evaluation of the classifier made use of testing data. In step 5, the Radial Basis Function administered the patients' drugs. The used network has three layers: input, hidden, and output. Information processing is not the duty of the input layer. Distribution of the input vectors to the hidden layer is the only duty of the input layer. There were several Radial Basis Function units in the hidden layer. 97% accuracy was found in the study. It was mentioned that the presented method might be expanded using the generalised regression neural network.

Syed Umar Amin, Dr. Rizwan Beg, and Kavita Agarwal have presented a hybrid approach using genetic algorithms and artificial neural networks to forecast cardiac illness based on risk factors. The backpropagation algorithm was employed to train the neural network. The backpropagation algorithm's two primary flaws have been identified. The first issue is that finding globally optimal beginning weights is essentially impossible. The backpropagation algorithm's slow convergence rate is the second issue. This issue was resolved by utilising a genetic algorithm to enhance the performance of an artificial neural network by optimising its connection weights. 12 input nodes, 10 hidden nodes, and 2 output nodes made up the neural network that was utilised in this investigation. The results show that the training accuracy is 96.2% and the acquired validation accuracy is 89% (Gavhane et al. 2018). A study by Jayshril S. Sonawane and D. R. Patil aims to use artificial neural networks to forecast heart illness. A multilayer perceptron neural network is used by the system. The suggested system consisted of two steps. 13 clinical data were received as input during the first phase, and the backpropagation algorithm was then used to train the network. The input, hidden, and output layers made up the network. There were 13 neurons in the input layer, matching the number of clinical data from a database of cardiac illnesses. To achieve low error and high accuracy, the hidden layer's neurons can be altered. There was only one neuron in the output layer that indicated the presence or absence of cardiac disease. 98% is the accuracy rate that the study found. Another study by Usman Qamar, Saba Bashir, and M. Younus Javed aims to predict cardiac disease. The suggested approach employs a hybrid model that combines Decision Tree, Support Vector Machine, and Naive Bayes. These three classifiers succeeded in obtaining the majority voting scheme. The suggested strategy included two steps. The first one produced the results of every three classifiers. The second included adding the choices together to create a new model using a majority vote system. The findings demonstrate that the study's accuracy rate is significantly higher than that of the competition. The study's findings for predicting heart disease included 74% sensitivity, 82% accuracy, and 93% specificity (Karayõlan and Kõlõç 2017).

To enable effective heart attack prediction, Patil and Kumaraswamy (2009) suggested removing notable patterns from the dataset. The clustering method K-Means was used. The weight of each item was calculated using the MAFIA algorithm. Based on the calculated weights, patterns with values above the threshold were considered for prediction. Using a data set of 15 features and data mining techniques, including ANNs, time series, clustering rules, and association rules, Soni (2011) report on heart disease prediction. The paper recommended the adoption of genetic algorithms to improve accuracy while reducing the amount of data.

To create models for heart disease prediction, Ananey-Obiri and Sarku (2020) used ML approaches such as linear regression, decision trees (DT), and Gaussian Naive Bayes (GNB). The models were created using a k-fold cross-validation method, and their effectiveness was evaluated using receiver operator characteristic (ROC) curves. The analysis for this study employed the UCI dataset of heart disease patients. DT classifier model accuracy was 79.31%, GNB accuracy was 76%, and LR accuracy was 82.75%.

15.3 Methodology and Data Set Analysis

Data mining is the process of extracting knowledge from secret data sets. Multiple sources are used to gather multi-dimensional data, which is then pre-processed and formatted appropriately. Then, these data are subjected to data mining techniques for additional classification.

The goal of machine learning is to develop effective software applications that can automatically access and utilise data. Machine learning is a type of system learning process that gives a system the ability to operate automatically or by itself with the help of correct training, improving system performance and experience effectively and without the need for human intervention. Machine learning will enable training on data sets using efficient learning techniques. The rules and obligations that result from these algorithms will be based on conclusions drawn from the data. The system may build many system models while being trained using various datasets and the same learning technique.

A. The variance in the neighbours within a class determines the class of a certain data point using the K nearest neighbours (KNN) classifier. Test scores with neighbours ranging from 1 to 20 are utilised to calculate test scores.
B. Based on the class values that will be given to each data point, the decision tree classifier creates a tree. One to thirty-point increments is used to count the features.
C. Each tree in the random forest splits out into a class prediction, and the model is made up of a collection of distinct decision trees. The bagging method is used in the Random Forest approach to add more randomness and diversity to the feature space. In other words, it randomly samples elements of the predictor space rather than looking greedily for the best predictors to generate branches.

This increases the variety and lowers the variance of the trees at the expense of an equal or larger bias. This procedure, also known as "feature bagging," is what results in a more reliable model.

D. Naive Bayes a statistical classifier makes no assumptions about the relationship between attributes. A supervised algorithm is the Naive Bayes classifier. It is a simple classification method that uses Bayes theorem. Strong (Naive) independence is assumed. among qualities. A Bayes theorem formula to calculate the percentage The Predictors are not connected to one another neither relate to another nor at least one another. All the attributes alone participate to the in order to maximise it. It identifies conditional independence, which is independent of the values when compared to the values of other characteristics but assumes an attribute value on a specific class.

E. An input layer, a number of hidden layers, and an output layer make up a regular neural network (NN). An input array is transformed by being placed in the input layer, passing through several hidden layers, and then receiving the prediction result from the output layer. Each layer of a NN is composed of a collection of neurons, with every layer being fully connected to every neuron in the layer before it. Consider a neuron with n inputs. Each input Xi is multiplied by the corresponding weight Wi, the sum is computed, the activation function is run, and the output is the result.

The human brain, which has remarkable processing power due to its network of interconnected neurons, serves as the model for artificial neural networks (ANN). ANNs are created utilising the perceptron, a type of fundamental processing unit. The single-layer perceptron algorithm handles issues that may be divided into linear segments. Multilayer Perceptron Neural Network can be used to solve issues that cannot be resolved linearly (MLP). Input, hidden, and output layers are among the many layers present in MLP. The multilayer perceptron neural network was used in the creation of the suggested heart disease prediction system. Three layers make up the planned ANN: the input layer, the hidden layer, and the output layer.

13 neurons were intended to be in the input layer. It was decided that the number of neurons would match the number of attributes in the data set. • Three neurons were intended to be housed in Hidden Layer. This sum was chosen as the starting point. By comparing their performances and then choosing the best one, the number was modified by going up one at a time until it reached the number of input layer neurons. This strategy is based on one of the principles of machine learning, according to which the number of neurons in the hidden layer should be equal to the average of those in the input and output layers. • Two neurons were included in the Output Layer's architecture. The created NN is a classifier that is now operating in machine mode.

F. The architecture of convolutional neural networks (CNNs) differs from that of traditional neural networks (NNs) in a number of ways. In each layer, a CNN first arranges its neurons in three dimensions (width, height, depth). Second, not every neuron in a layer is linked to every other neuron in the layer above.

The classifier and the feature extractor are two more of the CNNs' component parts. To find the features in the feature extractor section, the network will do a series of convolutional (by convolution layers) and pooling (by pooling layers) operations. On the extracted features in the classifier section, the fully connected layers (FCNet) function as a classifier. It also assigns a probability to the input array in order to illustrate what it predicts.

G. Support vector machines (SVMs) are supervised learning techniques that, unlike the C4.5 algorithm, do not rely on decision trees to complete tasks. The likelihood of classification errors is reduced when support vector machines are used.

15.3.1 Experimental Procedures

Significant supervised classification techniques include SVM. The target classes are classified using a hyperplane. Identification of the hyperplane separating one class from the other classes is the process of classification. Although the SVM takes a very long time to train, it is quite accurate at predicting the target classes.

15.4 Feature Engineering

To study the UCI machine learning archive classification workflow dataset for the Cleveland epileptic recognition dataset is performed. The training dataset and testing dataset are the two sets created from the data set. On the training data, associated feature engineering is carried out, and the resulting model is applied to the test data to make predictions. The following is the job description for the next position: Task Description: "Predict the value of patients with heart disease" by analysing the number of patients with heart disease consisting of 303 records. The attributes f_ca, f_thal, f_oldpeak, f_thalac and f_cp. attributes are selected based on the values that correlate with the target attribute num.

Data attributes are,

f_age—age data base feature is given in years.

f_sex—features in the Gender database are classified with a male value of 1 and a female value of 0.

f_cp—the scores for angina, atypical angina, nonanginal pain, and asymptomatic pain in the chest pain database are 1, 2, 3, and 4, respectively.

f_trestbps—this database characteristic attributes the resting blood pressure (BP) at the time the subject was admitted to the hospital, expressed in mmHg.

f_chol—this database characteristic is serum cholesterol expressed in mg/dl.

f_fbs—this database characteristic is the Fasting Blood Glucose >120 mg/dL attribute, which is digitized as 1, 0 for true and false.

f_restecg—this database feature is resting ECG results expressed as 0.1 values for normal and ST T-T wave abnormalities (T wave inversion and/or ST elevation or depression >0.05 mV) standard.

f_thalac—this database function applies to the patient's maximum heart rate.

f_exang—this database attribute pertains to exercise-induced angina and is numerically 1 and 0 for categorical values Yes and No.

f_oldpeak—this database feature relates to exercise-induced ST depression versus rest.

f_slope—this database characterization of ST segment slopes during peak exercise expressed in terms of ascending, flattening, and descending slopes with values of 1, 2, and 3, respectively.

f_ca—this database function counts the number of major vessels ranging from (0 to 3) by fluoroscopic staining.

f_thal—this database feature applies to the types of cardiac defects, with values of 3 for normal, 6 for fixed defects, and 7 for reversible defects.

f_num—this database feature is used to predict patients suffering from heart disease.

The 303 input tuples are divided into 91 tuples for the test data set and 212 tuples for the training data set. Equations 15.1 and 15.4 are used to create the training dataset, which is then performed in Python.

$$\text{X\_train} = \text{df.iloc}[\,0:212\,,\ :\,] \tag{15.1}$$

$$\text{y\_train} = \text{np.array}(\text{X\_train.iloc}[:, -1]) \tag{15.2}$$

$$\text{X\_test} = \text{df.iloc}[\ \text{int\_a:303}\ ,\ :\ ] \tag{15.3}$$

$$\text{y\_test} = \text{np.array}(\text{X\_test.iloc}[:, -1]) \tag{15.4}$$

$$\text{cls} = \text{svm.SVC}(\text{kernel} = \text{‘rbf’}, \text{gamma} = \text{‘auto’}) \tag{15.5}$$

$$\text{cls.fit}(\text{X\_train},\ \text{y\_train}) \tag{15.6}$$

The training set derived from Eqs. 15.2 and 15.3 is the variables used in the svm function in Eq. 15.4. Kernels used for splitting include linear, polynomial, sigmoidal, and radial values. Equation 15.4 above uses a kernel with linear values. Kernels with sigmoids are also used for additional analysis. Equations 15.5 and 15.6 are used to fit the training set.

15.4.1 Performance Analysis

Mean Absolute Error (MAE), Sum Squared Error (SSE), and Mean Squared Error are a few essential metrics used to evaluate a dataset's performance. The average absolute difference between an instance's actual value and expected value is referred to as "MAE." By adding the squares of the dataset's actual instance values minus its projected instance values, the SSE is determined. The MSE of a dataset is calculated as the average of the squares of the actual instance values less the projected values (Japp, 2016).

15.5 Predictive Analysis

Preprocessing the data and evaluating the missing attributes using the average of the attributes is done before analyzing the predictions. We then calculated the power scores MAE, SSE, and MSE for the entire heart disease data set as shown in Table 15.1 The global dataset considered 80, 70, and 60% of the training dataset. We found lower values for MAE, MSE, and RMSE for the 70% test data set compared to 80% and 60%. Therefore, we treat 70% of the training dataset as a shared dataset. From Table 15.2, it can be seen that MAE, MSE and RMSE are minimum when the f_ca values are greater than the mean f_ca value. As a result, when the value of f_ca values greater than the mean f_ca value a higher degree of accuracy. Table 15.3 shows that MAE, MSE and RMSE are lower when f_thal values lesser than mean or equal to mean f_thal values. Therefore, the f_thal values lesser than or equal to mean f_thal value is better at predicting. Now Table 15.4 shows that MAE, MSE and RMSE are minimized when f_oldpeak values lesser than or equal to mean f_oldpeak values, thus f_oldpeak values lesser than or equal to mean property gives better predictions when the f_oldpeak values greater than the range. Now, looking at Table 15.5, it can be seen that MAE, MSE and RMSE are low when f_thalac values greater than mean values. Thus f_thalac values greater than mean values are better in prediction accuracy. Table 15.6 shows that MAE, MSE and RMSE are equal for both the cases. Thus f_cp prediction for a given range is the same. Analysing the Tables 15.2, 15.3, 15.4, 15.5 and 15.6, we find that MAE, MSE and RMSE lowest value, f_thal values lesser than or equal to the mean f_thal value has the least value. Thus, this is considered as the attribute property for prediction in the heart disease data set.

Table 15.1 MAE, SSE and RMSE for overall test dataset

Error type	60% training dataset	70% training dataset	80% training dataset
MAE	0.9426	0.9120	1.0
MSE	2.254	2.120	2.311
RMSE	1.501	1.456	1.52

Table 15.2 MAE, SSE and RMSE for f_ca attribute

Error type	f_ca values lesser than or equal to mean f_ca	f_ca values greater than mean f_ca
MAE	0.9426	0.9120
MSE	2.254	2.120
RMSE	1.501	1.456

Table 15.3 MAE, SSE and RMSE for f_thal attribute

Error type	f_thal values lesser than or equal to mean f_thal	f_thal values greater than mean f_thal
MAE	0.5344	1.575
MSE	1.224	3.696
RMSE	1.106	1.922

Table 15.4 MAE, SSE and RMSE for f_oldpeak attribute

Error type	f_oldpeak values lesser than or equal to mean f_oldpeak	f_oldpeak values greater than mean f_oldpeak
MAE	0.6065	1.533
MSE	1.262	3.866
RMSE	1.123	1.966

Table 15.5 MAE, SSE and RMSE for f_thalac attribute

Error type	f_thalac values lesser than or equal to mean f_thalac	f_thalac values greater than mean f_thalac
MAE	1.355	0.4782
MSE	3.266	1.0
RMSE	1.8073	1.0

Table 15.6 MAE, SSE and RMSE for f_cp attribute

Error type	f_cp values lesser than or equal to mean f_cp	f_cp values greater than mean f_cp
MAE	0.9120	0.9120
MSE	2.120	2.120
RMSE	1.456	1.456

15.6 Conclusion

A major challenge for machine learning has been to effectively classify medical datasets. When associations and patterns are rapidly extracted from these complex health datasets, the diagnosis, prediction, and accuracy of cardiovascular disease outcomes can be improved. Processing cardiac information with raw healthcare data enables long-term lifesaving and early detection of heart disease abnormalities. Machine learning techniques were applied in this study to process the raw data and provide fresh and unique insights into heart disease. Predicting heart disease is difficult and important in the medical industry. However, early detection of the disease and taking preventive measures as soon as possible can significantly reduce mortality. More complicated models and model combinations are required to increase the accuracy of detecting the early start of heart illnesses. In this paper the hybrid support vector machine is used for a prediction made for the heart disease taking the UCI machine learning Cleveland data set repository. The attributes f_ca, f_thal, f_oldpeak, f_thalac and f_cp attributes are selected based on the values that correlate with the target attribute num. MAE, MSE and RMSE are calculated considering the data set attributes. We find that f_thal values lesser than or equal to mean f_thal value has the least value has a higher chance of prediction than the other attributes in a given range. In future other prediction methods such as artificial neural networks, genetic algorithms, decision tress and deep learning procedures will be utilized for prediction.

References

AliveKor [Online]. https://www.alivecor.com/how-it-works

Ananey-Obiri D, Sarku E (2020) Predicting the presence of heart diseases using comparative data mining and machine learning algorithms. Int J Comput Appl 975

Atallah R, Al-Mousa A (2019) Heart disease detection using machine learning majority voting ensemble method. In: 2019 2nd international conference on new trends in computing sciences (ICTCS), Amman, Jordan, IEEE, pp 1–6

Basha N, Ashok Kumar PS, Gopal Krishna C, Venkatesh P (2019) Early detection of heart syndrome using machine learning technique. In: 2019 4th international conference on electrical, electronics, communication, computer technologies and optimization techniques (ICEECCOT), Mysuru, India, 11th June, pp 1–5

Boudi FB (2016) Risk factors for coronary artery disease [Online]. https://emedicine.medscape.com/article/164163-overview

Gavhane A, Pandya I, Kokkula G, Devadkar K (2018) Prediction of heart disease using machine learning

Japp AG, Gulati A, Cook SA, Cowie MR, Prasad SK (2016) The diagnosis and evaluation of dilated cardiomyopathy. J Am Coll Cardiol. American College of Cardiology Foundation Published by Elsevier

Johnson-Coyle L, Jensen L, Sobey A (2012) Peripartum cardiomyopathy: review and practice guidelines. Am J Crit Care 21(2)

Karayõlan T, Kõlõç Ö (2017) Prediction of heart disease using neural network

Kohli PS, Arora S (2018) Application of machine learning in disease prediction. In: 2018 4th international conference on computing communication and automation (ICCCA, Greater Noida, India, 29th July, pp 1–4

Lin C-H, Yang P-K, Lin Y-C, Fu P-K (2020) On machine learning models for heart disease diagnosis. In: 2nd IEEE Eurasia conference on biomedical engineering, healthcare and sustainability 2020, Tainan, Taiwan, pp 158–161

Machine learning based decision support systems (DSS) for heart disease diagnosis: a review. Online, 25 March 2017. https://doi.org/10.1007/s10462-01

Maru A, Sharma AK, Patel M (2021) Hybrid machine learning classification technique for improve accuracy of heart. In: Proceedings of the sixth international conference on inventive computation technologies [ICICT 2021] IEEE Xplore Part Number: CFP21F70-ART. ISBN: 978-1-7281-8501-9, pp 1107–1110

Motarwar P, Duraphe A, Suganya G, Premalatha M (2020) Cognitive approach for heart disease prediction using machine learning. In: 2020 international conference on emerging trends in information technology and engineering (ic-ETITE), 27 April, Vellore, India, pp 1–5

Ngom F, Fall I, Camara MS, Alassane BAH (2020) A study on predicting and diagnosing noncommunicable diseases: case of cardiovascular diseases

Patil SB, Kumaraswamy YS (2009) Extraction of significant patterns from heart disease warehouses for heart attack prediction. IJCSNS 9(2):228–235

Purusothaman G, Krishnakumari P (2015) A survey of data mining techniques on risk prediction: heart disease. Indian J Sci Technol 8(12). Alexander CA, Wang L (2017) Big data analytics in heart attack prediction. J Nurs Care 6(393):1168–2167

Sebastian VB, Unnikrishnan A, Balakrishnan K (2012) Grey level co-occurrence matrices: generalisation and some new features. Int J Comput Sci Eng Inf Technol (IJCSEIT) 2(2):151–157

Sharma H, Rizvi MA (2017) Prediction of heart disease using machine learning algorithms: a survey

Soni J et al (2011) Predictive data mining for medical diagnosis: An overview of heart disease prediction. Int J Comput Appl 17(8):43–48