Asset Analytics
Performance and Safety Management
Series Editors: Ajit Kumar Verma · P. K. Kapur · Uday Kumar

Vikram Bali
Vishal Bhatnagar
Joan Lu
Kakoli Banerjee *Editors*

Decision Analytics for Sustainable Development in Smart Society 5.0

Issues, Challenges and Opportunities

Springer

Asset Analytics

Performance and Safety Management

The main aim of this book series is to provide a floor for researchers, industries, asset managers, government policy makers and infrastructure operators to cooperate and collaborate among themselves to improve the performance and safety of the assets with maximum return on assets and improved utilization for the benefit of society and the environment.

Assets can be defined as any resource that will create value to the business. Assets include physical (railway, road, buildings, industrial etc.), human, and intangible assets (software, data etc.). The scope of the book series will be but not limited to:

- Optimization, modelling and analysis of assets
- Application of RAMS to the system of systems
- Interdisciplinary and multidisciplinary research to deal with sustainability issues
- Application of advanced analytics for improvement of systems
- Application of computational intelligence, IT and software systems for decisions
- Interdisciplinary approach to performance management
- Integrated approach to system efficiency and effectiveness
- Life cycle management of the assets
- Integrated risk, hazard, vulnerability analysis and assurance management
- Adaptability of the systems to the usage and environment
- Integration of data-information-knowledge for decision support
- Production rate enhancement with best practices
- Optimization of renewable and non-renewable energy resources

Review Process

The proposal for each volume follows multi-pronged review process. The first level of review (single blind) is conducted by the series editors who may or may not decide to enlist the help of editorial board members for a second level of review. Proposals may also undergo a third level of peer review (double blind) if recommended by the Series Editors.
The series follows Ethics Statement found in the Springer standard guidelines here. https://www.springer.com/us/authors-editors/journal-author/journal-author-helpdesk/before-you-start/before-you-start/1330#c14214

More information about this series at https://link.springer.com/bookseries/15776

Vikram Bali · Vishal Bhatnagar · Joan Lu ·
Kakoli Banerjee
Editors

Decision Analytics for Sustainable Development in Smart Society 5.0

Issues, Challenges and Opportunities

Springer

Editors
Vikram Bali
IMS Engineering College
Ghaziabad, India

Joan Lu
Department of Computer Science
School of Computing and Engineering
Centre for Planning
Autonomy and Representation
of Knowledge
University of Huddersfield
Huddersfield, UK

Vishal Bhatnagar
Department of Computer Science
and Engineering
NSUT East Campus (Ambedkar Institute
of Advanced Communication Technologies
and Research)
Delhi, India

Kakoli Banerjee
Department of Computer Science
and Engineering
JSS Academy of Technical Education
Noida, India

ISSN 2522-5162 ISSN 2522-5170 (electronic)
Asset Analytics
ISBN 978-981-19-1688-5 ISBN 978-981-19-1689-2 (eBook)
https://doi.org/10.1007/978-981-19-1689-2

Preface

The world is encountering a vast technological shift from the traditional methodology to IoT-based sensor-enabled high-technology solutions which are easy, simple, and accurate in precision for providing the output. The inclusion of the IoT makes the usage easier and demands simultaneously usage of security aspect which may preserve the privacy of the users. The inclusion of the blockchain technology which is hash-based function implication helps in providing a unique identification to the owner which helps him get an ownership right which is a matter of concern in today's scenario. The usage of intelligent analytics in form of decision analytics helps in making an accurate and correct decision with usage of advanced deep learning, nature-inspired, and genetic algorithms.

Society 5.0 is a human-centred society that integrates economic advances such as increased demand for resources, longer life expectancy and increased international competitiveness, combining cyberspace and physical space with addressing social challenges through a mechanism such as GHG pollution reduction, cost mitigation associated with ageing society, and the promotion of sustainable industry.

This book aims at multidisciplinary research for sustainable development in smart society using decision analytics. The goal is to apply decision analytics principles to determine high-quality decision elements to improve delivering values using multiple statistical approaches. Decision analytics is the methodical quantitative and visual approach to address data science principle for sustainable development of smart societies. The digitization brings transformation acceleration of ecosystem by adaptation of innovation, value creation, and strategic market decisions towards creation of sustainable societies.

Using decision analytics, this book targets multidisciplinary analysis for sustainable growth in smart society. The aim is to apply decision analytics concepts to define high-quality decision elements to increase value distribution using various statistical approaches. Decision analytics is the methodical quantitative and visual methodology for discussing smart societies' theory of data science for sustainable growth. Digitization brings acceleration of ecosystem change by adapting creativity, value generation, and strategic business choices for sustainable communities.

Smart infrastructure's key aim is to achieve optimum resource utilization and boost efficiency. However, there are numerous challenges facing adoption of a smart infrastructure concept, particularly in developing countries such as location and lack of human skills, financing, and well-developed business models. Smart Healthcare is a groundbreaking idea, which inevitably affects multiple senses—Culture: encouraging healthcare facilities can clearly help to build a healthy society in which people are more concentrated on proper diet and physical activity. Government: implementing smart healthcare and patients can benefit from effective therapies at the lowest cost.

Smart manufacturing problems include security issues in smart manufacturing, device convergence, interoperability, safety in human–robot cooperation, multilingualism, and investment return in emerging technologies.

Smart cities consist of dynamic, networked assemblies of emerging technology, ICT networks, and IoT sensors to manage numerous city-wide services and structures and maximize resource use. Because smart cities' IoT devices need interconnection across various types of networks, they reveal the vast amount of security risks and there are more possible entry points for network-wide attackers. Smart city IoT devices consist of thousands or millions of interconnected IoT devices that capture and transmit data. As a result, malevolent intruders will exploit the smart city structure to build and distribute self-propagating malware that can propagate through several linked networks. Attackers can quickly manipulate private information, including users' financial healthcare and bank passwords. They are also able to conduct different forms of cryptography attacks with the intention of destroying: (i) secrecy to retrieve information and track system operation, such as illicit data collection by eavesdropping or monitoring message traffic; (ii) honesty to alter information and adjust system environment, such as unauthorized access to sensitive information. Smart cities require a range of modules, including universal sensor systems, heterogeneous networks, vast libraries, and efficient data centres to capture, move, store, and smartly process real-time information. Smart cities will deliver new applications and programs to improve residents' day-to-day decision-making, energy use, transit, housing, and education. Smart city application is responsible for providing people with a liveable, sustainable, climate-friendly, and engaging atmosphere. Smart energy seeks to manage energy and resource use effectively and expands the use of sustainable energy sources based on an integrated and resilient supply structure and insight-driven and pioneering strategic planning approaches. ICT plays an important role in the growth and activity of smart building, where real-time reaction to the ever-increasing data stream is focused on analytics and data management. In other words, the smart building's sensors, displays and controls can capture and interpret the data generated in real-time mode to keep it under control. One of smart building's key applications is to control energy appliances by automatically responding to internal policy and external signals to minimize energy usage and improve energy reliability. In everyday practices, smart people typically implement innovative and green technologies, develop innovative programs and goods, explore new ways of exchanging knowledge, and engage actively in decision-making. One of the major cities' greatest problems is developing an effective, smart traffic management scheme. Moreover,

traffic congestion triggers numerous problems such as wasted time, rising emissions and safety worries, and impacting economic growth and residents' quality of life. Smart traffic definition is a successful way to tackle this basic problem by taking a swift decision based on sensors, mobile devices, license plate readers, and CCTV cameras.

Smart culture will transform healthcare landscape. It will communicate and exchange knowledge between users of medical data and uses artificial intelligence to promote freedom. It further encourages the use of self-driving taxis and public transit buses in versatility. Sensors and Artificial Intelligence are used in smart city to manage highways, bridges, tunnels, and dams. In this, it facilitates cashless transactions and uses blockchain transmission technologies.

Distributed Denial-of-Service (DDoS) attacks arose as one of the main security threats for public web servers, where cloud computing vendors are responsible for delivering customer services. DDOS attacks also impact various facets of smart cities. Enabled devices and smart objects typically communicate in heterogeneous settings due to the variety of vendors and organizations developing these devices under diverse specifications, protocols, and technical needs. As a result, proposing an appropriate approach that can satisfy all smart city devices' protection standards, homogeneity, and interoperability criteria is very difficult. This is because end-to-end connectivity suffers from link, which also involves gateways or proxies.

This book covers sustainable development in smart society's 5.0 using data analytics. The data analytics is the approach of integrating diversified heterogeneous data for predictive analysis to accredit innovation, decision-making, business analysis, and strategic decision-making. The data science brings together the research in the field of data analytics, online information analytics, and big data analytics to synthesize issues, challenges, and opportunities across smart society 5.0. Accordingly, the book offers an interesting and insightful read for researchers in the areas of decision analytics, cognitive analytics, big data analytics, visual analytics, text analytics, spatial analytics, risk analytics, graph analytics, predictive analytics, and analytics-enabled applications.

Chapter 1 discusses about the standards of the healthcare industry affecting the economic performance and quality of life of a country. The new technological advancements and the services ensure patient's and practitioner's satisfaction. The remote monitoring, stress analysing, cancer detection, and proactive diagnose have been made possible with the evolution of new technologies. This chapter highlights various technological aspects in the healthcare industry and discusses about the related work presented by different researchers. The aim, applications, and challenges faced by the implementation of technologies like robotics and artificial intelligence, augmented reality, blockchain technologies, Internet of Things, and three-dimensional printing have been highlighted. The chapter presents a quick glance at the various intelligent healthcare systems developed based on presented technologies along with significance and challenges of data analytics in the healthcare industry.

Chapter 2 discusses about the fruit flies which are the common pests found in fruits and vegetables. They invade into the fruit causing severe damage to farmers. The fruits invaded by fruit flies are rejected in the export market. The performance of deep nets

is very high but the negative side of that are the problems associated with training the data. Associated with Deep nets is the complexity of the network and the vanishing gradient problem. These are overcome using Transfer Learning. Here, pretrained networks are used. These take lesser time since they are already pretrained and thus the time taken also reduces. This also has the advantage of delivering increased accurate results when compared with the traditional neural networks. It has been proven that Deep Networks give very efficient and accurate results when classifying images. In this project, we propose to train the gherkins dataset using the ResNets to identify fruit fly-infected gherkins. We have used Transfer Learning approach to train the classifier. We have also built a Flask web framework-based web application to help users upload gherkins images and get back predictions. In this chapter, we discuss a method which could help gherkin industries and farmers gain higher profits in their respective businesses for long-term sustainable development.

Chapter 3 discusses about the countries that are budging towards environment-friendly congestion; EVs are procuring supplementary reputation beyond the world. Unlike orthodox vehicles that could be replenished without complete fuelling setup at anticipated locations, some of the positive features of electric vehicles are that they can be refuelled at several places like homes, offices, shopping malls, parking areas, etc. However, appropriate and suitable charging arrangement is required to build at these locations. A widespread and definitely nearby charging network will be utmost important for huge adoption of electric vehicles. The development of a strong charging structure network is broadly known as a vital necessity for a huge-scale evolution to electromobility. This arrangement will not only give additional charging choices for drivers but would also encourage alertness and range self-confidence among approaching electric vehicle holders. As to acclimate towards progress of growing extent along with complex design of Electric Vehicle (EV) Fast Charging Station (FCS), also tendency of elegant city, this chapter elucidates an application that is based on the acceptance of cell phone Internet as well as mobile station to comprehend intellectual booking structure of electric vehicle fast charging station. Within this application, the structure recognizes user's vehicle details, shows nearby stations, helps to book slot as per his/her convenience, provides routing guidance, and starts the charging process.

Chapter 4 discusses about documents that contain a tremendous quantity of important human information. The use of automatic text classification is necessitated by the fast rise in the quantity of machine-readable documents for public or private access. In a standard text categorization system, preprocessing is one of the most important elements. This article will look at the influence of preprocessing on text classification from a variety of perspectives, including classification accuracy, text domain, text language, and dimension reduction. Text categorization is a key difficulty for knowledge-based technologies since it touches on all of artificial intelligence's old demons: the knowledge engineering bottleneck, scalability issues, simple transfer across various applications, and cost-effective system design. By specifying a document, information retrieval (IR) systems have typically avoided all of these difficulties. Message classification is a text categorization problem that has piqued machine learning's interest. The categorization of brief text messages is one element

of message classification that poses a special problem. The technique of categorizing documents into specified groups based on their content is known as text classification. It is the process of automatically categorizing natural language documents into preset groups. Text categorization is a prerequisite of both text retrieval systems, which retrieve texts in response to a user query, and text understanding systems, which change text in some way, such as creating new documents. Existing supervised learning methods for text classification require a large number of documents to learn correctly.

Chapter 5 discusses about the Data Owners who can store their documents on multiple distributed servers in distributed systems. It allows the use of data by sitting anywhere on any computer in and from multiple servers. The efficient hunt for distributed documents offering data protection on ownership documents is also a very demanding mission. One potential alternative is that of data protection indexing (PPI) for the current method. Documents are transmitted via many private servers collectively managed by the cloud/public server in this framework. Users wanting to provide such documents ask for public cloud, which then returns a list of applicants which is a private server list to their users. After a list is collected, users can access documents on a particular private server, but documents are stored on a private server in plain text form, which violates their privacy. But the system proposed improves the security and efficiency of this current system.

Chapter 6 discusses that traffic is increasing on the roads, and the accident cases are also increasing. For controlling this, the cameras are installed at various traffic signals or various road intersections in the smart cities. The image captured from these cameras will help to detect the rule violators. This has been done by using an Automatic Number Plate Recognition (ANPR) system which is an automated system and will be able to detect the various types of traffic rule violations. The number plate has been detected after identifying the vehicle. From the number plate, the vehicle owner's information has been collected and the e-challan has been generated. This generated e-challan will then be sent to the registered email ID for the information and the fine towards the traffic rule violation.

Chapter 7 discusses about the agriculture field, the presence of plant disease and organic deficiency causes major harm and loss to farmers. The conventional way of detecting plant pathogens is through visual checking. Visual checking is mostly done after major damage to the crop, so treatments will be of limited or no use. To prevent the plant from severe damage and to achieve sustainable development in the farming sector, farmers must detect an infection before it is visible. This is achieved by premature disease diagnosis. By considering this problem as a challenge, the proposed system provides a solution by using the concept of Internet of Things to achieve sustainability in the smart farming domain. The proposed system aims at the development of an automated tool to identify whether the plant is normal or diseased. The main goal of the research work is the design and development of an Internet of Things (IoT)-based handheld tool to detect pigeon pea diseases. The proposed system attempts to design and develop an automated tool that senses the presence of disease in the plants using various sensors like temperature, humidity, and colour. Identification of plant disease is done through sensor readings.

Chapter 8 discusses about smart city applications which collect massive amount of data from various types of IoT sensors, engines, and people. Most of the data generated is heterogeneous in nature. However, this data needs to be integrated with legacy applications based on SQL. Some of these applications also require migration to NoSQL for improved performance and fault tolerance. This paper addresses challenges of working in hybrid environments and migration issues from SQL to NoSQL databases. Rapid rate of growth in heterogeneous data and characteristics of NoSQL database like easy scalability, high availability, high performance, and low cost are the motivating factors to migrate towards NoSQL from relational databases, especially for applications requiring dealing with unstructured data. NoSQL database gives dynamic schema, adjustable data model, scale-out architecture, and allows storage of big data and access to it in an efficient manner. The relational databases store data in form of tables with fixed schema. Relational databases are structured and not capable of handling unstructured and big data. In relational databases, because of normalization, data is spread across multiple tables and expensive join operation is required to integrate data. Due to limitations of relational databases, most of the leading organizations are migrating towards NoSQL. NoSQL databases are analysed into four categories: key-value database, column-oriented database, document-based database, and graph database. Key research issues and challenges in migrating from relational to NoSQL include model transformation (including mapping and schema conversion), application integration, strategies related to perform join in different scenarios, use of indexes, storage issues, etc. Massive growth is likely in the area of NoSQL over cloud because of initiatives like smart city; however, these applications required integration with legacy applications built over SQL. Thus, need for migration as well as bridges for connection leading to requirements of our research area. This chapter tends to explore comparative study of Relational databases and NoSQL databases, classification of NoSQL databases, case study, popular approaches for migration, and some of the key issues and challenges in migrating from Relational to NoSQL databases. This chapter will be fruitful for students and researchers too.

Chapter 9 discusses that with research and development in full swing for super smart society, Collaborative Business Intelligence (CBI) has emerged as a key component in the whole ecosystem. Super smart society aims to meet the needs of members of society through the provision of the quality goods and services in adequate amounts to address their distinctive needs based on age, gender, religion, or language. This requires highly structured and secure collaborative business intelligence that would analyse data from various elements of super smart society like IoT, AI, and Big Data. This chapter aims to explore recent developments in the field of CBI. It further explores information security practices and related mechanisms to ensure safeguarding CBI processes. Finally, chapter covers a use case and design of data security framework for CBI that can boost sustainability of super smart society.

Chapter 10 discusses that a few months back all were living on with a unique life uninformed of how quickly things planned to change. The coronavirus is a worldwide upsetting influence that has constrained us all to meet up and attempt to stop it. While we are generally making changes and doing what we can to help, there is still significantly more that should be finished. We are building a site that gives data to

help battle this COVID-19 pandemic and its impact on sustainable development. This application is for all the clients who need to think about the current circumstance in regard to this pandemic time on the whole over world and if a client needs to look for data about COVID-19 cases in their territory. For utilizing this framework, one simply needs to have an advanced mobile phone or PC or framework with web access and some essential information on utilizing it. The expressed motivation behind this application is to spread familiarity with COVID-19 and to associate fundamental COVID-19-related well-being administrations to individuals and offer best practices and warnings. It is the following application that utilizes the gadget's area highlights to follow the COVID disease. With informing framework, it attempts to decide the danger if one has been almost a COVID-19-contaminated individual, by looking over a data set of known cases across India. Utilizing area data, it decides if the area one has a place with one of the contaminated zones dependent on the information accessible.

Chapter 11 discusses that Geographical information systems (GIS) are proven to be extremely useful in a variety of natural resource applications in Smart Society, including topography hydrology modelling for water harvesting. The watershed is the basic object in such systems, which is either physically drawn on topographic map sheets or computationally generated from digital elevation model (DEM) data. This article describes a method for automatically defining drainage and watersheds using a terrain's DEM. The efficacy of Digital Elevation Models in Morphometric Analysis of Rapti Sub-watershed in Gorakhpur District, Uttar Pradesh was investigated in this study. The Rapti sub-watershed, which covers an area of 20.8 km^2 in Gorakhpur, was selected and six different threshold levels of 0.05 km^2, 0.25 km^2, 0.50 km^2, 0.125 km^2, 1.0 km^2, and 2.0 km^2 were applied during the automization process using ESRI ArcGIS, open-source QGIS software, and freely available Digital Elevation and Surface Models of ASTER-GDEM, and ALOS-DSM ALOS-DSM displays streams with finer features than ASTER-GDEM at all stream thresholds. For ALOS-DSM, the correlation coefficient R2 between thresholds and stream length (Lu) is 0.8817, while for ASTER-GDEM, it is 0.9017. In the comparative assessment of ASTER-GDEM and ALOS-DSM which has been carried out has identified ALOS-DSM exhibits finer drainage networks in comparison with ASTER-GDEM proving ALOS-AW3D 30 DSM is the best alternate source available.

Chapter 12 discusses that nowadays Internet of Things (IoT) is important part in the present world, which fills in as a feature of our standard life exercises. Despite the fact that IoT gives benefits in many ways. Like through IoT the communication between the doctors and patients was possible only through texts, visits, and telecommunications. However, the advancement in technology and the evolution of IoMT in healthcare have significantly changed the scenario. In this chapter, we will see the role of IOT in health sector in every aspect like COVID-19 pandemic. Then we will see IOT healthcare networks along with its architecture. Along these lines, the requirement of a secure climate is indispensable to get the communicating information through IoT. Therefore, in this chapter, a secure plan is also recommended on utilizing steganography as an elective security system related to a healthcare sector

to get the communicated information from IP camera as the IoT gadget to different gadgets.

Chapter 13 discusses about Medium Access Control (MAC) protocols which are responsible for handling the interactions between different communication mediums. They perform tasks like frame delimitation, addressing of nodes, providing transparency to data communication, protection against errors, etc. In this paper, an improved MAC protocol design is proposed, which hybridizes MAC protocols like Time division multiple access (TDMA), bit-map-assisted (BMA) and sensor MAC (SMAC), and then adds multi-channel capabilities to it in order to improve the network Quality of Service (QoS) parameters. The proposed protocol also incorporates sleep scheduling capabilities to the hybridized multi-channel MAC protocol in order to further optimize the network lifetime and reduce routing computational complexity. The proposed protocol is compared with non-hybridized implementations of standard MAC protocols, and it is observed that a delay reduction of more than 20% is achieved, along with an improvement of 10% in the overall network lifetime, which can be used for sustainable development in smart cities that require low energy.

Chapter 14 discusses that in real-time processing, stream has to be processed as soon as it is generated. Data streams generated from IOT sensors are processed into a finite size window. A sheer window is considered for processing of data streams in a particular time stamp. In these sheer windows, compare reduce aggregate (CRA) algorithm is applied for determining linear relation for multiple feature vectors. A real-time inference pattern is determine using hash-based classification. In this chapter, Sheer window hash-based classification using binarized window analytic (SHCUBA) approach is proposed. This approach is beneficial for calculating linear relationship between sheer windows. SHCUBA approach is comprised of transformation and virtualization. Here, time and space complexity for this approach is and , respectively. This reduces latency and space requirements for various real-time use cases such as smart applications, sentiment analysis, IOT-based solutions, fraud detection and prevention, stock market prediction, etc. Data generated in smart societies can be correlated using SHCUBA approach for inferring useful decisions.

Chapter 15 discusses that quality of education is essential for the teachers and the learners, and it is also crucial for the analyst. The process of maintaining high education quality is critical and challenging too. Education is a very compound and complicated system which has a cause-and-effect relationship that is nonlinear. Thereby, present chapter explores the satisfaction of the students towards the higher educational institutions in Oman for sustainable society. The study was based on a primary data and a structured questionnaire has been designed to collect the data. The measurement model has been tested to measure the internal reliability. All the constructs have indicated high internal reliability as Cronbach's alpha, and composite reliability values have been higher than 0.70. When all the conditions for the measurement model have been met, then structural equation modelling is conducted. When all the conditions for the measurement model have been met, then structural equation modelling is conducted. The structural model has been analysed by way of bootstrapping methodology. The consistent PLS bootstrapping found that certification most

significantly impacts satisfaction followed by website satisfaction. The analysis of the bootstrapping exhibited that certification is the most significant variable which influence the satisfaction of the students followed by the website design. Along with the website and certification, it was suggested that the educational institutions need to concentrate on the placement policies which can allow the students to be placed.

Chapter 16 discusses that Big Data describes the form of data whose size is massive in volume, which turns out to be both structured and unstructured which aims to drive operations of a business. When we speak of Big Data it is not important that how much amount of data a business is concerned about. The most important thing matters with the huge amount are data is, how the organization is going to use the data or what they are going to do with that data. This huge amount of Big Data is analysed for insights that lead to efficient decision-making and helps to plan the strategic business moves. Big data is defined by its vastness, and it can comprise a variety of data types, such as urban data.

Chapter 17 discusses that tourism is a trending industry in the current era of the economic growth of the country. Tourism can be explained as travelling to a place which is different from your home city or country for numerous reasons. But the tourism industry has been facing problems like Identity theft, intermediaries, payment failures, fake reviews, transparency, settlement, overbooking, policy, and most importantly safety concerns due to COVID-19 pandemic. To overcome the above drawbacks after referring to some of the research papers, we have come up with a solution called blockchain. Blockchain is a chain of blocks where each block contains a set of transactions and the blocks are connected using cryptography. Mainly blockchain and tourism have the potential to show into a really advantageous combination as this technology will offer additional security, safety, eradication of middleman, data integrity, decentralized business models, immutability, transparency, and accurate contact tracing app for Tourism Industry in Future Smart Society. This paper gives the clear idea of how BLOBIM helps to overcome difficulties/challenges in tourism industries. Our future proposed architecture blockchain-based integrated model removes the third-party intervention, improves security, safety, trust, transparency, etc.

In Chap. 18, based on recent publications, it is hoped to gain a better understanding of the framework of Society 5.0 in commercial, as well as its latent for proactivity in the Sustainable Development Achievements, as well as possible consequences on sustainability and accounting reporting. Given the breadth of the deviations forced by the 4.0 Industry, it is very critical to comprehend the current effects and forthcoming opportunities of this progress with Cross Industry 5.0. All residents are expected to become dynamically involved in a Society 5.0, incorporating digital technology into a range of systems and speeding up their implementation. As a result, it is projected to expand on the notion and possibility of the specific technology interaction in the advancement of their own and society's quality of life, with an emphasis on sustainable development, sustainability, and reporting. As a result, it is a contribution that aims to contribute to this extremely important and vital conversation by offering its progression. It is determined that Industry 4.0, and more newly, Industry 5.0 are here to remain, putting artificial intelligence (AI), big data, robotics, and other

technologies by the provision of man in a world where everything will be connected and society will have to adapt. This position chapter aims to create a theoretically focused perspective on sustainable digital inventions in the framework of Society 5.0, as well as the underlying empirical potentials of this association.

Ghaziabad, India Vikram Bali
Delhi, India Vishal Bhatnagar
Huddersfield, UK Joan Lu
Noida, India Kakoli Banerjee

Contents

Editors and Contributors

About the Editors

Dr. Vikram Bali is Director of IMS Engineering College, Ghaziabad, India. He has more than 20 years of rich academic and administrative experience. He has published more than 50 research papers in international journals/conferences, and has edited and authored text books to his credit. Additionally, he has published five patents. Dr. Bali is on the editorial board and on review panel of many international journals, and is series editor for three book series of top international publishers. He was awarded the Green Thinker Z- Distinguished Educator Award 2018 for remarkable contribution in the field of Computer Science and Engineering at the Third International Convention on Interdisciplinary Research for Sustainable Development (IRSD) at the Confederation of Indian Industry (CII). His research interest includes cyber security, automata theory, CBSS, and ERP.

Dr. Vishal Bhatnagar holds B.E, MTech and PhD in the Engineering field. He has more than 21 years of teaching experience in various technical institutions. He is currently working as Professor in Computer Science & Engineering Department at Netaji Subhash University of Technology (East Campus) Formely Ambedkar Institute of Advanced Communication Technologies & Research, Delhi, India. His research interests include database, advance database, data warehouse, data-mining, social network analysis, Data Science and big data analytics. He has to his credit more than 120 research papers in various international/national journals, conferences and Book Chapters. He is currently working as Associate Editor of few Journals of IGI global and Inderscience. He has to his credit experience of handling special issues of Many Scopus, ESCI and SCIE Journals. He is also series editor of CRC book aimed towards Handling Data Science problems and solutions. He is also working as expert of allied regulatory and institutional bodies Like NBA, AICTE, UGC, IGNOU, NILEIT. Currently he is guiding 08 students for PhD and already guided 03 students. He can be reached through Email Id: vishalbhatnagar@yahoo.com, vishalbhatnagar@aiactr.ac.in.

Prof. Joan Lu is in the Department of Computer Science and is the research group leader of Information and System Engineering (ISE) in the Centre of High Intelligent Computing (CHIC), having previously been team leader in the IT department of Charlesworth Group publishing company. She successfully led and completed two research projects in the area of XML database systems and document processing in collaboration with Beijing University. Both systems were deployed as part of company commercial productions. Professor Lu has published seven academic books and more than 200 peer reviewed academic papers. Her research publications have 1388 reads and 185 citations by international colleagues, according to incomplete statistics from the research gate.

Professor Lu has acted as the founder and a program chair for the International XML Technology Workshop for 11 years and serves as Chair of various international conferences. She is the founder and Editor in Chief of International Journal of Information Retrieval Research and serves as a BCS examiner of Database and Advanced Database Management Systems, and is an FHEA. She has been the UOH principle investigator for four recent EU interdisciplinary (computer science and psychology) projects: Edumecca (student responses system) (143545-LLP-NO-KA3-KA3MP), DO-IT (multilingual student response system) used by more than 15 EU countries (2009-1-NO1-LEO05-01046), and DONE-IT (mobile exam system) (511485-LLP-1-2010-NO-KA3-KA3MP), HRLAW2016/3090/001–001.

Dr. Kakoli Banerjee is heading the Department of Computer Science and Engineering, JSS Academy of Technical Education, Noida, India. Prior to her 18 years in academia, she has three years of industry experience. Her papers have been published in international journals of repute indexed in Scopus, ESCI, and other eminent databases. Currently, she is working for a funded projects, sanctioned by the Collaborative Research and Innovation Program (CRIP) under TEQIP-3 by Dr. A. P. J. Abdul, Kalam Technical University Uttar Pradesh, Lucknow. She has been a reviewer and guest editor of several reputed journals and books. She has a Granted Patent to her name and one published. She has developed video content for subjects like Operating System and Database Management System for public universities and for government websites. She has published a book on Industry 4.0 by a leading international publisher.

Contributors

Afroz Dayananda Sagar University, Bengaluru, Karnataka, India

Afsa Dayananda Sagar University, Bengaluru, Karnataka, India

Mohit Agarwal School of Engineering & Technology, Sharda University, Greater Noida, India

Alankrita Aggarwal Panipat Institute of Engineering and Technology, Samalkha, Panipat, India

Manuj Aggarwal Department of Information Technology, Guru Nanak Dev Engineering College, Ludhiana, Punjab, India

Vikram Bali IMS college, Ghaziabad, India

Kakoli Banerjee Department of Computer Science and Engineering, JSS Academy of Technical Education Noida, Noida, Uttar Pradesh, India

Pankaj Bhambri Department of Information Technology, Guru Nanak Dev Engineering College, Ludhiana, Punjab, India

Pradnya S. Borkar Jhulelal Institute of Technology, Nagpur, India

Charu Chhabra The Northcap University, Haryana, India

Saloni Chinchmalatpure Computer Technology Department, Yeshwantrao Chavan College of Engineering, Nagpur, India

Sneha Choudhary The Northcap University, Haryana, India

Sachin Dhawan Panipat Institute of Engineering and Technology, Samalkha, India

Shubhra Dubey Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra, India

B. C. Girish Kumar Sharda University, Greater Noida, India

Vishal Goar Government Engineering College Bikaner, Bikaner, Rajasthan, India

Bhoomi Gupta Maharaja Agrasen Institute of Technology, New Delhi, India

Rashmi Gupta Netaji Subhas University of Technology, Delhi, India

S. L. Gupta Birla Institute of Technology, Noida, India

Sachin Gupta Maharaja Agrasen Institute of Technology, New Delhi, India

Amit Kanojia Department of Computer Science, Mata Jijabai Government Girls PG College, Indore, India

Pooja Kansra Mittal School of Business, Lovely Professional University, Punjab, India

Smita Kapse Computer Technology Department, Yeshwantrao Chavan College of Engineering, Nagpur, India

Bikram Pal Kaur Chandigarh Group of Colleges, Landran, Mohali, Punjab, India

Jaskirat Kaur Chandigarh Group of Colleges, Landran, Mohali, Punjab, India

Ramanpreet Kaur IKGPTU, Kapurthala, Punjab, India;
Chandigarh Group of Colleges, Landran, Mohali, Punjab, India

Kavyashree Dayananda Sagar University, Bengaluru, Karnataka, India

Rameshwari Kuhikar Prolifics Corporation Ltd., Hyderabad, India

Mukesh Kumar Mahant Bachittar Singh College of Engineering and Technology, Jammu, India

Rajender Kumar Panipat Institute of Engineering and Technology, Samalkha, Panipat, India

Sunil Kumar Electronics & Communication Engineering, Kalinga University, Raipur, India

Devesh Kumar Lal School of Computer Science & IT, Devi Ahilya University, Indore, India

Ashutosh Lanjewar Electronics & Communication Engineering, Kalinga University, Raipur, India

Latesh Malik Computer Science Engineering, Government Engineering College, Nagpur, India

Pranali Maraskolhe Computer Technology Department, Yeshwantrao Chavan College of Engineering, Nagpur, India

Sindhu P. Menon Dayananda Sagar University, Bengaluru, Karnataka, India

Nikhil Mishra Panipat Institute of Engineering and Technology, Samalkha, Panipat, India

Niraj Mishra Birla Institute of Technology, Ranchi, India

Parma Nand Sharda University, Greater Noida, India

Rashmita Nikhare Birlasoft The CK Group, Pune, India

Sakshi Ramgirwar Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra, India

Arun Kumar Rana Panipat Institute of Engineering and Technology, Samalkha, India

Sita Rani Department of Computer Science and Engineering, Gulzar Group of Institutions, Khanna, Punjab, India

Vinay Kumar Saini Maharaja Agrasen Institute of Technology, New Delhi, India

M. B. Santhosh Kumar Department of Civil Engineering, JSS Academy of Technical Education Noida, Noida, Uttar Pradesh, India

Supriya Sawwashere Computer Science Engineering, Kalinga University, Raipur, India

Sharad Sharma Maharishi Markandeshwar University, Mullana, India

Ramya Shetty Dayananda Sagar University, Bengaluru, Karnataka, India

Amrit Pal Singh Department of Information Technology, Guru Nanak Dev Engineering College, Ludhiana, Punjab, India

Harpuneet Singh Department of Information Technology, Guru Nanak Dev Engineering College, Ludhiana, Punjab, India

Parveen Singla Chandigarh Group of Colleges, Landran, Mohali, Punjab, India

Ruchi Pasricha Singla Chandigarh Group of Colleges, Landran, Mohali, Punjab, India

Punit Soni Chandigarh University, Chandigarh, India

Ugrasen Suman School of Computer Science & IT, Devi Ahilya University, Indore, India

S. Tanwani School of Computer Science & IT, DAVV, Indore, India

Rakshanda Tarekar Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra, India

Reena Thakur Jhulelal Institute of Technology, Nagpur, India

Supriya S. Thombre Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra, India

L. N. Tilak Department of Civil Engineering, JSS Academy of Technical Education Noida, Noida, Uttar Pradesh, India

Sampada S. Wazalwar Department of Information Technology, G. H. Raisoni College of Engineering, Nagpur, Maharashtra, India

Nikhil Wyawahare Electronics Engineering Department, G. H. Raisoni College of Engineering, Nagpur, India

Nagendra Singh Yadav Government Engineering College Bikaner, Bikaner, Rajasthan, India

Mohammad Yawar Waljat College of Applied Sciences, Muscat, Sultanate of Oman

Shivani Zulkanthiwar Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra, India

Chapter 1
Intelligent Systems for Sustainable Development of Healthcare Industry

Ramanpreet Kaur, Bikram Pal Kaur, Ruchi Pasricha Singla, Jaskirat Kaur, and Parveen Singla

Abstract The standards of the healthcare industry affect the economic performance and quality of life of a country. The new technological advancements and the services ensure patient's and practitioner's satisfaction. The remote monitoring, stress analysing, cancer detection and proactive diagnose has been made possible with the evolution of new technologies. This chapter highlights various technological aspects in the healthcare industry and discusses about the related work presented by different researchers. The aim, applications and challenges faced by the implementation of technologies like robotics and artificial intelligence, augmented reality, blockchain technologies, Internet of Things and three-dimensional printing have been highlighted. The chapter presents a quick glance at the various intelligent healthcare systems developed based on presented technologies along with significance and challenges of data analytics in the healthcare industry.

1.1 Introduction

The medical practitioners and the various smart facilities available for patient care focus on providing the effective and efficient healthcare system at the doorstep. The smart healthcare facilities aim at providing safe and secure environment to the patients as well as medical practitioners. The smart healthcare system not only provide customer satisfaction but also provide advanced patient care and focus on patient-centred experiences.

The smart healthcare technologies use smart, low-powered, miniature sized medical devices which are connected wirelessly through different wireless technologies. These devices are responsible for measuring different physiological parameters of the body like glucose level, body temperature, heart rate, blood pressure,

R. Kaur (✉)
IKGPTU, Kapurthala, Punjab, India
e-mail: ramanpreet.71@gmail.com

R. Kaur · B. P. Kaur · R. P. Singla · J. Kaur · P. Singla
Chandigarh Group of Colleges, Landran, Mohali, Punjab, India

V. Bali et al. (eds.), *Decision Analytics for Sustainable Development in Smart Society 5.0*, Asset Analytics, https://doi.org/10.1007/978-981-19-1689-2_1

oxygen level, body position, etc. The sensors used in the network are also capable of processing, capturing and security of medical data [1]. All the devices or the sensors used in the hospitals or clinics form "Internet of Medical Things". This is because different medical sensors placed on or implanted inside the body or embedded inside the hospital beds or medical equipment scan receive significant medical information of the remote patient and medical professionals act according to the situation and provide proactive solution to the problem.

The major objective of smart healthcare or Healthcare Industry 4.0 "is to allow for progressive virtualization in order to enable the personalization of health and care next to real time for patients, professionals and formal and informal carers" [2]. The major technological advancements in the field of healthcare as presented in Fig. 1.1 are robotics and artificial intelligence, augmented reality, blockchain technologies, Internet of Things and 3D printing [3–5].

The efficiency and effectiveness in the patient healthcare have improved due to the adoption of advanced technologies by various healthcare organizations. Before implementing the advanced technologies, the healthcare organizations must ensure the network requirements to handle critical medical data as the increasing demand may affect the performance of the network.

The chapter is organized as follows: Sect. 1.1 presents the introduction to the smart healthcare systems and their aim in accordance with the Healthcare Industry 4.0 guidelines. Section 1.2 discusses the advanced healthcare technologies like augmented reality, blockchain technologies, Internet of Things, robotics and artificial intelligence and 3D printing. This section also discusses the challenges and growth statistics of these technologies in past, present and future. The intelligent healthcare solution based on these technologies has been presented in Sect. 1.3. Section 1.4 discusses the role of data analytics and challenges in smart healthcare

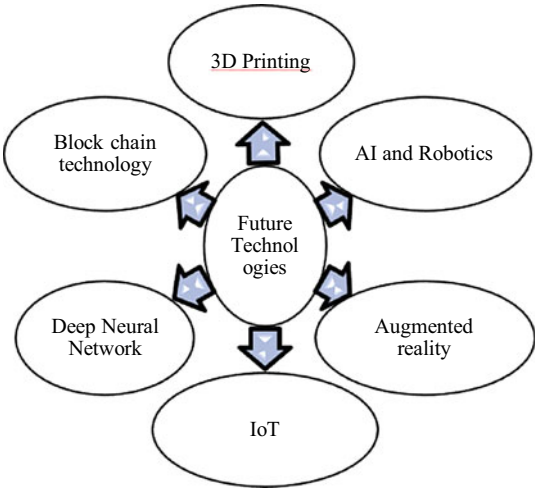


Fig. 1.1 Future healthcare technologies

industry. Section 1.5 concludes the chapter with future scope of these technologies in remote monitoring, surgical operations, customized prosthetics during normal and emergency circumstances.

1.2 Healthcare Technologies

The healthcare industry has seen tremendous growth with the advent and implementation of advanced technologies in surgical operations, telemedicine, remote patient monitoring, stress management and many more. The present section discusses about the advancements related to emerging technologies and the related work presented by different researchers. The objective and challenges faced by each technology have also been presented in Table 1.1 [5, 6, 11, 12, 18, 19, 25, 26, 29, 30].

Table 1.1 Outline of future healthcare technologies

S. No	Technology	Aim	Applications	Challenges
1	Augmented reality	Increase patient safety and decrease recovery time	Patient rooms, billing information, surgical operations, smart glasses, cafeteria design	Stress, overload, security, privacy, sometimes not user-friendly
2	Blockchain technology	Security, integrity, interoperability of data	Health insurance, drug supply, medical education, biomedical research	Latency, scalability, power limitations
3	AI and robotics	Perform human-like task and reduce patient discomfort	Recommend diagnosis, lab test, answer health-related queries, stress management, identifying cancer cells	Higher cost, fear of unemployment, lack of interoperability
4	IoT	Anytime, anyone, any place, anything, any network connectivity	Remote monitoring, interact with medical machines	Security, privacy, may cause accidental failures
5	3D printing	Additive manufacturing, treatment under direct supervision of experts	Bioprinting of tissues, organoids, 3D-printed models for surgery, surgical tools, customized prosthetics	Longer printing time, clinical investigations not accurate

1.2.1 Augmented Reality

It is one of the technologies helpful in providing virtual overlays to help the amateurs during any surgical operation and provide traditional learning environment. One of the applications is the use of smart glasses by both medical practitioners and helpers which not only view the medical statistics of the patient without consulting medical chart but also proved helpful in designing the patient rooms and providing billing information. It is also helpful in the patient insurance surgical operations and can also assist in cafeteria designing [6].

In another contribution by authors, the difference between virtual reality and augmented reality has been presented and it is stated that augmented reality is an extension of virtual reality only. The different biomedical applications of augmented reality like invasive surgery, mental healthcare have also been explained [7].

The authors in [8] have developed a mobile application named as Clinical Decision Support System (CDSS) based on augmented reality and deep learning concept for low-cost devices. This has been specifically designed for the patients suffering from bedsores. The decision has been taken based on the present situation of the bedsore. Also, three different situations of bedsore have been considered and two more stages have been added by CDSS. The three main functionalities proposed by the authors are bedsore classifier, its measurement and its time machine.

In another contribution, the authors have developed a prototype application based on augmented reality where virtual images of medical devices have been combined with real-time world images [9]. The presented work had been carried in three stages—assessment of user's need, development of the prototype and evaluation of the software. The work has been completed with the contribution of 11 nurses for focus group and 280 healthcare professionals for online available questionnaire. It has been concluded that augmented reality technology provides both safety and security to patients and improves the clinical support to the medical professionals. The major limitation of the work is that the app requires a prior information of the medical devices to be fed and updated regularly by professionals.

The authors in [10] have presented that the main aim of augmented reality is to enhance the patient's safety and decrease the recovery time. The authors have studied the impact of augmented and virtual reality in context to bone fracture for different stages like pre- and post-surgery including bone fracture healing, its diagnosis and rehabilitation. In [11], it is highlighted that both augmented and virtual reality technologies have the potential to be used for plastic surgery in preoperative applications which will help in improving clinical outcomes, operative time and cost-effectiveness.

1.2.2 Blockchain Technology

This is a technology behind Bitcoin and that aims to provide security, integrity and interoperability of data. The major advantage of this technology is that the data is not stored at a central location but is distributed all over the network participants [12]. Various industrial and manufacturing units like IBM and Deloitte have also developed products that ensure data integrity and accuracy using blockchain technology [12]. The major advantages of this technology are better and accurate diagnosis, provide effective treatment, patient satisfaction and enhanced security as data is immutable. The changes made in the data can be easily tracked by the experts. The blockchain technology has greater outcome related to data security and access by all network participants [13].

Data reports presented in Fig. 1.2 from BIS research estimate that blockchain technology is estimated to reach $5.61 billion by 2025, observing a double-digital growth during the period of 2018–2025 [14]. Blockchain technology apart from healthcare provides great help in e-commerce, education and manufacturing industry by allowing customer and company to keep online records for transactions and others.

In another contribution, the authors presented the various issues like delayed communications, lack of interoperability, availability and security of medical records stored in fragmented locations, in healthcare industry and application of blockchain technology to overcome these issues have been highlighted. The design considerations related to implementation of blockchain technology in healthcare have also been presented [15].

With the standardized regulations for protection of medical data of patients and healthcare data on cloud, various new options for managing health data have been developed using blockchain technology [16]. Various researchers have presented state-of-the-art of blockchain technology implemented in health insurance, drug supply, medical education and biomedical research [17].

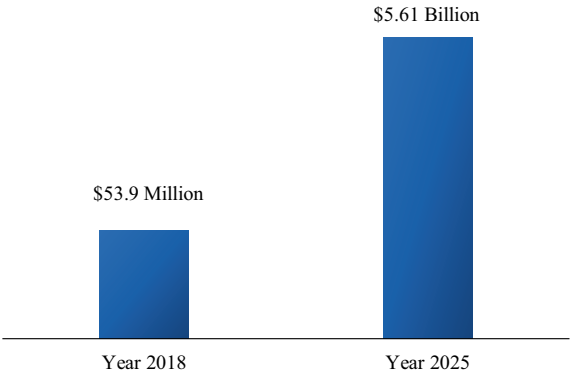


Fig. 1.2 Blockchain in healthcare industry [14]

1.2.3 Artificial Intelligence (AI) and Robotics

Robotics with artificial intelligence has proved to be the major technological trend in the health industry. Robotics machine like Da Vinci, Vasco Logic Venous Pro and many more help to perform the human-like tasks and has greatly reduced the patient discomfort [18]. This smart technology has also helped elders to find their best pal in case of loneliness and care. It has been found that robotics and AI have been accepted by both medical practitioners, patients, research as it provides advanced computer technology and can also recommend diagnosis perform lab test and answer any health-related question. The microphones, sensors and camera used in this technology allow the patient and the robot to talk to each other and provide a healthy communication between the two [19]. AI has been used to develop wireless powered sensor system which is used to send emergency messages to the medical practitioner to provide information about the patient suffering from ulcers and its ongoing pressure. In another work [20], machine learning algorithms have been used to classify the burns and ulcers of the patient using a support vector machine and pre-trained convolutional neural network.

Figure 1.3 shows that AI has significant role in healthcare. At the early stage, it helps to identify the illness and its algorithms help to assist improved decision-making procedure. With the help of previous records, it helps to provide better treatment plan. This gives a superior experience to patients. AI-based wearable provides a track of our daily routine link sleeping pattern, count of calorie, steps, etc. Robots with AI provided a better platform for old age patients for self-ruling life with the need of doctors.

Figure 1.4 shows the estimated and projected share of different countries of use of AI in healthcare industry. North America plays a key role of AI in healthcare industry due to existence on multinational companies like IBM, Microsoft, Intel and Google

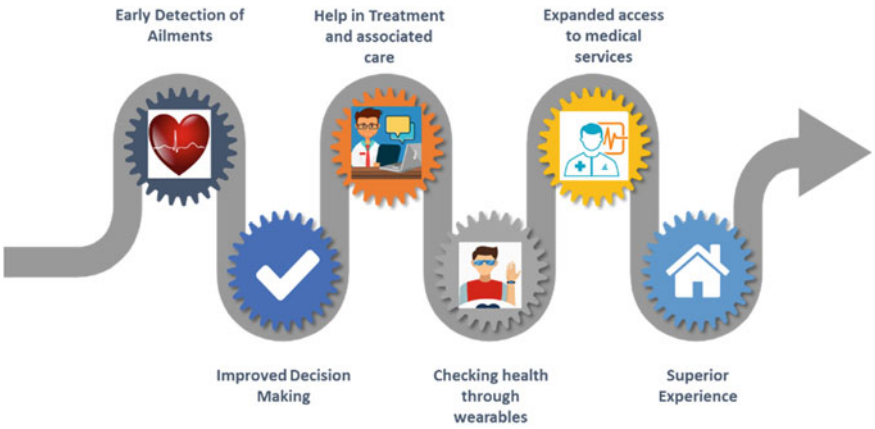


Fig. 1.3 AI in healthcare industry Source: https://www.botreetechnologies.com/blog/artificial-intelligence-in-healthcare-industry

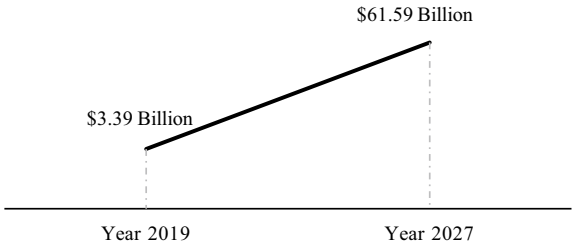


Fig. 1.4 AI in healthcare industry by region (USD Billions) [21]

[21]. United States, with greater number of hospitals, rely on AI for maintaining patient data and care. Due to increment in population and decrement in healthcare staff, demand of AI technology for treatment and diagnosis is obvious. Artificial intelligence has significantly helped in improving the different factors in healthcare industry. Using artificial intelligence, the cost of treatment is reduced, and quality of life is improved [21]. Results and decision taken from the patient data using AI have good response. AI provided better prediction and improved outcome. Besides good scaling of AI, its faster algorithm implementation process is lacking somehow. This is due to only patient security and for better tested accuracy.

Deep learning applications in medical domain have led to another drastic change that helps to automatically detect cancer cells. It has been proved by researcher's teams at University of California, Los Angeles, who built an advanced microscope that yields a high-dimensional data set used to train a deep learning application to accurately identify cancer cells [22]. Also, the estimation of stress can be analysed using a trained deep neural network [23].

1.2.4 Internet of Things

Internet of Things is another promising technology which has found applications in every field of life. This technology has found tremendous growth in health industry as it can interact with the machines like pacemakers and glucose monitoring systems implanted in the patient body. The miniaturization and tiny sensor nodes can detect diseases by regularly monitoring a patient and help paralysed persons to send signal through brain to different parts of the body. IoT systems are also helpful to the healthcare industry in terms of smart building technologies, remote monitoring and providing cost-effective solution to the patients [24].

The authors in [25] have presented a unique model as shown in Fig. 1.5 for IoT-enabled healthcare systems which can be applied to generalized systems as well as in specialized systems. The accessibility of medical data using Cloud has been stressed upon for future healthcare systems due to high processing speed despite limited resources provided by body sensors. But due to security and privacy factors associated while saving medical data on cloud, it has been stressed upon that IoT-based healthcare systems are best approach for immediate application.

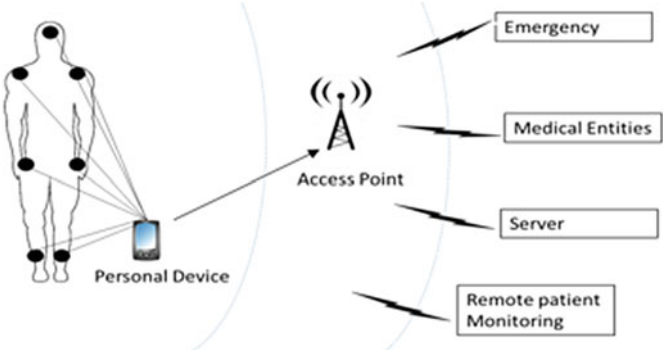


Fig. 1.5 Simple model for IoT-based healthcare system

The Internet of Things (IoT) is a concept that reflects a "connected set of anyone, anything, anytime, anyplace, any service, and any network". The basic characteristics of the IoT systems are mentioned as a real-time wireless solution in a global environment and monitoring of remote patients through tracking objects. But, IoT has helped in breaking geographic barriers, providing rapid clinical responses, medical consultation and communication links of medical images.

No doubt, the efficiency and effectiveness in the patient healthcare have improved manifolds. At the same time, the healthcare sector must have the ability to handle the increase in demand as latest technologies require more bandwidth that tangle the performance of the system. In such situation, the healthcare organizations must stay in contact with various network service providers to provide best possible bandwidth requirements for accurate diagnose to the patient. Another solution is the use of software-defined wide area networks (SD-WAN) that will support and manage multiple organizations on a cloud-based framework [26].

In 2018, IoT market size in healthcare industry was 147.1 billion and is anticipated to witness a Compound Annual Growth Rate (CAGR) of 19.9% over the forecast period as shown in Fig. 1.6 [27]. The abrupt increase in wearable technology, invest-

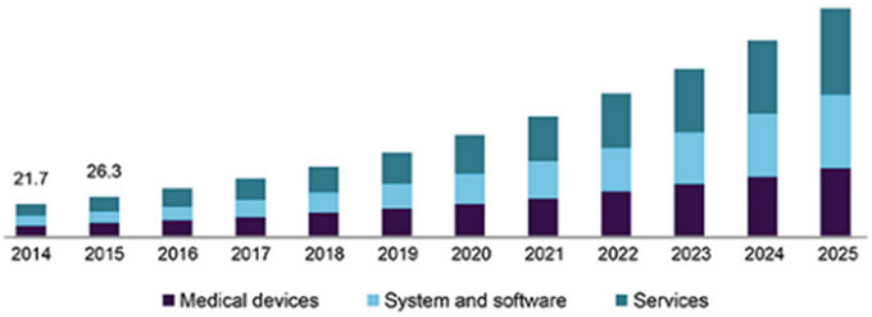


Fig. 1.6 IoT market in US (USD Billions) [27]

ments for establishing digital technologies in healthcare points and the emergence of connected care are the key factors boosting industry growth. Also, advancements in the technology and growing population day by day coupled with chronic conditions are also positively impacting the market expansion.

The major limitations that IoT-enabled health system to suffer from are computational limitations, memory limitations and energy limitations. Such devices face challenges related to the central processing unit (CPU), which is not very powerful in terms of its speed. In addition to this, these devices also face issues related to computationally expensive operations and low memory. Such devices are activated using an embedded operating system, system software and an application binary. A general IoT-based healthcare system includes various physiological sensors to measure body parameters and these sensors can conserve energy by switching on the power-saving mode when no sensor reading needs to be captured.

1.2.5 3D Printing

It has been predicted that the 3D printing technology in the healthcare field will grow by 17.7% and reach $2.5bn by 2026 in forecast 2021–2026 [28]. The major applications of 3D printing in medical fields are bioprinting of tissues and organoids, 3D-printed models for surgery, surgical tools and custom-made prosthetics [29].

The authors in [30] presented the way to explore 3D printing in personalized healthcare. The major challenges like longer printing time, accuracy in medical outcomes and applications of 3D printing have also been highlighted. It has been stated that drug 3D printing has also been the revolutionary word in the sustainable healthcare. The 3D printing technology helped in solving the issues related to complexity, sustainability in healthcare and in exploring possibilities offered by technology in healthcare variety management. An intelligent use of 3D printing technology in producing personalized medications and an industrial technological revolution of compounding pharmacy. The authors in [31] have presented a state of the art of 3D printing technology implementation in healthcare industry. 3D printing is also called as additive manufacturing in which the different materials are placed one over the other in layered architecture form. This technology has gained much recognition due to decline in the cost of 3D printers and enhanced printing accuracy and speed.

Figure 1.7 shows the market size of 3D printing technology in healthcare industry. It is expected to be 3700 million by 2025 and USD 1150 in 2021 showing more than 18% CAGR during the period [27]. Due to advancements in technology, the present-day hybrid 3D printing helps in more accurate diagnosis and surgery planning as compared to earlier available 3D printing technology which supported image modality only. The success of personalized scaffolds using different techniques has also grown with the advent of 3D printing technology. It has been popularly employed in making medical devices, human organs and for improving quality of medical treatment.

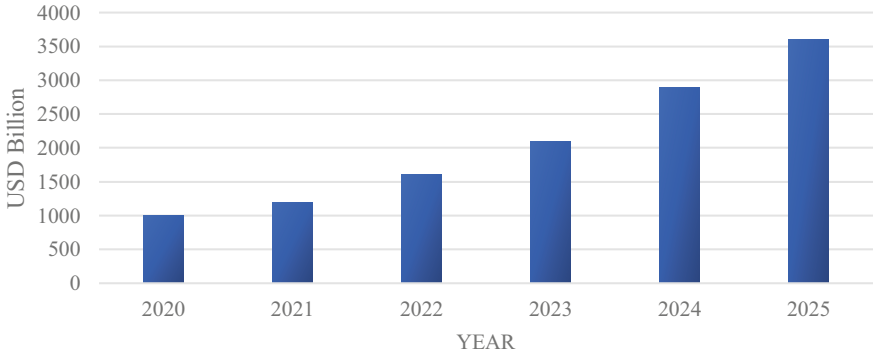


Fig. 1.7 Global healthcare 3D printing market size (USD Billions) [28]

Italy had developed 3D-printed respiratory valves for the Corona Virus Disease (COVID-19) which are effective for patients in their country [32]. Researchers in different countries have also developed 3D-printed artificial limbs, chest, abdomen, head, neck and other body parts required for anthroponomy training in medicals hospitals and institutes.

The authors in [33] have presented the various technologies, materials used in the 3D printing for different medical applications. The chapter also highlighted the advantages and disadvantages of each of these technologies. It has been concluded that 3D printing technology has provided the way to treat more patients with less time under direct supervision of medical experts.

Many authors have also worked upon the development of data fusion techniques to provide smart healthcare. The smooth combination of different data sets to make it representable, accurate and consistent for use is called data fusion. Data fusion techniques enhance the accuracy of individual data from different sources by integrating these data into single information. The authors in [34] have also proposed a technique using Complex Event Processing (CEP) technology which not only supports time-critical applications but also storage-limited IoT devices. A specialized beamforming antenna can also be used for transmission of data from various healthcare devices [35].

1.3 Intelligent Systems in Healthcare Industry

This section contributes towards the intelligent systems developed and implemented in healthcare industry based on the healthcare technologies. The augmented and virtual reality technology-based products have been developed by different countries. OxfordVR has been designed by United Kingdom based on virtual reality which helped the patients suffering from mental disorders and fears [36]. The system is

efficient enough to reduce the mental disorders by an average of 68% within few hours of treatment.

Israel has developed Augmedics [36] system which helped in providing guidance during surgical operations. The surgeons can see through patient's skin or tissue just like X-ray vision. Precision VR [36] has been developed for various neurosurgical procedures for pre-operation planning. A complete process based on VR scenario can be seen before the operation. Several smart systems based on blockchain technology have been developed. Some of them are Ethereum [37], Ripple [38] and Hyperledger Fabric [39]. These systems help in providing research and innovations through blockchain technology for commercialized solutions.

AI and robotics have given us Xenex Robot Disinfecting system [40] which has proved to control the infections acquired by the patients and reduced the risk to 70 percent. Another contribution in this domain is Riba [41], which helps in the movement of bed-ridden patients. The robot helps the patients to stand, sit and move in or out of bed to avoid any kind of bed sores.

3D printing helped in regeneration of tissues and skin damaged by any kind of burns, etc. The patients can get a new life through organ implant [42]. The orthopaedic implants developed in early 2000s got approval from Food and Drug Administration (FDA) to create organ implants. Germany developed surgical tool made from stainless steel alloy using 3D printing technique to remove the hip cups [43]. Copper3D NanoHack [44] mask is provided with an improved design and tested face mask to protect from COVID-19 virus. This mask has an input port for the flow of air and reusable filters hold by the screws. The design of the mask has also been modified with updated requirements.

Virtual wards and hospitals operated in many countries like Australia, UK and UAE during COVID-19 pandemic to provide remote monitoring and medical consultancy to the patients affected with chronic diseases and recovering from the COVID-19 virus infection [45]. A specialized thermometer, also called as connected inhaler, has been developed for the medical practitioners which provides a proactive approach to get in contact with the patients regarding appropriate use of inhaler and proper medication being taken.

1.4 Data Analytics and Challenges in the Healthcare

Data analytics is related to analysing present and past industrial data to improve and manage the industrial growth. In healthcare industry, healthcare analytics help in analysing the present technological trends, challenges in implementation and maintaining medical record. The analysis will improve the quality of service both for the patients and medical professionals. The healthcare industry will gain in terms of research and development, patient behaviour and their sentiment analysis, claims on medical cost, patient care and efforts, improving the conventional procedure [46]. Not only this, but data analytics in healthcare will also greatly help in detecting and

predicting frauds related to false insurance claims, unnecessary diagnose by hospitals and forged bills. The study and analysis of cyber-attack on medical data can also be predicted and cybersecurity can be provided as a precautionary measure.

1.4.1 Challenges in Healthcare Analytics

- The diverse nature of medical data makes it very difficult to compare, analyse and share the data. This is due to the use of vast varieties of medical software and their data formats. The interoperability of such record with different rules and regulations is very challenging.
- It was estimated by medical facility providers that 2314 exabytes of newer data will be generated by 2020 rising from the previous 153 exabytes of data in 2013 [46]. The storage of such a massive data is adding to the rise in cost of storage.
- Measurement of healthcare data need specialized data analysts which is again a factor that will add to the cost depending on the volume of the data.
- Other aspect of healthcare analytics is managing data related to donations, charity and grants. The optimization of this data is based on donor's previous contribution and retention.
- No doubt, with the advent of latest software tools and applications, the advantages of data analytics can never be questioned.

1.5 Conclusion and Future Scope

The healthcare ecosystem is becoming a technological hub to provide quality healthcare facilities at lower cost. This chapter directed about the various booming technologies like augmented reality, AI, robotics, blockchain technology, IoT and 3D printing in the healthcare industry. The intelligent systems developed based on various technologies have been discussed in separate section. The major aim, applications and challenges in implementing these technologies and impact of data analytics in healthcare industry are the key highlights of the chapter. It is concluded that the presented technologies can provide a proactive approach to many medical issues like remote monitoring, surgical operations, customized prosthetics during normal and emergency circumstances and have been globally accepted by researchers and medical practitioners for implementation. It is further concluded that data analytics in conjunction with the healthcare technologies will be beneficial for the mankind. Overall, there is a great interest and demand for these technologies to bring revolutionary changes in the healthcare industry.

Acknowledgements The authors would like to extend their sincere thanks and gratitude to the I.K.G. Punjab Technical University, Kapurthala for providing the opportunity to carry out the present study.

References

1. Movassaghi S, Abolhasan M, Lipman J, Smith D, Jamalipour A (2014) Wireless body area networks: a survey. IEEE Commun Surv Tutor 16(3):1658–1686
2. Thuemmler C, Bai C (2017) Health 4.0: how virtualization and big data are revolutionizing healthcare. Springer, Cham, Basel, Switzerland
3. Ravì D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, Yang GZ (2017) Deep learning for health informatics. IEEE J Biomed Health Inf 21(1):4–21. https://doi.org/10.1109/JBHI.2016.2636665.
4. Elhoseny M, Ramírez-González G, Abu-Elnasr OM, Shawkat SA, Arunkumar N, Farouk A (2018) Secure medical data transmission model for IoT-based healthcare systems. IEEE Access 6(c):20596–608. https://doi.org/10.1109/ACCESS.2018.2817615
5. Khezr S, Moniruzzaman M, Yassine A, Benlamri R (2019) Blockchain technology in healthcare: a comprehensive review and directions for future research. Appl Sci (Switzerland), 9(9):1–28. https://doi.org/10.3390/app9091736
6. O'Dowd E (2017) Practical applications of augmented reality in healthcare. HIT Infrastructure, May 17, 2017.
7. Sheng Bin, Saleha Masood, Younhyun Jung (2020) In biomedical engineering, biomedical information technology (Second Edition). Academic Press, pp 673–686
8. Orciuoli F, Orciuoli FJ, Peduto A (2020) A mobile clinical DSS based on augmented reality and deep learning for the home cares of patients afflicted by bedsores. Proc Comput Sci 175:181–188. https://doi.org/10.1016/j.procs.2020.07.028
9. Escalada-Hernández P, Ruiz NS, San Martín-Rodríguez L (2019) Design and evaluation of a prototype of augmented reality applied to medical devices. Int J Med Inf 128:87–92. https://doi.org/10.1016/j.ijmedinf.2019.05.004.
10. Negrillo-Cárdenas J, Jiménez-Pérez JR, Feito FR (2020) The role of virtual and augmented reality in orthopedic trauma surgery: from diagnosis to rehabilitation. Comput Methods Programs Biomed 191. https://doi.org/10.1016/j.cmpb.2020.105407.
11. Vles MD, Terng NCO, Zijlstra K, Mureau MAM, Corten EML (2020) Virtual and augmented reality for preoperative planning in plastic surgical procedures: a systematic review. J Plastic Reconstr Aesthet Surg. https://doi.org/10.1016/j.bjps.2020. 05.081.
12. Blockchain in healthcare: Patient benefits and more, Webpage, IBM. https://www.ibm.com/blogs/blockchain/2017/10/blockchain-inhealthcare-patient-benefits-and-more/
13. Blockchain: Opportunities for health care, Webpage, Deloitte. https://www2.deloitte.com/us/en/pages/publicsector/articles/blockchain-opportunities-for-health-care.html
14. https://bisresearch.com/industry-report/global-blockchain-in-healthcare-market-2025.html
15. Zhang P, Schmidt DC, White J, Lenz G (2018) Blockchain technology use cases in healthcare. Adv Comput Elsevier 111:1–41
16. Dimitrov DV (2019) Blockchain applications for healthcare data management. Healthcare Inf Res 25:51–56
17. Radanović I, Likić R (2018) Opportunities for use of blockchain technology in medicine. Appl Health Econ Health Policy 16(5):583–590
18. No longer science fiction, AI and robotics are transforming healthcare, PwC webpage. https://www.pwc.com/gx/en/industries/healthcare/publications/airobotics-new-health/transforming-healthcare.html
19. Robotics in Healthcare—Get Ready! The Medical Futurist. http://medicalfuturist.com/robotics-healthcare/
20. https://cnsi.ucla.edu/blog/2016/04/13/
21. https://www.reportsanddata.com/press-release/global-artificial-intelligence-in-healthcare-market. Accessed 20 Aug 2021
22. Kinjo Y, Sakuma Y, Kobayashi T, Sugimoto C, Kohno R (2019) Patient stress estimation for using deep learning with RRI data sensed by WBAN. In: International Symposium on Medical Information and Communication Technology, ISMICT 2019 May 1–4.https://doi.org/10.1109/ISMICT.2019.8743842

23. Dautov R, Distefano S, Buyya R (2019) Hierarchical data fusion for smart healthcare. J Big Data 6(19)
24. Baker SB, Xiang W, Atkinson I (2017) Internet of things for smart healthcare: technologies, challenges, and opportunities. IEEE Access 5:26521–26544. https://doi.org/10.1109/ACCESS.2017.2775180
25. Carnaz G, Nogueira VB (2019) An overview of IoT and healthcare. In: International conference of the IEEE engineering in medicine and biology society, pp 1455–1458
26. Rajput DS, Gour R (2016) An IoT framework for healthcare monitoring systems. Int J Comput Sci Inf Secur 14(5):451–455
27. https://www.grandviewresearch.com/
28. https://www.industryarc.com/Report/1268/3D-printing-in-healthcare-market-analysis.html. 3D Printing in Healthcare Market–Forecast (2021–2026)
29. Yan Q, Dong H, Su J, Han J, Song B, Wei Q, Shi Y (2018.) A review of 3D printing technology for medical applications. Engineering 4(5):729–742. https://doi.org/10.1016/j.eng.2018.07.021
30. Aimar A, Palermo A, Innocenti B (2019) The role of 3D printing in medical applications: a state of the art. Hindawi J Healthcare Eng. Article ID 5340616. https://doi.org/10.1155/2019/5340616
31. Aquino RP, Barile S, Grasso A, Saviano M (2018) Envisioning smart and sustainable healthcare: 3D printing technologies for personalized medication. Futures, 103:35–50. https://doi.org/10.1016/j.futures.2018.03.002
32. https://archinect.com/news/article/150190023
33. 3-D bioprinter to print human skin, press release, Science Daily, Jan 23, 2017. https://www.sciencedaily.com/releases/2017/01/170123090630.html
34. https://www.medicaldevice-network.com/features/3d-printing-in-the-medical-field-applications/ "3D-printing stats"
35. Singla P, Saxena J (2013) Modified optimum beamforming for smart antenna in WCDMA network. In: 2013 International conference on machine intelligence and research advancement, pp 155–161 (2013). https://doi.org/10.1109/ICMIRA.2013.37
36. https://hitconsultant.net/2020/06/29/augmented-reality-and-virtual-reality-companies-to-watch/#.YR-MeY4zbIU. Accessed 20 Aug 2021
37. Ethereum. https://www.ethereum.org/. Accessed 20 Aug 2021
38. Ripple. https://ripple.com/. Accessed 20 Aug 2021
39. Hyperledger. https://www.hyperledger.org/. Accessed 20 Aug 2021
40. https://news.crunchbase.com/
41. Mukai T, Hirano S, Nakashima H, Kato Y, Sakaida Y, Guo S, Hosoe S (2010) Development of a nursing-care assistant robot RIBA that can lift a human in its arms. In: Proceedings of International conference on Intelligent Robots and Systems (IROS), pp 5996–6001
42. 3-D bioprinter to print human skin, Press Release, Science Daily, January 23, 2017
43. https://amfg.ai/2019/08/30/3d-printing-in-healthcare-where-are-we-in-2019/
44. About NanoHack. https://copper3d.com/hackthepandemic/.
45. Wang L, Alexander, CA Big data analytics in healthcare systems. Int J Math Eng Manag Sci 4(1):17–26. https://doi.org/10.33889/IJMEMS.2019.4.1-002
46. https://www.statista.com/statistics/1037970/global-healthcare-data-volume/

Chapter 2
Fruit Fly Damage control—A Comprehensive Solution for Sustainable Development of Gherkin Industry

Sindhu P. Menon, Ramya Shetty, Afroz, Afsa, and Kavyashree

Abstract Fruit flies are the common pests found in fruits and vegetables. They invade into the fruit causing severe damage to farmers. The fruits invaded by fruit flies are rejected in the export market. The performance of Deep Nets is very high but the negative side of that is the problems associated with training the data. Associated with Deep nNets is the complexity of the network and the vanishing gradient problem. These are overcome using Transfer Learning. Here, pretrained networks are used. These take lesser time since they are already pretrained and thus the time taken also reduces. This also has the advantage of delivering increased accurate results when compared with the traditional neural networks. It has been proven that Deep Networks give very efficient and accurate results when classifying images. In this project, we propose to train the gherkins data set using the ResNets to identify fruit fly-infected gherkins. We have used Transfer Learning approach to train the classifier. We have also built a Flask web framework-based web application to help users upload gherkin images and get back predictions. In this chapter, we discuss a method which could help gherkin industries and farmers gain higher profits in their respective businesses for long-term sustainable development.

2.1 Introduction

Gherkin (Cucumis anguria), also called as bur gherkin or West Indian gherkin, is grown all over the world but more in locations having warm climate like southern Africa. They can be consumed in various forms like pickled in vinegar or raw or cooked. The pickled form has lot of demand in the commercial market. The small fruits have a greater demand in the export market. Hence, a lot of companies start exploring options of growing these plants. But it is not long and hard cucumber that one would find in grocery store and would put in a salad. These are a specific kind of

S. P. Menon (✉) · R. Shetty · Afroz · Afsa · Kavyashree
Dayananda Sagar University, Bengaluru, Karnataka, India
e-mail: sindhup-cse@dsu.edu.in

V. Bali et al. (eds.), *Decision Analytics for Sustainable Development in Smart Society 5.0*,
Asset Analytics, https://doi.org/10.1007/978-981-19-1689-2_2

Fig. 2.1 Fresh gherkins

cucumber. 'Gherkin' word originally comes from the Dutch gherkin which meaning 'small pickled cucumber'.

The gherkin fruit (Fig. 2.1) is similar to cucumber in terms of nutritional value. Cucumber and gherkins belong to the same species 'Cucumis sativus' but they have different gene structure [1]. Gherkin plant can be grown in any seasons throughout the year. This provides employment opportunities to both landless labourers in the village/rural areas and landholders. Palmately lobed leaves represent the structure of a gherkin leaf. The fruits have small thorn-like structures all over the body and the plant can grow up to 8 ft. in height [1, 2].

The name gherkin normally refers to the pickled cucumber which is very delicious. The fruit is plucked when the size of the same reaches 4–8 cms. Then they are kept in brine solution or vinegar and exported or bottled. Bottled gherkins have a great demand in the export market [1].

There are two forms in which these fruits are exported:

(1) Provisionally preserved (i.e. preserved in vinegar, acetic acid)

Provisionally preserved: In this method, the gherkins are exported in the bulk form which is a large high-density polyethylene drums, the capacity of which will be 220 L. Once this is completed, as per the requirements, the customers pack them into smaller jars or bottles [1].

(2) Preserved in vinegar:

The ready-to-consume ones are preserved in vinegar and later packed in jars or cans as per the need (Fig. 2.2).

2.2 Types of Gherkins

The biggest difference between dill (Fig. 2.3) and sour pickle (Fig. 2.4) gherkins is that the former includes fresh dill weed (and occasionally dill seeds or oil) for a boost of herby flavour. The biggest problem affecting the gherkin crops is fruit fly which is massive threat to gherkin industry. Fruit fly is the most destructive pest to gherkin

Fig. 2.2 Gherkin farming

Fig. 2.3 Dill pickle gherkins

Fig. 2.4 Sour gherkins

through the country. Adult fly punctures the early tender fruit tissues and lays its eggs and the eggs hatch into maggots. The infection of fruit fly in gherkins often reaches cent percent which leads to complete loss/destruction of the crop. Unfortunately, every year the fruit fly damages a large chunk of gherkins. Almost more than 50% of the cultivated crop becomes unsuitable or unfit for human consumption due to the fruit fly infestation. Gherkin industry is facing a very dangerous threat due to the fruit fly infestation.

In many countries across the globe like Europe and US, these fruits get completely rejected due to this infestation and the companies face huge claims for the same. At present, identifying these fruit flies is done manually. This is a very tedious task and

is prone to errors as humans get tired since it is very labour intensive (Fig. 2.2). In this method, the fruit flies are identified visually leading to errors, since the prick is very minute that it may not be visible to the human eye.

One well-known species of the fruit fly belonging to the family Drosophilae is the Drosophila melanogaster [3]. Studies have shown that when compared with other insects, the male insects of Drosophila melanogaster have the longest sperm cells on Earth. The structure of the eggs shows respiratory filaments present, which permit oxygen supply to reach the embryo. Also known as the fruit fly, its larvae feed on microorganisms which surface on decaying matter. In addition to this, it gets its nutrition from the vegetable matter too. A number of parameters contribute to its growth like the crowding, the temperature and the breeding substrate. The time needed for it to develop into a fully grown insect varies between 7 and 60 days [3].

The time at which the eggs are laid also contributes to the development of the insect. The eggs laid during the night have a higher chance of survival when compared to those laid during the day. During the day, the insects emit more larvae. They survive for 35–45 days. This nature of the fruit fly wherein the time of laying the egg determines the survival of the egg has a great reproductive advantage.

Hygiene plays a very important role in the reproduction of the fruit fly. Unhygienic places where food gets rotten and fermented are the areas where it can breed more. It could be homes, restaurants or any other place. The adult fruit fly has red eyes with a tanned body. The length is about 1/8th of an inch. Moist surfaces are areas where they lay eggs to breed. At any instance, the fruit fly can lay up to 500 eggs. This shows the high reproductivity of the fruit fly. The growth from egg to adult takes approximately 6–7 days. Since the larvae feed on the surface, the ripened or rotten area could be cut off to further spread of the developing larvae. These flies can also adulterate the food through bacteria and other organisms which causes diseases [4, 5].

Ripened fruits attract these flies as in Fig. 2.5 (gherkins). The other areas where they can breed are sites of garbage disposal, mops, empty cans and trash. They can enter homes through windows and doors after infecting the ripened gherkins. A brief survey of the work carried out in these areas is discussed in the next section.

Fig. 2.5 Fruitfly pricking the gherkin

2.3 Literature Survey

Wazir et al. [6] opined that weather parameters had a great impact on the survival of fruit fly Bactrocera cucurbitae (Coquillett). The traps used to determine the incidence of the fruit fly for the year 2017–2018 and further effect of weather parameters on the population dynamics of the fruit fly Bactrocera cucurbitae (Coquillett) were obtained. The trap count of the fruit fly was recorded at the weekly interval by emptying the traps and constituents were refilled. The result on seasonal incidence of the fruit fly was recorded on summer squash. The data was then recorded at weekly intervals. The maximum and minimum mean temperature along with humidity were calculated and recorded. Along with this the rainfall was also calculated and the analysis was presented in the form of correlation and regression [7].

Manoj et al. [8] were of the opinion that one of the most devastating pests was the cucurbit fruit fly, Bactrocera cucurbitae (Coquillett). Once it punctures the fruit, a liquid oozes out from it. This later becomes deeper and turns into a brown deposit. At times, it shows a false positive. The fruit looks punctured but there won't be any egg inside it. This leads to rejection in the export market. It was observed that out of the various shapes of the traps, the fruit flies were attracted more by the cylindrical, followed by spherical and conical shapes. Among the foods, it was found that fruit flies are more attracted to palm juice (padaneer) of 200 ml. Luring fruit flies using traps and baits played a major role in the management of fruit flies.

Fruits of bitter gourd are good source of vitamins, Gaddanakeri and Rolania [9] minerals, carbohydrate and proteins and have the richest nutrition value among cucurbits. Infested bitter gourd fruits were collected from the field. These infested fruits were kept in the tray consisting of 5–6-cm-thick layer of sieved soil, in insect-rearing cages. Subsequently, the rotted fruits were removed after 5–7 days the soil was sieved to collect pupa. This culture was utilized in carrying out biological studies in which maggot period, pre-pupal period, pupal period and adult longevity were recorded. The oviposition period, fecundity and sex ratio were also recorded. The freshly laid eggs were elliptical, pure white in colour, nearly flat on the ventral surface and slightly curved on other side [10].

During the year 2015, a study on population magnitude of melon fruit fly Bactrocera cucurbitae (Coq.) was undertaken at Hyderabad [11]. The study involved around three different crops, i.e. bitter gourd (Momordica charantia), bottle gourd (Lagenaria siceraria) and ridge gourd (Luffa acutangula) and were carried out at two different farms using baited traps. The traps are for males. The male lure baited traps were stacked at a height of 3 m and were restocked once every 15 days. L. siceraria recorded the lowest population of 104.725 and L. acutangula recorded the highest population of 124.86 at both the sites [12].

Doorenweerd et al. [13] concluded that it is very difficult task to rightly apply the scientific names. At times, the results could be misinterpreted, which could lead to the emergence of new species of pests. One of the most challenging tasks is to bifurcate the various groups when it contains a large number of organisms and a variety of species. The most commonly found species is the tribe Dacini which constitutes

about 20% of the entire population. Around ten percent of the pests are those that infect the fruits and vegetables [14].

Benjamin et al. [15] through their studies have come up with a conclusion that in the sub-Saharan Africa, the horticulture industry is facing a huge threat due to pests and the various diseases. Majority of the fruit fly population feeds on wild fruits and crops, thereby infecting them and causing huge damage. Most common among them are the Tephritid fruit flies. These along with their quarantine status cause huge damage to the crops. Some of the pests which are of economic importance are Ceratitis, Trirhithrum, Bactrocera and Dacus [16, 17].

Lu and Ariana [18] were of the opinion that during the process of pickling cucumber for production, fruit fly infestation could be a huge threat.

In many countries across the globe like Europe and US, these fruits get completely rejected due to this infestation and the companies face huge claims for the same. At present, identifying these fruit flies is done manually. This is a very tedious task and is prone to errors as humans get tired since it is very labour intensive (Fig. 2.2). In this method, the fruit flies are identified visually leading to errors, since the prick is very minute that it may not be visible to the human eye. Various types of images were used. Some of them were hyperspectral reflectance images and transmittance images. The former was of 450–740 nm and the latter was of 740–1000 nm. Mean of the images were taken and they were correlated with the size of the fruit using the diameter correction equation. Partial least square discriminant analysis for the infected and normal gherkins was performed. The accuracies for the three size classes and the mixed class were found to be between 82 and 88% for the reflectance mode but was higher for the transmittance mode (between 88 and 93%). The results did not show huge variations when the two modes were merged. The hyperspectral system showed better results, thereby proving the fact that hyperspectral images can be used for identification of infected gherkins [19].

Adedeji et al. [20] came up with the fact that almost all of the applications depend on the reflectance-based data acquisition. The reflectance-based hyperspectral imaging has its own benefits such as the detection of defects on surface of food commodities, which were caused by the mechanical injuries, fungal contamination or chemical degradation. However, the internal structure of these items cannot be identified clearly using hyperspectral reflectance imaging [21]. In this study, a hyperspectral transmittance imaging system is proposed to detect the internally contaminated food items non-invasively. The system is designed on the grid-based conveyor belt to get transmittance signals in the real time. 2×250 W halogen light bulbs are used to [22] illuminate the objects, which are positioned under the conveyor belt. The images are acquired by a line-scan hyperspectral camera with a spectrum range of approximately 400–1000 nm. The developed system is then tested on detecting Aspergillus niger-contaminated dried figs. Using transmittance signals, an automated classification system is developed. For classification, the data sets of 240 dried figs and their flattened counter parts are used. Feature selection is then done by forward feature selection up to 30 features. The features are classified by linear discriminant, support vector machine, logistic linear and discriminative restricted Boltzmann machine classifiers. Based on classification performed by two times twofold cross-validation, the

best result is obtained by linear discriminant classifier with 98.75% accuracy using only best six features among all the features [23].

Ekramirad [24] proposed that the damaged fruits and vegetables on the losses in agriculture sector. To sustain the quality of the crops, the requirement of the hour is to observe the crops and detect the infestation of insects. If they are detected early, then quarantine methods could improve the productivity. Hence, the current requirement is to develop cost-effective, quick and accurate system which could diagnose the fruit fly-infected fruits, thus resulting in a world advancement. The authors have also discussed the various strategies for detecting the damages caused by insects and their applications. It also throws light on various techniques like optical properties, sonic properties, thermal imaging, machine vision technique, magnetic resonance imaging (MRI), X-ray computed tomography and chemical chromatography which are conducted post harvest. In addition to this, the various challenges and limitations of these techniques were also highlighted [23].

2.4 Proposed System

We have come up with a system which could identify a fruit fly-infected gherkin whether or not it is a small prick. The info set for our proposed system may be a set of prime quality images taken employing a high-resolution camera. The subsequent step is to use various image processing techniques in order to induce a correct vision of the image if it is a small prick. The data set for our proposed system is a set of high-quality images taken using a high-resolution camera. The next step is to apply various image processing techniques in order to get a proper vision of the image.

Here we've collection of 1,348 of which 579 are good gherkin and 769 are infected gherkin. In this we have split the information set in the ratio 70:30 for training and test data. For collecting these images/data set, we have visited a company called Ken Agritech private limited which is mainly involved in the export of gherkins. The company has been into bulk and bottling. They have been in the industry for more than three decades. Established in the year 1997, Ken is one of the largest producers of gherkins (cornichons) in India [25]. It also supplies baby corn, jalapeno, cherry tomatoes, Capsicum annuum longum and many more. The greatest advantage with this crop is that it is not seasonal. It can be grown throughout the year. This organization is located in Hubli. It has growing areas around Hubli which have very good climate for growing gherkins. The company has an export market in Europe, USA, China, Russia and many other places. The core management team of Ken has great expertise in growing this crop [25].

The different stages of classification are discussed next. They include preprocessing and segmentation. These are achieved using Convolutional Neural Networks and ResNet Model. Activation function used is Softmax Classifier and Categorical Cross-Entropy Loss function is applied to reduce the loss.

2.4.1 Preprocessing

The next step is to preprocess the image. The raw images collected initially can't be used for processing as they may contain noise like lighting intensity and that they could also be of assorted shapes and sizes. These images have to be normalized and then used for classification. Here we use the concept of digital image processing. Various algorithms work on these images to transform them into a form suitable for processing. The main goal of digital image processing is to enhance the image features using computer models to improve the data by eliminating unwanted distortions in the form of noise. These images are 2D arrays ranging between 0 and 255. The various stages of image processing which will be applied are input and read the images, **Resize the images and later segment the resized images**.

Resize Images

This is the method of adjusting the image size without removing any part of it. Here the objective is to either reduce or increase the size of the file. This can be achieved by taking a large picture and then reducing the size. But a common problem which occurs when reducing the size is that the originality is lost. This can be achieved here by reducing or increasing the number of pixels, keeping the resolution intact. Certain images captured are of varying sizes. Hence, it is needed to establish a base size for our algorithm [26].

Resized (100, 100, 3)-(height, weight, RGB)

RGB: Here RGB represents the Red, Blue and Green colour. The pixels contain three integers whose numbers range from 0 to 255 [27].

2.4.2 Segmentation

Image segmentation is one of the most important phases of any image processing project. The objective here is to partition the image into multiple portions, where each portion is a set of pixels. The reason behind doing this is to simplify the image so that it can be easily analysed. During this process, the boundaries of the image like curves, lines, etc. will be identified. Every pixel in the images is labelled, so that pixels with the same image can be grouped together as they share similar characteristics as shown in Fig. 2.6. After this process, we get a set of images with similar characteristics like texture, colour or intensity.

Assume that 'X' is a set of images to be classified. Let 'Y' be a subset of 'X' which contains images from the set 'X'. Repeat steps 1–6 for each image in 'Y'.

1. One of the traditional filters is used to segment the input image. This is the first stage of the preprocessing step. Now the segmented image is the output.
2. Set a threshold 'T' for accuracy a. Next step is to test the model for the set accuracy. Assume that the accuracy is measured in the range of 0 to 1. if a $\geq$ T, go to Step 4. This is done because we have reached the target set else segment the

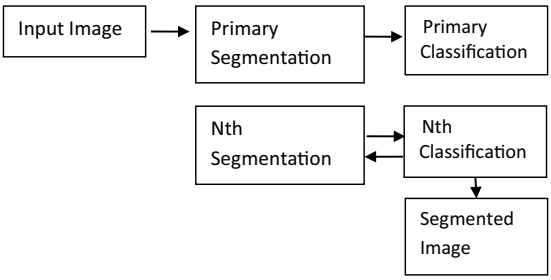


Fig. 2.6 Image Segmentation

image again using any other segmentation techniques. For instance, if technique 1 was used in the first case, then any other technique other than technique 1 must be used in the next iteration.

3. Use this segmented image and iterate the second step.
4. Observe and record the technique which resulted in better classification accuracy with the input image. For each input image in 'Y', let 'Z' denote the set which contains all the methods which resulted in good accuracy.

The sample images are represented as Y = {y1, y2, y3 …. yn} and the segmented images are represented as Z = {z1, z2, z3 …. zn} for all images in 'X'.

To get the image, we follow the procedure given below. The possibility of elements being repeated is more. Hence, we find the average of set Z called as 'Z-Average'. The segmentation technique which is present in many of the images of the data set will be seen in Z-Average. This technique is now used to segment the image in the set X [28–30].

This can be summarized as follows.

All the images to be segmented for classification will be represented by the set X. For each image in this set, various segmentation techniques are available. Each of these techniques is tested in sequence to look for the one which gives highest accuracy. All the segmentation techniques are averaged in the set Z-Average. This technique will yield better results than the individual methods (Fig. 2.7).

2.4.3 Feature Selection

This is the next stage for classification. Features are parameters on the basis of which learning can be done. More the number of features, better the learning. In neural networks, it is necessary to backpropagate the error so that features can be learnt effectively. Normally, autoencoders are used to reconstruct the errors. The difference between machine learning and deep learning is that in deep learning feature selection is done automatically. Features can be divided as low-level and high-level features. Low-level features like lines, edges, texture and colour-based features can be identified using image processing techniques whereas high-level features like object identification can be done using machine learning techniques. Initially, the

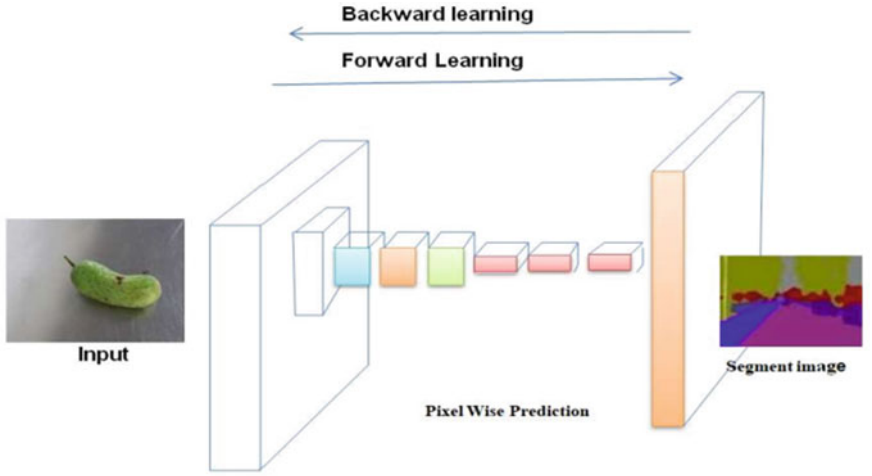


Fig. 2.7 Segmentation process

systems extract low-level features. Then high-level features are built upon this using these low-level features.

In this project, Convolution Neural Networks is used which identifies and extracts these high-level features automatically. The low-level features which can be considered here are colour, weight of the fruit and size of the fruit. Based on these features the model gets trained and predicts the infected gherkin.

2.4.4 Convolutional Neural Networks

A sequence of layers of neurons is the representation of a CNN shown in Fig. 2.8. Each layer consists of 'n' number of neurons. The main three parts of CNN are the convolution layer, pooling layer and the fully connected layer [31].

The fully connected layer is the final layer which takes its input from the pooling layer. These represent the last few layers of the network. The function of the pooling layer is to pool together all the features of the feature map and feed the same to the fully connected layer.

Convolution layer contains neurons connected along with the weights. These weights will be altered to backpropagate the error so that the network learns correctly. The image is converted into a pixel of values, normalized and then fed as input. This input is represented as a matrix of values. The weights are represented by 'w'. Based on the weights, summation of all the activations from input layer is computed [31].

$$(w_{0,0}a_0^{(0)}0) + w_{0.1}a_1^{(0)}1 + \ldots..w_{0,n}a_n^{(0)}1 \tag{2.1}$$

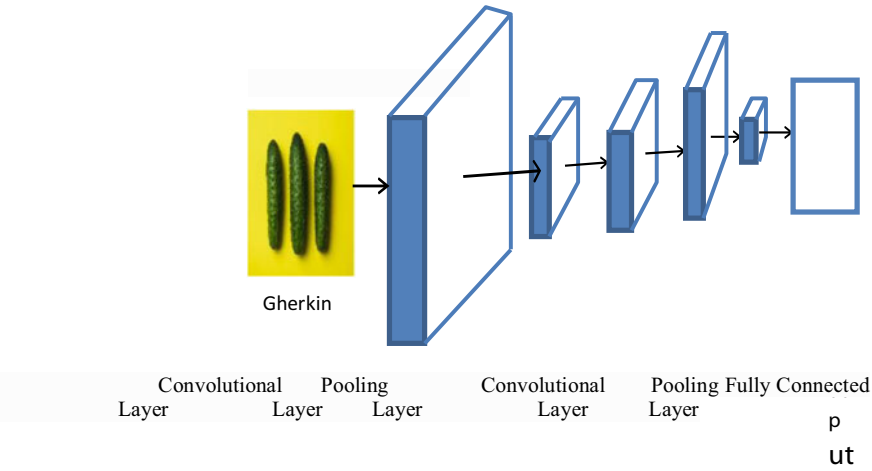


Fig. 2.8 CNN network

The activation function used is the sigmoidal function. This is used to squash values between 0 and 1. A suitable bias 'b' is also added. After the weights are altered, the activation function is applied and the output is fed to the next layer [31] (Fig. 2.9).

Neurons in the input layer are represented as

$$a_1^0 = \sigma(w_{0,0}a_0^{(0)}) + w_{0.1}a_1^{(0)} + \ldots..w_{0,n}a_n^{(0)} + b_0 \tag{2.2}$$

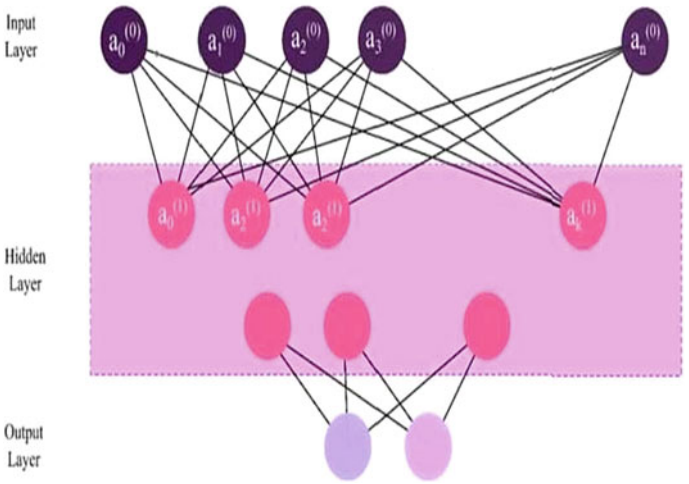


Fig. 2.9 Neurons and weights connected to represent a CNN

These values are then converted to a matrix form and fed to the activation function of the next layer [31]

$$\begin{bmatrix} w_{0,0} & \cdots & w_{0,n} \\ \vdots & \ddots & \vdots \\ w_{k,0} & \cdots & w_{k,n} \end{bmatrix} \begin{vmatrix} a_0^0 \\ \cdots \\ a_n^0 \end{vmatrix} + \begin{vmatrix} b_0^0 \\ \cdots \\ b_n^0 \end{vmatrix} = \begin{vmatrix} a_0^1 \\ \cdots \\ a_k^1 \end{vmatrix}$$

The activation maps are generated here by the computation of the scalar product.

The next step is dimensionality reduction. This task is accomplished by the pooling layer. It reduces the features using either average pooling or max-pooling techniques. In average pooling, the average of all pixel values is taken and thus the features are reduced whereas in max pooling, the maximum of all features are taken and thus the features are reduced as shown in Fig. 2.10. After this stage, all the features thus gathered are pooled and fed to the dense layer. This layer outputs the image and classifies it based on the loss.

Here, the proposed CNN model classifies as Good Gherkin or Infected Gherkin. The input layer has a input size of X*Y*3 which is equivalent to the size of the input image, where 'X' represents the height of the image and 'Y' represents the width of the image. Based on the features, the height varies from model to model. If the feature set contains height X = width Y = 100, next to the input layer, two convolutional layers will be formed with the matrix size of X * Y, and a kernel of 3*3. Dropout is assumed to be 0.5 and remains constant. Next a max pool layer of size 2*2 will reduce the features. After this, two more convolution layers are padded. The output of these layers is summed and fed into the max pool layer for dimensionality reduction. The activation function used in this is ReLU. The features obtained from this are reduced to a 1 vector and sent through a dense layer having a dropout of 0.5.

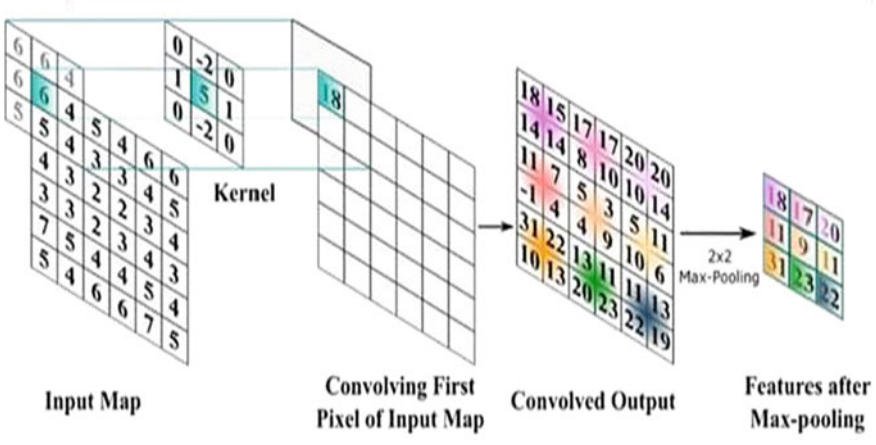


Fig. 2.10 Representation of convolving with the input and later application of max-pooling

The outputs of all these features are combined and normalized using the Softmax activation function to predict the final output. Adam optimizer is used to train the extracted features with a learning rate of 0.00001 with over 6 epochs and a batch size of 8 [20] [new].

6	4	5	4	6	6
5	4	3	3	4	5
4	3	2	2	3	4
3	3	2	3	4	3
7	5	4	4	5	4
5	4	6	6	7	5

Input 6*6

0	-2	0
1	5	1
0	-2	0

0-8+0
5+20+3
0-6+4

6x0+4x-2+5x0
5x1+4x5+3x1
4x0+3x-2+2x2

=

18

2.4.5 Resnet Model

The gherkins images will be trained through ResNet model. ResNet can be abbreviated as Residual Network. These were introduced by Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun in 2015 in their paper 'Deep Residual Learning for Image Recognition' [32]. These networks are well known as they construct like a pyramid skipping connections and jumping over the layers. Normally, these models will be constructed using many layers having nonlinearities and batch normalizations among them [3]. The weight vectors may be used to bypass the weights. In order to avoid the problem of vanishing gradient, skipping of weights is done so that the activation functions are reused till the learning completes in the adjacent layers. In the training phase, some of the layers above will be deactivated and the previous layer which was skipped will be amplified. This method will be successful when linearity is introduced in the intermediate layers.

When the process of skipping is performed, it deactivates a few layers in the initial phases. This will in turn reduce the speed as the number of layers have decreased. It also decreases the problem of vanishing gradient as there are few layers to pass through. As the network learns the features, it begins to activate the other layers too. After all the layers are activated, it starts learning faster [3].

2.4.6 Transfer Learning

Transfer learning is the method of storing knowledge gained from a particular problem and re-applying it to a new problem. For example, knowledge gained while learning to recognize cars could apply when trying to recognize trucks. In many real-world applications, it is expensive or impossible to recollect the needed training data and rebuild the models. Transfer learning is thus a way to create new models with very little data compared to the initial training.

2.4.7 Image Net

ImageNet7 is a large-scale image database. In 2009, ImageNet had over 3.2 million of images of over 5000 categories and has only expanded since. ImageNet is a popular database for collections of data sets to train neural networks and has since 2010 held the annual Large-Scale Visual Recognition Challenge8 which has participants from more than 50 institutions. The variety of images makes the database a great source to train a neural network. ImageNet is a common data set to have as a foundation when applying training techniques on a CNN because of variety of categories the database provides.

2.4.8 ResNet50

The problem associated with deep neural networks is the vanishing gradient. When more number of layers are added to the network, the efficiency of the network begins to degrade. As errors are propagated backwards and multiplied with the weights to make the model learn better, the gradient problem surfaces. Transfer Learning using ResNet50 Model is applied here to learn the networks. The last layers of this model are removed and replaced with the pretrained model. During the process of training, the weights are frozen and are not updated. The last layer will be used for classification of the gherkin by representing it as top = 'False'. The reason behind not training ResNet Layers is because we want the network to be trained from image net. Now the final layer is added using the Softmax Activation Function. This improves training and performance is achieved with less time.

Next step is to fine-tune the ResNet50 model. This can be done by freezing the first convolution blocks and unfreezing the last few layers. Using this many generic features will be learnt in the early stages. Initially, the layers of ResNet50 model were set to false and now they are trainable (Fig. 2.11).

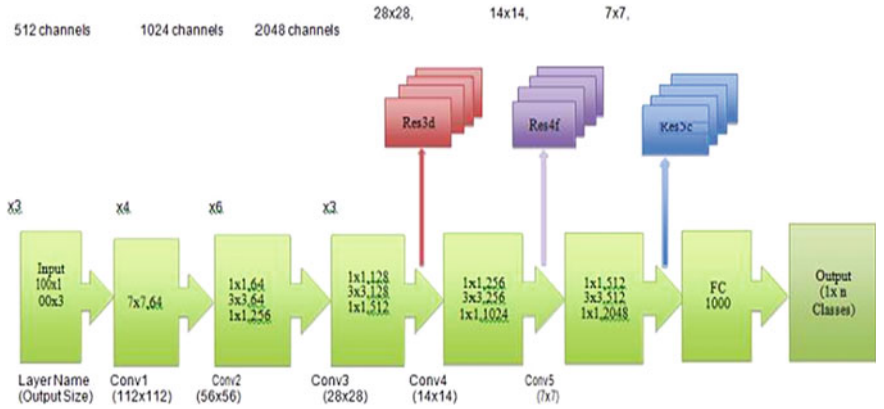


Fig. 2.11 ResNet model

2.4.9 Softmax

It is necessary to squash the output obtained between certain values to ease computation and to normalize its values. A number of activation functions exist. Softmax is one such function which squashes the value between 0 and 1 in the form of a vector and then adds up the elements to 1. Here the elements of a class can be represented using the probabilities of that class. It is applied to each value of s_i.

For a given class si, the Softmax function can be computed as follows:

$$f(s_i) = e^{Si} / \sum_{j}^{c} sj$$

where $\sum_{j}^{c} sj$ is the sum of all components and sj are the net values obtained for every class C.

2.4.9.1 Categorical Cross-Entropy Loss

In this work, loss is computed using the cross-entropy activation function as shown in Fig. 2.12. This is mainly used for multiclass classification. Since it has to classify as two separate classes, we use multiclass classification [33].

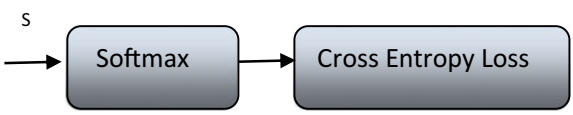


Fig. 2.12 Cross-entropy loss

When using multiclass classification, the labels keep only the positive term that is a good gherkin [34]. One element in the target will not be zero, hence we will have to discard that and then we write

$$\mathrm{CE} = -\log e^{\mathrm{Si}} / \sum_{\mathrm{j}}^{\mathrm{ct}} \log(\mathrm{f(s)}) \tag{2.3}$$

where Sp represents the positive class obtained using convolutional neural network.

f(s) = Softmax function;

CE = cross-entropy loss;

t = target vector, Prediction: good gherkin = 1 Bad gherkin = 0.

Good gherkin is indicated by a 1 value and 0 indicates a bad gherkin.

Softmax:

$g = e^1/e^1 + e^0$

$b = e^0/e^1 + e^0$

where e^1 = output of the neuron for good gherkin and e^0 = output of the neuron for bad gherkin.

$g = e^1/e^1 + e^0$

$g = 2.7183/(2.7183 + 1)$

$g = 0.7311$

$b = e^0/e^1 + e^0$

$b = 1/(2.7183 + 1)$

$b = 0.2689$

CE Loss = –(1*ln(g) + 0 *ln(b))

= –(1*ln(g) + 0)

= –(1*ln(0.7311)) = 0.3132

2.5 Results and Discussion

The model has been trained using 2863 images of good gherkins and 1974 images of infected gherkins. Total data of 4837 images was used for training. Testing on the Resnet model was done using 2887 images out of which 1674 images were of good

Table 2.1 Values of accuracy and loss during various epochs

Number of epochs	Accuracy	Loss
1	72.659	0.569
2	81.293	0.412
3	84.453	0.4899
4	89.210	0.395
5	90.891	0.328
6	94	0.1991
7	94.44	0.248
8	94.69	0.294
9	94.92	0.3728
10	95.32	0.3157
11	95.61	0.491
12	95.75	0.430

gherkins and 1213 images of infected gherkins. Table 2.1 displays the accuracy and loss during various epochs.

The accuracy obtained after 12 epochs is 95.75%. Figure 2.13 shows the accuracies for various epochs and it stabilizes between 10 and 12. Similarly, Fig. 2.15 shows the loss for various epochs. It begins at 0.5.

The graph in Fig. 2.14. Shows the accuracy versus loss scenario.

Based on the uploaded image you will get prediction as good or bad as shown in Figs. 2.16 and 2.17, if the uploaded gherkin image is good you will get prediction as good. These analytics help us to demystify the application of CNNs in prediction of gherkins.

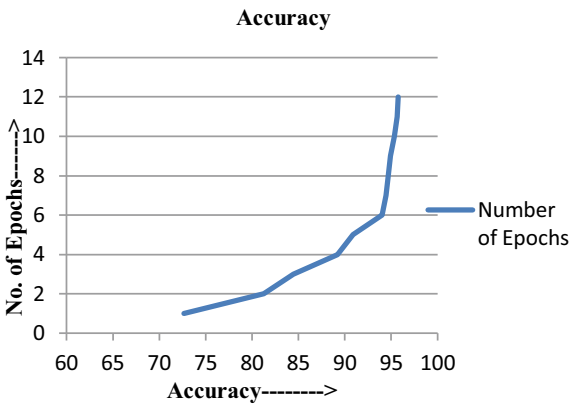


Fig. 2.13 Graph showing accuracy of prediction based on epochs

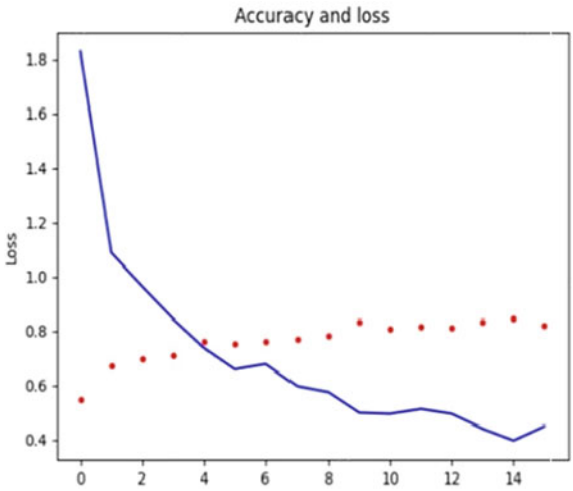


Fig. 2.14 Graph showing accuracy and loss 0.4

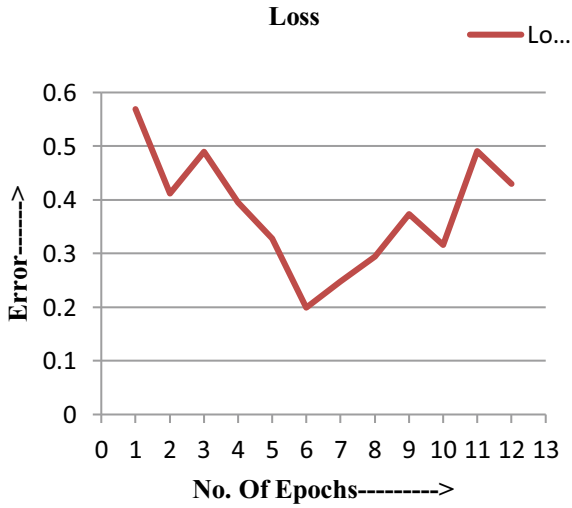


Fig. 2.15 Graph showing loss in prediction based on epochs

Fig. 2.16 Prediction of a good gherkin

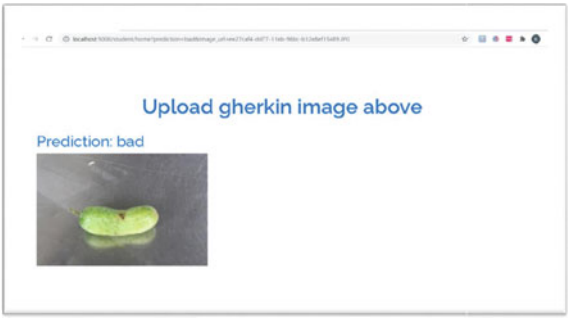


Fig. 2.17 Prediction of a fruit fly-infected gherkin

2.6 Conclusion and Future Work

2.6.1 Conclusion

This works aims at identifying the fruit fly-infected gherkins using images. This was developed using Transfer Learning and Deep Neural Network. The main motivation behind this work is to help farmers and industries reduce their loses and increase their profit margin. Farmers are the backbone of every nation and our effort is to help them in a small way. We developed a model which can predict whether the gherkin is good or infected. The results look very encouraging as an accuracy of 95.75% is obtained.

Reducing the rejections of gherkins is the need for farmers as it can cause major effect on farmers and industries.

Neural network contains several different algorithms and techniques for prediction. Different techniques and different algorithms of training can easily forecast the values.

2.6.2 Limitation

The drawback of this approach is that we could collect only a few thousand samples for training and testing. If this is increased, then the accuracy of the model may be increased. We have used only three features. More number of features may lead to better results.

2.6.3 Future Work

Overcoming the limitation becomes the future work. The same work could be experimented with a larger data set and the efficiencies could be compared. It could then be converted into a X-ray machine which could detect the infected gherkin as it passes through. This would be a great boon to the gherkin industry.

References

1. https://www.agrifarming.in/
2. https://www.britannica.com/
3. https://en.wikipedia.org/
4. https://www.yahoo.com/
5. http://www.howtogetridoffruitfliesfast.com/
6. Wazir ZA, Singh AK, Ramana N (2019) Seasonal incidence of fruit fly on Summer squash (Cucurbita pepo L.) and effect of weather parameters on population dynamics of fruit fly Bactrocera cucurbitae (Coquillett). J Entomol Zool Stud 7(5):167–170
7. Rahman MM, Sultana T, Uddin MM, Shahjahan M (2015) Impacts of temperature, relative humidity and total number of fruits on cucurbit fruit fly infestations under field condition. Progr Agric 26(1):1–5
8. Manoj AM, Sridharan S, Elango K, Muthukumar M (2018) use of traps and baits in the management of cucurbit fruit fly. Bactrocera cucurbitae (Coquillett)
9. Gaddanakeri S, Rolania K (2020) Biology and morphometrics of melon fruit fly, Bactrocera cucurbitae Coquillett on bitter gourd (Momordica charantia L.). J Entomol Zool Stud 8:994–998
10. Akter T, Jahan M, Bhuiyan MSI (2013) Biology of the Angoumois grain moth, Sitotroga cerealella (Oliver) on stored rice grain in laboratory condition. J Asiatic Soc Bangl Sci 39(1):61–67
11. Zain-Ul-Aabdin A, Baloch N, Khuhro NH, Qazi WA, Saeed NA (2017) Population densities of melon fruit fly Bactrocera cucurbitae (Coquillett) in vegetables agro-ecosystem in District Hyderabad, Sindh, Pakistan. Sarhad J Agric 33(2):331–337
12. http://www.researcherslinks.com/
13. Doorenweerd C, Leblanc L, Norrbom AL, San Jose M, Rubinoff D (2018) A global checklist of the 932 fruit fly species in the tribe Dacini (Diptera, Tephritidae). Zoo Keys 730:19
14. https://worldwidescience.org/
15. Benjamin BK, Kelvin BM, Kwame AN, Daniel OO (2012) Farmers' knowledge and perceptions of fruit fly pests and their management in Northern Ghana. Greener J Agric Sci 2:412–423
16. http://www.biopublisher.ca/
17. https://www.gjournals.org/
18. Lu R, Ariana DP (2013) Detection of fruit fly infestation in pickling cucumbers using a hyperspectral reflectance/transmittance imaging system. Postharvest Biol Technol 81:44–50
19. https://law.resource.org/
20. Adedeji AA, Ekramirad N, Rady A, Hamidisepehr A, Donohue KD, Villanueva RT, Parrish CA, Li M (2020) Non-destructive technologies for detecting insect infestation in fruits and vegetables under postharvest conditions: a critical review. Foods 9(3):927
21. Yan T, Xu W, Lin J, Duan L, Gao P, Zhang C, Lv X (2012) combining multi-dimensional convolutional neural network (CNN) with visualization method for detection of aphis gossypii glover infection in cotton leaves using hyperspectral imaging. Front Plant Sci 12:74
22. https://www.mdpi.com/
23. http://www.mafiadoc.org/

24. Ekramirad N, Adedeji AA, Alimardani R (2016) A review of non-destructive methods for detection of insect infestation in fruits and vegetables. Biosyst Agric Eng Faculty Publ
25. http://www.kenagritech.com/
26. http://www.blog.befunky.com/
27. http://www.towardsdatascience.com/
28. Chinajariyawong A, Kritsaneepaiboon S, Drew RAI (2003) Efficacy of protein bait sprays in controlling fruit flies (Diptera: Tephritidae) infesting angled luffa and bitter gourd in Thailand. Raffles Bulletin Zoology 51:7–16
29. (IAEA) International Atomic Energy Agency (2003) Trapping guidelines for area-wide fruit fly programmes. International Atomic Energy Agency, Vienna, Austria, p 47
30. Sarango VMG, Ekbom B, Ooi P (2009) Monitoring and pest control of fruit flies in Thailand: new knowledge for integrated pest management. Examensarbete 15:2–38
31. Rasheed J, Hameed AA, Djeddi C, Jamil A, Al-Turjman F (2021) A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images. Interdiscipl Sci Comput Life Sci 13(1):103–117
32. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
33. Submitted to Nanyang Technological University
34. Sanyal B, Mohapatra RK, Dash R (2021) Real-time embedded TSR using MDEffNet: An efficient compact DNN on varied multi-resolution images across multiple databases. Inter J Wavelets Multiresol Inform Proc 19(03):205

Chapter 3
Uprising of EVs: Charging the Future with Demystified Analytics and Sustainable Development

Pankaj Bhambri, Manuj Aggarwal, Harpuneet Singh, Amrit Pal Singh, and Sita Rani

Abstract Majority of countries are budging toward environmental-friendly congestion; EVs are procuring supplementary reputation beyond the world. Unlike orthodox vehicles that could be replenished without complete fueling setup at anticipated locations, some of the positive features of electric vehicles are that they can be refueled at several places like homes, offices, shopping malls, parking areas, etc. However, appropriate and suitable charging arrangement is required to be built at these locations. A widespread and definitely nearby charging network will be utmost important for huge adoption of electric vehicles. The development of a strong charging structure network is broadly known as a vital necessity for a huge-scale evolution to electromobility. This arrangement will not only give additional charging choices for drivers but would also encourage alertness and range self-confidence among approaching electric vehicle holders. As to acclimate toward progress of growing extent along with complex design of Electric Vehicle (EV) Fast Charging Station (FCS), also tendency of elegant city, this chapter elucidates an application that is based on the acceptance of cell phone Internet as well as mobile station to comprehend intellectual booking structure of electric vehicle fast charging station. Within this application, the structure recognizes user's vehicle details, shows nearby stations, helps to book slot as per his/her convenience, provides routing guidance, and starts the charging process.

Keywords Electric Vehicle Charging Station (EVCS) · Capital Expenditure (CAPEX) · Operating Expense (OPEX)

P. Bhambri (✉) · M. Aggarwal · H. Singh · A. P. Singh
Department of Information Technology, Guru Nanak Dev Engineering College, Ludhiana, Punjab, India
e-mail: pkbhambri@gndec.ac.in

S. Rani
Department of Computer Science and Engineering, Gulzar Group of Institutions, Khanna, Punjab, India

V. Bali et al. (eds.), *Decision Analytics for Sustainable Development in Smart Society 5.0*, Asset Analytics, https://doi.org/10.1007/978-981-19-1689-2_3

3.1 Introduction

EV charging location correspondingly titled electrically powered reviving socket or plug-in automobile stream apparatus is the groundwork that provisions rechargeable liveliness to refill the battery of power-driven automobiles.

Level 1 charging stations consist of 120 V, AC plug which in addition requires an out-and-out path, proposing round five miles of range after every hour of charging. Level 2 stations consist of a 240 V, AC plug and involve homegrown charging otherwise open charging apparatus to be mounted. Level 2 stations give 10–20 miles of array next each period of charging. Level 2 chargers remain the utmost used plus charge at roughly the equivalent haste as a household arrangement [1].

Level 3 chargers are likewise demarcated as DC fast chargers. Level 3 comprises a 480 V in addition to DC plug. They evade the involved charger besides offering DC energy to the battery operated from side to side of a superior charging harbor. DC chargers afford equal to 40 miles of array next each 10 min of charging, however are not used through all automobiles. Around decorum charging stations, like Tesla Supercharger, are constructed for expressively advanced rapid charging. As plea surges for additional freely reachable charging places, at hand is a countless necessity for apparatus that involves quicker charging at sophisticated energies in addition to currents that are not presently obtainable from housing. Universally, the quantity of rechargeable automobile systems is mounting to afford an arrangement of openly handy charging locations for EV reviving [2].

The developed application enables customers to charge their vehicles by offering them with three kinds of connectors—DC–001, CCS, and Type 2 (AC). DC–001 supports moderate charging and CCS supports fast charging whereas Type 2 (AC) supports both fast and slow charging. Customer can choose any of them according to his/her vehicle specification and can book a slot for charging [3, 4].

This chapter encapsulates the project proposing a brief outline of Android application and its need, characteristics, and services offered by charging stations; expected hurdles and solution to these hurdles; modules; and methodology used in the project. The new study supported the present research in building a new system to find available charging stations in the area and then helps authenticated customers reserve these stations according to their requirements. This final chapter explores everything from cost to market size to pricing, in order to make it clear how lucrative this software will be for everyone involved [5–9].

3.2 Electric Vehicle: A Sustainable Development Approach

In the current scenario, majority of people are advancing toward the EV environment as they prefer to drive sustainably. Trade of EVs is intensifying at more than 50% rate year after year. Hence, all EV drivers entail a station to charge their vehicles.

People having admittance to EV charging can straightforwardly lash out fossil fuels. Thus, magnification of EV charging alliance can fortify sustainable possibilities.

Removing automotive emissions using an all-electric vehicle improves the environment significantly. The use of EV driving contributes in renewing energy by bringing it into the fold. Charging stations function on renewable energy that strips away any odors, along with charging rates and energy output. The new-and-improved ambience is ensured by gathering and concentrating exhaled exhaust fumes at stations. This is a process that is more efficient than using cars with internal combustion engines. EVs are, thus, crucial to customers. Driving on energy is a double-whammy for the world and your wallet: you save gas money and you are doing something better for the environment.

3.3 Need for EV Charging Application

This EV charging application needed to be designed and developed based on the various significant issues as highlighted below:

- Difficulty in finding nearby available charging stations: Inaccurate mechanism for tracking the nearby available charging stations allows the end users to find it difficult for tracking the right location for electric charging purposes.
- Hassle-free booking of available charging slots: No accessibility to appropriate option for pre-booking of the charging slots availability at the concerned charging stations makes the hurdles to the traveling.
- Problems during long routes driving: No pre-booking options availability (while traveling on long routes) while utilization of electric vehicles put the journey into discomfort.
- Communication Gap: EV owners and charging station owners may be needed to communicate for the desired services. Miscommunication/delayed information between these parties create a lot of hassle.
- Difficulty in finding chargers according to make-in-model of electric vehicle: Electric vehicle chargers may be desired as per the make and model of the vehicle. No prior information for such mismatch builds a massive obstacle of driving the electric vehicle.

3.4 Features of EV Charging Application

To cope up with aforementioned challenges/difficulties, a real-time application for charging the electric vehicles is being developed and it encapsulates the following remarkable features:

- Enquire the end users for all desired details while signing up and thus to provide full access of the application only after verifying all necessary details.
- Display of the entire neighboring charging locations to the authenticated end users. It will display the complete information including the engaged and vacant charging slots/stations.
- Allows the end users to pre-book the vacant slots as per concerned requirements.
- As the customers reach the charging location, they can utilize the privilege of prior appointment based on the authenticity condition of sharing scanned QR code or entering OTP.
- Display the charging stations with crucial information like conveniences, comforts, and types of power channels. This information also include each minute charge plus distance starting from the existing position.
- Empowering the end users to anticipate with their involvement and make it noticeable to other users, too.
- To allow the users to enquire for adding their minute particulars while signing up. It also allows them to reserve the desired spot.
- To allow the users to pay in multiple ways and making the reward package thereafter.

3.5 Services Rendered from the Charging Station

Figure 3.1 depicts the list of service categories and typical services offered by them to the users [10]. All these services are provided by charging stations. Some of the service categories are management, maintenance and repair, and transaction handling. Elegant charging needs an EV car driver to recognize at the charging place. Identification links the electric car driver, charge point, and charging incident collectively. Electric car operators who are registered to charging amenity, identification is unproblematic for them as they immediately log in onto mobile phone application and commence charging. Unregistered clients can too utilize stations and make imbursement with smart cards. The charging affair itself occurs depending on settings and rates lay down by station proprietor [11–15].

Figure 3.2 depicts that there are three phases of charging process at a charging station. The very first phase commences with the user who will recognize himself/herself after reaching the station and this phase is known as identification. After recognition, the chief phase is charging in which vehicle gets charged with the help of connectors. Last one is the billing phase and user can pay through various modes via cards, Paytm, etc. [16–18].

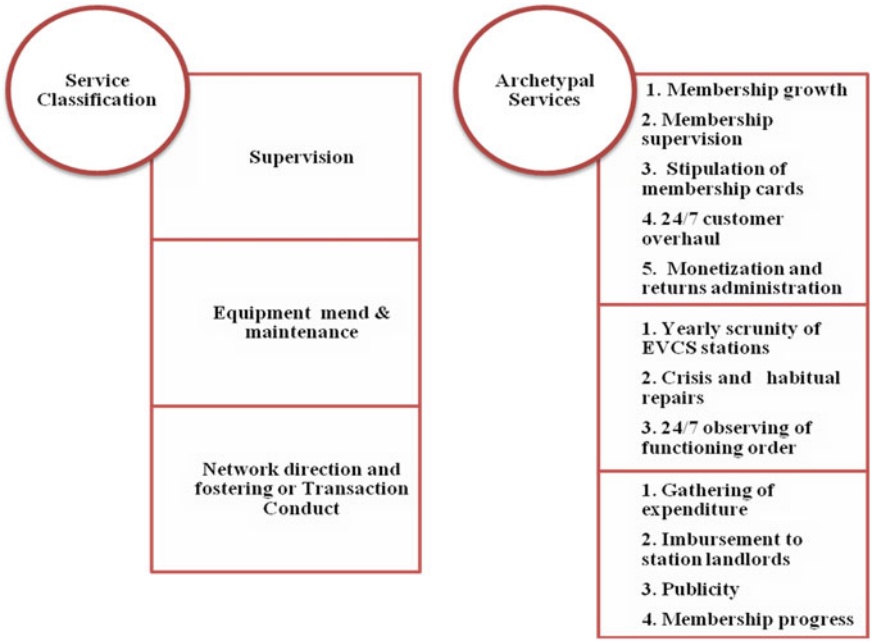


Fig. 3.1 Charging station services

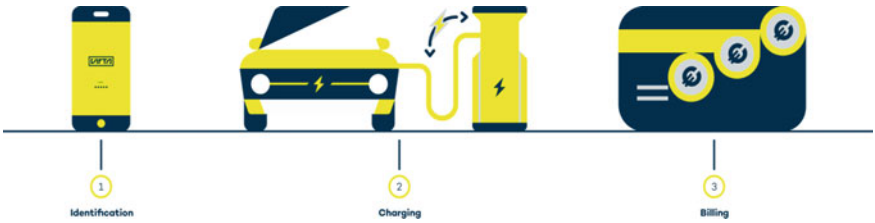


Fig. 3.2 Charging phases

3.6 Tribulations of Implementations

Many challenges for implementing the EV charging station application are expected as given below:

- Due to the lack of regular power supply, insufficient charging arrangements make the application failure.
- There is huge dependence on the battery-operated weights and on the introduced mechanisms and portions.
- Encouragements are catered to be interconnected with native engineering.
- Due to the novelty of implementation, nervousness amid users could occur.

- There is high value of EVs in the present scenario.
- There is deficiency of choices for high-performance EVs.
- Poor power sources in various parts of India could make it unrealistic to get it deployed everywhere.
- Absence of worth repairs and restoration routes could make the battlement for its implementation.
- This application may realize the threat of exaggeration by bigger vehicle production downturn.

3.7 Resolution of Encountered Tribulations

Remedial behavior for the encountered tribulations is essentially to be deployed in the design process of this application. These challenges are successfully defended during the design and implementation of this application with the following considerations:

- Android app which aims at providing information regarding all the nearby available charging stations from the user current location [19].
- Manage database of all the registered users and stations on application.
  - Adding a fresh plug-in automobile charge socket.
  - Getting driving directions from your location to a plug-in automobile charge socket.
- Providing all the necessary vehicle and user information for a smooth and reliable booking system.
  - Showing a list of nearby available charging stations in convenient manner.
  - Making hassle-free charging by booking appointments.
  - Making hassle-free payments and promote online payments (FASTag method).
  - Making profitable business and expand it.

3.8 Proposal Feasibility

A probability study governs whether the purposed system is achievable centered on the primacies of the necessities of an organization. A viability study concludes in a viability statement that suggests a resolution. It aids to calculate the rate efficiency of a purposed structure. A good viability study shows the strengths and shortcomings before the project is planned or budgeted for. The viability study is taken out to investigate if the anticipated structure is value being employed. After performing an investigation, collecting and inferring facts and figures relating to the task, a viability check is prepared which encompasses sequences of phases to form methodological, economical and operatable feasibility. This EV charging project is feasible. It is laidback to effort with and is very operator responsive. Technical feasibility puts the

emphasis on the technical aspects accessible to the organization. It supports organizations to decide whether the technical aspects fulfill volume and whether the technical crew is proficient of transforming the thoughts into operational schemes. Technical feasibility thus includes the assessment of the hardware, software, and new technical necessities of the proposed system. Software/hardware requirements include Android Studio (IDE), PC/laptop with any OS, Android Virtual Device (AVD), smartphones with different Android versions to run and test apps, other libraries and dependencies like Google Maps SDK, DirectionsAPI, and Firebase as Database and Storage [16, 20]. Economic feasibility characteristically encompasses a budget and profit analysis of the strategy, thus aiding establishments to adopt the sustainability, budget, and paybacks related with a plan beforehand monetary possessions are allotted. It furthermore assists as an autonomous venture assessment and fortifies plan trustworthiness assisting conclusion fabricators to conclude the optimistic monetary paybacks to the association that the anticipated venture offers. Operational feasibility encompasses task to examine and identify in addition how healthy the establishment's requirements can be seen by concluding the plan. It also studies in what way a task strategy accomplishes the necessities renowned in the desire scrutiny stage of structural advancement. The needs of building EVCS application are satisfied as follows:

- By showing a list of nearby available charging stations in convenient manner:
  - By getting driving directions from your location to a plug-in automobile charge socket.
  - By adding a fresh plug-in automobile charge socket.
- By making hassle-free charging by booking appointments:
  - By making hassle-free payments and promote online payments (FASTag method).
  - By making profitable business and expand it.

3.9 Classification of Various Modules

In the design and implementation of various modules of this application, numerous features are classified into two broader categories as User App or Operator App.

3.9.1 User App

Following are the significant operations as per titles implemented in the user app:

- Sign-Up/Sign-In through email or mobile number.
- User profile.
- To show EVCS location and ETA which is Interactive Map and Tracking.

- To show parking and amenities available.
- To book your slot now or later.
- To show traveling route in real time.
- To make built-in messenger.
- To show traveling route in real time.
- Cancel/alter booking.
- Payment sync using wallet/cash/card/auto-debit.
- To apply coupon code.
- Promotion and referral code.
- Ratings for EVCS with forms.
- To make charging info sharing to let close ones know that you're at a safe place.
- To list My Fuel, showing the charging history of the vehicle.

3.9.2 Operator App

Following are the significant operations as per titles implemented in the operator app:

- Alerts on booking requests.
- In-app payment details.
- Go online/offline status.
- Update bank accounts.
- Chat/call users.
- In-App navigation.
- Multiple charger categories.
- Tips to EVCS.
- Maintenance request and chat with Admin.

3.10 Methodology

This chapter summarizes a project report that focuses on agile methodology**.** This methodology for mobile app development is the best well-organized method to software growth business, it assurances a precise network of communication, which supports jointly the clients and App Developers bring out the anticipated mobile application or in fact any application.

The exact meaning of agile is to be quick and able to make speedy changes, similar works with the agile methodology. The value of Agile Methodology creates an easy job for mobile application growth so that the mobile app outcome is flexible after its announcement. This methodology, on the other hand, is really easy and permits for variations to be made in the project development necessities even after the early planning has been done.

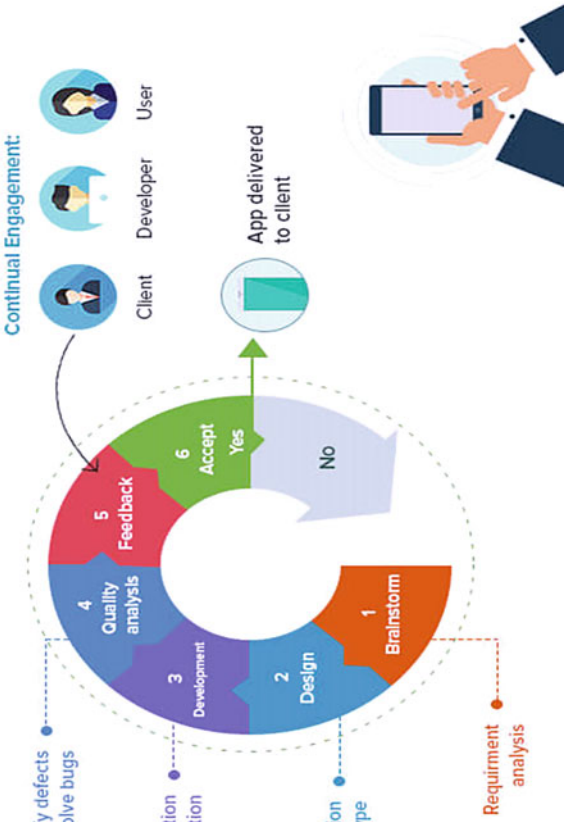


Fig. 3.3 Agile Methodology

Figure 3.3 displays the stages of Agile Methodology which are helpful in developing and designing an Android application [5, 6]. This approach consists of six steps which must be followed in a sequential manner. If the last stage is not fulfilled or the developed app is not accepted then the cycle restarts from phase 1.

3.11 Significance of Agile Methodology

Following description shows the significance of deployed agile methodology:

- The project growth steps such as development, designing, testing, etc. as chunk of the agile methodology, all trail an iterative development method.
- This will be very greatly valuable when requirement is for quick manufacture rather than the excellence of the product.
- The users will be capable to alter the scope of the project consequently when necessary.
- This gives clue of the final product will be comparable.
- Good for flexibility.

3.12 Prototype SDLC Model

This chapter highlights the software prototyping model used in the project. Prototype approach is described like a Software Enhancement prototype within which an archetype is constructed, tested, and after that revised when required up to a suitable archetype is obtained. Moreover, it builds a support to yield an ultimate structure. Software prototyping replica toils greatest wherever project's requirements are omitted. It is frequentative, cut, and try technique that occurs among contriver and clientele. Prototyping model has subsequent 6 SDLC (Fig. 3.4) levels.

This approach of SDLC implementation has numerous advantages as given below:

- The customers watch the fractional invention early in the life cycle. This safeguards a better level of user fulfillment and comfort.

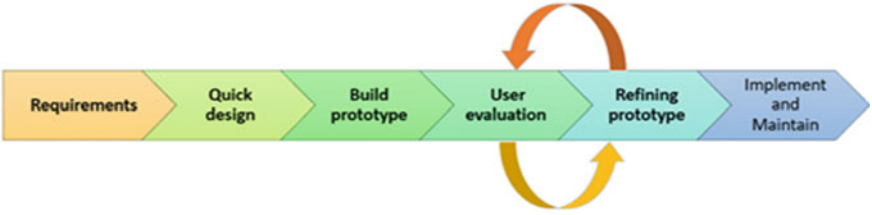


Fig. 3.4 Phases of prototype model

- First-hand necessities can be effortlessly accommodated as there is room for fine-tuning.
- Misplaced functionalities can be certainly engaged out.
- Faults can be spotted much effortlessly thus saving a portion of struggle and charge, above and beyond augmenting the value of the software.
- The developed prototype can be yet again used by the inventor for additional complex plans well ahead.
- Elasticity in design.

3.13 Cost–Benefit Analysis

Table 3.1 depicts Cost and Benefit Analysis of EVSE for 5 years. It shows the charger type and its connectors, voltage, vehicle type, tentative costs associated with it, and also the maximum power consumed by EVs per day. It also displays the estimated net revenue for the imminent 5 years [21].

3.14 Results and Discussions

The primary motive of this project was to implement an EV charging app using Android Studio with the aid of JAVA. The final application shows nearby charging stations based on the user's current location and organizes the stations based on proximity and open chargers [7, 16].

With the app, it's easier to locate and book charging outlets. It gives the ability to reserve slots in a swift and straightforward manner based on user desire. Once the booking process is over, the user is given a path showing their location and the location of the booked station, which aids the user in finding the station. User is able to evaluate prior or upcoming reservations. It gives customers the freedom to edit their profiles and cancel any reservation before the allotted time frame, even after signing up.

The app allows users to charge their vehicles quickly and safely which will help in saving environment. Further it supports the businesses to scrutinize charging proceedings tenuously and keeps control on maximum power utilization. If the customer has any questions, they can get in touch with the customer service team by emails, calls, or even WhatsApp.

3.15 Conclusion and Future Scope

The approaching era is anticipated to be an era of entirely electric vehicles. India has the ability to expand from prevalent embracement of e-mobility. With "Make

Table 3.1 Cost–benefit analysis

CAPEX

Kinds of chargers	Joints of charger	O/P Volt	Connector guns	EV category	Hesitant cost	Max. power consumed (per day kWh)
Rapid	CCS,CHADEMO,TYPE- 2 AC	200–750 or in between or above	1	2,3,4 wheelers	15,75,000	4500,3000,2490
Sedate/Gentle	Bharat DC-001	48, 72 or above	1	2,3,4 wheelers	4,80,000	720
Connection, Transformer, etc					7,50,000	
General jobs					2,50,000	
EVSE Supervision Software					40,000	
Camera Surveillance					30,000	
Overall CAPEX					3,125,000	

Anticipated returns

Depiction	1-Year—CUF-15 percent	2-Year—CUF-25 percent	3-Year—CUF- 40 percent	4-Year—CUF- 65 percent	5-Year—CUF- 85 percent	Total in 5 years
Fringe on power levy-A	16,32,150	27,20,250	43,52,400	70,72,650	92,48,850	2,50,26,300
Fringe on power levy in all five years-B	19,58,580	32,64,300	43,52,400	70,72,650	73,99,080	2,40,47,010
OPEX	6,36,000	6,67,800	7,01,190	7,36,250	7,73,062	35,14,301

(continued)

Table 3.1 (continued)

Anticipated returns						
Depiction	1-Year—CUF-15 percent	2-Year—CUF-25 percent	3-Year—CUF- 40 percent	4-Year—CUF- 65 percent	5-Year—CUF- 85 percent	Total in 5 years
EVSE Supervision Software at 15%	2,44,823	4,08,038	6,52,860	10,60,898	13,87,328	37,53,945
Sum of OPEX-A	8,80,823	10,75,838	13,54,050	17,97,147	21,60,389	72,68,246
EVSE Supervision Software at 15%	2,93,787	4,89,645	6,52,860	10,60,898	11,09,862	36,07,052
Total OPEX-B	6,36,000	6,67,800	7,01,190	7,36,250	7,73,062	71,21,353
Net Revenue-A	7,51,328	16,44,413	29,98,350	52,75,503	70,88,461	1,77,58,054
Net Revenue-B	10,28,793	21,06,855	29,98,350	52,75,503	55,16,156	1,69,25,657

in India" motive program, the share of industrialized sector amplifies India's GDP. Furthermore, EVs diminish operating as well as safeguarding cost and boost performance, sales, growth, and drivability. People dependency on fuel will be zero after approaching of EVs. These vehicles also contribute majorly in reduction of pollution and thus results in safe, bright, and green future. Thus, electric vehicles contribute significantly to achieving sustainable development goals. Figures 3.5, 3.6 and 3.7 reflect the front end for the utilization of application with specification of home activity, slot booking, and real-time tracking of speed, distance, and remaining time, respectively. This application has the following future scope statements:

- Sign-In with Google and Facebook.
- Swipe Gestures Support.
- Voice Assistant to help user run any feature of this app.
- Dark Mode to save battery.

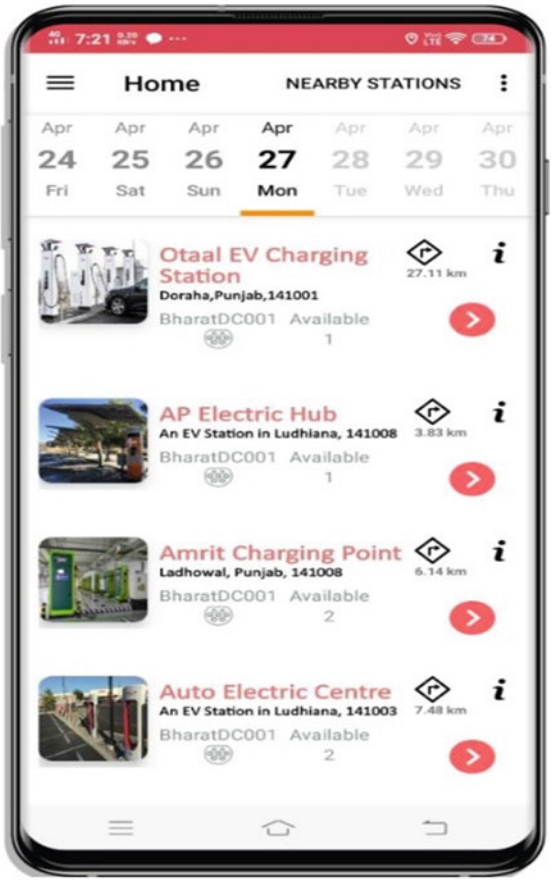


Fig. 3.5 Home activity

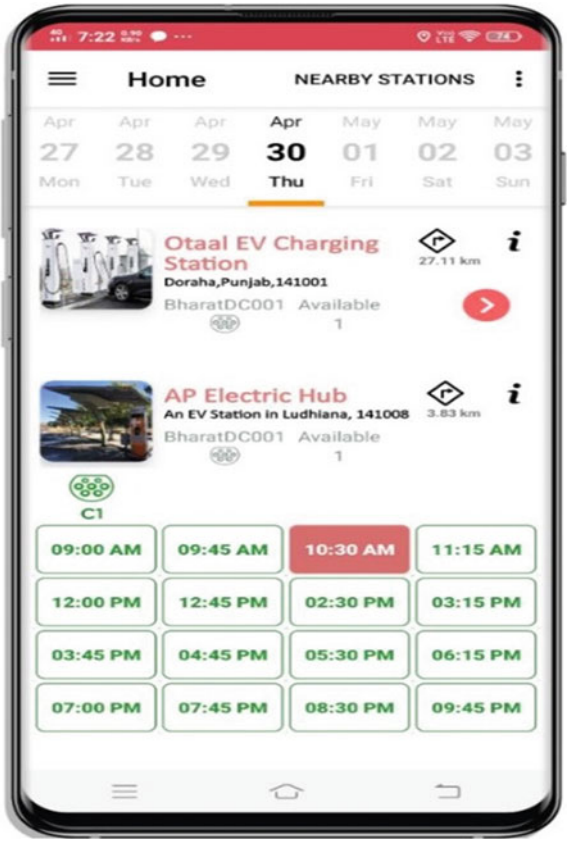


Fig. 3.6 Slot timings

- Chat-Bot that helps to demeanor an online chat discussion through text or text-to-speech, in lieu of giving direct interaction with an alive human agent.
- Use of QR codes instead of OTP for quick charging.
- Use of AI for automatic verification of vehicle documents.
- Rate and bookmark most liked or disliked charging stations.
- Showing available charging stations while traveling or en-routing.
- In-built payment methods for pre-payment services.
- Options to manage and block regular notifications.
- Facility to add and manage more than electric vehicle for same user.

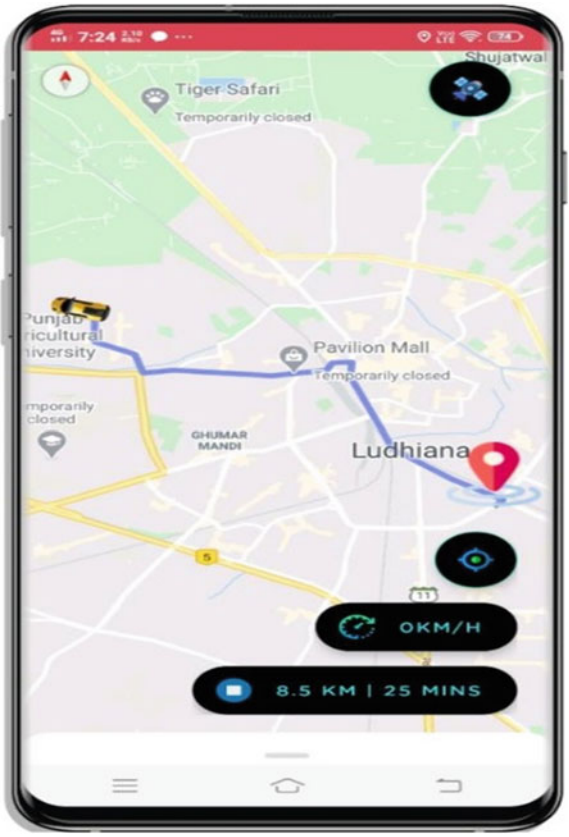


Fig. 3.7 Tracking

References

1. GATS India Ltd. (2021) GATS, SCS (Smart charging station)-EV charging stations model [Online]. http://www.gats.in/scs.html
2. Coruscate Solutions (2021) Scope & features of building EV charging stations [Online]. https://m.coruscatesolution.com/ev-charging-station-app-ev-charging-station-software-development/
3. Wikipedia (2021) App classification "types of app" [Online]. https://www.en.wikipedia.org
4. Java Platform and OOPS (2020) About java platform and OOPS related concepts [Online]. https://www.java-questions.com
5. DiMarzio JF (2021) Beginning android programming with android studio, 4th edn [Online]. https://www.pdfdrive.com/beginning-android-programming-with-android-studio-e39898073.html
6. Drongelen M (2021) Android studio cookbook: design, test, and debug your apps using android studio, 3rd edn [Online]. https://www.pdfdrive.com/android-studio-cookbook-design-test-and-debug-your-apps-using-android-studio-e185748368.html
7. Google Firebase docs (2021) Authenticate with firebase on android using phone number [Online]. https://firebase.google.com/docs/auth/android/phone-auth?authuser=0

8. Andriod App Testing (2021) Android app testing [Online]. https://www.softwaretestinghelp.com
9. Andriod App (2021) Scope of android apps [Online]. https://www.guru99.com
10. Singh AP, Aggarwal M, Singh H, Bhambri P (2020) Sketching of EV network: a complete roadmap. In: Singh H, Cheema PPS, Garg P (eds) Lecture notes in civil engineering, Springer, pp 431–442
11. Pradeep K, Pankaj B (2015) To design an algorithm for text watermarking. SIJ Trans Comput Sci Eng Appl 3(5):62–67
12. Bhambri P, Gupta OP (2013) Design of distributed prefetching protocol in push-to-peer video-on-demand system. Int J Res Advent Technol 1(3):95–103
13. Bhambri P, Gupta OP (2014) Dynamic frequency allocation scheme of mobile networks using priority assignment technique. Int J Eng Technol Innov 1(1):9–12
14. Vishal P, Pankaj B (2013) Edge detection-fuzzy inference system. Inter J Manag Inform Technol 4(1):148–155
15. Pankaj B (2014) Kaur pradeep a novel approach for zero watermarking for text documents. Inter Jo Ethics Eng Manag Educ 1(1):30–33
16. Google Firebase docs (2021) Add realtime database to app [Online]. https://firebase.google.com/docs/database/android/start?authuser=0
17. Google Firebase docs (2021) Get google maps sdk api key [Online]. https://developers.google.com/maps/documentation/embed/get-api-key
18. App Building (2021) App building strategy [Online]. https://www.buildfire.com
19. Murphy ML (2021) Android programming tutorials, 4th edn [Online]. https://www.pdfdrive.com/android-programming-tutorials-e10189445.html
20. Google Firebase docs (2021) Add firebase to android project [Online]. https://firebase.google.com/docs/android/setup?authuser=0
21. Research Gate (2021) CAPEX Cost-benefit analysis of EVCS [Online]. http://www.researchgate.net/publication/338187639_2019_Global_Conference_for_Advancement_in_Technology_GCAT

Chapter 4
Assessing Intelligence Text Classification Techniques

Charu Chhabra and Sneha Choudhary

Abstract Documents contain a tremendous quantity of important human information. The use of automatic text classification is necessitated by the fast rise in the quantity of machine-readable documents for public or private access. In a standard text categorization system, preprocessing is one of the most important elements. This article will look at the influence of preprocessing on text classification from a variety of perspectives, including classification accuracy, text domain, text language, and dimension reduction. Text categorization is a key difficulty for knowledge-based technologies since it touches on all of artificial intelligence's old demons: the knowledge engineering bottleneck, scalability issues, simple transfer across various applications, and cost-effective system design. By specifying a document, information retrieval (IR) systems have typically avoided all of these difficulties. Message classification is a text categorization problem that has piqued machine learning's interest. The categorization of brief text messages is one element of message classification that poses a special problem. The technique of categorizing documents into specified groups based on their content is known as text classification. It is the process of automatically categorizing natural language documents into preset groups. Text categorization is a prerequisite of both text retrieval systems, which retrieve texts in response to a user query, and text understanding systems, which change text in some way, such as creating new documents. Existing supervised learning methods for text classification require a large number of documents to learn correctly.

Keywords Text classification · Mining text · Support vector machines · Machine learning · Software modeling

4.1 Introduction

In the time today, numerous digital documents are generated and made available on a daily basis. However, sorting them into useful categories, for instance, essential or

C. Chhabra (✉) · S. Choudhary
The Northcap University, Haryana, India
e-mail: charu.machinelearning@gmail.com

V. Bali et al. (eds.), *Decision Analytics for Sustainable Development in Smart Society 5.0*, Asset Analytics, https://doi.org/10.1007/978-981-19-1689-2_4

minor, spam and without any spam would take a long time and a lot of human effort. Text the classification of documents and natural resource management are all tasks that must be completed. Depending on the topic, language text document is a result of previous work and assigned classification is a result of essential component that lots of data organized and management related work which can be well performed. The phrase categorization refers to the process of classifying newly found things or writings. The material must be categorized according to the previously identified kind. Elaborating the ML techniques, classification algorithms are created digitally utilizing some knowledge gained out of the previous activities so that classification may be formed in pre-standardized categorized training records. The text files or documents have been moved contrary to the dictionary's selected Word document (Word overlaps). Each term should appear at least once in the input document. Natural heritage text materials cannot be read and readable using traditional algorithms. Categorization of the text document classifiers and algorithms depicting assimilation never converts text documents into text form. In software development, software maintenance tasks (such as tests and reviews) are critical, and effort must be allocated efficiently. As a result, various software development support tools have been employed in software development, such as static code analyzers. There are several techniques that make it feasible for developers to concentrate their efforts on fault-prone modules and eliminate flaws at an early stage, resulting in benefits such as reduced work and improved software quality. RE is focused with the gathering, analyzing, defining, and validating of user requirements, which are typically written in plain language. The software requirements specification (SRS) document is the item produced by the RE process [1]. The performance of the documentation which we term as SRS, which happens to be the relative input to be it design, coding, and testing phases, is extremely priority and important to the success of a software project. This study takes a novel approach to the problem of automatic quality evaluation of textual requirements, specifically by employing the Natural Language Processing (NLP) text categorization technique. The impact of a software failure is described as severity. It represents the degree to which a defect's severity impacts functionality and the degree to which a flaw can impair software development or operation. As a result, the defect severity may be defined as the impact of a flaw on a component or piece of software. It emphasizes the defects criticality. An automated extraction and analysis of language from issue reports has been used to avoid a testing engineer from assigning incorrect severity levels to a defect, to save time in the case of an urgent-time-bounded application, and to ensure that no high-severity problems have been ignored. Previous researches used logistic regression and machine learning approaches to analyze the precision of defect proneness prediction for object-oriented design metrics in order to examine the accuracy of defect prediction while taking defect severity into consideration. Their findings show that at various severity levels, there is a statistical relationship between design metrics and fault-proneness of classes. Many software development tasks include learning a portion of an Application Programming Interface rapidly (API). API learning materials are essential for helping developers learn an API, but knowledge relevant to a certain topic of interest may be spread over several publications, making locating the

information they need more difficult. Automated text categorization into specified categories has long been seen as a critical approach for managing and processing a large number of digital documents that are widely distributed and growing [2]. Documents, conference materials, publications, journals, editorials, and web pages are examples of this type of online content, also known as digital or electronic information.

The chapter is organized into two following parts. In Sect. 4.2, we focus on text classification in software modeling and also elaborated several functionalities of the same. Section 2.1 emphasized on knowledge engineering for text classification and text categorization. Section 4.3 explained the techniques for intelligent text classification which elaborated the ML techniques used for the same like Naïve Bayes, Support Vector Machines (SVM). Section 4.4 elaborated the review on text classification with respect to the machine learning models for sustainable development goals. Section 4.5 concluded the entire chapter with the review of the text classification techniques and further scope for the functionalities of the same.

4.2 Text Classification in Software Modeling

Security is still a low rated concern in many businesses, as seen by the amount of broad scale software security breaches. Breach can be averted even before the program is published if security is considered during the design and execution of the product. One method to security assurance is to examine developer conversations as a means of investigating security during software development [3]. The fact that these software implementation are specific to certain projects, thereby limiting their implementation by other projects or organization, is an issue that is well known. The main focus must be on developing a general classification model that can be used to a variety of tasks. To gather security keywords from relevant security sources and classify them into four categories: asset, attack/threat, control/mitigation, and implicit. It is the matter of the fact that security breaches have become regular occurrences, with devastating consequences and costs to organizations and society [4]. Content-based document management systems have risen to prominence in the fields of computer and information systems engineering and computer science during the previous 15 years. The popularity of content-based management systems can be attributed to two factors. Text Categorization (TC) is being employed in a number of scenarios right now. Document filtering, automated meta-data generation, word sense disambiguation, and the population of hierarchical catalogues of web resources are just a few of the applications [5].

4.2.1 *Knowledge Engineering for Text Classification and Text Categorization*

Categorization of text is now utilized in a number of settings. Whether we keep into account the filtering of documentation, automated generation of the meta-data, and population of web resource catalogues in a hierarchical structure are just a few of the applications [6]. An SLR seeks to find, analyze, interpret, collate, and generate the summary, the most current and relevant data on a research subject or phenomena of interest. The use of classifying text in medicine to help SLR processes is becoming more widely accepted. Classification of text which is also known as document categorization is the process of assigning semantic labels to texts based on a set of predefined meaningful categories or classes. Text mining (TM), information retrieval, and natural language processing are all applications of the text classification problem. Sentiment analysis, genre analysis, language identification, content suggestion (text recommendation), and literacy evaluation are only a few examples [7]. Text classifiers can be built using supervised machine learning (ML) algorithms, statistical techniques, or handwritten rules based on data. Linguistic expertise machine learning techniques are very useful and popular because the classifier may be built automatically from corpus (also known as text sets) [8, 9]. There are several appropriate corpus that are publicly available for such purposes.

4.2.2 *Dimensionality Reduction*

Dimensionality reduction is the process of translating data from a high-dimensional space to a low-dimensional space in such a way that the low-dimensional representation retains some essential aspects of the original data, ideally close to its intrinsic dimension. Working with high-dimensional spaces can be problematic for a variety of reasons: raw data is often sparse due to dimensionality's curse, and data processing is typically computationally challenging. Signal processing, voice recognition, and other disciplines that elaborate the working with humongous numbers of observations and/or variables frequently use dimensionality reduction. Dimensionality reduction decreases the course of twisted words to their naïve form, allowing several variants of the same term to be grouped together [10]. Different approaches may be used to implement stemming algorithms. The removal of letters which are terminating with the words based on a list of suffixes which were already defined is a basic variation.

4.3 Techniques for Intelligent Text Classification

To enhance text document categorization into specified categories, information which was based on indexing or citation and content abiding the structure (e.g., title,

abstract) might be integrated [11]. The automatic categorization of text into predefined categories has sparked a lot of attention in recent years due to the growing number of digital documents and as a result of the desire to arrange them orderly. Traditional content-based classifiers, on the other hand, have a history of underperforming when documents are noisy. It may be identified semantically coherent word groups by clustering words represented in latent semantic vector space using LSA partitioned into numerous categories with comparable ideas. To deal with the problem of expensive data labeling, a learning approach with unlabeled data called self-confident sampling was developed. Vote entropy is the information-theoretical criteria for assigning a label to an unlabeled document. Text classification with machine learning entails applying a machine learning approach to assign a set of predefined classes to a text content [12]. In most cases, categorization is done using text documents and chosen documents and characteristics. The extant of such tremendous amount information which survived desultorily is the reason text categorization needs are increasing. A machine learning technique is required to categorize this data.

4.3.1 Naïve Bayes Classification

For identification of the pattern and categorization that comes within distinct variants in pattern recognition and classifiers for the fundamental probability, the Nave Bayes method is employed. The Naive Bayes classifier is a basic classifier that divides data into categories based on event probability. It is frequently used in text categorization. Despite being a basic method, it performs well in a variety of text classification tasks. The Bayes theorem is the one which creates a group of algorithms which works on classification known as Naive Bayes classifiers. It is a that category of algorithms that share a similar premise, such that every pair of characteristics which is categorized is certainly not dependent on the others [13].

4.3.2 Support Vector Machines for Classification

When conducting text categorization using machine learning, there is variety of machine learning methods to select from. Support vector machines is one of them (or SVM). The categorization of documents has previously been extensively researched. In fact, some research examined the performance of different algorithms, while others compared feature selection approaches or feature space transformation [14]. Following the recent surge in interest in the support vector machine, several researches have revealed that SVM outperforms alternative classification techniques.

4.3.3 Decision Tree Induction

Classification of the text is an ongoing text mining research topic in which documents are categorized into preset categories. Letters, technical reports, conferences, and journal papers are examples of mostly text documents. Document filtering, which is likewise based on a classification algorithm, extracts relevant documents from a group of documents linked to a certain topic [15]. Text classification tasks can be divided into two types: supervised document classification, in which an external mechanism (such as human feedback) provides information on the correct classification of documents or defines classes for the classifier, and unsupervised document classification, in which no external mechanism (such as human feedback) provides information on the correct classification of documents or defines classes for the classifier [16]. Figure 4.1 shows the text classification flowchart.

The most commonly used inductive learning approaches are decision trees. The learning of decision trees from labeled training materials is known as decision tree classification. ID3 and its successors C4.5 and C5 are two of the most well-known decision tree algorithms. A decision tree is a structure which resembles like a family tree such that it looks like a flowchart, with each internal node depicting a document test and each branch representing an outcome [17].

Their capacity to learn disjunctive phrases and their resilience to noisy data appear to make them appropriate for document categorization. Decision trees are straightforward to comprehend and interpret. They just need a little amount of information and

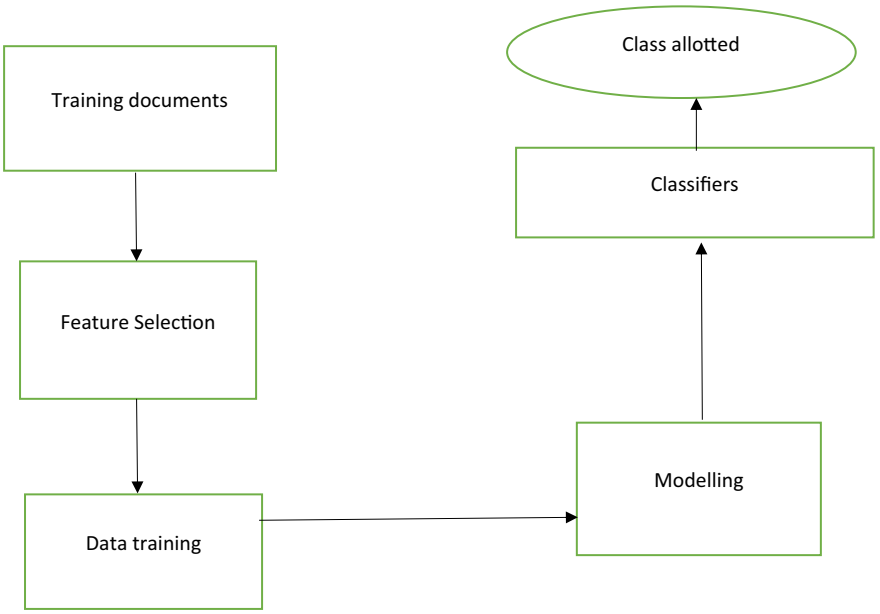


Fig. 4.1 Text classification paradigm

can handle both numerical and category data. Even when there are different amounts of training instances, this approach scales effectively. Heuristic algorithms, such as the greedy algorithm, are used in decision tree learning methods, where choices are made locally at each node. Such algorithms can't promise that they'll give you the best decision tree in the world. Overfitting can occur in a fully developed tree if certain branches are overly particular to the training data. As a result, most decision tree learning systems incorporate a way for developing the trees. The number of online papers has exploded as the use of the Internet has increased dramatically. The assignment of text documents to one or more preset groups based on their content is known as classification of text which is a key component in many information management activities.

4.4 Text Classification Machine Learning Models for Sustainable Development Goals: A Review

In order to analyze the remote sensing data, the system deliberately requires the analysis which includes statistical approach which can be effectively utilized to create measures of the environment, agriculture, and sustainable development has grown in popularity, resulting in high-frequency interaction and emphasis between the earth science and other domains containing statistics. In light of this, there are numerous industries which require intelligent risk assessment and classification technique to propagate system on statistical machine learning algorithms widely used to analyze remote sensing data, CSR, intelligent decision-making systems, and many more to the count. The essay discusses the issues of risk assessment and long-term development. The categorization of risk assessment methodologies can easily be stated out using the novel classification techniques of the machine learning. The fundamentals of sustainable development are examined. Literature review is the approach that was used. It is possible to make a decision based on the results of the investigation.

4.5 Conclusion

Text mining is the process of extracting useful but previously unknown information from a text document. Text categorization (also known as text classification) is an important study topic in the field of text mining. The significant growth in the amount of digital content available has created a difficulty in managing this online textual data. As a result, classifying/categorizing huge texts (documents) into particular classes has become required. Text classification categorizes a text document into one of several categories. This article provides classifier architecture and text classification application, as well as several text classifications approaches. The simplest technique for determining the class of unlabeled documents is the K-nearest

neighbors algorithm (KNN), which is a common non-parametric method. However, because the computing time for high-dimensional documents rises, this approach is not suited for such documents. When dealing with multi-dimensions, SVMs and neural networks perform considerably better.

Large sample sizes are necessary for SVMs and neural networks to obtain maximum classifier accuracy. As a result, we'll require a lot of storage space for both training and testing materials. Naive Bayes, on the other hand, may demand a smaller dataset and less storage space. Because Naive Bayes has a large bias and does not give effective results when words (or terms) are linked, the term graph model is chosen to preserve the connection between adjacent words. KNNs, or Keras neural networks, are typically regarded noise-intolerant, whereas association-based classification and decision trees are considered noise-resistant because to their pruning techniques that prevent overfitting. In comparison to other classifiers, SVM outperforms them in terms of accuracy, learning speed, classification speed, and tolerance to irrelevant characteristics and noisy input.

SVMs also have the unique virtue of being able to learn regardless of the dimensionality of the feature space. However, it is impossible to suggest one approach over another.

References

1. Hoch R (1994) Using IR techniques for text classification in document analysis. SIGIR '94, 31–40. https://doi.org/10.1007/978-1-4471-2099-5_4
2. Kowsari K, Jafari Meimandi K, Mendu S (2019) Text classification algorithms: a survey. Information 10(4):150. https://doi.org/10.3390/info10040150
3. Kim S-B, Han K-S, Rim H-C, Myaeng SH (2006) Some effective techniques for Naive Bayes text classification. IEEE Trans Knowl Data Eng 18(11):1457–1466. https://doi.org/10.1109/tkde.2006.180
4. Yang SM, Wu XB, Deng ZH, Zhang M, Yang DQ (n.d.) Relative term-frequency based feature selection for text categorization. In: Proceedings of international conference on machine learning and cybernetics. https://doi.org/10.1109/icmlc.2002.1167443
5. Kadhim AI (2019) Survey on supervised machine learning techniques for automatic text classification. Artif Intell Rev 52(1):273–292. https://doi.org/10.1007/s10462-018-09677-1
6. Zhang X, Li B, Sun X (2010) A k-nearest neighbor text classification algorithm based on fuzzy integral. In: 2010 sixth international conference on natural computation. https://doi.org/10.1109/icnc.2010.5584406
7. Aggarwal CC, Zhai CX (2012) A survey of text classification algorithms. Mining Text Data, 163–222. https://doi.org/10.1007/978-1-4614-3223-4_6
8. Lilleberg J, Zhu Y, Zhang Y (2015) Support vector machines and word2vec for text classification with semantic features. In: 2015 IEEE 14th international conference on cognitive informatics and cognitive computing (ICCI*CC). https://doi.org/10.1109/icci-cc.2015.7259377
9. Shah K, Patel H, Sanghvi D, Shah M (2020) A comparative analysis of logistic regression, random forest and KNN models for the text classification. Augment Human Res 5(1). https://doi.org/10.1007/s41133-020-00032-0
10. Zulqarnain M, Ghazali R, Mohmad Hassim YM, Rehan M (2020) Text classification based on gated recurrent unit combines with support vector machine. Int J Electric Comput Eng (IJECE), 10(4):3734. https://doi.org/10.11591/ijece.v10i4.pp3734-3742

11. Thatha VN, Babu AS, Haritha D (2019) An enhanced feature selection for text documents. In: Smart intelligent computing and applications, pp 21–29. https://doi.org/10.1007/978-981-32-9690-9_3
12. Lydia EL, Kannan S, SumanRajest S, Satyanarayana S (2020) Correlative study and analysis for hidden patterns in text analytics unstructured data using supervised and unsupervised learning techniques. Int J Cloud Comput 9(2/3):150. https://doi.org/10.1504/ijcc.2020.109373
13. Ghiassi M, Olschimke M, Moon B, Arnaudo P (2012) Automated text classification using a dynamic artificial neural network model. Expert Syst Appl 39(12):10967–10976. https://doi.org/10.1016/j.eswa.2012.03.027
14. Gargiulo F, Silvestri S, Ciampi M, De Pietro G (2019) Deep neural network for hierarchical extreme multi-label text classification. Appl Soft Comput 79:125–138. https://doi.org/10.1016/j.asoc.2019.03.041
15. Raza M, Hussain FK, Hussain OK, Zhao M, ur Rehman Z (2019) A comparative analysis of machine learning models for quality pillar assessment of SAAS services by multi-class text classification of users' reviews. Futur Gener Comput Syst 101:341–371. https://doi.org/10.1016/j.future.2019.06.022
16. Pavlinek M, Podgorelec V (2017) Text classification method based on self-training and LDA topic models. Expert Syst Appl 80:83–93. https://doi.org/10.1016/j.eswa.2017.03.020
17. Ye Z, Geng Y, Chen J, Chen J, Xu X, Zheng SH, Wang F, Zhang J, Chen H (2020) Zero-shot text classification VIA REINFORCED SELF-TRAINING. In: Proceedings of the 58th annual meeting of the association for computational linguistics. https://doi.org/10.18653/v1/2020.acl-main.272

Chapter 5
Automated Traffic Rule Violation Detection with E-Challan Generation for Smart Societies

Supriya S. Thombre, Sampada S. Wazalwar, Rakshanda Tarekar, Sakshi Ramgirwar, Shivani Zulkanthiwar, and Shubhra Dubey

Abstract Nowadays as the traffic is increasing on the roads, the accidents cases are also increasing. For controlling this the cameras are installed at various traffic signals or various road intersections in the smart cities. The image captured from these cameras will help to detect the rules violators. This has been done by using an Automatic Number Plate Recognition (ANPR) system which is an automated system and will be able to detect the various types of traffic rules violations. The number plate has been detected after identifying the vehicle. From the number plate, the vehicle owner's information has been collected and the e-challan has been generated. This generated e-challan will then be sent to the registered email-id for the information and the fine toward the traffic rule violation.

5.1 Introduction

5.1.1 Overview

Nowadays the volume of traffic increases due to increasing the number of vehicles. This may lead to traffic violations. It may cause terrible damage to the assets and also the humans will be prone to accidents [1]. For avoiding these, automated traffic rules violation systems are essential. It should enforce the traffic regulations all the time and after identifying the violator it will issue the e-challan to those who don't follow. So, an increase in traffic will correspond to the implementation of an automated system for road safety and which will efficiently detect rule violations faster than humans [2].

S. S. Thombre (✉) · R. Tarekar · S. Ramgirwar · S. Zulkanthiwar · S. Dubey
Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra, India
e-mail: supriyathombre@gmail.com

S. S. Wazalwar
Department of Information Technology, G. H. Raisoni College of Engineering, Nagpur, Maharashtra, India

V. Bali et al. (eds.), *Decision Analytics for Sustainable Development in Smart Society 5.0*, Asset Analytics, https://doi.org/10.1007/978-981-19-1689-2_5

A user-friendly graphical user interface was developed to make it simple to operate. This system is capable of detecting the three types of real-time traffic violations which are a signal violation, a parking violation, and a wrong direction violation. The system can also monitor the traffic and will take the action against the violations of traffic rules [3]. The copy-paste forgery can also be done on images so, for this system designing the real-time images captured from the roadside cameras is used [4].

5.1.2 Problem Definition

Number Plate Recognition is the most efficient and intelligent system for finding the number plate's location by analyzing image/video data for an ANPR system. Videos or images are captured by a camera placed at every intersection of the road. For identification, the super-resolution techniques are to be applied to the image so that the license plate will get easily identified with the help of the Optical Character Recognition (OCR) technique. The nature and texture of an image should be maintained during the super-resolution of an image. After getting the vehicle number it will initiate the activity to generate the e-challan and send it to the vehicle owner using email.

5.1.3 Objective

The aim is to develop a system that will automatically track the rule violators by identifying the vehicle's number plate and information retrieval from the same. The e-challan will then be issued to the owner of the vehicle for a traffic rules violation. This will help to reduce the traffic rules violations which will reduce road accidents.

5.2 Literature Survey

M. Chaitanya Sai, Deepesh Chandwani, and Saravana bhava proposed a method that uses the Otsu Thresholding scheme. The image was firstly segmented by the proposed methodology and license plate and object detection have been carried out using a double binarized approach. The blurred or noisy image has been used during the implementation and multi-object detection is used for License Plate Detection (LPD) and License Plate Recognition (LPR). The implementation has been carried out using multiple stages like the acquisition of an image, gray processing, binarization of an image, recognition of number plate, and Optical Character Recognition (OCR) is used for template matching. For checking the performance of an algorithm, the authors have used various real-time test images for extracting the number plates

from an inputted image. The IoT can be used along with the proposed approach for enhancing the tracking systems [5].

The two approaches have been used by an Automated License Plate Recognition (ALPR) for performing LPD and LPR. The LPD uses segmentation of characters and OCR for identifying vehicle location and license plate detection. As an alternative to using a general object detector method like Faster-RCNN or SSD300, the authors have used the segmentation-free OCR evaluation approach for developing a precise system for detecting license plates. The detection in the feature map pyramid is performed by object detectors for detecting objects at different scales. As the deep network needs multiple samples, the new data augmentation technique has been developed which has the potential to train and evaluate [6].

The authors have identified the problems during image capturing. The vehicle license plate detection from an image is itself a complex problem as the image is captured can contain noise. For considering complex scenarios, the application demands robust and generalized license plate detection techniques. The new technique has been proposed by the authors which mainly focuses on the problems such as variations in illumination and color pattern of an image. They have used a static camera for capturing an image. The proposed method can be used for the detection of license plates in each frame of a video sequence, detecting license plates partially and the detection of license plates using moving cameras and vehicles. For object detection, region proposal (RCNN) with convolutional neural networks and its inheritors (Fast-RCNN and Faster-RCNN), and the exemplar-SVM have been used [7, 7].

Yogiraj Kulkarni, Shubhangi bodhke, et al. consider moving objects for LPD. The proposed method uses a neural network with an improved OCR. This method uses K-Nearest Neighbors (KNN) for classifying the features retrieved from region properties in an image. This system uses a decision tree classifier with Ada-boost for detecting a partial and complete helmet. For achieving accuracy, they have also used CNN for the classification of the pre-trained model. For subtracting the background from a video for getting moving objects, various features of computer vision have been used like Hough transform, LBP, HOG, SIFT, etc. For identifying geometric shapes like circles, ellipses in an image Hough transform are used. The distribution of edge directions or intensity gradients is used for describing the shape and its appearance of a local object within an image. For describing the local texture structure and for capturing key points in an image the LBP and SIFT have been used, respectively. The author gets an accuracy of 98.72% for traffic videos which was earlier as 96.36% using Tesseract OCR on the test images [9].

Praveen and Arihant have significantly shown the usage of deep learning techniques for simplifying the process of ALPR which also allows the inclusion of all the abnormal features present in the license plates. The proposed method uses a recursive training method for the character segmentation network and for detecting the license plate. The YOLO framework has been trained using the same data. This network has been proven for fast object detection in a real-time scenario [10].

Region Growing Algorithm (RGA) uses some heuristic approach with a recursively region-growing procedure. This also focuses on the symbols present on the license plate. The background of the license plate consists of region dimensions, and

position, width, and height (aspect ratio), and color and typestyle size. The proposed algorithm is used for identifying the Region of Interest (ROI). This algorithm has been tested on 350 images containing light and heavy vehicles [11].

The License Plate Recognition (LPR) algorithm is based on license plate region extraction, plate character segmentation, and each character recognition. Image acquisition is a very challenging task due to a variety of number plate formats and non-uniform lighting conditions. Taking this into consideration many approaches works for limited conditions such as vehicle speed limit, fixed lightning, selected routes, and stationary backgrounds. Immobile images or video sequences have been used in various techniques developed for LPR. The authors have carried out the work categorizing and assessing them by considering the time required for processing, the computational power, and the recognition rate [12].

5.3 Work Done

5.3.1 Overview

ANPR is a system that itself capable of license plate detection without human intervention. It works on four steps: character segmentation and recognition, capturing vehicle image, number plate detection. It uses an image captured with supporting illumination [13].

Figure 5.1 shows the flowchart of the process. When an image is considered for license plate detection it undergoes four factors, such as gray scaling and blurring, background subtraction, binary threshold, dilution, and find the contour. After which the vehicle classification and detection are done where the rule violating vehicles are classified. For number plate recognition, super-resolution and OCR classification is used. Once the license plate is detected for a particular vehicle, an email is generated and sent using the SMTP service to the registered email-id.

The System consists of two main components:

- Vehicle detection model and
- A graphical user interface (GUI).

At every road intersection nowadays the CCTV cameras are installed. The vehicles are detected from the CCTV camera footage mounted at the roadside. This footage will help in identifying the license place of a vehicle that violates the traffic rules. Figure 5.1 describes the overview of the system and its working.

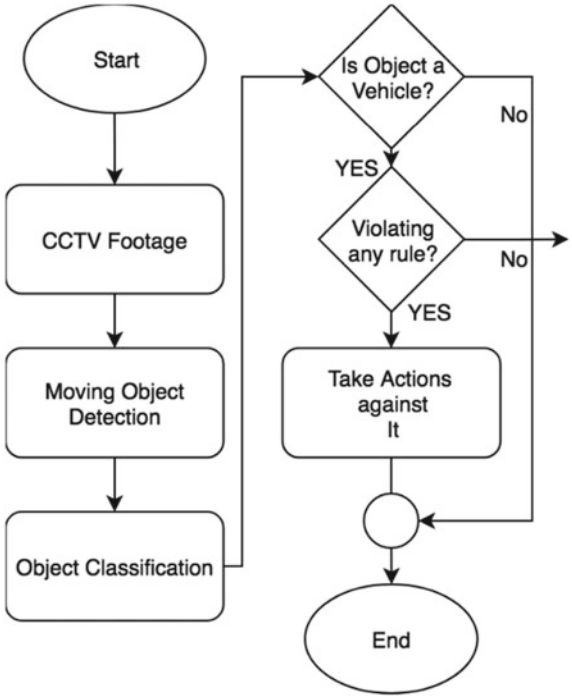


Fig. 5.1 Overview of proposed methodology

5.3.2 Software Used

5.3.2.1 Python 3.9

This latest version provides the Python 2 backward compatibility layers which help the project maintainers for adding the Python 3.9 support and removing Python2 support.

5.3.2.2 Imutils = 0.5.3

OpenCV and both Python 2.7 and Python 3 are easier for image processing functions like translation, resizing, rotation, skeletonization, sorting contours, detecting edges, etc. This resizing function uses the keywords height and width for maintaining the aspect ratio and ensuring the dimensions of an image. By using this function the developer need not compute the above parameters explicitly. An inter keyword has been used for specifying the interpolation method.

5.3.2.3 PyQt5 = 5.12.3

For developing the GUI application the PyQt5 has been used. It is a cross-platform GUI toolkit consisting of a set of Python bindings for Qt v5. Using this library, an interactive desktop application can be developed. A tool called 'QtDesigner' is used for designing the front-end of any application.

The installation process for this library using the command prompt or the terminal using the commands is as follows:

$$pip\ install\ pyqt5$$

One can verify the successful installation by using the command:

$$>>>\ \text{import}\ PyQt5$$

The number of tools like QtDesigner was provided by PyQt5. For doing this execute the following command:

$$pip;\ install\ PyQt5 - tools$$

5.3.2.4 OpenCV = 4.3.0.38

OpenCV has a vital role in real-time operations on computer vision, machine learning, image processing, etc. For identifying the objects, faces, human handwriting, etc. by processing an image or any video using OpenCV. The Python is capable of processing the OpenCV array structure used for analysis once integrated with various libraries or NumPy.

5.3.3 Methodology

5.3.3.1 Image Processing

(a) Grayscaling and Blurring: As part of pre-processing the input frame got from the CCTV footage, the image is grayscaled and blurred with the gaussian blur method.
(b) Background Subtraction: The background subtraction method is used to subtract the current frame from the reference frame to get the desired object's area. The equation shows the method. $dst(I) = saturate(|scr1(I) - scr2(I)|)$.
(c) Binary Threshold: The binarization method is used to remove all the holes and noises from the frame and get the desired object area accurately.

(d) Dilation and find the contour: After getting the threshold image, it is dilated to fill the holes, and the contour is found from the image and it is denoted by the rectangle box.

5.3.3.2 Vehicle Classification

For classifying the moving objects into various classes such as car, motorbike, or a non-vehicle. The moving objects were extracted after pre-processing an image. MobileNetV1 neural network architecture has been used for building the classifier model.

For training the model with the dataset, the transfer learning approach has been used. Per class, the dataset consists of 500 images. Table 5.1 mentions the training parameter.

5.3.3.3 Violation Detection

Three violation cases arise after detecting the vehicles:

- Signal violation: The signal violation has been detected if there is a red signal and the vehicle crosses the predefined line.
- Parking violation: The parking violation has been detected if the vehicle stands still in the NO Parking zone for a predefined time.
- Direction violation: The direction violation has been detected if the vehicle moving in the wrong direction. The current position and previous few positions are used for determining the directions of the vehicle.

5.3.3.4 Number Plate Recognition

The number plate recognition consists of the following steps:

1. **Super-Resolution**: Super-resolution uses a combination of a low-resolution sequence of images of a scene for generating a high-resolution image. The high-resolution image can be reconstructed from the low-resolution images. The depth-based image super-resolution algorithm can also be used where the image registration and correspondence between the left view and right view of an image needs to be calculated [14].

The Fast-Super-Resolution Convolutional Neural Network (FSRCNN) approach for super-resolution has been used. This approach has mainly five steps:

Table 5.1 Training hyperparameter

Parameter name	Value
Learning rate	0.01
Training steps	100

a. Feature extraction: Adopts the original low-resolution image as input
b. Shrinking
c. Mapping
d. Expanding
e. De-convolution: gives the higher resolution image

The proposed approach will generate a high-resolution image in a lesser time than the existing methods. This method can also be implemented in real-time images.

2. **Optical Character Recognition (OCR)**: It is a process in which an electronic device is used for checking the characters, shapes by detecting the color pattern present on the paper and translating it into machine-encoded text. This is generally used for converting the text image into text. While generating a high-resolution image from its corresponding low-resolution image it may produce multiple errors due to a lack of high-frequency image details in low-resolution text images which may lead to incorrect information retrieval. For achieving higher OCR accuracy, super-resolution pre-processing on inputted text images should be carried out.

5.3.3.5 E-challan

The challan will be generated based on violations by vehicles and an e-challan will be sent via email to the violator using SMTP service. To achieve this the Simple mail Transfer Protocol (SMTP) Server can be used. By using this SMTP, the system can easily send or receive emails. It acts as a relay between the sender and receiver of the email.

5.3.3.6 Database Structure

The SQLite database with python has been used to manage the whole data of the application. BCNF of five tables has been used in the relational database. The tables are as follows:

1. Cars
2. Rules
3. Cameras
4. Violations
5. Groups

Figure 5.2 shows the various tables of the database along with their parameters.

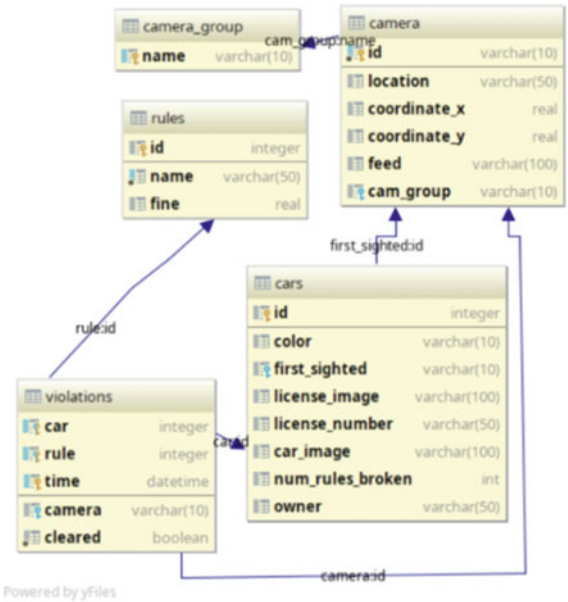


Fig. 5.2 Details of each table

5.3.4 Implementation

5.3.4.1 Image Processing and Computer Vision

For image processing purposes, an OpenCV computer vision library is used. Tensorflow machine learning framework is used for implementing the vehicle classifier.

5.3.4.2 Graphical User Design

The GUI will be helpful for all the administrative and debugging purposes so that it will be more users friendly. For adding the sample cars or cameras in the database, the Menu item is useful as shown in Fig. 5.3.

Initially, the administrator needs to add a camera using a Menu item. The location of the camera can also be added by using the feed file for the camera which is needed to be installed using the camera module over the internet. The Linux file sharing pattern has been used for getting the video from the camera where it will feed the given file to the server, and then it will take to process and detect a violation. The X and Y coordinates are required while using the map for locating the cameras. These coordinates will be entered by the admin. The admin needs to specify some rules for

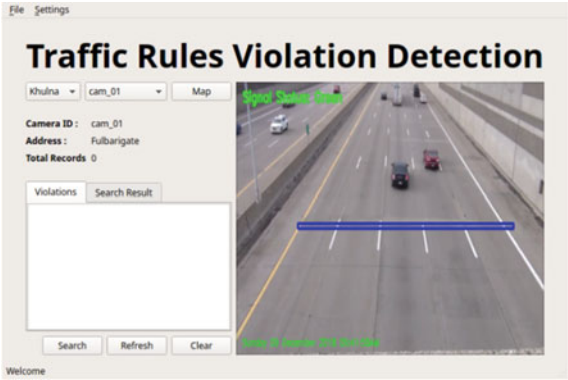


Fig. 5.3 Overall user interface view

the cameras using a JSON file. The rule can be for crossing the road during the red light, i.e., red light violation or it can be used for wrong parking detection.

Figure 5.4 shows an interface adding the camera and then the software will automatically start detecting the violations of traffic rules as shown in Fig. 5.5.

Figure 5.6 gives an interface to add the following objects into the database.

1. Camera
2. Car
3. Rule
4. Violation

Figure 5.7 elaborates an interface from where the user can see the car details that violate the rules. A user can be able to file a report or sending the ticket to the car owner after clicking on the detail button.

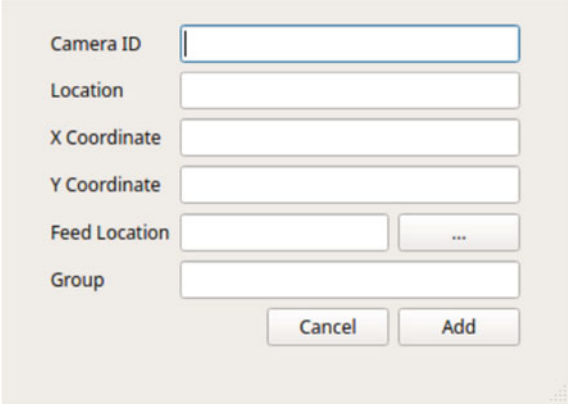


Fig. 5.4 Interface for adding camera entity

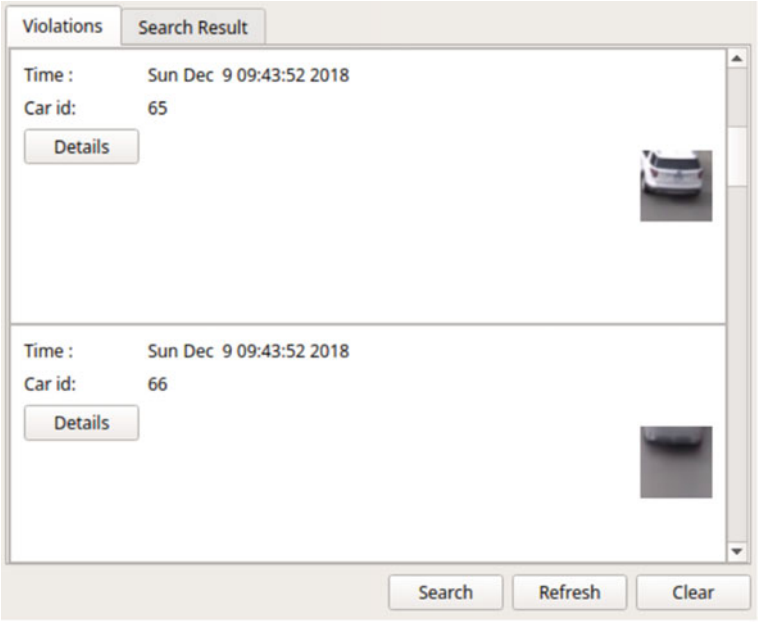


Fig. 5.5 A list view of violation records

The admin can have a right to delete the records if it is a false positive. The deleted record can be recovered from the archive.

Figure 5.8 shows that by using the license number, vehicle color, or date of a rule violation the user can search for a vehicle. The license number has a constraint for the text prediction so only an existing license number will be detected.

5.3.4.3 Rules Violation Representation in UI

Figure 5.9 shows that the car is violated the traffic rule by crossing the road during the red light (a blue line in the above figure indicates the red signal). A picture of that car is registered in the database along with some environmental values.

Figure 5.10 denotes the parking violation. The restricted area for parking was shown by the outer dark blue rectangle. If there is a vehicle in the rectangle for more than a predefined time, then an image with other environmental values is being registered to the database.

Figure 5.11 detects the direction violation, some lines are drawn to divide into regions. The direction of the car is being measured when a car moves from one region to another. The car will get registered if the direction is wrong.

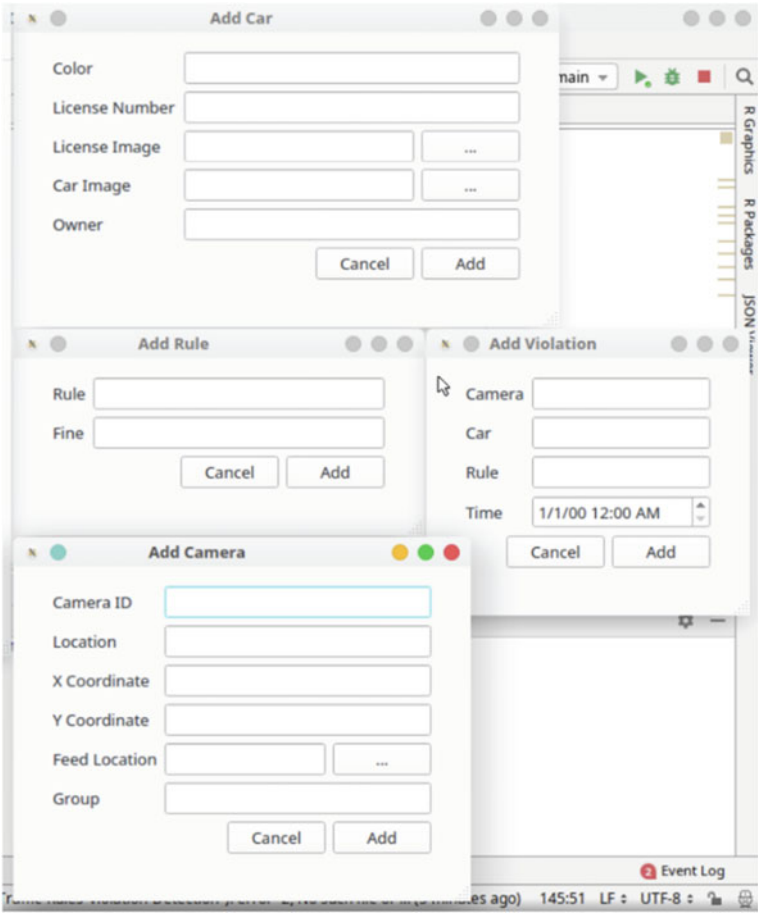


Fig. 5.6 Adding items interface

5.4 Results and Discussions

The goal is to automate the traffic rule violation system and also to monitor the traffic and also for taking the action against the rule violators. After identifying the violator details the e-challan will automatically generate and it will get sent to a registered email-id of the vehicle owner. The system detects various violations like a signal violations, parking violations, and wrong direction violations. The user-friendly GUI was designed to operate as well as to monitor the traffic and also for taking the action against the rule violators (Figs. 5.12 and 5.13).

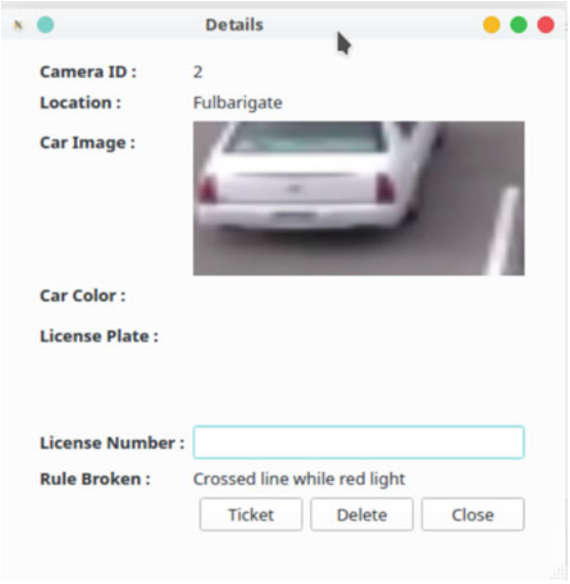


Fig. 5.7 Details of rule violation

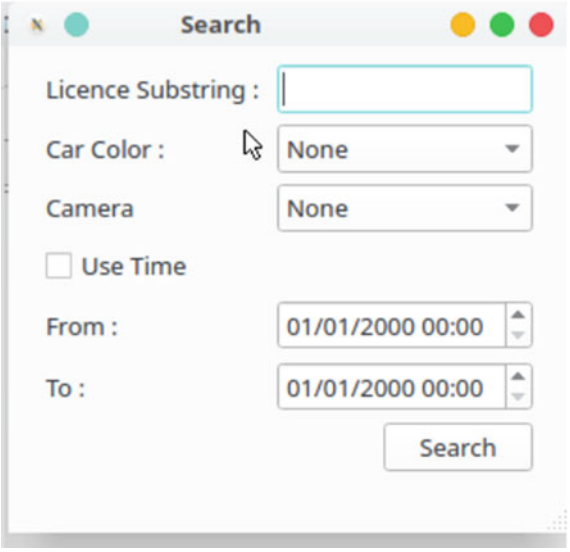


Fig. 5.8 Searching a car or rule violation

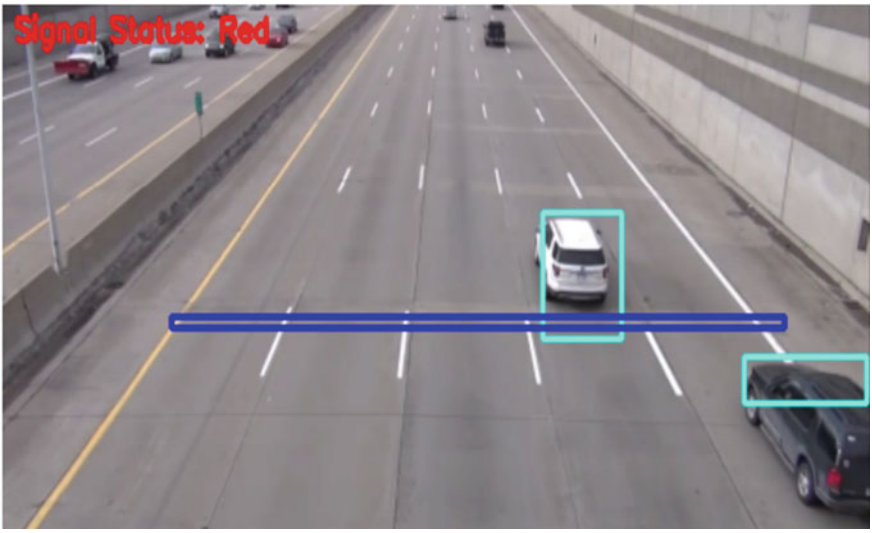


Fig. 5.9 Signal violation camera representation

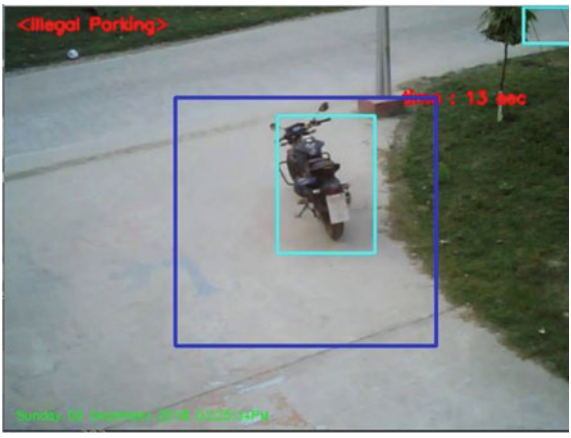


Fig. 5.10 Parking violation camera representation

5.5 Conclusion and Future Scope

5.5.1 Conclusion

In smart cities, every road intersection has cameras installed. It has been observed that due to the increase in traffic the rules violators are also increasing. This system takes the images captured by the cameras installed at various intersections for identifying

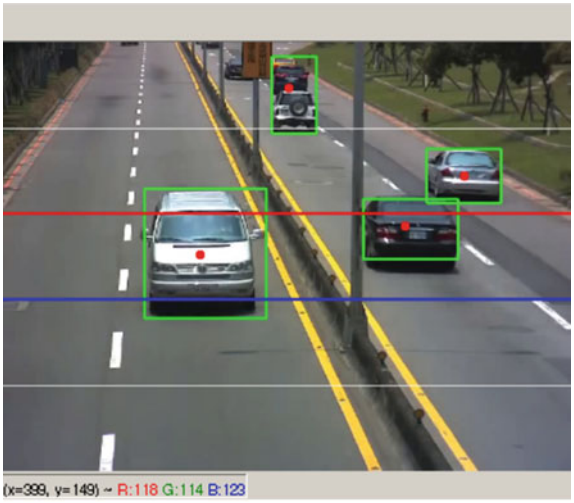


Fig. 5.11 Direction violation camera representation

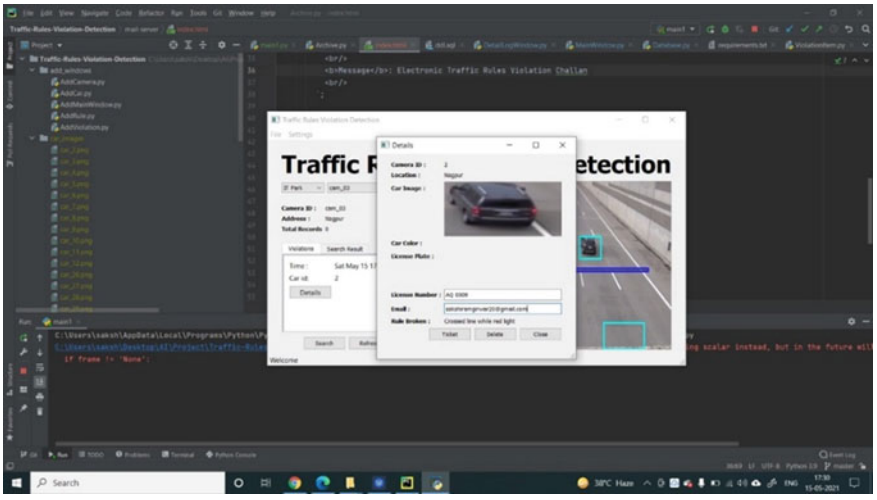


Fig. 5.12 Output

the rule violators. It is an automated system that will directly identify the vehicle that violates the traffic rules. Three types of violations have been considered like signal, parking, and direction violations. After detecting the vehicle and identifying the vehicle's number plate the e-challan will get generated and will be sent to the owner's registered email-id. This automated system will help in reducing the accidents as the rule violators will be issued with the e-challan. This system will contribute to the sustainable development of smart cities.

Fig. 5.13 E-challan

5.5.2 Future Scope

1. The system can be implemented for identifying the variable size number plates.
2. Developing new technologies will contribute to improving character recognition which will lead to more accuracy in violator detection in the smart city.
3. This system will also work for theft detection.

References

1. Artan Y, Bulan O, Loce RP, Paul P (2015) Passenger compartment violation detection in HOV/HOT lanes. In: IEEE winter conference in applications of computer vision, Steamboat Springs, CO, USA
2. Zhang X, Tian Y, Liang L, Huang T, Gao W (2012) Macro-block-level selective background difference coding for surveillance video. In: IEEE international conference on multimedia and expo, Beijing, China
3. Ou G, Gao Y, Liu Y (2012) Real-time vehicular traffic violation detection in traffic monitoring stream. In: IEEE/WIC/ACM, Beijing, China
4. Tembe AU, Thombre SS (2017) Survey of copy-paste forgery detection in digital image forensic. In: International conference on innovative mechanisms for industry applications (ICIMIA 2017). IEEE
5. Singh A, Gupta AK, Singh A, Gupta A, Johri S (2018) Vehicle number plate detection using image processing. Int Res J Eng Technol (IRJET), March 2018
6. Chuang C, Tsai L-W, Deng M-S, Hsieh J-W, Fan K-C (2014) Vehicle license plate recognition using super-resolution technique. In: 11th IEEE international conference on advanced video and signal-based surveillance (AVSS). https://doi.org/10.1109/AVSS.2014.6918703
7. Li H, Yang R, Chen X (2018) License plate detection using convolutional neural network and visual features. In: 3rd IEEE international conference on computer and communications, 19 November 2018
8. Lin NH, Aung YL, Khaing WK (2019) Automatic vehicle license plate recognition system for smart transportation. In: IEEE international conference on Internet of Things and intelligent system (IoTAIS), 7 January 2019. https://doi.org/10.1109/IOTAIS.2018.8600829
9. Chong L, Tianhua C, Linhao J (2013) Vehicle license plate recognition based on edge detection algorithm. In: International conference on intelligent information hiding and multimedia signal processing, May 2013. https://doi.org/10.1109/IIh-mSP.2013.105

10. Chaitanya Sai M, Chandwani B (2019) Advanced vehicle monitoring system with multi-object automatic number plate detection. In: Global conference for advancement in technology (GCAT), 18–20 October 2019. https://doi.org/10.1109/gCAT47503.2019.8978321
11. Ravirathinam P, Patamari A (2019) Automatic license plate recognition for indian roads using faster-RCNN. In: International conference on advanced computing (ICOAC), Chennai, 7 May 2020, 20 December 2019. https://doi.org/10.1109/ICoAC48765.2019.246853
12. Naren Babu R, Sowmya V, Soman KP (2019) Indian car number plate recognition using deep learning. In: International conference on intelligent computing, instrumentation and control technology (IEEE), 15–16 July 2019. https://doi.org/10.1109/ICICICT46008.2019.8993238
13. Jichkar N, Kapse S, Thombre S, Meshram A (2016) Survey: on vehicle number plates detection techniques. Int J Eng Innov Res 5(2). ISSN: 2277-5668
14. Chavanke P, Thombre S (2016) Depth based super-resolution for multiview images. In: IEEE sponsored world conference on futuristic trends in research and innovation for social welfare (WCFTR'16)

Chapter 6
Internet of Things Based Pigeon Pea Disease Detection Tool to Achieve Sustainable Development in Smart Farming

Smita Kapse, Nikhil Wyawahare, Rameshwari Kuhikar, Rashmita Nikhare, Pranali Maraskolhe, and Saloni Chinchmalatpure

Abstract In the agriculture field, the presence of plant disease and organic deficiency causes major harm and loss to farmers. The conventional way of detecting plant pathogens is through visual checking. Visual checking is mostly done after major damage to the crop, so treatments will be of limited or no use. To prevent the plant from severe damage and to achieve sustainable development in the farming sector, farmers must detect an infection before it is visible. This is achieved by premature disease diagnosis. By considering this problem as a challenge, the proposed system provides a solution by using the concept of Internet of Things to achieve sustainability in the smart farming domain. The proposed system aims at the development of an automated tool to identify whether the plant is normal or diseased. The main goal of the research work is the design and development of an Internet of Things (IoT)-based handheld tool to detect pigeon pea diseases. The proposed system attempts to design and develop an automated tool that senses the presence of disease in the plants using various sensors like temperature, humidity, and color. Identification of plant disease is done through sensor readings.

6.1 Introduction

India's principal source of income is agriculture. The most common type of disease attacks on the plants is a variant of bacterial and fungal diseases. The growing rate

S. Kapse (✉) · P. Maraskolhe · S. Chinchmalatpure
Computer Technology Department, Yeshwantrao Chavan College of Engineering, Nagpur, India
e-mail: smitarkapse@gmail.com

N. Wyawahare
Electronics Engineering Department, G. H. Raisoni College of Engineering, Nagpur, India

R. Kuhikar
Prolifics Corporation Ltd., Hyderabad, India

R. Nikhare
Birlasoft The CK Group, Pune, India

V. Bali et al. (eds.), *Decision Analytics for Sustainable Development in Smart Society 5.0*,
Asset Analytics, https://doi.org/10.1007/978-981-19-1689-2_6

of population and changing climatic conditions also cause plant diseases. The major challenge of sustainable development is to reduce pesticide use, lower production costs, conserve the environment, and improve product quality. Early detection of plant disease can help to limit the use of pesticides.

One of the most crucial peas harvested in semi-arid tropical regions is pigeon pea. Fresh seeds of it are eaten like a vegetable or dried and eaten as a pulse. Pigeon pea is used widely. Diseases are a primary biological limitation to production, and pigeon peas can be infected by more than 60% of pathogens.

Premature disease detection of crop health leads to the control of diseases. By considering the importance of sustainability in the smart farming domain, the goal of the proposed system is to develop a device that will help in the early detection of disease and will also suggest some fertilizers and pesticides for the plant. In this device, the detection can be done by using a leaf or by using its roots on the basis of color detection, temperature, and humidity values received from sensors. Cure suggestions will be given considering the color, temperature, humidity, and soil moisture level of the field.

The remainder of the chapter is ordered as follows: Sect. 6.2 mentioned a brief idea of previous work done in this research domain. Section 6.3 gives an idea about methodology and design, i.e., software use, database use, design diagram, etc. Section 6.4 describes the experimental results obtained. In Sect. 6.5, conclusion and future scope are discussed.

6.2 Literature Review

The term Internet of Things is a recent evolving study area where many researchers contributed from the last decade to achieve sustainable development in a diverse fields. This technology adds the third dimension to the Internet service, i.e., anything connection to the user in addition to anyplace and anytime connection [1]. Application of this area is widely spread in every aspect of a human being like smart city, smart farming, smart energy management, smart traffic management system, smart industry, and many more. This section describes a literature survey carried out by us in the area of the Internet of Things, smart farming, and plant disease detection.

In [2], early disease detection of tomato crops is proposed. This system makes use of a convolution neural network for the detection of tomato plant disease.

A literature study of [3] was also done using a convolution neural network to detect plant disease at an initial stage. It considers plants like corn, strawberry, grape, tomato, and potato plants.

In [4], the system is developed for the detection of rice leaf diseases using artificial intelligence techniques. The proposed system in [4] classifies the healthy and infected leaves of rice plants.

Nisar et al. [5] gives study of various researches done using image processing techniques for plant disease detection.

In [6], an IoT-based intelligent streetlight management system is developed which facilitates the saving of non-renewable sources of energy. This system also deals with the controlling flow of traffic and the pollution levels at various places by monitoring air quality which contributes to the sustainability of the smart city.

In [7], automatic detection and image recognition of the citrus plant is suggested. The researchers use YOLO (You Only Look Once) algorithm which is an object detection model to detect and recognize the diseases from citrus leaf images.

In [8], bibliographic analysis is done to show how deep learning makes smart farming smarter.

In [9], the authors developed an Internet of Things-based smart farming application for soil testing based on measuring and observing soil parameters. It aims at reducing the probability of soil degradation and mainlining crop health.

In [10], the study is carried out via an improved segmentation technique using a combination of thresholding and Morphological operations. This system uses a deep neural network for classification.

In [1], how the Internet of Things technology is a bliss to farmers is discussed. It elaborates how sensors are helpful in agriculture.

In [11], US patent is for detecting plant disease via image processing techniques. The proposed system in [11] uses a Bayes filter on a color-normalized version of the image which significantly reduces the amount of data to be processed, which allows the use of less powerful hardware or to free up resources of the system for another task.

The patent in [12] is for detecting plant disease and pest detection methods based on support vector machine learning.

The patent in [13] is for the detection of Fusarium species infecting corn using the polymerase chain reaction.

As discussed in the above literature survey, most of the research was done for plant disease detection using machine learning algorithms and image processing techniques. The problem with this is farmers do not have direct access to the system. So to solve this problem, the proposed system aims at providing a handheld tool that can be operated by the farmers to detect diseases of pigeon pea plants. By performing an extensive literature survey, we can claim that first-time research is done using IoT to detect pigeon pea plant diseases.

6.3 Methodology and Design

This section describes the hardware design diagrams, database use in the proposed research work.

For the implementation of the proposed work, Arduino IDE is used. For designing hardware module number of sensors like TCS3200 color sensor, DHT11 temperature and humidity sensor, and moisture sensor are mounted on node MCU.

6.3.1 Disease Database

The following disease database is used in the proposed module (Table 6.1).

Output Samples Values (Table 6.2):

6.3.2 Design Diagram

Following use case diagram of pigeon pea disease detection demonstrates the role of the various users in the proposed module (Fig. 6.1).

User Hardware Module

This is the model which the farmer will have with him. In this model, the disease name, as well as fertilizer suggestion, will be displayed on the LCD (Fig. 6.2).

Field Hardware Module

This is the second module that will be present in the field. This module will give us current values of temperature, humidity, and soil moisture of the field. The controller will send data to a specific IP address of the receiver's IoT board only.

The device will analyze all these values and work accordingly (Fig. 6.3).

6.3.3 Working of the Proposed System

The proposed system will have two modules. One will be installed in the field and another one will be on the user side.

Disease detection workflow:

1. User side module will have the feature of disease detection. It will have a slot to insert the leaf or root of the plant.
2. On pressing the switch button, it will start analyzing the colors of the leaf based on the database the device will detect the disease.
3. After detecting the disease, the device will also suggest fertilizer according to the disease.
4. Disease, as well as fertilizer, will be shown on an LCD display at the user's side.

Fertilizer suggestion workflow:

1. It will continuously receive the field's temperature, humidity, and soil moisture values.
2. Controller will send data to specific IP address of receiver IoT board.
3. After getting all these values (RGB, humidity, temperature, soil moisture), the device will analyze it and suggest a proper cure according to the database. With

Table 6.1 Disease database [13]

Disease name	Symptoms	Suggestion	Images
Yellow mosaic	Mungbean yellow mosaic disease is categorized by a bright yellow mosaic on the leaves of infected plants and leads to loss of mungbean	(1) Spot treat with least toxic, natural extermination products, like Safer Soap, Bon-Neem, and diatomaceous earth, to decrease the number of infection-carrying insects (2) Harvest-guard row cover will help keep insect pests off vulnerable crops/transplants and should be installed until bloom	
Eafminer liriomyza	This disease basis on damage to plants. Direct damage occurred by the larvae mining the leaf tissue, which leads to dryness, early leaf fall, and cosmetic destruction. This disease leads to the burning in fruit such as tomato and melon mostly in tropical and subtropical regions	(1) Spraying of Fenazaquin 10 EC (Magister) at 1 mL/L of water on 45 and 60 DAS (2) Scoundrel out the infected plants in the primary stages	
Wilt Fusarium Udum	Sudden yellowing of leaves; death of leaves; plant death; blackened tissue at the base of stem; symptoms may be present on only one side of plant	(1) Carbendazim seed treatment at 2 g/kg of seeds and Trichoderma viride at 2.5 kg/ha in FYM at 50 kg/ha recorded lowest mean wilt incidence of 11.38 percent with a highest mean yield of 969.18 kg (2) Trichoderma viride alone	
Anthracnose colletotrichum	Spots or blotches on the leaflets, pods, and seeds. Here on grown-up leaflets, the spots are dark brown, definite, somewhat angular, and usually delimited by veinlets. The spots are most conspicuous on the superior surface of the leaflet	(1) Seed treatment with Trichoderma viride at 10 g/kg of seed or Thiram (2 gm) + Carbendazim (1 gm)/kg of seed (2) Soil application—T. viride—2 0.5 kg ha and 50 kg of well-decomposed FYM or sand at 30 days after sowing	

(continued)

Table 6.1 (continued)

Disease name	Symptoms	Suggestion	Images
Sterility Mosaic	Sterility Mosaic Disease (SMD) is a Foliage symptom type of disease induced as a result of genotype–virus interactions. From left to right: a healthy pigeon pea plant	(1) Spraying with fenazaquin at 1 ml/liter rapidly when the disease appears and, replicates every after 15 days if required (2) Wintergreen oil 30EC (2%) was effective by recoding 19.0% SMD incidence with 58.7% disease reduction	

Table 6.2 Test sample values

Sample	Sample values								
	R	G	B	R	G	B	R	G	B
Healthy	50	234	47	50	228	55	47	228	59
Dark	7	228	128	7	237	128	–	–	–
Yellow	149	184	10	155	184	10	140	190	10
Spotted yellow	71	288	24	100	221	10	30	243	55
Whitish	42	221	78	–	–	–	–	–	–
Black dot	111	187	32	85	193	63	88	193	59

the help of node MCU, the problem of connectivity will get resolved as the communication will be over Wi-Fi.

6.4 Experimental Results

6.4.1 The Output of the User Module

This module is with the farmer. The farmer checks the quality of the leaf by inserting the leaf into this module. After inserting, our device will display RGB values of the leaf, associated disease if any (if the plant/leaf is healthy, it displays the message "Healthy leaf"). The proposed system also suggests the suggestion number by referring to a manual appropriately for that suggestion, farmers understand how to get rid of that disease by applying suitable fertilizer, etc. Figure 6.4 shows the output of the user module.

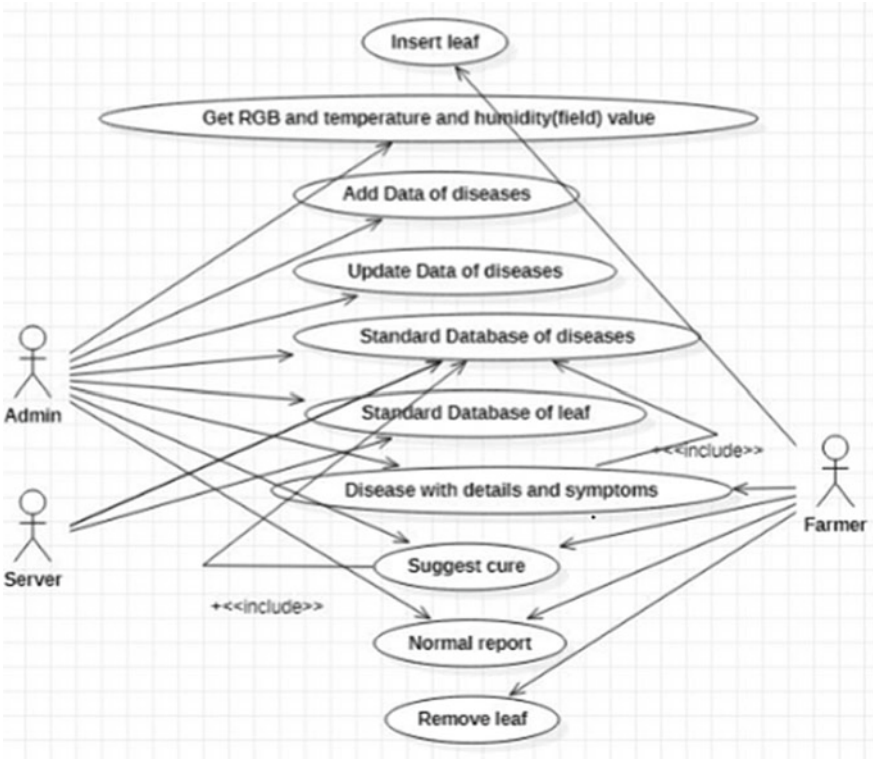


Fig. 6.1 Use case diagram of the proposed system

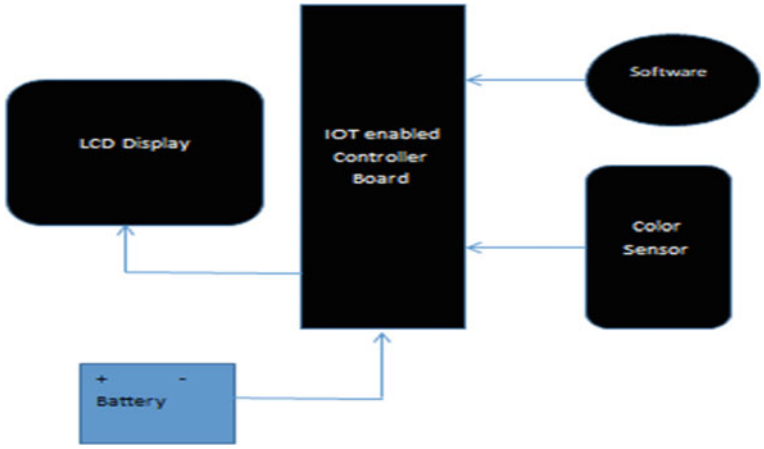


Fig. 6.2 User hardware module of the proposed system

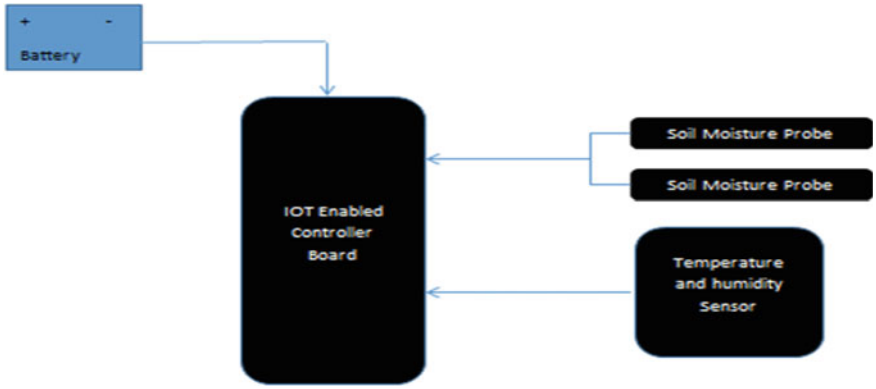


Fig. 6.3 Field hardware module of the proposed system

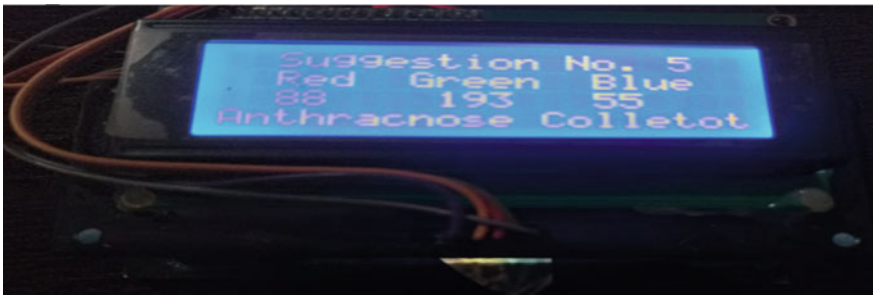


Fig. 6.4 Output of the user module

6.4.2 The Output of the Field Module

This module is placed at the field, this continuously monitors the temperature, humidity, and soil moisture level of the field. Figure 6.5 shows the output of the field module.

6.4.3 Cure Mechanism

After detecting a disease particular disease, the following suggestion number will be displayed on the LCD board of the user module. Accordingly, the user will apply a cure mechanism (Table 6.3).

Fig. 6.5 Output of the field module

Table 6.3 Disease management [14]

(1) Yellow mosaic
1. In order to decrease the amount of infection-spreading insects, spot treats can be done 2. The infection can also be blowout through activity of human, gardening tools, etc. Regularly washing hands and disinfecting garden tools (1:4 ratio of bleach to water) minimize the risk of infection 3. Use of tobacco around vulnerable plants should be avoided because these crops are infected first, then blowout the virus
(2) **Sterility mosaic**
1. Use of Fenazaquin 10 EC (Magister) at 1 mL/L of water on 45 and 60 DAS 2. Remove the diseased plants at an primary stage 3. Crops can be planted in turn by turn basis with non-host crops
(3) **Wilt fusarium**
1. For seed treatment, Trichoderma viride and Carbendazim can be applied 2. After seeding fertilizers are applied to the soil after 30 days 3. Multiple crops can be grown with sorghum 4. Remove droop plants 5. Do not give excess or less water to plants 6. Modification of soil can be done with oil cakes and trace elements can be used. Applying green leaf manure
(4) **Anthracnose colletotrichum**
1. At the starting of spring when leaves start to grow, spraying of sulfur powders and liquid copper every after 7 days should be done and continue it through the growing season. Always spray in the morning, and evade use during hot weather. Before planting, seeds can be treated 2. Apply neem oil at the early stage or schedule it for every 7 days until the problem is solved 3. Avoid manure of diseased part of the plant and carefully sanitize farm areas in the autumn, after yield, so as to minimize the fungal spores from over-wintering 4. If fungal infection is common problem then do not save your seeds from planting 5. To mitigate the effect of disease spreading, clean the garden tools regularly after use
(5) **Leafminer liriomyza**
1. During the growing season, the maximum temperatures in the first and second years were recorded as 29.4 and 27.3 °C in April and 40.2 and 35.0 °C in May, respectively

6.5 Conclusion and Future Scope

This chapter discussed the design and implementation of the Internet of Things-based handheld device for detection of disease accurately at a very early stage of the pigeon pea plant. Use of this tool results in the reduction of damage of the crop in the field of the farmer. The proposed model is designed to detect the quality of leaf by using various sensors such as color sensors, temperature, humidity, soil moisture, etc. The proposed tool detects whether the leaf under consideration is healthy or diseased and gives the solution or suggestion (e.g., fertilizer to be used) for that particular disease.

The proposed system is a dedicated design for the detection of pigeon pea plant diseases. In the future, this model can be made generic for multiple plant disease detection. This system can also be equipped with an automated soil moisture system [9].

References

1. Sen S, Madhu B (2017) Smart agriculture: a bliss to farmers. Int J Eng Sci Res Technol 197–202
2. Paymode AS, Magar SP, Malode VB (2021) Tomato leaf disease detection and classification using convolution neural network. In: Proceedings of international Bhurban conference on applied sciences and technologies, pp 346–351
3. Radha N, Swathika R (2021) A Polyhouse: plant monitoring and diseases detection using CNN. In: Proceedings of the international conference on artificial intelligence and smart systems, pp 966–971
4. Shahidur Harun Rumy SM, Jahan F, Tanvin T, Ishan Arefin Hossain M (2021) An IoT based system with edge intelligence for rice leaf disease detection using machine learning. In: Proceedings of International IOT, electronics and mechatronics conference
5. Nisar N, Awasthi A, Chhabra M, Abidi AI (2020) Image based recognition of plant leaf diseases: a review. In: Proceedings of fourth international conference on I-SMAC (IoT in social, mobile, analytics and cloud), pp 373–378
6. Alkari V, Telharkar S, Pathak N, Katkamwar T, Kapse S (2018) An intelligent automated street light controller and lively emission control system. Helix Sci Explor 3950–3953
7. Song CC, Wang CL, Yang YF (2020) Automatic detection and image recognition of precision agriculture for citrus diseases. In: Proceedings of 2nd IEEE Eurasia conference on IOT, communication and engineering, pp 187–190
8. Unal Z (2020) Smart farming becomes even smarter with deep learning—a bibliographical analysis. IEEE Access 105587–105609
9. Kapse S, Kale S, Bhongade S, Sangamnerkar S, Gotmare Y (2020) IOT enable soil testing NPK nutrient detection. JAC: J Compos Theory, May 2020, 310–318
10. Arifur Rahman Md, Mukitul Islam Md, Shahir Mahdee GM, Wasi Ul Kabir Md (2019) Improved segmentation approach for plant disease detection. In: Proceedings of 1st international conference on advances in science, engineering and robotics technology
11. Alexander J (2017) System and method for detecting plant diseases. US Patent, November 17, 2017
12. Landwehr VR (2012) Plant disease and pest detection method based on SVM (support vector machine) learning. CN Patent, September 20, 2012
13. Mach R (2003) Technische Universität Wien. Detection of fusarium species infecting corn using the polymerase chain reaction. Patent April 3, 2003
14. Reddy MV, Raju TN, Sharma SB, Nene YL, McDonald (1993) Handbook of Pigeon pea diseases. In: International crops research institute for the semi-arid tropics Patancheru, Andhra Pradesh, India

Chapter 7
IT Infrastructure for Smart City: Issues and Challenges in Migration from Relational to NoSQL Databases

Amit Kanojia and S. Tanwani

Abstract Smart city applications collect massive amount of data from various types of IoT sensors, engines, and people. Most of the data generated are heterogeneous in nature. However, this data need to be integrated with legacy applications based on SQL. Some of these applications also require migration to NoSQL for improved performance and fault tolerance. This paper addresses challenges of working in hybrid environments and migration issues from SQL to NoSQL databases. Rapid rate of growth in heterogeneous data and characteristics of NoSQL database like easy scalability, high availability, high performance, and low cost are the motivating factors to migrate toward NoSQL from relational databases, especially for applications requiring dealing with unstructured data. NoSQL database gives dynamic schema, adjustable data model, scale out architecture, and allows storage of big data and access to it in an efficient manner. The relational databases store data in form of tables with fixed schema. Relational databases are structured and not capable of handling unstructured and big data. In relational databases, because of normalization, data is spread across multiple tables and expensive join operation is required to integrate data. Due to limitations of relational databases, most of the leading organizations are migrating toward NoSQL. NoSQL databases are analyzed into four categories- Key-value database, Column-oriented database, Document-based database, and Graph database. Key research issues and challenges in migrating from relational to NoSQL include model transformation (including mapping and schema conversion), application integration, strategies related to perform join in different scenarios, use of indexes, storage issues, etc. Massive growth is likely in the area of NoSQL over cloud because of initiatives like Smart City, however, these applications required integration with legacy applications built over SQL. Thus need for migration as well as bridges for connection leading to requirements of our research area. This chapter tends to explore the comparative study of Relational databases and NoSQL

A. Kanojia (✉)
Department of Computer Science, Mata Jijabai Government Girls PG College, Indore, India
e-mail: kamit.ind@gmail.com

S. Tanwani
School of Computer Science & IT, DAVV, Indore, India

V. Bali et al. (eds.), *Decision Analytics for Sustainable Development in Smart Society 5.0*,
Asset Analytics, https://doi.org/10.1007/978-981-19-1689-2_7

databases, classification of NoSQL databases, case study, popular approaches for migration, and some of the key issues and challenges in migrating from Relational to NoSQL databases. This chapter will be fruitful for students and researchers too.

7.1 Introduction

A smart city comprises a major amount and variety of IoT sensors, engines, and people. A smart city endeavor to upgrade everyday operation. The Smart City idea is used to improve various metropolitan services with the use of ICT. In the Smart City, ICTs are used for improving the consistency, performance, and interaction of urban services. It reduces cost and resource utilization. Numerous smart city projects have already been initiated with new ideas, like smart health care, smart transportation, smart waste management, smart parking management, etc [1]. A smart city relies on data and uses large amount of data based on big data collected by IoT devices everywhere. Data generated in such type of applications are unstructured and heterogeneous, which cannot be managed effectively by legacy SQL applications. Thus, there is a need of migration from SQL to NoSQL and integration with legacy applications developed using SQL for sustainable development in smart society. Smart city projects use different ICT tools to make communications effective and innovative [2].

In this section, basic ideas of relational databases, NoSQL databases and their characteristics have been introduced. Carlo Strozzi in year 1998 coined the term NoSQL. NoSQL means "Not Only SQL". The huge amount of heterogeneous information generated from the web applications and business applications is difficult to deal with by RDBMS, which basically manages structured data. An alternative data model handling both structured and unstructured data was felt by business community. The demand led to development of NoSQL. NoSQL data sets are characterized as disseminated, evenly versatile, and open source [3].

RDBMS possess static schema and data are inserted strictly according to schema due to which data accessing rate becomes slower and difficult. RDBMS remains the top data storage technology, however, several industries wish to decrease operating cost to make scalable application that uses cloud computing technologies [4]. NoSQL databases do not have predefined schema that helps in making major application transformations in real time. It makes faster enhancement and more feasible to access data easily. NoSQL systems are high performance, scalable systems [5]. Various reasons for motivation to utilize NoSQL databases are high availability, scalability, distributed architecture, varied data structure, adaptable schema, and consistency and fault tolerance. NOSQL databases have many benefits such as faster read and write access and support massive storage at low cost. The scaling approach easily supports adding multiple servers. Data structure used by NoSQL is in form of key-value pairs [6].

NoSQL has conquered many restrictions of the relational databases [7]. Considering the above, many business applications presently on RDBMS technology intend to migrate over NoSQL. Therefore, in order to get a consistent and effective alternative to change the schema and migrate data to NoSQL from the legacy relational database applications, effective strategies and guidelines are required. The RDBMS is based on structured data model and uses structured query language (SQL). NoSQL offers a adaptable structure and doesn't adhere to standard SQL interface [3]. NoSQL databases are grouped into four most important categories (i) Key-Value Databases (ii) Document-oriented databases (iii) Column-oriented databases and (iv) Graph-oriented databases. Relational Databases follow ACID property whereas CAP theorem is used by NoSQL databases.

Researchers and industrialists exploring and implementing big data provide better opportunity to transform data into an appropriate form. They store and process the data and finally analyze it. With the rapid increase of data, issues occur about the organization, querying, and storage. Relational database management systems have limitations when data is accessible in different formats (such as semi-structured and unstructured data). NoSQL database system provides a streamlined database, supports multiple data models, and provides a schema-free architecture.

Data migration is the technique of relocating data between systems or storage units. For making this efficient, the data extraction, transformation, and loading methods should be implemented. The techniques will help in mapping data from existing system to the latest system which is being implemented. Database migration is the way toward moving the business logic, schema, physical data, and database dependencies from an existing system to a latest system. Database Migration is utilized when there is a necessity to move one database vendor to other. There are various reasons for this, such as cost, capabilities, functionalities, and requirements. Generally data migration occurs during an upgrade of existing hardware or transferring to a totally new system [6].

The rest of the chapter is organized into eight sections. After introducing the main concepts of relational databases and migration from relational to NoSQL database we proceeded further to discuss the related work in Sect. 2. Section 3 describes smart city requirements, Sect. 4 includes a comparative study of relational and NoSQL databases. Section 5 describes the classification of NoSQL databases. Various migration approaches with some examples have been discussed in Sect. 6. The challenges and some of the key issues are described in Sect. 7. Section 8 illustrates case study on MongoDB database. Section 9 concludes the chapter and outlines future work.

7.2 Related Work

The most proficient strategy to migrate a database relies upon the application availability and the computational resources accessible for it [8]. The technique also helps to choose a correct migration strategy to utilize and the metrics that can be gathered to more readily assess the exhibition of a migration. The paper not only

recognizes the current migration methods but also recommends receiving the most effective strategy for migrating relational databases to NoSQL databases. Results show that migration can be successfully performed, if semi-automatic migration approach including manual intervention is the most competent methodology.

Outline of method for mapping relational databases with NoSQL databases is explained in [9]. Different methods for data conversion and middle layer solutions are summarized. In their work, they introduced some research on new research on data migration from relational databases to NoSQL databases.

Methods of Migration the relational model to the document-oriented model is presented in [5]. A migration methodology to MongoDB is proposed, beginning with a technique for extracting a data model from a relational database through metadata. Next, it describes on how to carry on with schema and data migration preserving the integrity constraints. The paper can be further extended to introduce an object modeling in the NoSQL databases, which will affect the best approach for data storage, selection, insertion, and modification time. A prototype is created to show the feasibility and efficiency of the proposed approach.

Reference [8] Gives comparison of the performance of the three migration methods, using metrics that provides with detailed perspective of resource utilization as well as details about migration is being verified. This study shows how techniques of migration from relational databases to NoSQL and their related metrics can be utilized. It is concluded that it is consistently important to approve resources inaccessibility as prerequisite of the projects and not the migration itself [6]. Database migration explains migration process in a phased manner. Migration, the intensive facts of what is expected for the proper design of migration can predict the potential problems that can occur during migration and may decrease the possibility of failure.

Reference [3] Proposed an approach of model transformation and migration of data from RDBMS to MongoDB. The proposed work is separated into four parts (i) considers the query characteristics and data characteristics (ii) designs a model transformation algorithm based on description tags and action tags (iii) automatically migrates the data into MongoDB based on the result of model transformation (iv) develop an useful tool. Experiments demonstrate that utilizing this methodology can accomplish a better performance. They give a tool which enable users to handily perform transformation and migration. Utilizing this tool, users can pick appropriate NoSQL database to replace relational database. The choice of appropriate mid-model is important for the effectiveness of the model transition and data migration between Nosql databases and relational databases [7]. Data and query features are introduced into mid-model. This model, if implemented as a tool can also serve as an interface for DBA.

Reference [10] Girts suggest an approach for faster migration of data from relational database to NoSQL document-oriented databases. Two logical levels using automated means allowing user intervention to further update the logical level are developed over physical data. The data migration function is implemented in the DigiBrowser relational database browser. The proposed method helps to get proof of ideas for a new document-oriented database solution in two to three days.

In [11], Hassan, Rondik J. et al. research show that the Internet of Things has proven to be the biggest promoter and promoter of smart city plans. This is a key step in the transformation of traditional public facilities into smart services and the creation of new services. Big data and cloud computing play a central role in advancing these new projects because they can expand existing work by providing ever-expanding information and computing power. The data generated in smart city projects is heterogeneous and unstructured in nature. NoSQL can easily store unstructured data due to its flexible model. Reference [12] describes numerous NoSQL benchmarks made by the scientific community and mentions HBase's ability to provide citizens with energy bills reinvented in smart city services. The effect confirms that Cassandra and HBase are generally diagnosed as appropriate general actors and can be followed without much concern. The structural design provided has passed ETL verification and predicts big data in the smart city environment. The architecture proposed in this chapter, NoSQL, now plays a vital role in the Internet and mobile applications.

7.3 Smart City Requirements

Smart cities can be defined as developed urban areas that create sustainable economic development and high-quality life through outstanding performance in many key areas such as economy, mobility, environment, people, life, and government. Smart cities use various technologies to improve the infrastructure, energy, learning, well-being, and water supply efficiency of their residents. Generally speaking, a smart city must have:

(i) Qualified supply consisting of public services such as water, electricity, solid waste, sanitation and sewage treatment, and related government services.
(ii) Supply–demand matching system ground transportation services, providing no Congested roads and shortest waiting time for public transportation.
(iii) Active surveillance in the city to provide the public safety that citizens urgently need.
(iv) Provide reliable emergency services as needed, such as ambulances, fire safety, etc.

7.3.1 Smart City Applications

1. ***Smart Traffic Management***

Smart Traffic Management includes modern technologies and services to enhance and innovate the urban traffic experience, solve traffic problems such as traffic congestion, and reduce injuries. The evolution of the scale and trends of urban traffic has raised problems that affect the well-being of cities. This type of transport has a great impact

on the atmosphere of the city: a large number of vehicles that use energy cause noise and air pollution, which causes climate change and damage to the atmosphere.

2. ***Cost-Effective Energy Efficient System***

The clean and green energy digital grid is one of the most attractive prospects for smart cities. For example, smart solar or wind power plants can become an important part of the smart city ecosystem.

3. ***Improving Accessibility of Healthcare Systems***

Smart health care is a combination of technologies designed to maximize the survival rate and quality of life of the population. Smart healthcare networks use mobile, Internet of Things, and computing technologies to ensure accurate diagnosis and transform health care.

4. ***Efficient Parking System***

The parking problem is some worldwide problems that famous urban residents often face, but it is also difficult to solve. Smart city technology can help detecting number of empty parking slots immediately, and reduce the time and confusion that usually comes with parking.

5. ***Smart Irrigation Systems***

In traditional irrigation systems, due to inefficiency, most of the water and other resources are wasted. Smart Irrigation Systems uses IoT sensors to monitor the schedule and run times according to needs. Smart Irrigation sensors monitor plant water, weather, and soil conditions to automatically adjust the watering schedules.

6. ***Smart Homes***

It includes Smart lighting, appliances, gas detectors, etc. Smart lighting systems control lighting of homes remotely via mobile or web applications. Smart appliances give status information of appliances to the users remotely. Smoke detectors use optical detection, ionization, or air sampling techniques to detect the smoke. Gas detectors can detect harmful gases and raise alerts to the users.

7. ***Weather Monitoring***

Weather monitoring systems provide information about weathers. In such systems data collected from several sensors (temperature, humidity, pressure, etc.) and this data is sent to cloud-based applications and storage back-ends.

8. ***Air Pollution Monitoring***

Air pollution monitoring systems are the systems that monitor emission of harmful gases like carbon dioxide, carbon mono oxide, nitrous oxide, etc. Factories and automobiles use gaseous and meteorological sensors. These systems require Integration with a single-chip microcontroller, several air pollution sensors, GPRS-modem, and a GPS module.

7.4 Comparative Study of Relational and NoSQL Databases

NoSQL database is distinctive in numerous perspectives from conventional databases such as complexity, schema, transaction methodology, and dealing with accessing big data. The IoT applications in smart city projects usually have storing big data. With rapid growth in the clients and sensor data, it is not easy for legacy SQL system to handle big users. SQL systems store data in form of static tables and data is spread across multiple tables. To access the relevant information, expensive join operation is needed. NoSQL databases have been introduced to conquer this problem.

7.4.1 Characteristics of NoSQL Systems

Following are the characteristics of NoSQL Systems (Table 7.1).

7.4.2 Advantages and Disadvantages of NoSQL Databases

Advantages of NoSQL Database [13]

- In relational database data is structured and spread across different tables, due to this, there may be a possibility of complex joins and affect the performance of the

Table 7.1 Characteristics of relational and NoSQL databases Source: [32]

Characteristics	Relational databases	NoSQL
Data model	Concept of set and relations in mathematics	Key-value stores, graph, and document data model
Scalability	Vertical scalability	Horizontal Scalability
Transactional reliability	Fully support ACID	Range from BASE to ACID
Handling big data	Complex	Designed to handle big data
Crash recovery	Guarantee crash recovery via recovery manager	Depend on replication as backups to recover from crash
Datawarehouse	Performance degrade due to big data	High performance, scalability, availability in storing big data
Complexity	Rises because structure of DB could be quiet complex, difficult and slow working	Have capabilities to store unstructured, structured, or semi-structured data
Security	Adopted very secure mechanism	Has shortage in security because focus is on other purposes than security

system. NoSQL databases can store structured, semi-structured, and unstructured data format and provides high flexibility and better performance.
- NoSQL databases are well suited for cloud-based application.
- NoSQL has faster speed, easy scalability, high efficiency, and flexibility.
- Regarding performance and retrieving Big Data, NoSQL plays a major role by giving techniques to deal with Big Data.

Disadvantages of NoSQL Database

- Security is one of the important concerns of NoSQL databases. Due to lack of standardization across various vendors and models, it is difficult to provide a unified security solution.
- As compared to relational databases, maintaining NoSQL databases is complex.
- Relational database has SQL as its own standard query language, but NoSQL does not.
- Some NoSQL database vendors and models do not comply with ACID properties.

7.4.3 Advantages and Disadvantages of Relational Databases

Advantages of Relational Databases

- Security methods are well established. However, security methods are in evolving stage in NoSQL database systems.
- When contrasted with NoSQL, RDBMS is very simple to utilize and easy to manipulate data and maintain data integrity with reduced redundancy and replication.
- As compared to NoSQL, the data independence across logical and physical level is better in Relational databases.
- Backup and Recovery mechanisms are sound in relational database system.

7.5 NoSQL Databases Classification

NoSQL databases are classified as

(i) Document-oriented database.
(ii) Key-value database.
(iii) Column-oriented database.
(iv) Graph-oriented database.

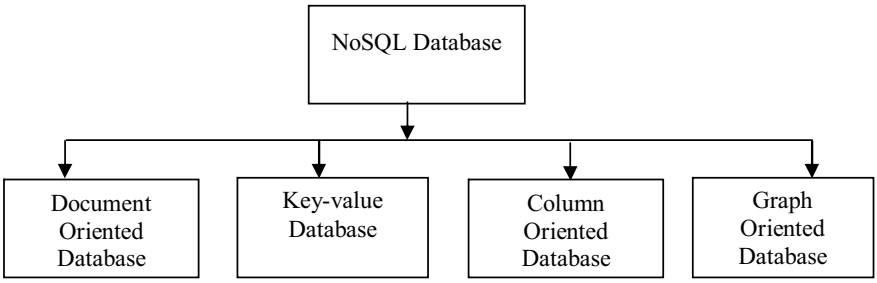


Fig. 7.1 Classification of NoSQL Databases Source: [27]

7.5.1 Document-Oriented Databases

Document-Oriented databases resemble Key-Value databases with the refinement that values are perceptible and can be queried. Document-oriented datasets utilized JSON or XML data format for storing data. Unlike key-value storage, it provides indexing and value-based queries. These databases store your data in the form of documents. The documents here are perceived by a unique set of keys and values, which are relatively the same as those in the key-value database. Document-based databases are schemaless and changeable in nature [10, 14]. The different characteristics of document-oriented storage are horizontal scalability and fragmentation between cluster nodes. Examples of document-oriented storage include MongoDB, Amazon DynamoDB, CouchDB, CouchBase, and so on [10].

7.5.2 Key-Value Databases

The simplicity of Key-value stores gives an incredible and effective methodology contrasted with different types of NoSQL database systems. Key-value database is a grouping of two components: Key and Values. Key is a unique identifier to particular information. The Value is a type of information indicated by the key [14]. Key-value store efficiently performs fundamental CRUD (Create, Read, Update, and Delete) operations. Compared to relational databases, the Key-value data model provides many features, such as highly scalable distributed data, big data support, high concurrency, and data retrieval speed. They also provide sharding across cluster nodes. Fragmentation is a horizontal division procedure used to divide a large amount of information into convenient, small, and effective pieces. However, the key-value database is not very flexible when querying and indexing complex and connected data. Queries in this category are usually based on keys rather than values. Examples of key-value storage are Redis, Memcached, Riak KV, Hazelcast, Ehcached, OrientDB, Aerospike, Amazon basic DB, etc. [10].

7.5.3 Column-Oriented Databases

Column-oriented databases are important when there is a need to deal with massive amount of data. Column stores in NoSQL are basically mixed row-column stores, not at all like pure relational column databases. In column-oriented database, each key is associated with one or more attributes. Column-oriented database stores information in such a way that columns that are needed and accessed at the same time are stored in one location in storage so that they can be accessed quickly with less I/O. It focuses on high scalability in data storage. One of the significant qualities is horizontal scalability. Examples of Column-oriented databases are Hbase, Accumulo, Hypertable, Google Cloud Bigtable, Sqrrl, ScyllaDB, and MapR-DB [10, 14].

7.5.4 Graph-Oriented Databases

Graph databases are based on the Graph theory. It deals with entities and their relationships in form of nodes and edges respectively. The graph is composed of vertex and edges, where nodes are objects and edges are the relationship among the objects. In graph databases, the relational table is replaced with structured relational graphs of interconnected key-value pairs. The graph consists of attributes related to nodes. It uses an index-free adjacency technique.

In this method every node comprises an immediate pointer which points to the adjacent node results in traversing millions of records. Graph databases focus on the relation established between data using pointers and offer schema less and well-organized storage of semi-structured data. Graph databases are faster than relational databases especially when identifying missing links and connections because in Graph databases queries are expressed as traversals. Graph databases support ACID property and support rollback [14].

Every real-world scenario can be corresponding to graphs and can also be modeled on graph database, because graphs have expressive and powerful modeling characteristics. Querying the graph database is more cost-effective because the graph query language does not require dense connections [10].

7.6 Methods of Data Migration

F. Matthews and C. Schulz describe the term data migration as follows: “Tool-supported one-time process which aims at migrating formatted data from a source structure to a target data structure whereas both structures differ on a conceptual and/or physical level”. There are two important steps in data migration: first, restructuring the source data to meet the requirements of the target system, and second, moving the data from the source to the target database. The document provides

several methodologies to manage the following steps: schema transformation, meta model approach, ETL (Extract, Transform, and Load), program transformation, model migration, automatic migration data transfer [10]. Following are some popular methods of data migration.

7.6.1 Mid-Model Approach [7]

The model proposed depends on two basic ideas: Data feature and query feature. The most fundamental element of the model is its generic nature, which can be modeled for any NoSQL database on demand. This model not only migrate data from relational to NoSQL but also maintains the integrity of data which was present in relational databases. This model not only migrates relational data to NoSQL, but also maintains the integrity of data already present in the relational database. Two strong points of Mid-model: (1) The data feature contains some characteristics of the relational database while (2) the query feature specifies the queries The principle challenge to mid-model design was to discover the distinctions that are present in the models of Relational and NoSQL databases. The middle model has an object because the base model is like an entity in a relational database that has a number of properties that describe the characteristics of the entity and its relationships with other objects.

7.6.2 NoSQLayer Approach [15]

NoSQLayer ensures that data is migrated from relational databases to NoSQL without modifying application code. The framework acts as an interface between the database and the application. The proposed solution is based on the use of two phases: migration and mapping. The framework contains two modules; The migration module ensures a consistent migration of data from relational databases to NoSQL databases using the metadata information contained in the data dictionary and the mapping module, providing the basis for running MySQL queries so that the legacy application can continue without modifications. It preserves the semantics of the source database and allows programmers to easily query the database by allowing them to code with a query on a relational database. The queries coming from the application are captured by NoSQLayer and converted to a specific NoSQL database format, the results generated by the NoSQL database are again captured by the framework and converted to tools specific to the application format. Mediator is an open source tool used to communicate between the application and the database. The results of the evaluation show that the framework is viable. However, the cost brought by the layer is very large when low volume data is involved in the operations, but it decreases

as the amount of data increases and thus NoSQLayer gives efficient results. The proposed framework works best with huge volumes of data correlating with MySQL databases. Compare the execution time of different queries with the evolution of data volumes and the proposed system is considered to be an efficient solution for data migration from relational databases to relational.

7.6.3 Data Adapter Approach [9]

Data Adapter system is mainly used for utilizing hybrid database. Data Adapter framework integrate RDBMS and NoSQL databases using Map Reduce process. Result demonstrates that the data adapter takes very less time to parse each query to get required data and information. Test data of Amazon Elastic Map Reduce is utilized as their relational database information source.

7.6.4 Automatic Mapping Framework [9]

It is a framework that provides automatic mapping of relational databases to MongoDB. MongoDB has a key and value pair. To map a relational database to a MongoDB collection, rows from the relational database are mapped as documents. Mapping a relational database to NoSQL involves thinking about the relationships that exist between different database tables. The easiest way to map in a 1:1 relationship includes embedding. Embedding and linking are commonly used in a 1:M relationship, while one-way or two-way embedding is used in an N:M relationship. The framework uses source information metadata from the database whether the relation contains table names, data types to be stored and access privileges, etc. The mapping steps include: creating a MongoDB database, creating a new table in the database using relational database metadata, and relating one table to another through primary keys and foreign keys.

7.7 Issues and Challenges in Migration

The use of Internet of Things (IoT) sensors is dramatically increased in smart city projects because they are cheaper, consume less power, are smaller, and are easier to use. IoT sensors typically collect large amounts of data and send it to a remote server. Data collected by sensors is heterogeneous i.e., contains images, videos, sound, fingerprints, Iris scan, text, etc. This data must be analyzed and accessible for various purposes. Legacy SQL systems are not capable of handling unstructured, heterogeneous, and big data generated in smart city projects. Thus migration as well as bridges for connection with legacy SQL systems is needed. At the moment, it

is difficult to find an appropriate process and strategy for migration from relational databases to non-relational databases effectively, due to following major issues and challenges [3].

7.7.1 Model Transformation

Since no special tools exist, most model transformation techniques depend on the experience of the DBA. Therefore, the physical model of NoSQL is often designed by individuals based on existing relational databases. Very inconvenient when the tables are involved and complex with a large number of attributes. Likewise, it is difficult to choose which tables should be integrated with each other and which tables should use references correctly. From a certain point of view, if we integrate all related tables, it can improve performance while causing data redundancy. Again, NoSQL does not support join operations. If we use a per-table reference, NoSQL will issue multiple queries when reading related documents. Reading related documents through different queries can lead to poor performance. A performance compromise and data redundancy technique works best. In addition, different applications may require different methodologies when converting models. Sometimes manual intervention is required to ensure that the new model is exactly what it is intended to be and that it will meet the requirements.

7.7.2 Data Migration

Currently, data migration tools use basic methodologies, such as moving each database table into a collection. Currently, no existing job can automatically migrate data based on model information. Semi-automatic tools with partial manual intervention are available. These tools provide a straight forward technique to migrate relational databases to NoSQL databases. The Extract-Transform-Load engine implements an interface with several mainstream NoSQL database systems to perform data migrations, such as MongoDB, Hbase, and Cassandra. The only downside is that these ETL systems are therefore unable to map the original database to the target database. ETL tools can only complement the information migration process, forcing users to define their own mapping strategies between different databases.

7.7.3 Schema Conversion

Wu-Chun Chung and Chongxin Li suggest two thoughts on independent modification of the database schema in HBase. The first idea uses the map-reduce framework. This framework converts each table in the database into an HBase table. After the

conversion, each table turns into a column family of HBase. The second reflection uses three rules to handle schema conversions. Related tables are nested and converted to an HBase table. However, it did not take into account multi-level nesting. To convert relational databases to NoSQL databases, some focus on transformation and NoSQL models, while others focus on data migration. Very few studies are based on the combined approach [3].

7.7.4 Strategies to Perform Join

In a relational database, data is stored as static table. Due to normalization, data is spread across multiple tables and while the user wants to access the relevant data as per the requirement, expensive join operation needs to be performed. Join is bulky operation that incurs cost and affects the performance adversely, if not implemented in proper manner. We need to think about how to minimize the cost and maximize the performance when performing join operations. It is evident that different applications and scenarios may require different strategies to execute join operation during migration process. In NoSQL databases, no join operation is required. When migrating from relational to NoSQL, we must define strategies to handle join. There may be a variety of scenario, for each of these, strategy needs to be defined.

7.7.5 Use of Indexes

In any database, the index is an important modifiable performance factor and then an important factor in schema design. NoSQL supports various types of indexes like binary, hash, unique, compound, array, sparse, etc. which are used to increase the data retrieval efficiency. Therefore, while migrating toward NoSQL, a strategy regarding which type of index can be used is a vital consideration. Index selection is based on the type and frequency of the queries used in applications. As with all databases, indexing is not free it incurs overhead cost and forces overhead and causes an overload of writing and resource usage (disk and memory).

MongoDB creates an index on the document's primary key _id field. All indexes are characterized by users as secondary indexes. Any field can be used for the secondary index, including fields inside sub-documents and Arrays. For optimizing performance, query optimizer chooses the index imperially by periodically executing substitute query plans and choosing the arrangement with the best response time [16].

7.8 Case Study on MongoDB

7.8.1 Comparison—SQL and MongoDB

MongoDB is a NoSQL database management system introduced in 2009. It stores data as JSON type documents with dynamic schemas. Its main purpose is to support big data handling. MongoDB is focused on four factors: flexibility, power, speed, and simple use. MongoDB has schemaless document-oriented data model whereas SQL supports relational data model. Relational databases have a standard language SQL though MongoDB supports API calls.

MongoDB has aggregate functions, a built-in map-reduce function that can be used to aggregate large amounts of data. The integrity model used by relational databases is ACID, while MongoDB uses BASE. MongoDB provides consistency, durability, and conditional atomicity. Relational databases provide integrity options that MongoDB does not, such as transactional, isolation, revision management, and referential integrity. On the distribution side, MongoDB and relational information are horizontally scalable and support data replication. While MongoDB supports sharing, relational databases do not. MongoDB and SQL are both cross-platform DBMSs. MongoDB is a free database system, while a license is required to use relational databases [17].

7.8.2 Features of MongoDB

- MongoDB provides high performance.
- Support basic CRUD operations and provide Aggregation.
- MongoDB gives High accessibility and auto-Replication feature. Data is restored through backup (replica) just in case of server failure.
- Provides automatic failover mechanism.
- A database in MongoDB is set of collections, which in turn holds collection of documents. A document is equivalent to record in database.
- Sharding is an important feature supported by MongoDB which provides horizontal scalability.

7.8.3 Advantages of MongoDB

- Installing and Setting up MongoDB is easy.
- MongoDB possesses schemaless structure and provides easy scalability.
- MongoDB supports dynamic queries.
- In MongoDB, no complex joins are required.
- MongoDB provides faster data access.

- As compared to relational databases performance improvement is done easily in MongoDB.
- MongoDB supports horizontal scaling whereas Relational databases support vertical scaling.

7.8.4 CRUD Operations in MongoDB

This section describes the basic CRUD operations in MongoDB. Two databases, one using SQL and one in MongoDB, were created to compare how data will be created, selected, inserted, and deleted in the two databases [18]. MongoDB provides all major CRUD operations and provides aggregation functionality.

Table 7.2 describes the main CRUD operations in MongoDB.

Smart city application stores records in the form of normalized tables. If it is required to retrieve relevant data, expensive join operation needs to be performed.

Table 7.2 CRUD operations in MongoDB Source: [40]

Operations	SQL	MongoDB
Create table	CREATE TABLE Accounts (first_name‘ VARCHAR(50) NULL, ‘last_name‘ VARCHAR(50) NULL, PRIMARY KEY (‘id‘));	db.accounts.insert({ name:”abc”, age:26, address:”indore”})
Delete a table	Drop table accounts;	db.accounts.drop()
Insert	Insert into accounts(name, age, address) VALUES (“abc”, 26, “indore”)	db.accounts.insert({ name:”abc”, age:26, address:”indore”})
Select	Select * from accounts	db.accounts.find()
Select fields	Select first_name, last_name from accounts	db.accounts.find({}, { first_name: 1, last_name: 1})
Conditional select	Select * from Accounts where dep_wid = ”D” and balance > 5000	db.accounts.find({dep_wid:”d”, balance:{$gt:5000}})
Ordered select ascending	Select * from accounts order by user_id asc	db.accounts.find({}).sort({user_id: 1})
Ordered select descending	Select * from accounts order by user_id desc	db.accounts.find({}).sort({user_id: -1})
Select with count	Select count(*) from users	db.users.count()
Update	Uupdate table student set section = “F” where marks <30	db.Student.update({marks:{lt:30}}, {$set:{Section:”F”}})
Delete	delete from Student	db.Student.remove()
Delete with condition	delete from Student where section = ”a”	db.student.delete({section:”a”})

Instead if this application is being migrated to NoSQL MongoDB, one record can be centralized. Other records may be kept on server. As data grows, more servers can be added which provide better scalability and availability. MongoDB supports database sharding feature which allow data packets are stored across different machines to ensure the system does not fail as volume rise. Data stored in smart city application is heterogeneous such as text, image, and biometric data (fingerprints, Iris, face recognition). MongoDB can efficiently store huge volumes of biometric data and images, whereas many other management systems, such as SQL, are less suited for image storage. In MongoDB multiple data can be stored in single collection in the form of embedded documents. It provides better flexibility in accessing data as no join is required. It reduces cost by avoiding multiple joins and also data access speed is increased. In smart city applications, biometric data need to be stored to enhance verification and security mechanism. MongoDB stores fingerprints and takes less time to access it as compared to relational databases.

7.9 Conclusion and Future Work

Relational databases are structured and have limitations when required to handle unstructured and big data. In relational databases, because of normalization, data is spread across multiple tables and expensive join operation is required to integrate data. Due to limitations of relational databases, most of the leading organizations are migrating toward NoSQL. NoSQL databases are well suited for the large volume, reliable and high availability web applications. Scale out approach and performance favors NoSQL.

This chapter tends to explore comparative study of Relational and NoSQL databases, classification of NoSQL databases, case study, popular approaches for migration and some of the key issues and challenges in migrating from Relational to NoSQL databases. The chapter is useful for students and researchers to investigate migration strategies for migrating from Relational to NoSQL databases. Smart city applications collect massive amount of data from various types of IoT sensors, engines, and people. Most of the data generated are heterogeneous in nature. However, this data need to be integrated with legacy applications based on SQL. Some of these applications also require migration to NoSQL for improved performance and fault tolerance. This paper addresses challenges of working in hybrid environments and migration issues from SQL to NoSQL databases.

Key research issues and challenges in migrating from relational to NoSQL include model transformation, data migration, schema conversion, application integration, strategies related to perform join in different scenario, use of indexes, storage issues. In model transformation, we need to consider how to make compromise among performance and data redundancy while transforming the schema. Moreover, various applications may require different strategies at model transformation. Semi-automatic strategies requiring manual intervention are required to tweak the model after it is transformed so as to exactly suit the application requirements. We

need to think about how to minimize the cost and maximize the performance when performing join operations. It is evident that different applications and scenarios may require different strategies in performing join during migration process. There may be a variety of scenario, for each of these, strategy needs to be defined. Indexes are the important performance factor in any database and are therefore crucial for schema design. So while migrating toward NoSQL, strategies regarding which type of index can be used are a vital consideration. On the basis of key issues and challenges addressed in this chapter, strategies for migration from relational to NoSQL for the potential applications of NoSQL databases can be proposed. A fully automatic methodology is proposed and under development using machine learning techniques, which will be based on keeping history data of past manual interventions and selecting one of the best interventions on the basis of application and database parameters. Massive growth is likely in the area of NoSQL over cloud because of initiatives like smart City, smart society however, these applications required integration with legacy applications built over SQL. Thus need for migration as well as bridges for connection leading to requirements of our research area.

References

1. Li F et al (2019) Toward semi-automated role mapping for IoT systems in smart cities. 2019 IEEE international smart cities conference (ISC2). IEEE
2. Sallow AB, Sadeeq M, Zebari RR, Abdulrazzaq MB, Mahmood MR, Shukur HM et al, An investigation for mobile malware behavioral and detection techniques based on android platform. IOSR J Comput Eng (IOSR-JCE) 22:14–20
3. Jia T, Zhao X, Wang Z, Gong D, Ding G (2016) Model transformation and data migration from relational database to MongoDB. In: 2016 IEEE international congress on big data (BigData Congress), pp 60–67. IEEE
4. Kuderu N, Kumari V (2016) Relational database to NoSQL conversion by schema migration and mapping. Int J Comput Eng Res Trends 3(9):506
5. El Alami A, Bahaj M (2017) Migration of relational databases to NoSQL: the way forward. In: 2016 5th international conference on multimedia computing and systems (ICMCS), pp 18–23. IEEE
6. Soni P, Koushal DS, Survey paper on data migration techniques of RDBMS to NoSQL
7. Liang D, Lin Y, Ding G (2015) Mid-model design used in model transition and data migration between relational databases and nosql databases. In: 2015 IEEE international conference on smart city/SocialCom/SustainCom (SmartCity), pp 866–869. IEEE
8. Oliveira F, Oliveira A, Alturas B (2017) Migration of relational databases to NoSQL-methods of analysis. In: 7th international conference on human and social sciences, pp 121–128. Richtmann Publishing
9. Ghotiya S, Mandal J, Kandasamy S (2017) Migration from relational to NoSQL database. In: IOP conference series: materials science and engineering, Vol 263, No 4, p 042055. IOP Publishing
10. Karnitis G, Arnicans G (2015) Migration of relational database to document-oriented database: Structure denormalization and data transformation. In: 2015 7th international conference on computational intelligence, communication systems and networks, pp 113–118. IEEE
11. Hassan RJ et al (2021) State of art survey for iot effects on smart city technology: challenges, opportunities, and solutions. Asian J Res Comput Sci, 32–48

12. Costa C, Santos MY (2016) Reinventing the energy bill in smart cities with NoSQL technologies. Transactions on engineering technologies. Springer, Singapore, pp 383–396
13. Sharma Y, Sharma Y (2019) Case study of traditional RDBMS and NoSQL database system. Int J Res GRANTHAALAYAH 7(7):351–359
14. Abraham SM (2016) Comparative analysis of Mongodb deployments in diverse application areas. Int J Eng Manag Res (IJEMR) 6(1):21–24
15. Rocha L, Vale F, Cirilo E, Barbosa D, Mourão F (2015) A framework for migrating relational datasets to NoSQL. Procedia Comput Sci 51:2593–2602
16. A MongoDB White Paper, RDBMS to MongoDB Migration Guide, considerations and best practices, June 2018
17. Boicea A, Radulescu F, Agapin LI (2012) MongoDB vs Oracle--database comparison. In: 2012 third international conference on emerging intelligent data and web technologies, pp 330–335. IEEE
18. Hecht R, Jablonski S (2011) NoSQL evaluation: a use case oriented survey. In: 2011 international conference on cloud and service computing, pp. 336–341. IEEE
19. Ghule S, Vadali R (2017) Transformation of SQL system to NoSQL system and performing data analytics using SVM. In: 2017 international conference on trends in electronics and informatics (ICEI), pp 883–887. IEEE
20. Hsu JC, Hsu CH, Chen SC, Chung YC (2014) Correlation aware technique for SQL to NoSQL transformation. In: 2014 7th international conference on Ubi-media computing and workshops, pp 43–46. IEEE
21. Potey M, Digrase M, Deshmukh G, Nerkar M (2015) Database migration from structured database to non-structured database. Int J Comput Appl 975:8887
22. https://www.couchbase.com/comparing-couchbase-vs-oracle Moving from relational to NoSQL: How to Get Started
23. Gayathiri NR, Jaspher DD, Natarajan AM (2019) Big Data retrieval techniques based on hash indexing and Mapreduce approach with NoSQL Database. In: 2019 international conference on advances in computing and communication engineering (ICACCE), pp. 1–8. IEEE
24. Das TK, Kumar PM (2013) Big data analytics: a framework for unstructured data analysis. Int J Eng Sci Technol 5(1):153
25. Eckerstorfer F (2011) Performance of NoSQL databases
26. Faraj A, Rashid B, Shareef T (2014) Comparative study of relational and non-relations database performances using Oracle and MongoDB systems. Int J Comput Eng Technol (IJCET) 5(11):11–22
27. Farooq H, Mahmood A, Ferzund J (2017) Do NoSQL databases cope with current data challenges. Int J Comput Sci Inform Secur (IJCSIS) 15(4)
28. Franco M, Nogueira M (2011) Using NoSQL database to persist complex data objects. In: Conference, Instituto de Informá tica Universidade Federal de Goias (UFG)
29. Heripracoyo S, Kurniawan R (2016) Big Data Analysis with MongoDB for decision support system. TELKOMNIKA (Telecommun Comput Electron Control) 14(3):1083–1089
30. Li Y, Manoharan S (2013) A performance comparison of SQL and NoSQL databases. In: 2013 IEEE Pacific Rim conference on communications, computers and signal processing (PACRIM), pp. 15–19. IEEE
31. Mapanga I, Kadebu P (2013) Database management systems: a nosql analysis. Int J Modern Commun Technol Res 1(7):265849
32. Mohamed MA, Altrafi OG, Ismail MO (2014) Relational vs. nosql databases: a survey. Int J Comput Inform Technol 3(03):598–601
33. Nayak A, Poriya A, Poojary D (2013) Type of NOSQL databases and its comparison with relational databases. Int J Appl Inform Syst 5(4):16–19
34. Swaroop P, Gupta V, Singh KR, Rajan SN (2016) NoSQL paradigm and performance evaluation.Scient Soc Adv Res Soc Change 3
35. Zaki AK (2014) NoSQL databases: new millennium database for big data, big users, cloud computing and its security challenges. Int J Res Eng Technol (IJRET) 3(15):403–409

36. Zvarevashe K, Gotora TT (2014) A random walk through the dark side of NoSQL databases in big data analytics. Int J Sci Res 3(6):506–509
37. Priyanka A (2016) A review of nosql databases, types and comparison with relational database. Int J Eng Sci Comput 6(5):4963–4966
38. Győrödi C, Győrödi R, Pecherle G, Olah A (2015) A comparative study: MongoDB vs. MySQL. In: 2015 13th international conference on engineering of modern electric systems (EMES), pp. 1–6. IEEE
39. Simanjuntak HT, Simanjuntak L, Situmorang G, Saragih A (2015) Query response TIME comparison Nosqldb Mongodb with Sqldb Oracle. JUTI: Jurnal Ilmiah Teknologi Informasi 13(1): 95–105
40. Truica CO, Boicea A, Trifan I (2013) CRUD operations in MongoDB. In: 2013 international conference on advanced computer science and electronics information (ICACSEI 2013), pp 347–350. Atlantis Press
41. Cattell R (2011) Scalable SQL and NoSQL data stores. ACM SIGMOD Rec 39(4):12–27
42. Barmpis K, Kolovos DS (2014) Evaluation of contemporary graph databases for efficient persistence of large-scale models. J Object Technol 13(3):3–1
43. Gu Y, Shen S, Zheng G (2011) Application of nosql database in web crawling. Int J Digital Content Technol Appl 5(6):261–266
44. Leavitt N (2010) Will NoSQL databases live up to their promise? Computer 43(2):12–14
45. Padhy S, Kumaran GMM (2019) A quantitative performance analysis between Mongodb and oracle NoSQL. In: 2019 6th international conference on computing for sustainable global development (INDIACom), pp 387–391. IEEE
46. Singh A (2019) Data migration from relational database to MongoDB. Global J Comput Sci Technol

Chapter 8
Data Security in Collaborative Business Intelligence for Sustainable Super Smart Society

Vinay Kumar Saini, Sachin Gupta, and Bhoomi Gupta

Abstract With research and development in full swing for super smart society, collaborative business intelligence (CBI) has emerged as a key component in the whole ecosystem. Super smart society aims to meet the needs of members of society through the provision of the quality goods and services in adequate amounts to address their distinctive needs based on age, gender, religion or language. This requires highly structured and secure collaborative business intelligence that would analyze data from various elements of super smart society like IoT, AI, and Big data. This chapter aims to explore recent developments in the field of CBI. It further explores information security practices and related mechanisms to ensure safeguarding CBI processes. Finally chapter covers a use case and design of data security framework for CBI that can boost sustainability of super smart society.

8.1 Super Smart Society: Introduction

Super smart society has been projected as the next area of research and development firstly by Japanese government in 2016 by introducing Super Smart Society 5.0 policy that aims to develop an ecosystem to facilitate commercial projects using technologies like robots, Internet of Things and Artificial Intelligence [1]. Japan has always been proactive for research and development involving information technologies, automation, and machine communication capabilities. Super smart society has the potential to solve societal and demographic problems [2, 3]. References [4–6]. Super smart society aims to achieve following:

- Leveraging sensing technology to develop autonomous societal operations;
- Adoption of artificial intelligence to facilitate human less analytical and decision-making ability;
- Facilitating societal services at flexible and usable via mobile technology; and

V. K. Saini (✉) · S. Gupta · B. Gupta
Maharaja Agrasen Institute of Technology, New Delhi, India
e-mail: vinay.kr.saini@gmail.com

V. Bali et al. (eds.), *Decision Analytics for Sustainable Development in Smart Society 5.0*, Asset Analytics, https://doi.org/10.1007/978-981-19-1689-2_8

- Providing users oriented services based on effective collaboration among service providers and consumers.

Super smart society emphasizes on an innovative growth model to solve social issues for building a better future. Some of the key areas [7] under super smart society are

- Smart management of water quality, forests, and biodiversity.
- Monitoring of meteorological data to resolve climate change issues.
- Bringing synergetic compatibility among the city smartness factors including safety, economic efficacy, and resident convenience.
- Bringing together corporate, academia, and all other stakeholders to create an innovative universal ecosystem providing smart education and safe information access.
- Building resilient infrastructure and promoting sustainable industrialization.
- Creating smart agriculture utilizing the Internet of Agri-Sensors, Artificial Intelligence-based crop management, and Big Data-based processing of biotechnological data.
- Smart management of infectious diseases by real-time health center data monitoring in conjunction with global parameters for early warning against epidemics and pandemics.
- Creation of a smart distribution grid toward a sustainable electric energy demand and supply management system.

Super smart society focuses on building an ecosystem to facilitate business and consultancy opportunities. Sustainability of such an ecosystem is of paramount importance as any risk may have disastrous impacts on the whole ecosystem. Following are some of the areas under super smart society [8] that need to be covered by information security practices to ensure sustainability of operations.

- Endpoint Security: End point security covers security of mobile devices and services at user's end. End point is the most vulnerable point in the whole system and hence needs robust security mechanisms to mitigate security threats.
- Security of smart city infrastructure: This includes physical and environmental security mechanisms to safeguard hardware and software that is part IT infrastructure in super smart city.
- Data privacy: This includes personal identification information of users/customers that is captured for facilitating service delivery under various super smart society systems. Exposure and theft of such information are referred as data privacy breach. Service providers must ensure secure and fair use of personal data of users/customers.
- Logical security: This includes security mechanisms like access control and algorithms and protocols security to ensure secure operations.

As super smart society is a network of various society processes operating on top of IoT and facilitated by AI and BigData, hence related security risks, operational risks, and environmental risks need to be addressed to make super smart society

sustainable. There is an imperative need for Smart Sovereignty [9]. The concept focuses on ensuring security and privacy concerns associated with the users and consumers of super smart society.

Therefore the research work and innovations are on the rise [10] in the following areas.

- *Cyber security*: Connected systems and remote control are inherent to systems under super smart society and hence there is a risk of vulnerabilities getting exploited by cyber threats. Cyber security aims to address this problem by providing a five-layer defense mechanism by exercising identification, prevention, detection, response, and recovery practices against security threats. Cyber security innovations aim to ensure the security of data and communication as highlighted by NIST cyber security framework [11] shown in Fig. 8.1.
- *IoT system architecture technology*: This involves multi-layer technological components weaving together collaborative low cost sensing and communication devices with the use of network programming leading to massively scalable information architectures.

  Four layers of components for any IoT system are depicted in Fig. 8.2.

  – First we have networked electronic devices including wireless sensors and actuators play the active role of capturing data from real-world functions.

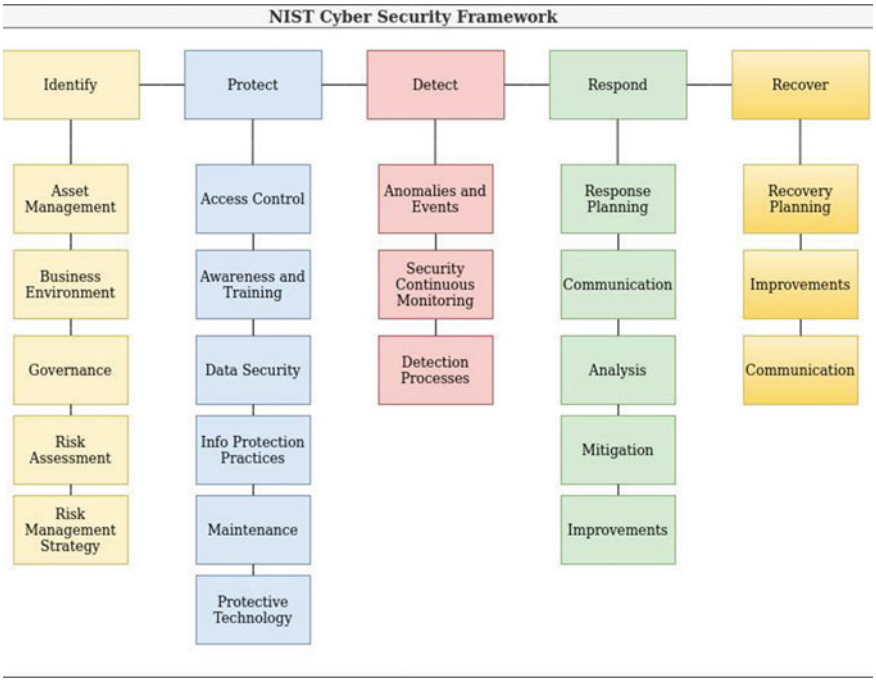


Fig. 8.1 NIST cyber security framework [11]

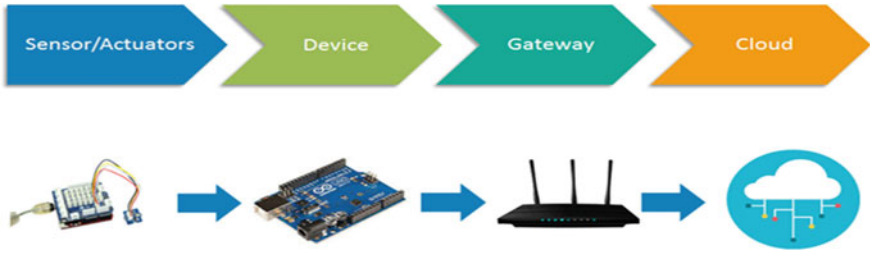


Fig. 8.2 IoT architecture layers [12]

  - Second layer comprises data aggregation systems that receive data from sensors and undergo analog-to-digital data conversion so as to prepare it for further processing.
  - Third layer associates the system with edge computing components to ensure that only preprocessed information is uploaded to persistent storage space on cloud or data center.
  - The last layer analyses the preprocessed data for converting information into intelligence for consumption by the smart systems and stores it on back-end systems.

- *Big data analytics*: A super smart society shall have a multitude of electronic devices with a rapid and voluminous data generation. We are already witnessing a variety of such data forms being generated from sources like social media platforms and networks. The data being generated varies from absolutely raw unstructured, to semi-structured data or fully structured in nature. To understand the terms better, we may visualize structured data as a definite format with rows and columns with singular values. Document files and emails are considered semi-structured, while the multimedia including picture files of digital audio and video is deemed unstructured data. The advent of big data analytics has opened up exciting avenues for discovery of useful insights from unstructured data, uncovering unconventional hidden correlations between trends across almost all domains. The resultant analytical procedures are paving the way for easier prediction and forecasting, prevention of economic and cyber offences and overall betterment of the decision-making process. It is expected to foster innovation to deliver value by understanding material insights from unstructured information.
- *Artificial Intelligence*: Artificial intelligence (AI) can bring smart computation with automation of redundant and time-consuming tasks with the help of smart applications and a smart computing infrastructure that can replace human beings and can make systems more effective. AI can back IoT to create smarter systems for super smart society.
- *Edge computing*: Edge computing is a specialized computing approach based on distributed computing that aims to put computation and data storage near to the location of business processes, to improve performance. Super smart societies can leverage this concept to build efficient IoT systems.

- *Collaborative Business Intelligence* aims to give impetus to interactions between stakeholders and subject matter experts to facilitate effective decision-making to address problems related to super smart society. The idea behind CBI is that when people work collaboratively multi-dimensional perspectives based on areas of expertise emerge to solve problems. The underlying foundation of using collaborative business intelligence is to expand the horizons of decision-making currently limited to the business process by including relevant data points covering concerns of all stakeholders of the super smart society.

8.2 Collaborative BI for Super Smart Society

It is evident from the previous section that the underlying opportunity with collaborative BI involves coverage of real-time data pertaining to all stakeholders in decision-making processes with the help of smart IT components. The potential maximization of benefits by effective real-time monitoring toward better decision-making may be realized in the decentralized business scenarios only by innovating the data processing approaches and system architectures. The intuitive approach of integrating data warehouses using customized extract, transform, and load mechanisms is imperative for collaborative BI [13].

Collaborative BI is the process that involves amalgamation of data analytics and technologies like IoT, Big Data, and AI to facilitate decision-making. Collaborative BI can help in developing business dashboard, to monitor and control various processes that are part of super smart society. Such a dashboard can help users in understanding insights of business processes. The combining of user domain knowledge with the above insights conclusively leads to decision-making that fulfills business objectives covering all stakeholders needs. Further, feedback in the form of comments, and ratings assigned by users can add valuable data to enhance results of Collaborative BI [14].

The primary focus of Collaborative BI program is to integrate various applications that are involved in managing processes of super smart society. The integration promotes real-time decision-making processes, eventually leading to better efficiency. To consolidate the gains realized, decision validation along with analytical assessment of potential impact is also feasible. Such analysis may help create effective mechanisms for feedback, which can potentially improve the overall performance leading to development of new business best practices while placing the most impactful data affecting this scenario under constant monitor [15].

Collaborative BI depends heavily on the following three aspects:

- Knowledge Discovery: This includes identification of material data items related to processes, operations, and user's experiences. System under a super smart society should facilitate identification of key parameters by careful examination of operations and interactions among various stakeholders of the ecosystem. The focus should be on identifying dependencies and relationships among various data items.

- Knowledge Retention: Operations in super smart society involve a variety of data including structured data like numerical readings from sensors, semi-structured data reports from subsystems, or unstructured data that may include images, audio, and video feeds. To facilitate integration of varied data forms, and effectively disseminate information insight, usage of collaborative software is recommended.
- Knowledge Proliferation: The pervasive use of professional social networking and increased reliance on personal social media is leading to the generation of huge volumes of data carrying massive amounts of valuable business information from both organizational and individual social networks.

Collaborative BI can add value to super smart society with the following benefits,

- Faster and more informed decision-making—Collaborative BI works on top of integrated data from all related processes and can facilitate expert decision-makers to quickly analyze the reports to foster quick decisions.
- Knowledge Sharing—Collaborative BI encourages participation of diverse groups of people to work on a problem to gather ideas based on different perspectives that may not be observed by decision-makers.
- Employee Satisfaction—Collaborative BI can present additional critical inputs and can empower employees to analyze the business scenarios to deliver valuable services. This can help in improving the effectiveness and productivity of employees. Inputs from Collaborative BI bring more clarity and give a multi-dimensional perspective about business processes, as business processes are often related to each other, and focusing only on one without understanding the relationship with others can lead to inappropriate understanding and wrong decisions that may affect the effectiveness of the system.
- Participative Governance—the time domain flexibility offered with the help of collaborative BI tools ensures constant participation by all stakeholders irrespective of their schedule forced differences based on different time zones, or by the choice of convenience.
- Immaculate Book-keeping—complete trails of meetings and conversations leading to decisions based on them, and the decisions themselves may be recorded in persistent storage for later references and cross verifications. This also ensures ample savings on time otherwise wasted on duplicate discussions with the help of documented conversations thereby reducing decision churn.
- Integrated Authorization—collaboration among multiple parties invariably raises security concerns, which are handled in the collaborative BI approach by integrating highly granular authorization checks across reviews, edits, and other operations securely and efficiently.

8.3 Information Security Practices for CBI Processes

Information security practices fundamentally aim to safeguard information assets for any information system. This may include control mechanisms to address security

objectives like confidentiality, integrity, availability, non-repudiation, authentication, and authorization. The most valuable asset for CBI processes is the transactional data collected from a network of IoT devices. Hence credibility of data sources with integrity of data and security practices around its usage is paramount [16]. Connected devices under super smart society may have exposure to various security attacks [17]. It's imperative therefore, to ensure that all the collected data, either processed or that which is intended to be processed and eventually saved to persistent storage in the collaborative business intelligence systems is protected with state-of-the-art mechanisms based on stringent security policies [18]. This makes the security and reliability of the system sustainable for a long-term operation. Following are the factors that justify the need of information security practices for collaborative business intelligence for super smart society.

- *Security equals reputation*: The damaging cost of insecure Collaborative BI is the reason for all paranoia surrounding cyber security provisions of any business system. The financial and reputational damages as per documented corporate history in the cases of cyber security breach far outweigh the other usual suspects including corporate espionage. Recently conducted surveys and analysis by IBM shows that the average cost of Data breach is $3.86 M and the worst hit sector in this domain is health care.
- *Maintaining public trust*: Super smart society involves interactions of various society systems that includes connected devices and networks. Maintaining public trust is imperative and assurance about cyber security is critical to support Collaborative BI over systems of super smart society. For a process of super smart society to ethically deliver better products and services, it is essential to provide all stakeholders with absolute confidence in the cyber safety of the system and about protection of their own private information. Even the slightest breach causes dent in this confidence and shakes the most loyal customers leading to attrition and general mistrust for the years to come.
- *IoT as a data origin for business intelligence*: Systems under super smart society have capabilities of controlling machinery in the background and in the foreground showing data and controls. Business intelligence from IoT carries valuable data that is vulnerable due to the nature of IoT devices [19]. The devices constituting IoT are the least secure in terms of physical security due to their deployment and are most vulnerable to attacks owing to their low resources [12]. These characteristics make their deployment a potential security hazard for any organization, but this tradeoff is still tolerated owing to the benefits outweighing the risks. The use of IoT merits exploration of innovative solutions specific to the domain like manufacturing industry, which may help keep not only the IoT secure from attacks, but also protects the collaborative BI using the same [20].

Now as it is evident from the above section that robust information security practices are imperative to ensure assurance in collaborative business intelligence for sustainability of super smart society. Now the following list highlights control mechanisms to address security objectives for robust information security practices.

- *Confidentiality*: Confidentiality focus on ensuring secure communication between sender and receiver and involve use of cryptography that encrypts the data sent by the sender and on receiver side decryption takes place that converts coded data into original form so that receiver can access and use it. The confidentiality is achieved by secure possession of keys at sender and receiver end. Cryptography is not a new concept, but to implement cryptography in computerized environment concepts of modern cryptography is used [21]. Modern cryptography has following three mechanisms:
  - *Symmetric cryptography*: This technique involves use of a secret key to encrypt the data and decrypt the data. The sender and receiver both should have this secret key for secure communication. The sender uses the secret key to encrypt the data and then this coded data is sent to the receiver. Receiver receives the coded data and decrypts it with the use of the same secret key to get original data as shown in Fig. 8.3. Hence confidentiality is maintained as only these sender and receiver can understand the data as only these two have the possession of the secret key.
  - *Asymmetric cryptography*: This technique was primarily proposed to solve the problem associated with symmetric cryptography. Though symmetric cryptography is good enough to ensure confidentiality but it has a problem with key transfer. Secure transfer of key to the sender is a must to ensure secure communication. But the problem is how to transfer the key to the receiver. Here comes the use of asymmetric cryptography, which uses a key pair that includes private key and public key. The private key is kept by the receiver and the public key is transferred to the sender. Sender can use the public key of the receiver to encrypt and send symmetric session key (that was supposed to be sent to receiver under symmetric cryptography). The encrypted data having symmetric session key can only be decrypted by receiver as he holds the private key. The private key and public key are reciprocal in nature, if one is used to encrypt the data the other would be used to decrypt the data presented in Fig. 8.4.

    Asymmetric cryptography can also be used to achieve authentication and is widely used to implement digital signatures where a private key is used by the sender to sign the document and public key can be used to validate the

Fig. 8.3 Symmetric cryptography [21]

Fig. 8.4 Asymmetric cryptography for secure symmetric key transfer [21]

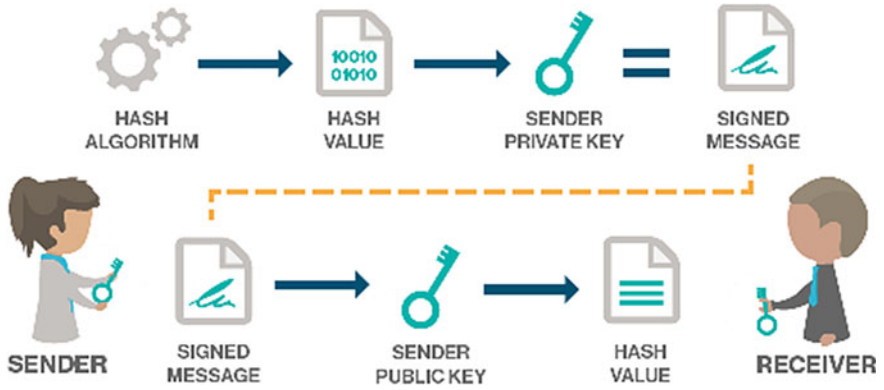


Fig. 8.5 Asymmetric cryptography for digital signature for user authentication [21]

authenticity of the document and identity of the sender who has signed the document as shown in Fig. 8.5.

– *Hashing*: This cryptographic technique is different from the other two techniques discussed in previous sections. Hashing is used to facilitate data integrity check during any data exchange or online communication. Hashing is a type of cryptography that is Irreversible unlike the previous two where decryption results in original data. In hashing input data is converted into a fixed length code also called message digest, which is unique for input data. Hash function generates the same code every time for the same input data. Message digest is a kind of summary that can be used to check correctness or to detect any tampering of data during exchange or transmission. Sender computes message digest and sends message digest along with data in encrypted form to the receiver. Receiver decrypts the coded data and generates message digest of received data and matches the new message digest to the one received from sender as shown in Fig. 8.6. If both don't match that means data has been changed during transmission.

- *Integrity*: Integrity focuses on ensuring protection against unauthorized tampering of data. This can be achieved in many ways and hashing has been an effective

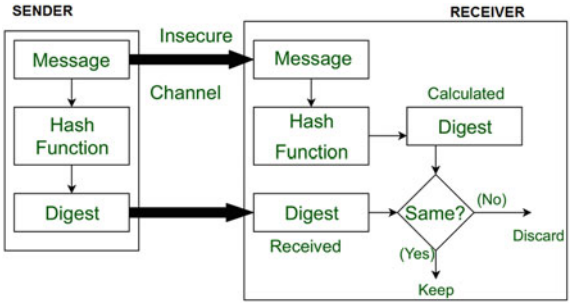


Fig. 8.6 Hashing for data integrity

control mechanism to ensure data integrity as described in previous sections as an irreversible cryptographic technique.

- *Availability*: This security objective focuses on ensuring the availability of system or service as and when is requested. Super smart society includes critical information assets or resources that may face threats in the form of malicious attacks for data theft or denial of service. Data theft attacks can be addressed using control mechanisms related to confidentiality. But denial of service may halt processes and may have a huge impact on all stakeholders. To ensure continuous availability various control mechanisms can be practiced like firewall, load balancing, and business continuity infrastructure. These control mechanisms can help prevent, detect, or correct incidents related to risks that may affect availability of services in super smart society.
- *Non-repudiation*: This control mechanism aims to facilitate accountability of the users. Users cannot deny any transactions or operations performed by them. This is implemented by mapping user identity with the transaction or operation. Such mapping may include identity parameters associated with the user like user ID, password, fingerprint, One-time password (OTP), and digital signature. This helps in defining user accountability in the operations.
- *Authentication*: This control mechanism helps in defining the validation and verification mechanism to confirm identity of users interacting with the system. This control helps in preventing unauthorized access to the system. There can be various authentication methods based on various characteristics of users or parameters associated with them. Three factors that are used for authentication are as follows:

  – Something user your knows (like username, password, pin, or passkey),
  – Something user has (like id card, ATM card, OTP through Mobile)
  – Something the user is (finger print, retina or facial scan, and other biometric properties).

  It is always recommended to have at least two-factor authentication to ensure reasonable protection against unauthorized access although three-factor authentication is the best authentication mechanism.
- *Authorization*: This control mechanism focuses on defining privileges and access rights so that users are only allowed to perform the operations they are responsible

for, based on their job profile. This can help in preventing unauthorized access and un-approved changes or transactions within the system.

Security requirements are one of the most critical concerns to ensure sustainability of smart city systems. As smart city systems involve network of IoT devices that provide relevant data to be analyzed by collaborative business intelligence. Many security mechanisms have been discussed in previous sections to implement security controls. But recent advancements in emerging technologies have presented blockchain as a hybrid approach to develop robust systems. The blockchain technology is one of the most competent front runners among smart city solutions development due to its desirable properties of immutability, transparent transactions, privacy preservation, and decentralization leading to auditability of the controlling authorities [22, 23].

While traditional systems have favored a central repository of the city control data, with replicated copies as backup, they remain vulnerable to hacking and ransomware attempts or may leak sensitive private information of the individuals. In extreme cases it is feared that the management of the centralized data falling in the wrong hands may even cause catastrophic events to the city leadership [24]. The situation as described above can be conveniently avoided by using a decentralized storage and management of city control data, and blockchain offers the technology just right for the maintenance with its immutable ledgers, working on a Peer-to-Peer (P2P) network [25]. Blockchain technology is fundamentally based on the following three concepts:

- Distributed ledger Technology: A distributed ledger is a transaction storage scheme supported by multiple players agreeing to common rules. The ledger is shared and updated in real time across all the players distributed over geographies and is accessible by all involved stakeholders. The player at each node of the network owns an identical copy of distributed ledger. Whenever any new transaction is added to the ledger it is uniformly shared to all players instantly. A distributed ledger is more secure and robust, whereas centralized systems are more prone to cyber-attacks and fraud, as have a single point of failure.
- Hashing: Blockchain uses hashing that includes a mathematical function that takes input data involved in the transaction of any type and any length and transforms it into a fixed length data. The fixed bit length can vary depending on the hash function which is being used. The fixed length output is called a hash.
- Public key Infrastructure: The public key infrastructure (PKI) security method is used to implement strong authentication to verify identity of the participants in blockchain network.

Blockchain powered by cloud infrastructure can be used to provide robust backend storage and transaction management for IoT systems in smart societies as shown in Fig. 8.7. Blockchain can certainly address security challenges of IoT-based systems in smart societies through its following listed features [26].

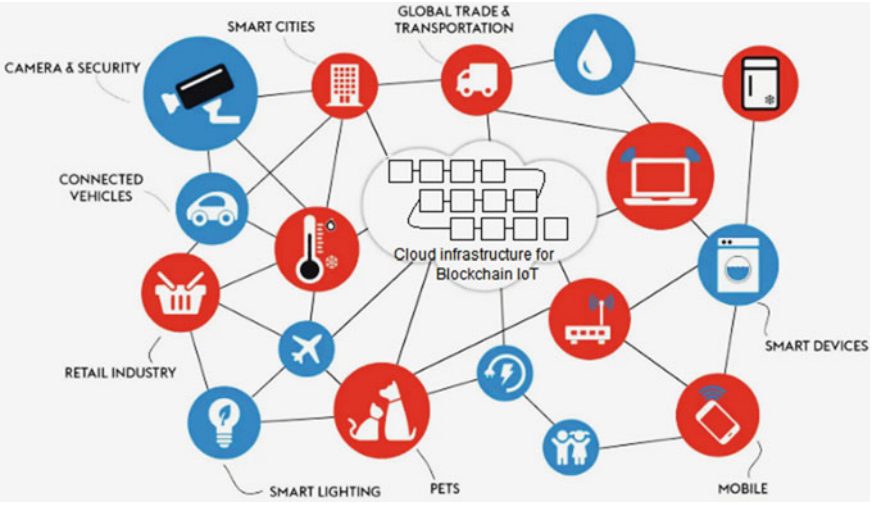


Fig. 8.7 Cloud computing driven Blockchain-based IoT systems

- Auditability: Auditability feature of blockchain can provide a record of events and operations throughout the processes of smart society systems and can facilitate event monitoring to ensure traceability in the processes of smart society systems.
- Immutability: Immutability feature of blockchain focus of addressing risk of data tampering. This is achieved by time stamping and computing hierarchical hash code to make transaction data tamper-proof.
- Smart Contract: Smart contract feature of blockchain provides real-time rule-based verification of multiparty confirmations that facilitates cost-effective adaptation to business environment changes to ensure flexibility in the supply chain.
- Distributed: the distributed feature of blockchain allows direct interaction with a trusted digital signature-based peer-to-peer interactions that ensure supply chain communication risk reduction and trust building in the supply chain ensuring stakeholder management.

Further, the blockchain-based systems incur minimized overall security monitoring cost and provide security against adversaries trying to gain access to personal information or control over the entire system.

8.4 Use Case: Retail Store in Super Smart Society

Retail market is one of the key segments in super smart society having cut throat competitive work space that must focus on increasing sales and stay ahead. A Collaborative BI solution can facilitate collecting material data and analyzing that to build

business development strategies. Data collection across multiple departments poses a serious challenge which is compounded by multiplicity of storage formats. With collaborative BI-based solutions, centralized data aggregation for trend analysis, time series management, and other useful information become streamlined. Collaborative business intelligence can facilitate substantial benefits as highlighted below:

- *Improved customer experience*—Collaborative BI can present valuable insights about customer expectations and their preferences, and these inputs can enable brands to ensure that potential customers can be targeted with best services to meet their expectations.
- *Predictive modeling*—Considering the continuous changes in a competitive environment retail businesses need to expand their exploration through CBI. They have to think beyond measuring their current data and explore predictive analysis and forecasting. Having invested in a software allowing them predictive analytics on retail trends, the business has potential to use present data as a weapon toward projecting fresh sales opportunities besides unforeseen trends. Recommendation engines running such stores create explicit models of frequently bought together items leading to increased sales.
- *Flexible demand-based pricing*—the documented case studies of uber and ola shared cabs surge pricing model effectively demonstrate that collaborative business intelligence has the potential to ensure real-time flexibility in commodity pricing based on basic economic indicators of demand and supply along with predicted trends.

8.5 Data Security Framework for CBI

Collaborative BI can significantly foster sustainability in super smart society. But cyber security challenges and rise in cyber threats is a serious concern that needs assessment and control mechanisms. This section highlights proposed data security framework for collaborative BI and takes into consideration critical inputs from Industry best practices and standards related to cyber security like cyber security framework by NIST.

The data security framework has the following three components as shown in Fig. 8.8.

- Best Practices and Information Security Standards: Best practices can facilitate risk assessment so as to focus on critical and material risks that may impact the system. These can provide significant inputs for defining and managing information security policies and procedures. These must address primary data security objectives like confidentiality, integrity, and availability. Various standards and best practices that may be considered include NIST cyber security framework and ISO 27000 series of standards for information security management.
- Risk Assessment: Risk assessment is a two-step process focused on ranking or risks associated with the system under consideration based on stated objectives

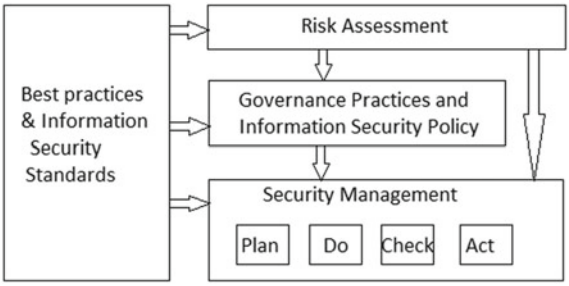


Fig. 8.8 Data security framework for Collaborative BI

and scope of risk assessment. First step is the identification of the risk and second step is to evaluate the risk on the basis of probability of occurrence of risk and impact of the risk. Risk ranking can help in identifying critical risk that needs to be addressed on a priority basis. Risk assessment can provide critical inputs to develop adequate governance practices and information security policy. Risk assessment can also facilitate security management by highlighting areas of high concerns.

- Governance Practices and Information Security Policy: Governance practices focus on evaluation, direction, and monitoring of operations under a system. This involves setting up policies to facilitate effective resource utilization and risk optimization. Business process owners must understand opportunities and threats. These can be critical inputs to design business strategy to take on the correct action against risks that could cause the processes some degree of failure.
- Security Management: Security management includes identification and classification of key data devices and the assets associated with information security. Once information assets are identified, a well thought and thoroughly analyzed security policy documentation is presented to the key stakeholders for approval and implementation for the organization. This document outlines the do's and don'ts of procedures based on the approvals granted in the officially adopted security policy document. These actions should be supported by risk assessment and selecting an appropriate risk response that may include avoid, accept, mitigate, or transfer. Risk assessment should evaluate probability of occurrence of risk and magnitude of the risk.

Periodic review and update of information security practices are imperative for ever evolving information security threats.

8.6 Case Studies

- **CareTech—Fukuoka, Japan:** Fukuoka a popular city in Japan has established a system for monitoring of elderly people suffering from dementia using IoT networks [27]. The system involves crowdsourcing of data by various service

providers, such as electricians, gas or water providers, newspaper deliverers, and postal workers. These users contribute content to the system about individual residents at their homes and neighborhood suffering from dementia. The project further involved Door-to-door visits to collect data on old age people. The system includes services like daily check-ups, calls and text messages, and advance search facility for seniors who get lost, and various IoT-based monitoring solutions. Collaborative business intelligence can facilitate better and smart care for elderly people suffering from dementia by analyzing and monitoring data in real time. System may be vulnerable at various points specifically end-points hence proper authentication check can make system robust. Every service provider's identity must be verified and approved before enabling his role in the system. Data Privacy and data security of data related to elderly people are serious concerns that need to be addressed with risk assessment and adequate controls to mitigate these concerns.

- **Harnessing City Data—Singapore:** Singapore is using collaborative business intelligence over government and private-sector data in a citywide data platform to facilitate better decision-making and planning in health and medicare [27]. Project titled Connected Life is a system that offers home monitoring solution that provides insights to health providers, the insurance industry, and government. The solution combines smart technology (IoT, such as motion and sound sensors, data analytics, and AI), 24/7 personal assistance, and customized insurance coverage. Collaborative business intelligence is used to monitor real-time information and allows families to protect their elderly members. This system uses DEX, a decentralized citywide data exchange which combines private- and public-sector data. DEX involves data exchange across various systems and can be point of exposure of sensitive data. Hence appropriate security controls ensure data privacy and data security concerns.
- **Smart Dubai Happiness Meter:** It is a live sentiment capture tool, implemented to measure happiness among city experiences across thousands of touch points [28]. Phased implementation of project covered 172 public- and private-sector entities in all customer interaction channels including websites, mobile apps, and physical service centers recording millions of Happiness votes from 4,400 touch points over 172 entities. Collaborative business intelligence can be implemented to improve the quality of service to the customers. This system carries information of customers and data privacy concerns needs to be addressed by using appropriate controls to ensure secure operations.
- **Bhopal Plus-India:** PWC is working on project Bhopal Plus, an e-governance initiative to facilitate smart and effective services to citizens [29]. This online platform enable citizens to make content contribution to ensure sustainability and continuous improvement. This platform includes services like public utility bill payment, municipality services (property tax payment, water tax payment, application for birth/death/marriage certificate), and smart maintenance services. This platform further offers services like tourism, bus route planner, book donations, sell or donate scrap/waste, city directory, and emergency SOS services and report an incident, city-related notifications where citizens can contribute content that is

published after verification and validation. The platform involves processing and transferring of sensitive data of users and government agencies' databases. This can be analyzed using collaborative business intelligence to offer better quality of service. But information security and data privacy concerns need to be addressed using adequate security mechanisms as highlighted in previous sections.

8.7 Conclusion and Future Scope

Super smart society has promising prospects and heavily depends on integration of various technologies like IoT, AI, and Big Data. Collaborative business intelligence can further add value and make super smart society sustainable. As CBI is a data-driven process, hence information security threats pose a serious risk that affects sustainability. This chapter has covered various security objectives and control mechanisms to address these concerns. But these control mechanisms can only be implemented effectively if standard operating procedures are defined based on inputs from process owners, best practices, and information security standards. Hence a framework has been proposed and discussed that covers all the mentioned concerns to facilitate implementation of adequate control mechanisms to address security risks for CBI in super smart society.

The future scope for this work lies in the identification of suitable use cases and implementing proposed framework with customization considering requirement of the use case. Monitoring dashboards can also be developed to ensure control on use of CBI in super smart society to deliver assured results.

References

1. Carin H (2020) Technological innovation and building a 'super smart' society: Japan's vision of society 5.0. J Asian Public Policy, https://doi.org/10.1080/17516234.2020.1749340
2. Cifaldi G, Serban I (2018) Smart Cities-smart societies. In: Karwowski W, Ahram T (eds) intelligent human systems integration. IHSI 2018. Adv Intell Syst Comput 722. Springer, Cham. https://doi.org/10.1007/978-3-319-73888-8_108
3. Deguchi A et al (2020) What is society 5.0?. In: Hitachi-UTokyo Laboratory(H-UTokyo Lab.) (eds) society 5.0. Springer, Singapore. https://doi.org/10.1007/978-981-15-2989-4_1
4. Kerlooza Y, Setiawan A, Asrianto R (2018) Towards smart society: a study on multi-channel and public participation-based system architecture for civil registration and population data in Indonesia. Int Conf ICT Smart Soc (ICISS) 2018:1–5. https://doi.org/10.1109/ICTSS.2018.8549956
5. Nugraha Y (2020) Building a smart city 4.0 ecosystem platform: an overview and case study. 2020 international conference on ICT for smart Society (ICISS), pp 1–7, https://doi.org/10.1109/ICISS50791.2020.9307538
6. Foresti R, Rossi S, Magnani M, Guarino Lo Bianco C, Delmonte N (2020) Smart society and artificial intelligence: big data scheduling and the global standard method applied to smart maintenance. Engineering 6(7):835–846, ISSN 2095–8099, https://doi.org/10.1016/j.eng.2019.11.014

7. Foresti R, Rossi S, Magnani M, Guarino Lo Bianco C, Delmonte N (2020) Smart society and artificial intelligence: big data scheduling and the global standard method applied to smart maintenance. Engineering 6(7): 835–846, ISSN 2095–8099, https://doi.org/10.1016/j.eng.2019.11.014
8. Ismagilova E, Hughes L, Rana NP et al (2020) Security, privacy and risks within smart cities: literature review and development of a smart city interaction framework. Inf Syst Front. https://doi.org/10.1007/s10796-020-10044-1
9. Aldabbas M, Teufel S, Teufel B, Pasquier V (2020) Smart Sovereignty: the security shield for smart society 5.0, smart sovereignty: the security shield for smart society 5.0. Int J Digital Soc (IJDS) 11(2):1619–1626
10. Záklasník, Martin & Putnová, Anna. (2019). Digital Society—Opportunity or Threat? Case Studies of Japan and the Czech Republic. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis. 67. 1085–1095. https://doi.org/10.11118/actaun201967041085.
11. NIST Cyber security Framework, https://www.nist.gov/cyberframework. Accessed 22 March 23, 2021
12. Khanna A, Kaur S (2020) Internet of things (IoT), applications and challenges: a comprehensive review. Wirel Pers Commun 114:1687–1762. https://doi.org/10.1007/s11277-020-07446-4
13. Dessì N, Garau G (2014) A case study for a collaborative business environment in real estate. In: Catania B et al (eds) new trends in databases and information systems. Adv Intell Syst Comput, vol 241. Springer, Cham. https://doi.org/10.1007/978-3-319-01863-8_37
14. Mola L, Rossignoli C, Carugati A, Giangreco A (2015) Business intelligence system design and its consequences for knowledge sharing, collaboration, and decision-making. Int J Technol Human Interact 11:1–25. https://doi.org/10.4018/IJTHI.2015100101
15. Raghavan S, Simon BYL, Lee YL, Tan WL, Kee KK (2020) Data integration for smart cities: opportunities and challenges. In: Alfred R, Lim Y, Haviluddin H, On C (eds) computational science and technology. Lecture Notes in Electrical Engineering, vol 603. Springer, Singapore. https://doi.org/10.1007/978-981-15-0058-9_38
16. Meng Y, Zhang W, Zhu H, Shen XS (2018) Securing consumer IoT in the smart home: architecture, challenges, and countermeasures. IEEE Wirel Commun 25:53–59
17. Rachit, Bhatt S, Ragiri PR (2021) Security trends in internet of things: a survey. SN Appl Sci 3, 121, https://doi.org/10.1007/s42452-021-04156-9
18. Jurcut A, Niculcea T, Ranaweera P et al (2020) Security considerations for internet of things: a survey. SN Comput. Sci. 1:193. https://doi.org/10.1007/s42979-020-00201-3
19. Dustdar S, Nastić S, Šćekić O (2017) Smart cities, the internet of things, people and systems. Springer International Publishing, ISBN - 978-3-319-60030-7,https://doi.org/10.1007/978-3-319-60030-7
20. Phuyal S, Bista D, Bista R (2020) Challenges, opportunities and future directions of smart manufacturing: a state of art review. Sustain Futures 2:100023, ISSN 2666–1888, https://doi.org/10.1016/j.sftr.2020.100023
21. Qadir AM, Varol N (2019) A review paper on cryptography,In: 2019 7th International Symposium on Digital Forensics and Security (ISDFS), pp. 1–6. https://doi.org/10.1109/ISDFS.2019.8757514
22. Bhushan B, Khamparia A, Martin Sagayam K, Kumar Sharma S, Abdul Ahad M, Debnath NC (2020) Blockchain for smart cities: a review of architectures, integration trends and future research directions. Sustain Cities Soc 61:102360, ISSN 2210–6707, https://doi.org/10.1016/j.scs.2020.102360
23. Majeed U, Khan LU, Yaqoob I, Ahsan Kazmi SM, Salah K, Seon Hong C (2021) Blockchain for IoT-based smart cities: recent advances, requirements, and future challenges. J Netw Comput Appl 181:103007, ISSN 1084–8045, https://doi.org/10.1016/j.jnca.2021.103007
24. Wang T, Zheng Z, Rehmani MH, Yao S, Huo Z (2019) Privacy preservation in big data from the communication perspective—a survey. IEEE Commun Surv Tutor 21(1):753–778. https://doi.org/10.1109/comst.2018.2865107
25. Novo O (2018) Blockchain Meets IoT: an architecture for scalable access management in IoT. IEEE Internet Things J 5(2):1184–1195. https://doi.org/10.1109/jiot.2018.2812239

26. Saini V, Gupta S (2019) Blockchain in supply chain: journey from disruptive to sustainable. J Mech Contin Math Sci 14:498–508
27. Community Paper, “Smart at Scale: Cities to Watch 25 Case Studies”, AUGUST 2020, Global Future Council on Cities and Urbanization, World Economic Forum
28. Okan Geray, Case study: smart dubai happiness meter, Dubai, UAE, October 2019, https://www.itu.int/dms_pub/itu-t/opb/tut/T-TUT-SMARTCITY-2019-2-PDF-E.pdf
29. PWC India Project, ‘Smart’ implementation of technology for Bhopal, https://www.pwc.in/about-us/analyst-relations/smart-implementation-of-technology-for-bhopal.html

Chapter 9
An Analytical Approach for Sustainable Development in Smart Society 5.0 Using Swasthya Sahayak Application

Rajender Kumar, Punit Soni, Alankrita Aggarwal, Mukesh Kumar, and Nikhil Mishra

Abstract A few months back all were living on with a unique life uninformed of how quickly things planned to change. The coronavirus is a worldwide upsetting influence that has constrained us all to meet up and attempt to stop it. While we are generally making changes and doing what we can to help, there is still significantly more that should be finished. We are building a site that gives data to help battle this Covid-19 pandemic and its impact on sustainable development. This application is for all the clients who need to think about the current circumstance in regard to this pandemic time on the whole over world and if a client needs to look for data about Covid-19 cases in their territory. For utilizing this framework one simply needs to have an advanced mobile phone or PC or framework with web access and some essential information on utilizing it. The expressed motivation behind this application is to spread familiarity with Covid-19 and to associate fundamental Covid-19-related wellbeing administrations to individuals and offer best practices and warnings. It is the following application that utilizes the gadget's area highlights to follow the Covid disease. With informing framework, it attempts to decide the danger if one has been almost a Covid-19 contaminated individual, by looking over a data set of known cases across India. Utilizing area data, it decides if the area one is in has a place with one of the contaminated zones dependent on the information accessible.

Keywords Coronavirus · Questionnaire · Latest news · Register · Notify me · Self-assessment · User-status · Ayurveda · Confirmed · Recovered · Active · Deceased · FAQs

R. Kumar (✉) · A. Aggarwal · N. Mishra
Panipat Institute of Engineering and Technology, Samalkha, Panipat, India
e-mail: rajenderkumar.cse@piet.co.in

P. Soni
Chandigarh University, Chandigarh, India

M. Kumar
Mahant Bachittar Singh College of Engineering and Technology, Jammu, India

V. Bali et al. (eds.), *Decision Analytics for Sustainable Development in Smart Society 5.0*, Asset Analytics, https://doi.org/10.1007/978-981-19-1689-2_9

9.1 Introduction and Foreword

The coronavirus is a worldwide unsettling influence that has constrained us all to meet up and attempt to stop it. While we are on the whole making changes and doing what we can to help, there is still much more that should be finished and creating a site gives information to battle this Covid-19. This will apply to all the clients who think about the present circumstance in regard to this pandemic time on the whole world and if a client needs to look for data on Covid-19 cases in a particular territory. To utilize this framework one simply needs to have a smartphone or PC or framework with web access and some fundamental information on utilizing it. It is the following application that utilizes the phone's area highlights to follow the Covid contamination. With informing framework, it attempts to decide the danger on the off chance that one has been almost a Covid-19 contaminated individual, by looking over an information base of known cases across India. Utilizing area data, it decides if the area is in the contaminated territories dependent on the information accessible. The Coronavirus Covid-19 is affecting all countries and territories around the world. We have developed a web application that tracks the spread of the virus. The Swasthya Sahayak is developed to help see how the virus is spreading and who is most at risk. The Swasthya Sahayak is a web application that tracks Covid-19 cases around an individual to keep him informed if they have come across Covid-19. During this pandemic, you can easily track the Covid-19 cases around the country [1–3].

9.1.1 Objectives of the Work

1. The main objective is to design a system that resolves the problem of users facing difficulties in Covid-19 crisis [4, 5].
2. Gaining knowledge of web applications using the Django web framework. Learning the frontend development for making an interactive user interface. Learning and practices of the backend development using Python [6].
3. Functionalities of the System: The proposed system has some core, enhanced, and some special functionality. So, let's discuss the chapter one by one and use this functionality in this system [7].

9.2 Literature Review

There are a lot of websites like corona info, etc. which show the list of Covid-19 in different regions. Even government also launches an app name "Aarogya-Setu" for the same. The developer wants to develop a website like this. Also, wants that the user will feel free to use the website and it will be user-friendly, not a messy thing, and also a site without ads. The entire thing which developer thinks that it is new

because none of the systems exists nowadays so in the future it may be possible that it is directly useful and implemented by the government [8].

V. Kachroo presented a conceivable method of transmission, definitions, some essential advice, determination, treatment, and the executive's convention being continued in India. The authors also characterized the Corona cases as suspect cases, probable causes, and confirmed cases. Further, the clinical disorder related to the Covid-19 contaminations were additionally examined as straightforward sickness, mild pneumonia, severe pneumonia, ARDS as a mild, moderate, and septic shock [9].

R. Gupta et al. introduced the work for the effect of different estimates taken with the assistance of exponential and polynomial regression displaying forecast of more than 80000 cases at the end of 2020. Toward the end, procedures for continuation for a lockdown have been examined and introduced. It gave the idea that solitary fundamental administrations ought to be open for the residents of India and the public lockdown ought to be carried on next 2–4weeks in India [10].

P. K. Jena researched the important effects of Covid-19 in Higher Educational Institutes (HEIs) in India. Estimates taken by HEIs and specialists of India offering instructions of assistance during the emergency are analyzed. The Covid-19 situation gave us new methods of learning, views, rise of new patterns for encouraging the learning as a new alternative of the scenario. Some productive proposals are additionally highlighted to do instructive exercises during the pandemic circumstance. UGC and MHRD have dispatched numerous virtual stages with layout storehouses, digital books, and other internet instructing/learning materials. A mix of the customary advancements like radio, TV, landline telephones with portable and web innovations to a solitary stage would upgrade the openness and adaptability to instruction [11].

A. Brodeur et al. investigated financial matters of Covid-19 delivered in limited time and also produced Covid-19 cases and deaths due to measurement error and testing capacity. Estimates from GPS phones utilized to measure social distancing and survey held for social distancing, its determinants, and the spread of Covid-19 consistency of the disease is measured [12].

N. Aggarwal et al. explored the impact of Covid-19 on the daily life of low-income households across India and resulted from researches to understand the gravity of situation and also did a panel survey conducted in three rounds by Dvara research. Rounds of survey conducted between 23April–7 May and 2 June–6 July, 2020 by collected evidence from 347 microfinance borrower households from different parts of the country on (i) visible signs of distress, (ii) access to essential services and Government welfare schemes, and (iii) coping mechanisms [13].

S. Prinja et al. developed a mathematical model to predict health outcomes comprising air restrictions and air travel restrictions along with 8-week lockdown. In addition, they estimated the effectiveness of an 8-week lockdown along with varying levels of effectiveness by susceptible-exposed-infectious-recovered (SEIR) [14].

R. Pal et al. discussed and summarized the present scenario of the pandemic in India and challenges faced due to Covid-19 situation in India, and witnessed a massive surge of cases in the last 3 weeks. As of April 30, 33 610 confirmed cases and 1075 deaths have been reported from 32 states & union territories. After nationwide

lockdown testing rate is increased and strengthened the health care sector to combat Covid-19. The authors also have in their fight against Covid-19 [15].

S. Udhaya et al. introduced the designing of study, analysis, interpretation of data and drafting of the research and also reviewed current scenario of Covid-19 in India that till May 8th, 2020, 56,342 positive cases have been reported. In return Indian government implemented a 55-day lockdown throughout the country started on March 25th, 2020 to reduce transmission of the coronavirus [16].

M. Pathak suggested that Indian researchers as well research institutions collaborated with China, USA, England, and 67 other countries in doing research for COVID-19. The research productivity is changing every day and India is among the top 10 countries in COVID-19 research and using databases are required to find out new imminent in novel research required from Indian researchers [17].

A. F. Siddiqui et al. investigated the impact of COVID-19 through viewing and observation technique in India as an important health system to detect the social, political, and economic outcomes. The research was done sequentially where data was collected between 30 January 2020 and 12 June 2020. For analysis mathematical models have been used to find the hopeful and gloomy situation of the lockdown and passengers screening system [18].

9.3 Research Methodology

9.3.1 Problem Description

The user gets completely satisfied after getting self asses to report. To build this web application, the analyst would need to experience some key points of the difficulties during the improvement like to implement ideas and thoughts linked which is the notable web structure of Python Language [19].

9.3.1.1 Technical Challenge

The challenges that would come in the way of the researcher while working on the technology area are as follows:

- Django Framework: Django is a Python-based web structure that permits you to rapidly make web applications without the entirety of the establishment or reliance issues that you ordinarily will discover with other systems. When you're constructing a site, you generally need a comparative arrangement of segments: an approach to deal with client validation (signing up, signing in, signing out).
- Web Platform: The absolute first thing that we might want to specify here is the information on the web platform. Since the framework will be created on the web platform and we are partially thought about this innovation, of course, we have

begun learning it yet we need to learn numerous ideas of the web application for fruitful improvement of this task.

- Equipment Fragmentation: There are a large number of assortments of gadgets that sudden spike in demand for the android stage and those different gadgets exist with various properties like memory, CPU speed, and graphical goals. There is additionally a discontinuity conceivable on the working framework level.
- Programming Fragmentation: Since there are such a large number of Django forms that are accessible running into the gadgets. Along these lines, I as an engineer can't zero in just on the single or most recent adaptation because it is difficult for clients to overhaul their working framework, and accordingly the majority of the clients should not have updated. Hence, the designer should realize which working frameworks are generally famous and create given the most recent broadly received adaptation [20].
- Web Services: Since, the web application needs to speak with the far off data set to amend the application to carry the significant information, in this manner designer should have to have profound information on a web administration as it will much of the time be utilized to bring information from or suit information to far off data set. The special features of the system offer the location of Covid-19 positive cases at the nearby area, alert message circulated to the user's device, provides health advice, and running social messages which aid the user to engage with the app and deliver info concerning safeguard for the Covid-19 virus.

The core functionality provides us the Dashboard for Social Groups and Sign-in features. Using these features users can first register themselves to the system with important credentials. Then after sign-in, the system's Covid-19 features will help the user in self-assessment, updates regarding Covid-19 cases, and state-wise list of cases (Table 9.1).

The Enhanced functionality delivers a Graphical and List view of Covid-19 cases to the users for better understanding. Users can use the guidelines of the application efficiently (Tables 9.2 and 9.3).

Table 9.1 Core feature (Dashboard for social groups)

S. no	Functionality	Description
1	Register	User registration
2	Sign-in	Users have email and password to log in
3	User status	Risk of getting Covid-19 to user
4	Self-assess	Users identify Covid-19 symptoms and risk
5	Covid-19 updates	Updated local and national Covid-19 cases
6	Maintained list of state-wise cases	User view list of Covid-19 cases by state names

Table 9.2 Enhanced features

Description	Description	Description
1	Graphical view	User access data of cases in interactive graphical view
2	List view	User view data for updates of Covid-19 cases in the list
3	User guide	Permits a user to use the application efficiently

Table 9.3 Special features

S. no	Functionality	Description
1	Getting location from the device	Enhanced features for user's location to show current Covid-19 positive cases nearby
2	Alert system	The person tested positive a message is circulated at the user's device to stay alert
		Feature allows alert user necessary action taken during the pandemic time for maintaining immune system helps to fight the Covid-19 virus
3	Health advice	Risk of getting Covid-19 to user
4	Running social message	Allow running social messages to help users to engage the app and take precautions for the Covid-19 virus

9.3.2 Academic Research

Before building up the proposed framework, the engineer led scholarly exploration by perusing research papers, diaries, and so forth to comprehend the genuine need of the Swasthya Sahayak site for the overall clients and its points of interest how it will be helpful to the clients. It will likewise assist the designer with refining the functionalities to be remembered for the framework.

9.3.2.1 Software Methodology and Software Engineering Concepts

To convey a client fulfilled framework inside time, the engineer needs to investigate the technique which best suits the site.

9.3.2.2 Human–Computer Interaction Principles

The designer should be clear with communication between the human and smartphones so the site is created remembering the execution of every one of those principles which justifiable and effective usage of the site.

9.3.2.3 Domain Research

Domain research is essential as the task space differs in different cases and whenever this is done for a specific module, at that point we can deal with examination and planning of that specific module gives simplicity of its utilization. During the phases of exploration, the specialist recognized that there are sufficient issues looked at by the client in getting appropriate and usable data. The application's domain area is centred on the python-based site and mobile computing. To have profound information on client prerequisites and significant parts of the framework, the analyst needs to execute the innovation to diminish issues looked at by the client at the hour of getting data.

9.3.2.4 Web Application Architecture

Users: A user is an individual who uses a PC or organization administration. Clients of PC frameworks and programming items for the most part come up short on the specialized skill needed to completely see how they work.

Front-end: The frontend of a product program or site is everything with which the client associates. From a client's stance, the front of the feature end is inseparable from the UI. From a designer's stance, it is the interface plan and the programming that makes the interface work. Alternately, the backend incorporates capacities and information handling that happens in the background. One of the essential objectives of frontend improvement is to make a smooth or "frictionless" client experience. At the end of the day, the front finish of an application or site ought to be natural and simple to utilize.

Back-end: In the computer world, the "backend" alludes to any piece of a site or programming program that clients don't see. It appears differently about the frontend, which alludes to a program's or site's UI. In programming wording, the backend is the "data access layer," while the frontend is the "presentation layer."

9.3.2.5 Product Analysis for Users

The analysis can be done by users over the product can be done by the user as shown in Table 9.4.

9.3.2.6 Evaluation of Literature Review

For the improvement of the proposed site to be used by the customer, all features are required by the end customer to overhaul the information gathering environment. The composing review makes the establishment to finish further investigation giving an idea with respect to the proposed system. The systems like the proposed structure have been investigated to make a norm for the system; what gives the designer can

Table 9.4 Qualitative analysis for users

S. no	Functionality	Description
1	Registration	Yes
2	Sign-In	Yes
3	Change password	Yes
4	User status	Yes
5	Self-assess	Yes
6	Updates	Yes
7	QR scanning	No
8	Graphical view	No
9	Remote database	No
10	Notification	No

face, what are the features that can't be executed using any and all means. While researching the composition, region, particular and academic assessments have been finished to get them through and through data on the features to be executed in the proposed structure and to examine the advancement to be used in completing the system. Accordingly, the helper assessment is finished guaranteeing review by considering other's work which will deal with the architect toward developing the proposed structure.

9.3.2.7 Critical Success Factor

Push Notification highlights the projected framework and it will give its clients the basic functionalities of this task. The engineer needs to finish all the highlights recognized during the writing audit; an auxiliary exploration that is to be executed to convey a proficient framework to the end client.

9.3.2.8 Critical Success Factor

The Backup Facility in Remote information base, Push Notification highlight the proposed framework will give its clients are the basic functionalities of this task. The engineer needs to finish all the highlights recognized during the writing audit; an auxiliary exploration that is to be executed to convey a proficient framework to the end client.

9.3.3 Research Methods

9.3.3.1 Primary Research

Primary research is data that is obtained first-hand. This means that the researcher conducts the research themselves or commissions the data to be collected on their behalf. Primary research means going directly to the source, rather than relying on pre-existing data samples. It is depicted as social event information or the data about the space of the assessment from this current reality straightforwardly, for no ifs, ands or buts the initial go through about a given subject unmistakably from this current reality. The originator can utilize this evaluation to move further in the improvement of the undertaking. The principal examination procedures utilized by the designer are overview and meeting which are depicted under.

9.3.3.2 Questionnaire

A questionnaire is for research for the succession of inquiries to gather data from responses and can be considered as a sort of composing meeting. It can be done by asking close and personal questions on phone, Google sheets, or post and is given to the respondents who are relied on to investigate and get a handle on the solicitations and record the fitting reaction in the space gave. Since the objective clients of the design are gigantic in number, and it is difficult to connect with every one of them autonomously to collect their necessities and ideas. With study, it is feasible to reach to masses at the same time self-governing of their topographical zone repetitive. Since the real strategies applied for assessment of the review are proficient which is done in terms of outlines and charts for accurate measurement. Clients are allowed to resolve questions to clear fundamentals and it's a sharp approach to overall be educated to the people or be scattered as a paper overview.

9.3.3.3 Questionnaire for User

The researchers are compassionately mentioning you to save a couple of moments to experience the inquiries and assist us with finishing this study by uncovering these underneath questions and these have been cautiously scheduled by adopting the honest procedures for the investigation effort of Kurukshetra University. The title of the task is "Swasthya Sahayak" to discover the issues looked at by the client while searching for an appropriate substance and data about Coronavirus which is a significant issue now daily and how to defeat this issue. This survey will assist the designer with thinking about the different inclinations of the client. The consequence of this assessment will be exceptionally helpful in getting the essential subtleties for building up the application.

This questionnaire will assist the specialist with assessing the undertaking research expectations.

Significantly, you answer all appropriate inquiries altogether.

Section A: This segment is elective and contains your own and expert particulars.

Section B: To check utmost appropriate choices from inquiries beneath and can be checked more alternative dependent on necessities.

Section C: Allowed to distribute viewpoints/propositions and help scientists with understanding the necessities.

Meeting: Interview is a kind of basic evaluation strategy wherein a ton of solicitations are facilitated and are introduced to the interviewee eye to eye. The solicitations are throughout open-wrapped up. It is proposed to accumulate data about a subject. For the proposed structure, those solicitations are introduced in the get-togethers whose answers require a bit of definite portrayal to see the value in the fundamental. The interviewees for the proposed framework are a designer. The point was to get a specific perspective on the clients to accumulate the most restricted data about the design fundamentals. It gives direct data so more chances of getting set up realities and quality rich information, new experiences, and closures. It will help in friendly event unmistakable data about the functionalities of the proposed structure as it is best for breaking down issues.

Interview Questions for Developers

1 Questionnaire on the theme: An examination to dissect the current circumstance in regards to the pandemic occasions all over India.
2 Which sort of System might you want to use for getting Coronavirus-19 updates and why?
3 How the information on the web will be stored? What will occur if the information will be lost?
4 Do you believe that the "Swasthya Sahayak site" will be going to help you if yes kindly indicate how?
5 What is the productivity level of reinforcement in the event that we consider email and cloud innovation for information reinforcement?
6 What are the likely issues that can come during the advancement period of the application?
7 Have there been any issues with the product? On the off chance that indeed, portray issues?

9.3.3.4 Secondary Research

Secondary research is an approach utilizing previous information and uses research material distributed in research reports and comparable records. These reports are accessible by open libraries, sites, and information gathered from effectively filled overviews. It is commonly coordinated going before fundamental investigation. It gathers information from distributed books, journals, diaries, research papers that are outcast confirmed sources. It is acquired from blunders have submitted on the

associated space to acquire from other's misunderstandings saving time instead of rehashing a similar slip-up.

9.3.3.5 Analysis

The questionnaires were circulated among the clients like a student, administration man, manager, and so on. This sort of essential examination which is a sort of review is helpful for the designer in knowing the client's perspectives about the issues looked at by the client while getting the ticket at the railway counter and how this can be upgraded in the railway environment. The overview performed utilizing the survey assisted the developer with having total information and comprehension of the necessity of the clients while getting the ticket. Along these lines, their sentiments assisted the developer with including the functionalities liked by such clients. The work was created to accumulate assessments of the designer about the application to be created. This helped the designer in distinguishing issues looked at by the clients when booking tickets in a disconnected mode in a railway environment. The designer of the proposed framework gained information about the necessary functionalities, advantages, and disadvantages of every module simply from the genuine end-users who will execute the framework. Consequently, the developer focused on the client who will go by train and need to get the overall ticket online where such a framework could be executed. Along these lines, through this exploration, the designer will want to build up the proposed application in a beneficial way, by coordinating some new highlights and modules like QR Scanner, Add the equilibrium into the TTEs account, e-wallet, and so on.

9.3.3.6 System Design

The design stage begins when the developer is finished with the research and examination stage. The design stage incorporates the change of the client's determinations and programming segments into the product interface to fabricate a stage for building up the framework. It covers the structure of blueprints of the framework that incorporate actual demonstrating utilizing programming techniques. In this stage the design is accepting the necessities as the information and will create a rule for the usage as yield. More design components were created for every module in wake of finishing the examination and investigation part which will be taken care of as a contribution to the design.

9.3.3.7 UML Diagram

The following diagrams are drawn for the creation of UML diagrams

Use Cases: A utilization case required by a specialist to track the movement of functionalities referred in the proposed system to support the plan and complexity

level of individual modules could be assessed without any issue and will be helpful in testing the structure forward and sorting it out.

Activity Diagrams: Activity Diagrams used to track movement of control program modules from one activity to another helping to show concurrence similarly as portions of control in the proposed system.

Sequence Diagrams: Sequence Diagrams help the engineer to demonstrate the sequence of relations winning among clients and framework/modules.

Class Diagrams: Class diagrams for the UML class give a self-sufficient portrayal of the things used in structure helpful in the execution stage. The plan can be used as a sort of viewpoint to complete at any stage for improvement climate.

9.3.4 Implementation

Besides, a rundown of issues is referenced and what were the likely arrangements that were given as shown in Table 9.5.

9.3.4.1 Implementation Plan

The implementation steps need to generate the registration and login of app so user be able to login into the system and perform functionality with creating an about us page for users can know about us. The next step is to make a news section for which aware of new news related to Covid-19 4. We developed the feature called self-assessment for users can assess their risk of viruses. After creating self-assessment, we made a section for user status in which user can view their status of risk, i.e., low, medium, or high. The next step is to develop functionality that will let one user have to send a notification to their loved ones for Covid cases in their respective area. The next step is to provide a complete FAQ list to the user. The next step is to provide some health tips in terms of ancient Indian Ayurveda tips to make. Last but not least users can sign out from the website.

Table 9.5 Tools used for implementation

Tools	Purpose
Window 10	Operating system to run all tools for the development of the system
Star UML	To draw UML diagrams
Microsoft word 2010	Used to the documentation of the project
Django	A web framework used for the development of web application
Microsoft Power Point 2010	Used to prepare the presentation slide
SQLite	Remote database (on the server-side)

9.4 Results and Discussion

The web application is installed to https://covidbuddy2021.herokuapp.com/ to know the Covid-19 cases in India.

9.4.1 Covid Update

In this module updation of the Covid-19 system is done because without updation no one can view the data. The web application shows the updated data according to the latest time. Covid update features to update the record of confirmed cases, recovered cases, active cases, and the number of deaths state-wise in India. This web application fetches the latest data so that users can get up-to-date information about the Covid cases. This application includes the most recent information and updates consequently (Fig. 9.1).

9.4.2 Login

This is the module from which users can log in to the application as shown in Fig. 9.2.

9.4.3 Register

In this module, the users can log in to the application as in Fig. 9.3.

9.4.4 Latest News

This module of the system is important because users are aware of the latest news related to Covid-19 as shown in Fig. 9.4.

9.4.5 Self-Assessment

In this module of the system because the user can take a self-assessment test as in Fig. 9.5.

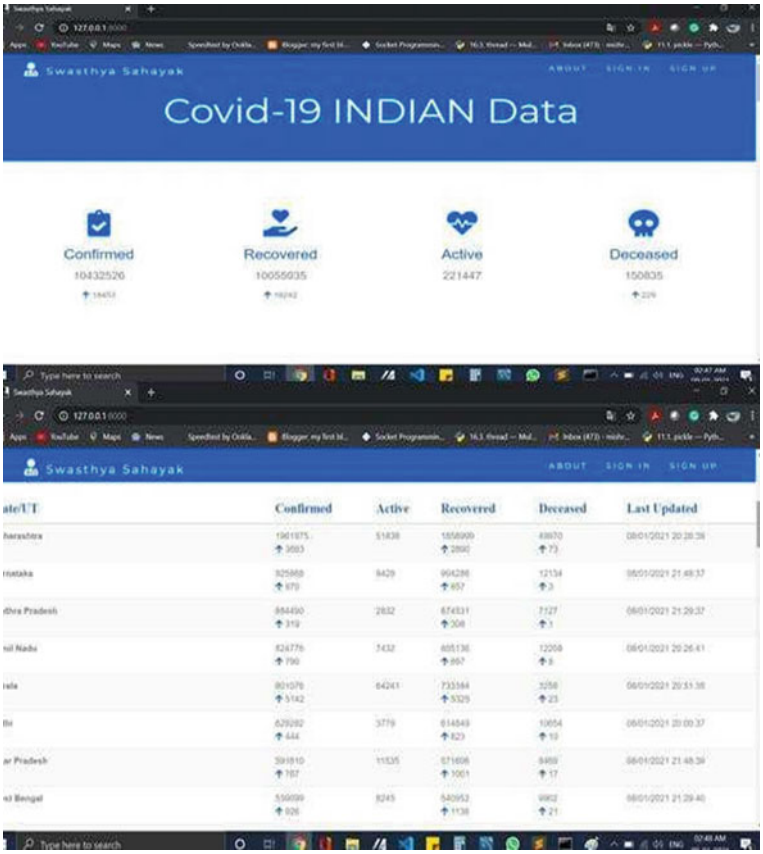
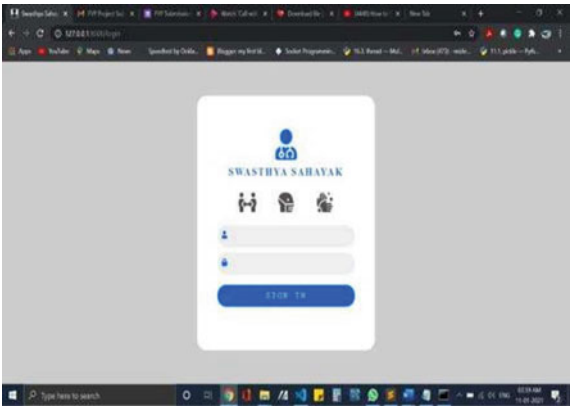


Fig. 9.1 Covid update module

Fig. 9.2 Login module



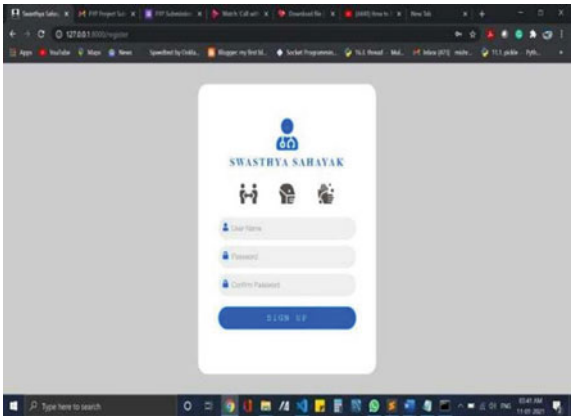


Fig. 9.3 Register module

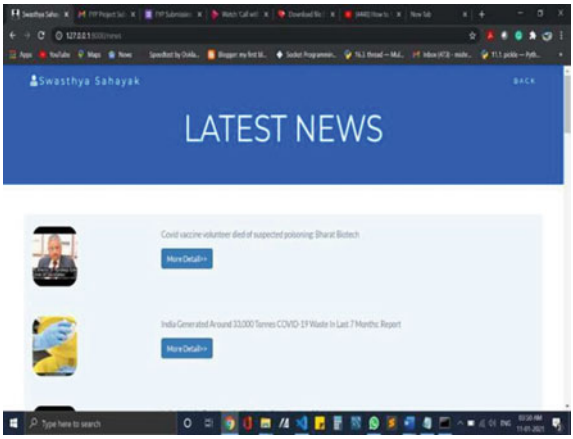


Fig. 9.4 Latest news module

9.4.6 User Status

This is the most important module of the system where the user can view status as shown in Fig. 9.6.

9.4.7 Notify Me

This is a feature by which user can send notifications for Covid-19 to their email as depicted in Fig. 9.7.

Fig. 9.5 Self-assessment module

9.4.8 FAQs

This is a feature in which users get frequently asked questions (Fig. 9.8).

9.4.9 Ayurveda

This is a feature in which users get health advice in terms of Ayurveda is a science of life, transmit gifts of nature to lead a healthy life. Ayurveda's widespread knowledge based on preventive care adopted from principles of "Dinacharya"—daily management and "Ritucharya"—chores to creating a healthy life (Fig. 9.9).

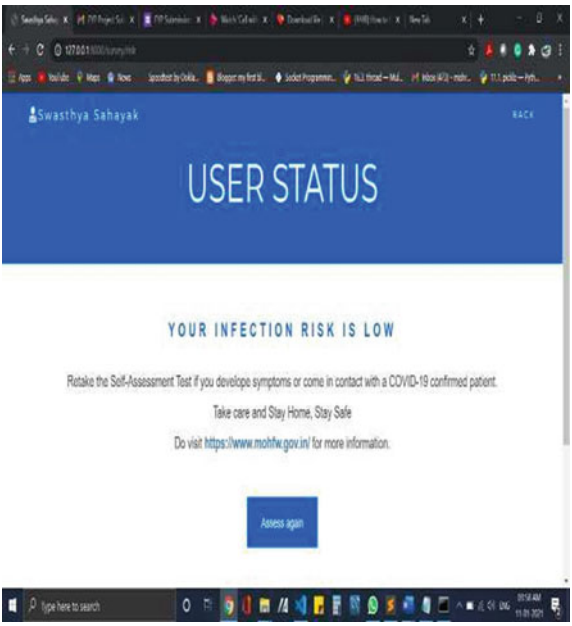


Fig. 9.6 User status module

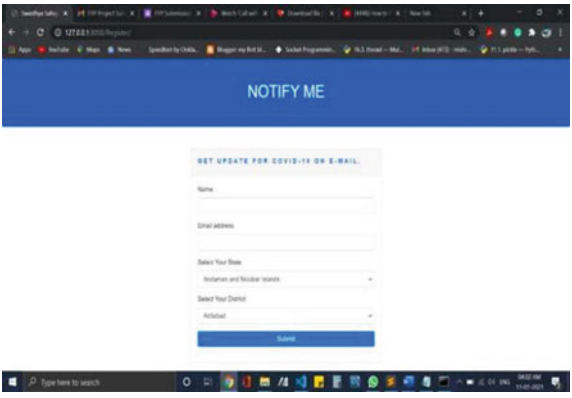


Fig. 9.7 Notify me module

9.5 Testing

Testing is the way toward working on programming to discover and change mistakes and to uncover different classes of slip-ups with a proportion of time and effort. To give a significantly satisfactory and botch-free site, the site should have to go up against the testing framework and appraisal of each and every module and value. This importance the testing stage is done to check and support the completed site. With the objective that it will execute and presented how it was ordinary and all of the functionalities fill in as was described in the endeavor specific construction.

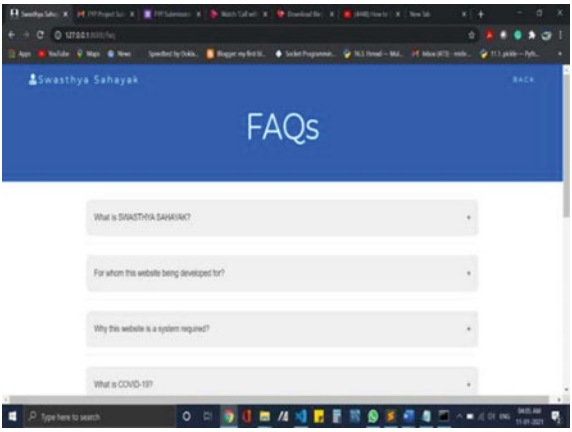


Fig. 9.8 Snapshot of FAQs to be asked

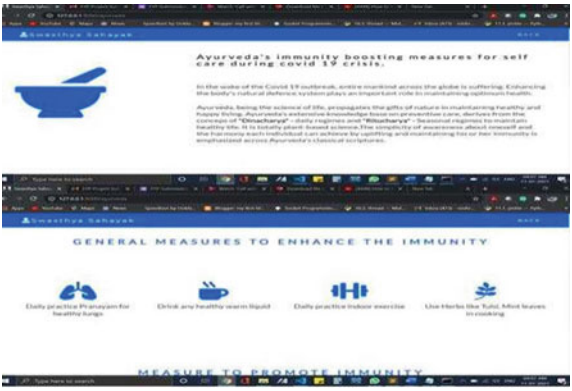


Fig. 9.9 Ayurveda module snapshot

The testing needn't bother with that the specialist needs to test portion of code to check errors and recover from slip-ups and messes with that are accessible in it yet it will likely uncover the conditions that could influence the site or the projector in making the undertaking ineffectual. Thusly, testing is done to redesign the idea of the site conveyed so it is sufficient by the normal customer with satisfaction. Programming progression estimates ordinarily revolve around avoiding bungles, recognizing and redressing the item faults that do occur, and anticipating enduring quality after progress. The test connection, methods, and gadgets are tremendous allies of effective and capable testing and quality confirmation. Diverse testing strategies to be executed for the proposed site are given in the accompanying region.

Figure 9.10 shows Covid-19 confirmed cases in India giving exact information about the number of cases at an instant time and date.

Figure 9.11 shows the data helpful to find out the normal state of health of the Indian population.

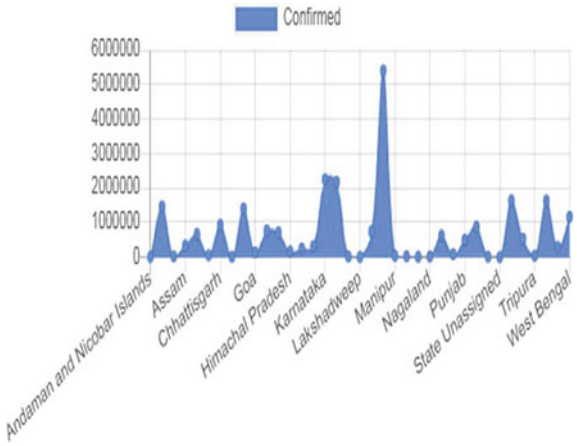


Fig. 9.10 Analysis of confirmed cases state-wise in India

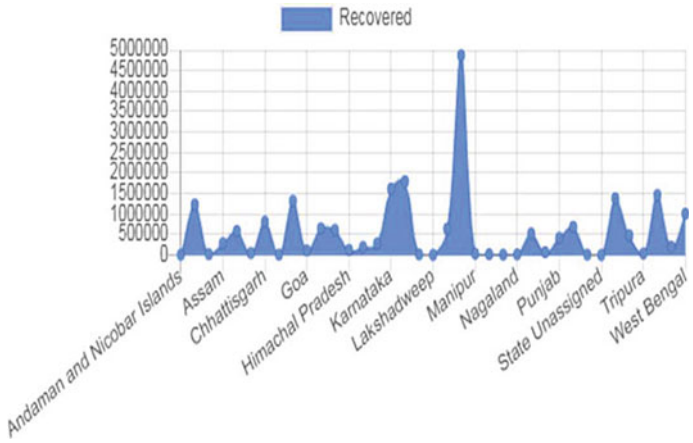


Fig. 9.11 Analysis of recovered cases of Covid-19 state-wise and India

Figure 9.12 depicts the formula for finding the active cases resulting a number of confirmed cases minus the number of recovered cases and deaths and these are cases that are still considered to be transferable.

Figure 9.13 shows the number of death cases in various states of India and prevention is required to fight this dangerous pandemic by apprising of facts and taking proper precautions. As we all are aware the second wave, as well as many variants, are rising of Covid-19 in India proving to be more hazardous. If the rate of deaths has to be minimized, preventive measures and follow-ups are needed as guided by the government.

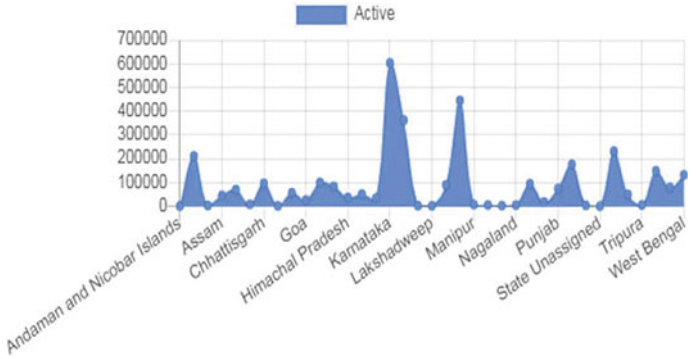


Fig. 9.12 State-wise active cases depicting active cases across India

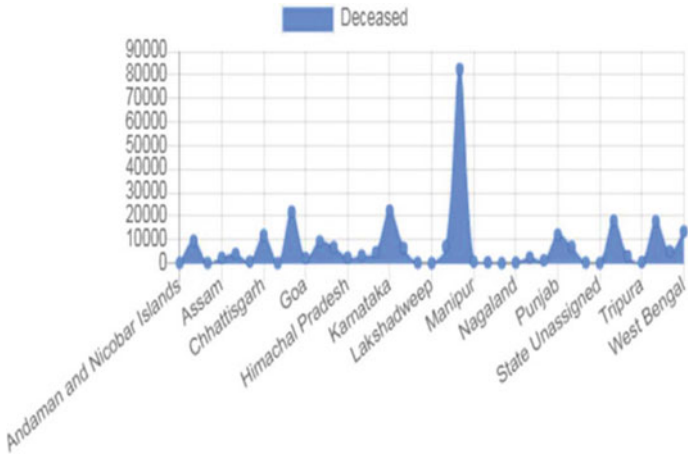


Fig. 9.13 Analysis of death cases in states of India

9.6 Conclusions

The proposed work presented a complete analysis of the COVID-19 outbreak situation in India has been designed to be helpful to users having a smartphone or computer and Also justified issues of saving time and can track cases of COVID-19 around the country at single platform. The intention behind the app is to find the person is at risk of being diagnosed with Covid-19. In addition, there is also a possibility of self-assessment tests to staying healthy. As user requirements keep on changing with time so the app may not solve all the issues. The survey of the research shows how the app-building went through various phases during its development by conducting a study in which users faces a lot of problems during the pandemic.

9.7 Implications of Future Work

The system is being developed for only web-based applications not having any mobile application facility. Users have to access themselves, from time to time to track their health status. Users will need a good internet connection for accessing the web application. The self-assessment feature is not completely accurate because it is just an idea for implementing the best practice for the development of a website as a project for learning. Manual accessing is slow because users are asked several questions; this consumes more time and effort. This web application is not available in offline mode. So it provides the online view of the Covid-19 pandemic to the users. The framework will work upon the present issues and will beat the issue that it has at present. The key characteristic of the site is self-assessment test based on the guidelines of ICMR, risk status of a person, updates, health advisory, and best practices, Covid-19 statistics based on geo-location, nationwide and emergency COVID-19 statistics, and contact helpline along with the ICMR. The cases are rising very fast and there is a need for control strategies from the administrative unit, frontline health workers, researchers, and scientists. This application helps to improve the process of manual accessing, gives knowledge about the symptoms, maintaining a list of cases, and is reliable as this application is not available in offline mode. The future enhancements in an existing system can be done like developer will be developing an android application of dashboard by adding features of online appointments from doctors and notifying on health issues and keeping dashboard live on the server.

References

1. Mankar V, Naravane M, Chakole S (2021) The rise and impact of Covid-19 in India: Aarogyasetu App. Europ J Molec Clin Med 8(1)
2. Gupta N, Bhatnagar T, Rade K, Murhekar M, Gangakhedkar R, Nagar A (2020) Strategic planning to augment the testing capacity for Covid-19 in India. Indian J Med Res 151(2)
3. Kumar SU, Kumar DT, Christopher BP, Doss CGP (2020) The rise and impact of Covid-19 in India. Front Med 7(250)
4. Smith RD (2006) Responding to global infectious disease outbreaks: lessons from SARS on the role of risk perception, communication and management. Soc Sci Med 63(12):3113–3123
5. Peeri NC, Shrestha N, Rahman MS, Zaki R, Tan Z, Bibi S, Baghbanzadeh M, Aghamohammadi N, Zhang W, Haque U (2021). The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned?. Int J Epidemiol 49(3):717–726
6. Aggarwal A, Dhindsa KS, Suri PK. Performance-aware approach for software risk management using random forest algorithm. Int J Softw Innov (IJSI) 9(1):12–19
7. Usage patterns and implementation of random forest methods for software risk and bugs predictions. Special Issue 8(9S):927–932, 2019. https://doi.org/10.35940/ijitee.i1150.0789s19
8. Kumar R (2012) ICT in health sector: role of ICT. LAP LAMBERT Academic Publishing, Lexington Kentucky(KY), USA
9. Kachroo V (2020) Novel coronavirus (COVID-19) in India: current scenario. Int J Res Rev 7(3)

10. Gupta R, Pal SL, Pandey G (2020) A comprehensive analysis of COVID-19 outbreak situation in India
11. Jena PK (2020) Impact of Covid-19 on higher education in India. Int J Adv Educ Res 5(3):77–81
12. Brodeur A, Islam A, Gray D, Bhuiyan SJ (2020) A literature review of the economics of Covid-19. IZA DP no. 13411. Institute of labour economics. Discussion paper series, pp 1–63
13. Aggarwal N, Ashraf H (2020) Covid-19 impact on daily life (CIDL) survey. Dvara research, pp. 1–20
14. Prinja S, Bahuguna P, Chugh Y, Vassall A, Pandey A, Aggarwal S, Arora NK (2020) A model based analysis for Covid-19 pandemic in India: implications for Health systems and policy for low and middle-income countries. medRxiv, pp 1–25
15. Pal R, Yadav U (2020) Covid-19 pandemic in India: present scenario and a steep climb ahead. J Primary Care Commun Health 11
16. Kumar SU, Kumar DT, Christopher BP, Doss CGP (2020) The rise and impact of Covid- 19 in India. Front Med 7(250):1–7
17. Pathak M (2020) Covid-19 research in India: a quantitative analysis. Indian J Biochem Biophys 57:351–355
18. Siddiqui AF, Wiederkehr M, Rozanova L, Flahault A (2020) Situation of India in the Covid-19 pandemic: India's initial pandemic experience. Int J Environ Res Public Health 17(8994):1–18
19. Shivani Gaba DK, Singla S (2019) A genetic improved quantum cryptography model to optimize network communication. Special Issue 8(9S):256–259
20. Gaba S, Aggarwal A, Nagpal S (2021) Role of machine learning for ad hoc networks. Cloud IoT Based Vehic Ad-Hoc Netw 269

Chapter 10
Effectiveness of Digital Elevation Models in Morphometric analysis Using Remote sensing and GIS Approach for Smart Society

Kakoli Banerjee, M. B. Santhosh Kumar, and L. N. Tilak

Abstract Geographical information systems (GIS) are proven to be extremely useful in a variety of natural resource applications in Smart Society, including topography hydrology modeling for water harvesting. The watershed is the basic object in such systems, which is either physically drawn on topographic map sheets or computationally generated from digital elevation model (DEM) data. This article describes a method for automatically defining drainage and watersheds using a terrain's DEM. The efficacy of DEMs in morphometric analysis of Rapti sub-watershed in Gorakhpur District, Uttar Pradesh, was investigated in this study. The Rapti sub-watershed, which covers an area of 20.8 km^2 in Gorakhpur, was selected and six different threshold levels of 0.05, 0.25, 0.50, 0.125, 1.0, and 2.0 km^2 were applied during the automization process using ESRI ArcGIS, open-source QGIS software, and freely available DEMs and surface models of ASTER-GDEM, and ALOS-DSM displays streams with finer features than ASTER-GDEM at all stream thresholds. For ALOS-DSM, the correlation coefficient R2 between thresholds and stream length (Lu) is 0.8817, while for ASTER-GDEM, it is 0.9017. In the comparative assessment of ASTER-GDEM and ALOS-DSM, carried out has identified ALOS-DSM exhibits finer drainage networks in comparison with ASTER-GDEM proving ALOS-AW3D 30 DSM is the best alternate source available.

Keywords ASTER-GDEM · ALOS-DSM · Morphometry · Thresholds

K. Banerjee (✉)
Department of Computer Science and Engineering, JSS Academy of Technical Education Noida, C-20/1, C Block, Phase 2, Sector 62, Noida, Uttar Pradesh 201,309, India
e-mail: kakoli.banerjee@jssaten.ac.in

M. B. Santhosh Kumar · L. N. Tilak
Department of Civil Engineering, JSS Academy of Technical Education Noida, C-20/1, C Block, Phase 2, Sector 62, Noida, Uttar Pradesh 201,309, India
e-mail: santhoshkumar@jssaten.ac.in

L. N. Tilak
e-mail: tilakln@jssaten.ac.in

V. Bali et al. (eds.), *Decision Analytics for Sustainable Development in Smart Society 5.0*, Asset Analytics, https://doi.org/10.1007/978-981-19-1689-2_10

10.1 Introduction

Topographic mapping is becoming more important as a national effort, and precise watershed features defined based on topographic maps are now regarded essential national geospatial data infrastructure components. Because drainage networks and elementary catchments which represent the feature of terrain and aggregation structure with a hierarchical structure, automating the acquisition of their spatial representation is critical. This structure connects multiple hydrographic information aggregation levels. It may be used for automated generalisation as well [27]. Watershed mapping's enormous contribution to national development continues to be recognized by national, state, and municipal governments all over the world. The extraction of catchment and drainage networks has long relied on digital elevation models (DEMs) obtained from remote sensing data. Grid DEMs are digital files that store the elevation values of the landscape at the nodes of a regular square grid [20, 21, 23]. The whole region delivering runoff to, and maintaining part or all of the streamflow of, the main stream and its tributaries is referred to as a drainage basin or a watershed. Quantitative analysis of a drainage network within a basin or watershed can reveal the form and dominant processes within it, information that is essential for watershed management, floods, soil erosion, mass movement, neotectonic activities, surface and groundwater, hydrograph derivation, and understanding the discharge characteristics of ungauged streams. ecology, etc. Since then, data gathered have been used as the foundation for a variety of related studies. Instead of using empirical and traditional methods for information processing, the quickly growing spatial data innovation, Geographical Information system (GIS), Geographical Positioning System (GPS), and remote sensing, have effective tools and devices to conquer the majority of logical questions, recognizing land and water assets [29–37]. The satellite-derived data may be used for quantitative evaluation and precise delineation of watersheds, in addition to other morphometric variables [1, 2, 11, 13]. Availability of a plethora of new satellite-based geological data sets has opened up new avenues for hydrologic and geomorphologic research, such as surface morphology and channel network structure analysis [42]. The main application is currently in widespread use in the automated extraction of the channel drainage network. Several research have been done to automate the extraction of drainage networks due to the growing accessibility of grid DEMs. Setting up physically based distributed hydrologic models necessitates DEM-based catchment delineation [16, 25, 26]. A successful GIS-based automated catchment delineation depends on the quantity and accessibility of data sources, the kind and distinguishing features of the modeled region, and the accuracy and competence of the data analysts [2, 7, 12, 14]. DEM-derived fundamental geomorphic or topographic features frequently fill in as contribution for the models available. DEMs and associated algorithms thus give the basis for an almost infinite number of logical questions regarding space and terrain [24, 27]. The main problems for extraction, delineation of drainage networks in watersheds, and countless other investigations are the accuracy levels of DEM and the algorithm's concept [8, 22]. Delineation of watershed and drainage network required for floodplain alteration, water resource

management, geomorphological and hydrological modeling, surface water mapping and flood risk management. Because high-resolution DEMs are frequently unavailable, open-source goods must be evaluated. In all kinds of hydrological research, morphometric analysis is a quantitative assessment of the form characteristics of the earth's surface. This refers to the capacity and algorithm-based analysis of the earth's surface and landform units. Morphometric analysis expresses physiographic information such as channel length, drainage divide location, channel network configuration, watershed slope, and geomorphologic parameters such as circulatory ratio, bifurcation ratio, shape factor, relative relief, and drainage density for watershed prioritization and implementation of soil and water conservation measures [3, 15, 34, 49]. There are two types of variables that might influence the behavior of a watershed: (1) fixed drainage characteristics such as drainage density and (2) transitory or changeable features such as precipitation and land use [49]. Prioritizing sub-watersheds in a river basin based on morphometric features is one of the simplest ways to construct watersheds for use and management. In India, there has been a lot of study on drainage morphometric analysis-based sub-watershed prioritization and management of watersheds in small- and medium-sized rivers [10, 32, 41]. Similar studies in large river basins, on the other hand, are few. The bulk of published research on the Indian continent has used conservative techniques to evaluate drainage and morphometric features of watersheds. Traditional research methods used topographic maps and/or field surveys to examine morphometric features. In recent years, however, numerous scientists and researchers all around the globe have employed GIS and RS approaches to achieve automated drainage network and feature extraction [3, 15, 34, 49]. As a result, the quality and accuracy of freely available DEM and DSM were examined for hydrologic research in this work by using automated drainage extraction from ASTER-GDEM and ALOS-DSM to better comprehend the drainage basin and topography of the watershed.

10.2 Study Area

The Rapti sub-watershed, with an area of 20.8 hectares, is located in the Gorakhpur district of Uttar Pradesh. The region's major lithological elements are Khadar (new alluvial) and Bhangar (old alluvial). The climate in this area is mild. It has an average temperature of 25.680 °C, with a maximum temperature of 31.950 °C, a minimum temperature of over 19.570 °C, and an annual rainfall of more than 119.2 cm. This watershed is located in the Rapti River's catchment and covers a total area of 2080.15 hectares. It is located between the latitudes of 26°57′06" N and 26°59′54" N, and the longitudes of 83°15′ 23" E and 83°17′53" E (Fig. 10.1).

Data Used: For the present study, we have used freely available date as mentioned in Table 10.1.

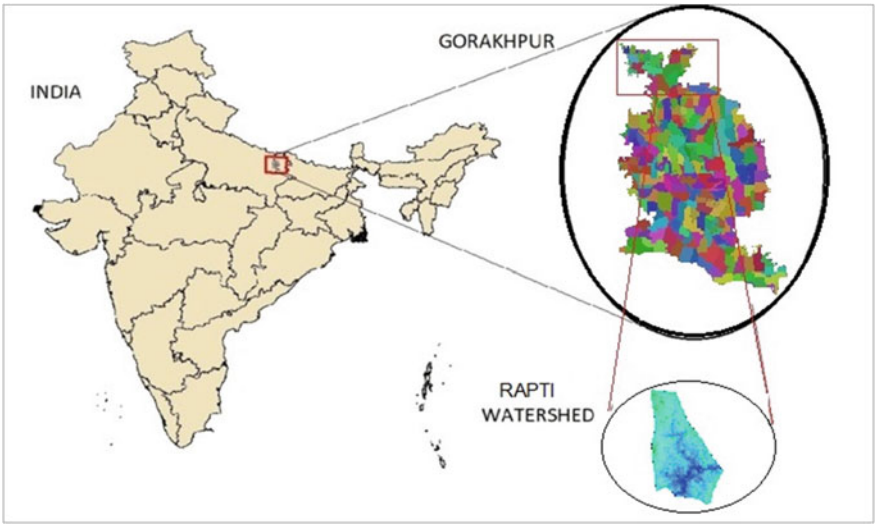


Fig. 10.1 Location Map of Rapti Sub-watershed (Gorakhpur, India) Source [4–6]

Table 10.1 Details of elevation and surface models used Source [4–6]

S no	Operator	DEM	Spatial resolution (m)	Site to download	DEM format
1	NASA	ASTER-GDEM	30	https://earthexplorer.usgs.gov/	WGS 84, Geo-Tiff
2	JAXA	ALOS	30	http://global.jaxa.jp/projects/sat/alos/index.html	WGS 84, Geo-Tiff

10.3 Methodology

The strategy embrace to assess the drainage network extracted from DEM and DSM comprises involve three significant advances: Digitization of stream network from Topographic map, Automated Delineation from ASTER-GDEM and ALOS-DSM by GIS software module, morphometric investigation and correlation. Stream network of the location has been digitized from topographic map of 1:50,000 scale in GIS environment after geometric transformation in Universal Transverse Mercator coordinate system.

The region of interest was obtained in Geo-TIFF format using geographic coordinates from the freely accessible ASTER-30 m GDEM and ALOS World 3D-30 m (AW3D30) of one arc-second (arcsec) grid, then projected to Universal Transverse Mercator (UTM). To extract drainage channels, a hydrological tool in GIS software.

Starting with a full DEM, the automated technique for defining streams went through a number of phases (Fig. 10.2). There are no depressions in a DEM that has

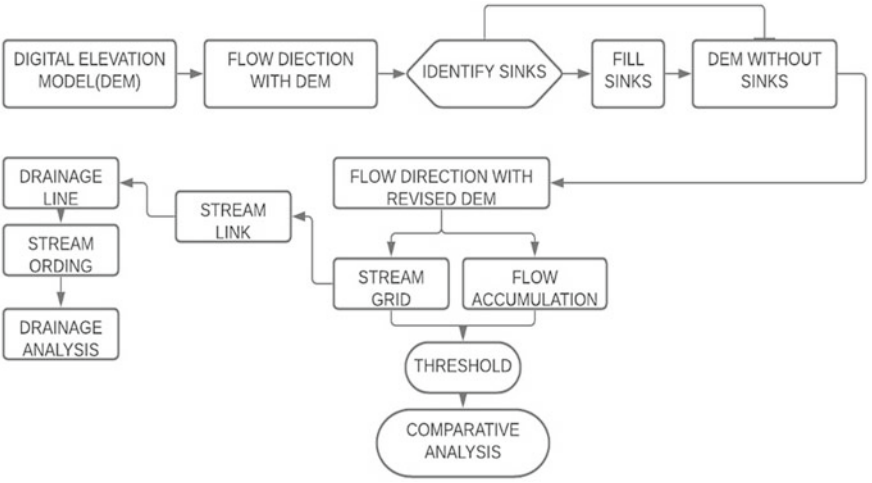


Fig. 10.2 Flow chart representing method adopted for delineation of drainage network Source [4–6]

been filled. To prepare the sink, we used the fill function in the Arc hydro toolset of ArcGIS software to fill up the depressions. The flow direction raster was created in the next step. The direction of water flowing out of each cell of a filled DEM may be determined using a flow direction raster. This GIS method assigns a cell's flow direction to one of the eight adjacent cells with the greatest distance weighted gradient [29, 36, 37, 48]. At that point, we utilized flow accumulation function in the Arc hydro tools for the extraction of flow accumulation raster. Stream network and stream link delineation were performed in the following phases utilizing the stream network and stream link delineation tools. This procedure is dependent on the accumulation value's threshold. The morphometric indices of sub-watershed was computed using standard mathematical expressions as mentioned below (Table 10.2) and relation between threshold and stream length was extracted to justify the threshold influence.

10.4 Results and Discussions

10.4.1 Extraction of Rapti Sub-Watershed Drainage Network

Georeferenced Survey of India (SOI) toposheets were used to automatic delineation of the Rapti sub-watershed with a spatial resolution of 30 m from the ALOS-DSM & ASTER-GDEM data. Various geoprocessing approaches in a GIS context were used to extract the contributing basin area. The Rapti sub-watershed basin's morphometric characteristics have been computed, and the findings are shown in Table 10.2.

Table 10.2 Standard mathematical formulae for calculation of morphometric parameters Source [4–6]

S. no	Morphometric parameters	Formula\Relatrionship	References
1	Stream order (u)	Hierarchical order	Strahler [47]
2	Stream length (Lu)	Lenguth of stream	Horton [18]
3	Bifurcation ratio (Rb)	Rb = Nu\Nu + 1, Nu = total number of stream segments of order "u", Nu + 1 = number pf stream segments of the next higher order	Schumm [39]
4	Drainage density (Dd)	D = Lu\A, where A is the total area of the basin $(Km)^2$, Lu is the total stream length of all orders	Horton [17]
5	Stream frequency (Fu)	Fs = Nu\A, where Nu is the total number of streams of all order, A is basin area in km^2	Horton [17]
6	Form factor (Ff)	Ff = A\Lb2 is the square of the basin length (km), A is the basin area in km^2	Horton [17]
7	Elongation ratio (Re)	Re = 2sqrt (A\p)Lb, where A is the area $(km)^2$ and Lb = basin length	Schumm [39]
8	Ruggedness number (Rn)	Rn = Bh Dd Where, Bh = Basin relief Dd = Drainage density	Schumm [39]
9	Constant chanel maintenance (C)	Inverse of drainage density	Schumm [39]

Dendritic drainage pattern were observed and is controlled by the area's overall terrain, geology, and rainfall conditions. ASTER-GDEM & ALOS-DSM are used to autonomously prepare drainage networks.

The following subsections demonstrate the importance of various morphometric metrics. Additional values for these parameters are acquired using techniques given by other researchers for the subject region and are listed in the corresponding descriptions. Using GIS software, the lengths of the streams, areas of the watershed, stream ordering , linear aspects, areal aspects, relief aspects drainage density, and frequency distribution of the watershed region were determined [17, 18, 33–39, 39, 40, 40–46].

10.5 Linear Aspects

10.5.1 Stream Order (U)

The technique developed by Strahler was used to rate streams in the current study. In the Rapti sub-watershed, stream orders are classed up to fourth orders. The stream in the Rapti sub-watershed might be classified as a fourth-order stream (Table 10.3). For first-order streams, the highest stream order frequency is recorded, followed by

Table 10.3 Morphometric parameters of the sub-watershed extracted from ASTER-GDEM and ALOS-DSM at different thresholds Source [4–6]

	ASTER-GDEM							ALOS DEM						
Sub water shed	Stream order	0.05	0.125	0.25	0.5	1	2	Stream order	0.05	0.125	0.25	0.5	1	2
Stream number	1	764	550	345	286	175	68	1	911	486	365	214	162	174
	2	443	290	197	91	60		2	378	266	178	203	106	28
	3	251	60	60				3	192	157	106	28	28	
	4	60						4	106	28	28			
								5	28					
Strem length	1	26.24	19.36	12.17	10.22	6.092	2.335	1	32.06	17.22	12.63	7.255	5.536	6.14
	2	16.67	10.57	7.12	3.371	2.328		2	13.94	9.774	6.401	7.48	3.875	1.23
	3	9.188	2.328	2.328				3	7.493	5.88	3.895	1.23	1.23	
	4	2.317						4	4.178	1.242	1.229			
								5	1.276					
Stream length ratio		0.64	0.55	0.58	0.33	0.38			0.43	0.57	0.51	1.03	0.7	0.2
		0.55	0.22	0.33					0.54	0.6	0.61	0.16	0.32	
		0.25							0.56	0.21	0.32			
									0.31					
Bifurcation ratio		1.57	1.83	1.71	3.03	2.62			2.3	1.76	1.97	0.97	1.43	4.99
		1.81	4.54	3.06					1.86	1.66	1.64	6.08	3.15	
		3.97							1.79	4.73	3.17			
									3.27					
Area (km^2)		20.8	20.8	20.8	20.8	20.8	20.8		20.8	20.8	20.8	20.8	20.8	20.8
Perimeter		20.6	20.6	20.6	20.6	20.6	20.6		20.6	20.6	20.6	20.6	20.6	20.6

(continued)

Table 10.3 (continued)

	ASTER-GDEM							ALOS DEM					
Mean bifurcation ratio		2.451	3.186	2.384	3.032	2.617			1.985	2.719	2.262	3.526	2.29	4.992
Mean stream length ratio		0.479	0.383	0.455	0.329	0.382			0.51	0.46	0.477	0.59	0.5	0.2
Basin length (km)		0.36	4.76	5.27	5.48	6.31	8.9		2.6	2.7	2.7	2.8	3.15	3.58
Drainage density dd (km/km^2)		2.61	1.55	1.03	0.65	0.4	0.11		2.77	1.64	1.161	0.76	0.51	0.35
Stream frequency		72.98	43.26	28.94	18.12	11.29	3.26		76.29	45.04	32.54	21.39	14.23	9.172
Ruggedness bumber (rn)		0.199	0.117	0.079	0.05	0.031	0.009		0.047	0.028	0.02	0.013	0.009	0.006
Elongated ratio		0.458	0.42	0.379	0.365	0.317	0.224		0.769	0.74	0.741	0.71	0.63	0.55
Form factor		1.09	0.91	0.74	0.69	0.522	0.262		3.077	2.85	2.853	2.65	2.09	1.62
Constant of channel maintance		0.38	0.64	0.96	1.53	2.47	8.91		0.36	0.61	0.86	1.3	1.95	2.82
Releief ratio (rh)		17.41	15.96	14.42	13.86	12.04	8.53		6.53	6.2	6.29	6.07	5.39	4.75

second-order streams. If a result, as the stream order increases, there is a drop in stream frequency, and vice versa.

10.5.2 Stream Length (Lu)

Stream lengths define the total lengths of stream segments of each of the successive orders in a basin and resemble a linear geometric series in which the first term represents the first-order stream's average length [17, 18]. The bedrock's hydrological characteristics and drainage area are measured by the stream length. In a well-drained watershed with porous bedrock and formations, only a few comparably longer streams develop, whereas when the bedrock and formations are less permeable, a large number of smaller streams emerge [41]. The results of order-wise stream length in the Rapti sub-watershed are shown in Table 10.3. As can be observed, the cumulative stream length is greater in first-order streams and decreases as the stream order increases.

10.5.3 Stream Length Ratio (RL)

According to Horton's law (1945), the mean stream length segments of each of the succeeding orders of Automatic extraction of Streams utilising ASTER-GDEM & ALOS-DSM are computed in Table 10.3. This movement could be attributed to changes in slope and topography, indicating that the Kanhar basin's streams are in a late stage of geomorphic evolution [43, 50].

10.5.4 Stream Number (Nu)

The number of stream channels in each order is referred to as stream order. "In a given drainage basin, the number of streams of various orders tends to approach as an inverse geometric series, with the first term equal to unity and the ratio equal to the bifurcation ratio." [17, 18]. According to this rule, the number of streams counted for each order is displayed on a logarithmic scale on the y-axis vs order on an arithmetic scale on the x-axis. GIS platforms are used to count and tabulate the number of streams of various orders, as well as the total number of streams in the basin. The number of streams reduces as the stream order increases; the size and order of tributary basins are mostly determined by the physiographical, geomorphological, and geological aspects of the region. The various values obtained in both elevation models are listed in Table 10.3.

10.5.5 Bifurcation Ratio (Rb)

Rb was thought to be a measure of relief and dissection [18], but only a little variation was observed for different places with varying conditions [38], particularly where strong geological control was present. The bifurcation ratio (Rb) is the proportion of one order's stream segments to the next higher order's stream segments. It's a one-dimensional metric that measures how well streams of different orders are integrated in a drainage basin. Table 10.3 shows how the Rb for the Rapti sub-watershed was determined using two elevation models. When the influence of geological factors on the drainage network is minimal, the mean bifurcation ratio (Rbm) for a basin typically ranges from 3.0 to 5.0. The higher Rb values in the Kanhar basin indicate that structural disturbances have a strong influence on the drainage pattern, while the lower values indicate that structural disturbances have a lesser impact on the sub-basins [9, 46, 50].

10.6 Basin Geometry

10.6.1 Length of the Basin (Lb)

The basin length is the basin's longest dimension parallel to the major drainage line. Table 10.3 shows the length of the basin for various thresholds computed using ASTER-GDEM and ALOS-DSM [39].

10.6.2 Basin Area (A)

Another essential characteristic, like the length of the stream drainage, is the watershed's area. Schumm [39] discovered an intriguing relationship between total watershed areas and total stream lengths supported by contributing regions. Table 10.3 shows how ASTER-GDEM and ALOS-DSM were used to determine the basin area for the research region.

10.6.3 Basin Perimeter (P)

The outside edge of the watershed that encloses its area is known as the basin perimeter. It is a measure of the distance between watersheds that may be used to determine the size and form of a watershed. Table 10.3 shows how ASTER-GDEM and ALOS-DSM were used to compute the basin perimeter for the research region.

10.6.4 Elongation Ratio (Re)

The elongation ratio (Re) is the ratio of the diameter of a circle with the same area as the basin to the maximum basin length [39]. It is a measure that is used to determine the shape of a river basin and is impacted by climatic and geologic variables. In comparison to an extended basin, a circular basin releases runoff more effectively [42, 43]. The re value of Rapti sub-watershed in the ASTER-GDEM varies from 0.45 for 0.05 thresholds to 0.22 for 2.0 criteria, whereas it was 0.76 for 0.05 thresholds and 0.56 for 2.0 thresholds in the ALOS-DSM (Table 10.3). Lower Re values imply considerable erosion and sediment load susceptibility, whereas higher elongation ratio values suggest good infiltration capability and limited runoff [37].

10.6.5 Form Factor (Rf)

The form factor is the ratio of the basin's size to the square of the basin's length [18]. For a completely round basin, the form factor would always be larger than 0.78. The smaller the form factor, the longer the basin will be. Table 10.3 shows the Rf value of the Rapti sub-watershed determined using ASTER-GDEM and ALOS-DSM. As a result, the Rapti sub-watershed is a lengthy one.

10.7 Drainage Analysis

10.7.1 Stream Frequency (Sf)

The total number of stream segments of all kinds per unit area is known as stream frequency (Sf) [17]. Low stream frequency Sf indicates the presence of a porous underlying material and low relief [37]. Although enumerating channel segment counts for unit areas is difficult [33], a census of Rapti sub-watershed stream frequency has been attempted (Table 10.3). The lithology of the basin determines stream frequency, which reflects the drainage network's structure. The basin's stream frequency (Fs) has a positive connection with the region's drainage density, implying that stream population grows as drainage density rises. Channel frequency density is a tool for detecting the erosional processes at work in a particular location; more specifically, it is used in conjunction with stream orders and their characteristics to provide information on relief sequences and the degree of roughness in the area [33].

10.7.2 Drainage Density (Dd)

The ratio of a basin's total stream length to its total area is known as drainage density (Dd) [47]. Drainage density is influenced by the variables that affect the watershed's characteristic length. Valley density, channel head source area, relief, climate and vegetation (Moglen et al. 1998), soil and rock attributes [19], and landscape creation processes are all linked to drainage density. The drainage density estimated from ASTER-GDEM and ALOS-DSM is shown in Figs. 10.3a and 10.4a. A basin with weak and impermeable underlying material, minimal vegetation, and high relief is associated with higher drainage density. Low drainage density results in coarse drainage texture, whereas high drainage density results in fine drainage texture, large runoff, and the possibility for erosion in the basin. Strahler is a fictional character who appears in the film Strahler [47]. Table 10.3 shows the drainage density of various research area criteria.

10.7.3 Constant of Channel Maintenance (1/D)

The constant denotes the amount of basin surface area necessary to build and sustain a 1 km long channel. The channel maintenance constant denotes the relative size of landform units in a drainage basin and has a genetic meaning [38, 39] Table 10.3 shows the channel maintenance constant for the Rapti sub-watershed computed using two elevation models.

10.8 Relief Characteristics

10.8.1 Ruggedness Number (Rn)

The roughness number calculated is the combination of basin relief and drainage density, and it effectively combines slope steepness and length [39]. Watersheds with a low ruggedness score are less prone to soil erosion and have inherent structural complexity in relation to relief and drainage density. Table 10.3 shows the Ruggedness Number (Rn) for the Rapti sub-watershed computed using two elevation models.

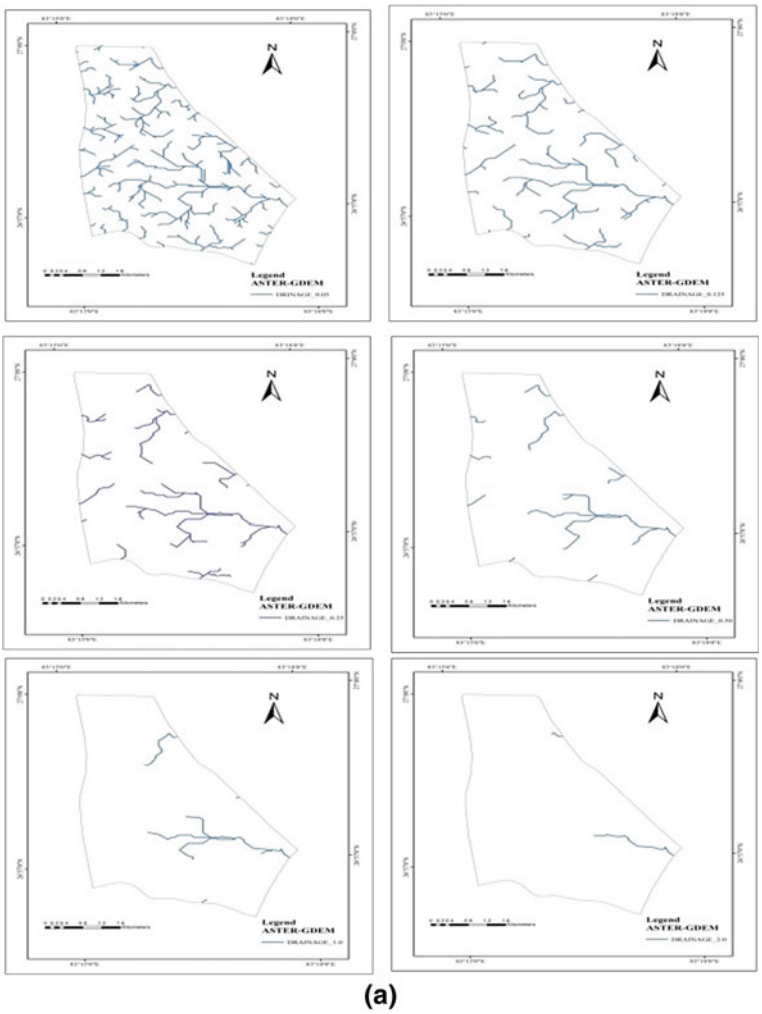
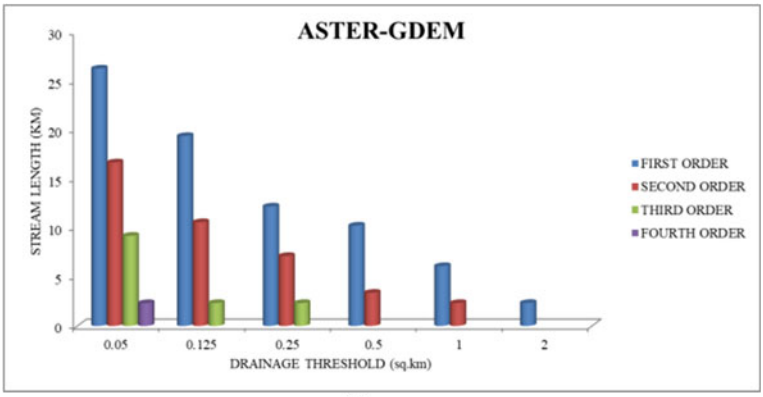


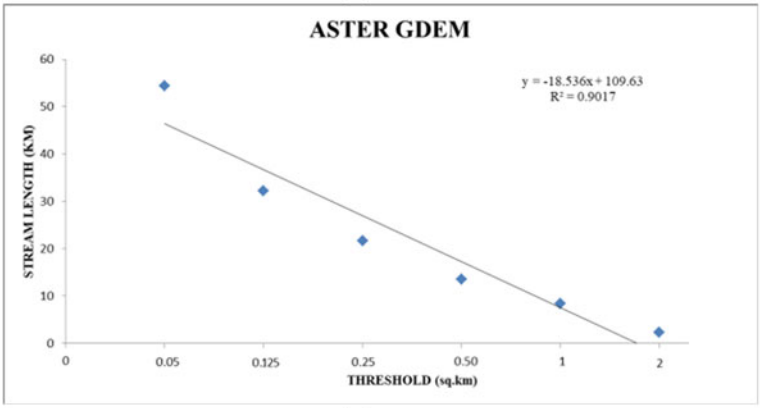
Fig. 10.3 **a** Drainage networks of ASTER GDEM at different thresholds. **b** Chart representing Stream length of ASTER GDEM at different thresholds. **c** Graph representing R^2 for ASTER GDEM Source [4–6]

10.8.2 Relief Ratio (Rh)

The relief ratio is defined as the maximum relief to horizontal distance over the basin's longest dimension parallel to the primary drainage line [39]. The difference in elevation between the highest point in the basin and the lowest point on the valley floor is known as the total relief of a river basin. The low value of relief ratios is due to the refractory basement rocks of the basin, as well as the low degree of slope [25].



(b)



(c)

Fig. 10.3 (continued)

Table 10.3 shows the relief ratio (Rh) for the Rapti sub-watershed computed using two elevation models (Fig. 10.5).

10.9 Conclusions

The drainage networks of ASTER-GDEM (Fig. 10.3a, b) and ALOS-DSM (Fig. 10.4a, b), recovered from stream thresholds of 0.05, 0.125, 0.25, 0.50, 1.0, and 2.0 km^2, are shown in the results of this study. The research found that at the lowest stream area of 0.05 km^2, the first-order stream length was 26.24 km (ASTER-GDEM) and 32.06 km (ASTER-GDEM) (ALOS-DSM). The percentage decline in

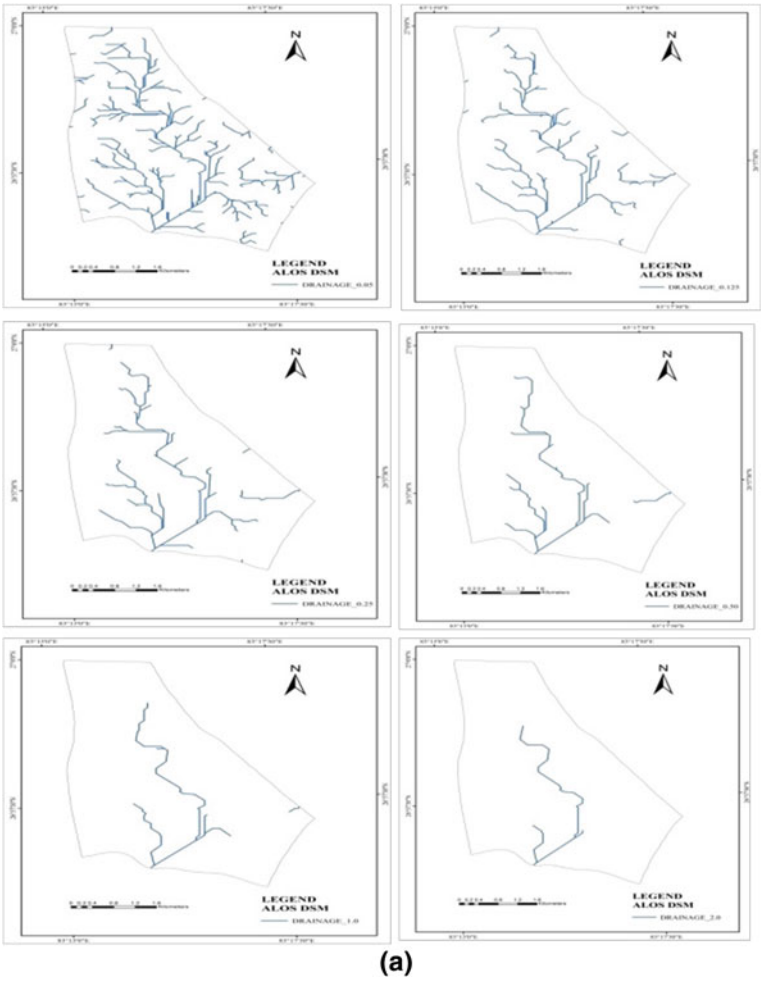
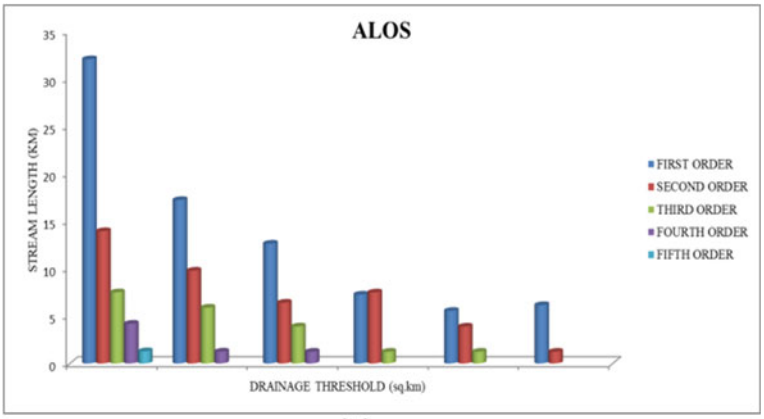


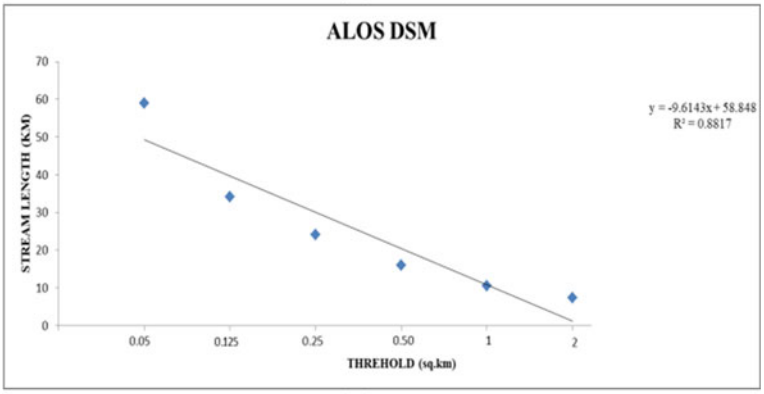
Fig. 10.4 **a** Drainage networks of ALOS DSM at different thresholds. **b** Chart representing Stream length of ALOS DSM at different thresholds. **c** Graph representing R^2 for ALOS DEM Source [4–6]

stream length from first to second order was 68.75% (ASTER-GDEM) and 43.48% (ASTER-GDEM) at 0.05km^2 of stream threshold (ALOS-DSM).

ASTER-GDEM exhibits highest total stream length of 19.365 km (first order), 10.22 km (first order), and 6.092 km (first order), where ALOS-DSM exhibits 17.221 km (first order), 7.255 km (first order), and 5.536 km (first order) when drainage network was delineated at a threshold of 0.125 sq.km, 0.5 sq.km, and 1.0 sq.km, respectively. ALOS-DSM exhibits highest total stream length of 6.14 km



(b)



(c)

Fig. 10.4 (continued)

(first order) compared to ASTER-GDEM (2.335 km), when drainage network was delineated at a threshold of 2.0 km^2 with a rise of second order stream length was noticed.

ALOS-DSM exhibits one higher stream order stream at all the selected threshold values which was not visible in the ASTER-GDEM. When compared to reference topographic map during visual interpretation, it was observed total number of streams was more in the drainage network extracted by DEM and DSM as represented in the Fig. 10.5a, b. The graph of R2 was extracted and their relation between the stream length was noted as 0.9017 (ASTER-GDEM) and 0.8817 (ALOS-DSM). ALOS-DSM (Fig. 10.4c) shows better performance in delineation of drainage network when compared to ASTER-GDEM (Fig. 10.3c). The results of Morphometric analysis are tabulated in Table 10.3 demonstrating the performance of elevation models

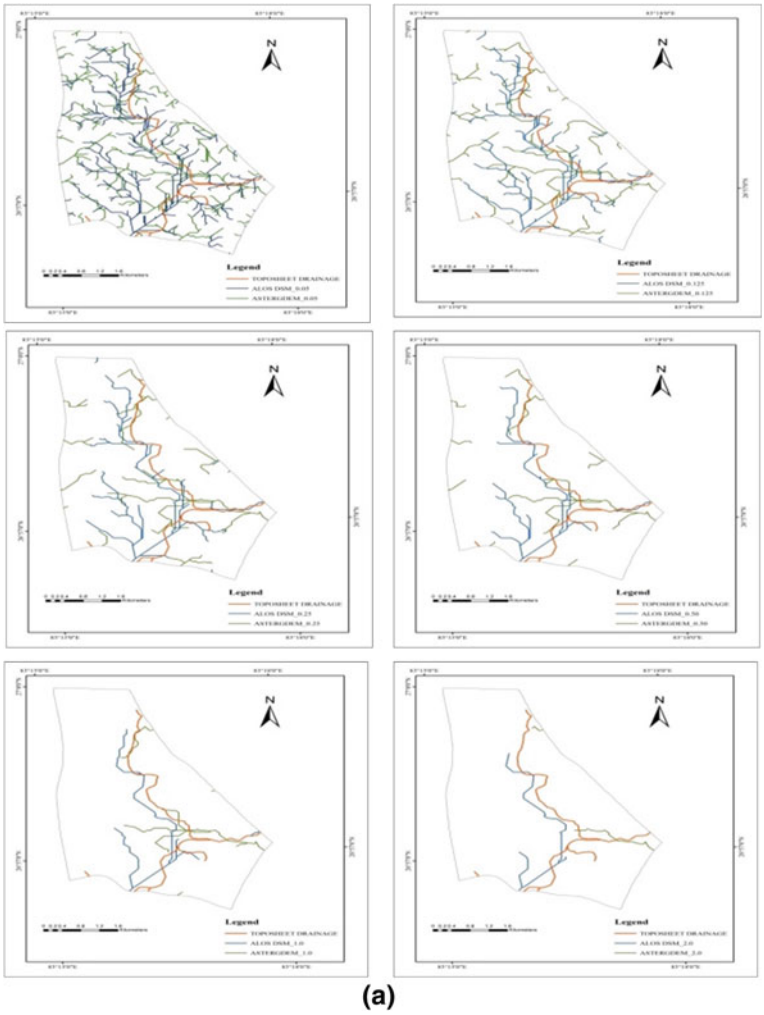


Fig. 10.5 **a** Comparison of Drainage networks extracted with toposheet (1:50,000 scale). **b** Chart representing different stream lengths extracted Source [4–6]

and surface model at lower threshold of 0.05 km^2 particularly ALOS-DSM shows exceptional performance which provides finer drainage network when compared to elevation models of same resolution.

Watershed delineation and hydromorphological applications have benefited from remote sensing capabilities linked with GIS. The study's goal was to do morphometric research on the Rapti sub-watershed in Gorakhpur, Uttar Pradesh, India. According to the findings of this study, morphometric indices show that ALOS-DSM works better at lower thresholds and that it is the best method for extracting finer drainage features at lower thresholds. In comparison to ASTER-GDEM, the current study shows that

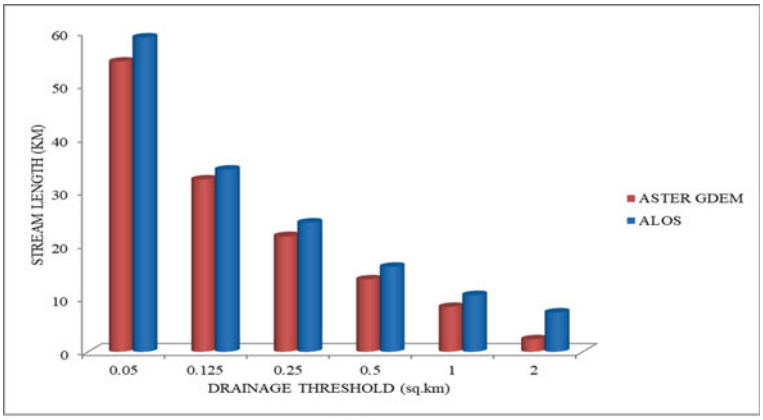


Fig. 10.5 (continued)

greater stream order was more consistent in ALOS-DSM with all thresholds used. The findings will aid policymakers in better planning and management of the research field. The findings may be extended to a number of purvanchal areas in Uttar Pradesh that have comparable morphological traits.

Conflict of interest None, Authors Express No Conflict of Interest.

References

1. Ahmed SA, Chandrashekarappa KN, Raj SK, Nischitha V, Kavitha G (2010) Evaluation of morphometric parameters derived from ASTER and SRTM DEM—a study on Bandihole sub watershed basin in Karnataka. J Indian Soc Remote Sens 38(2):227–238
2. Akram F, Rasul MG, Khan MMK, Amir MSII (2012) Automatic delineation of drainage networks and catchments using DEM data and GIS capabilities: a case study. 18th Australasian Fluid Me- chanics Conference Launceston, Australia, p 4
3. Arulbalaji P, Padmalal D (2020) Sub-watershed prioritization based on drainage morphometric analysis: a case study of cauvery river basin in South India. J Geol Soc India 95:25–35
4. Banerjee K, Santhosh Kumar MB, Tilak LN (2021) Delineation of potential groundwater zones using Analytical hierarchy process (AHP) for Gautham Buddh Nagar District, Uttar Pradesh, India, Materials Today: Proceedings, Volume 44, Part 6, Pages 4976–4983, ISSN 2214–7853, https://doi.org/10.1016/j.matpr.2020.12.917.
5. Banerjee K, Santhosh Kumar MB, Tilak LN, Vashistha S (2021). Analysis of Groundwater Quality Using GISBased Water Quality Index in Noida, Gautam Buddh Nagar, Uttar Pradesh (UP), India. In: Choudhary A, Agrawal AP, Logeswaran R, Unhelkar B (eds) Applications of Artificial Intelligence and Machine Learning. Lecture Notes in Electrical Engineering, vol 778. Springer, Singapore. https://doi.org/10.1007/978-981-16-3067-5_14
6. Banerjee K, Bali V, Nawaz N, Bali S, Mathur S, Mishra RK, Rani S (2022) A Machine-Learning Approach for Prediction of Water Contamination Using Latitude, Longitude, and Elevation. Water, 14(5):728

7. Biswas S, Sudhakar S, Desai VR (1999) Prioritization of sub watersheds based on Morphometric analysis and drainage basin: a remote sensing and GIS approach. 1999 J Indian Soc Remote Sens 27(3):155–166
8. Chandra V, Gidde MR (2019) Morphometric analysis of Baltira Watershed using QGIS Platform. Int J Recent Technol Eng (IJRTE) ISSN: 2277–3878, vol 8, no 2
9. Chopra R, Raman DD, Sharma PK (2005) Morphometric analysis of sub-watersheds in Gurdaspur district, Punjab using remote sensing and GIS techniques. J Indian Soc Remote Sens 33(4):531–539
10. Dikpal RL, Renuka Prasad TJ, Satish K (2017) Evaluation of morphometric parameters derived from Cartosat-1 DEM using remote sensing and GIS techniques for Budigere Amanikere watershed, Dakshina Pinakini Basin, Karnataka. India Appl Water Sci 7:4399–4414
11. Donker N (1992) Automatic extraction of catchment hydrologic properties from digital elevation model. ITC J 3:257–265
12. Forkuor G, Maathuis B (2012) Comparison of SRTM and ASTER derived digital elevation models over two regions in ghana-implications for hydrological and environmental modeling. Studies on environmental and applied geomorphology. Dr. Tommaso Piacentini (ed.), ISBN: 978–953–51–0361- 5
13. Gardener TW, Sasowsky KC, Day RL (1991) Automatic extraction of geomorphic properties from digital elevation data. Z GeomorpholSuppl 80:57–68
14. Gayen S, Bhunia GS, Shi PK (2010) Morphometric analysis of Kangshabati-Darkeswar Interfluves area in West Bengal, India using ASTER DEM and GIS techniques. Geol Geosci 2(4):1–10
15. Gopinath G, Swetha TV, Ashitha MK (2013) Automated extraction of watershed boundary and drainage network from SRTM and comparison with Survey of India toposheet Saudi Society for Geosciences. https://doi.org/10.1007/s12517-013-0919-0
16. Gottschalk LC (1964) Reservoir sedimentation in handbook of applied hydrology. McGraw Hill Book Company, New York (Section10.7–10.1)
17. Horton RE (1932) Drainage basin characteristics. Trans Am Geophys Union 13:350–361
18. Horton RE (1945) Erosional Development of streams and their drainage basins, hydrological approach to quantitative morphology. Bull Geophys Soc Am 56:275–370
19. Kattimani JM, Prasad TJR (2016) Cartosat Digital Elevation Model (DEM) to drainage extraction techniques of Vrushabhavati basin of Karnataka, India using remote sensing and GIS techniques. Int J Adv Res 4(7):2008–2013
20. Kienzle S (2004) The effect of DEM raster resolution on first order, second order and compound terrain derivatives. Trans GIS 8(1):83–111
21. Kumar R, Kumar S, Lohani AK, Nema RK, Singh RD (2000) Evaluation of geomorphological characteristics of a catchment using GIS. GIS India 9(3):13–17
22. Lin WT, Chou WC, Lin CY, Huang PH, Tsai JS Automated suitable drainage network extraction from digital elevation models in Taiwan's upstream watersheds. https://doi.org/10.1002/hyp.591
23. Luo Y, Su B, Yuan J, Li H, Zhang Q (2011) GIS techniques for watershed delineation of SWAT model in plain polders. Procedia Environ Sci Part C 10:2050–2057
24. Maathuis BHP (2006) Digital elevation model based hydro-processing. Geocarto Int 21(1):21–26
25. Mahadevaswamy G, Nagaraju D, Siddalingamurthy S, Lakshmamma MSL, Nagesh PC, Rao K (2011) Morphometric analysis of Nanjangud taluk, Mysore District, Karnataka, India, using GIS techniques. Int J Geomat Geosci 1:179–187
26. Mark DM (1984) Automatic detection of drainage networks from digital elevation models. Cartographia 21:168–178
27. Mokarram M, Hojati M (2015) Comparis of digital elevation model (DEM) and aerial photographs for drainage. https://doi.org/10.1007/s40808-015-0050-1
28. Nikolakopoulos KG, Kamaratakis EK (2006) SRTM vs ASTER elevation products. Comparison for two regions in Crete, Greece. Int J Remote Sens 27(21):4819–4838

29. Obi Reddy GB, Kumar N, Sahu N, Singh SK (2018) Evaluation of automatic drainage extraction thresholds using ASTER GDEM and Cartosat-1 DEM: a case study from basaltic terrain of Central India 2018. G.P.O. Reddy et al. / Egypt J Remote Sens Space Sci 21:95–104
30. Ozdemir H, Bird D (2009) Evaluation of morphometric parameters of drainage networks derived from topographic maps and DEM in point of floods. Environ Geol 56:1405–1415
31. O'Callaghan JF, Mark DM (1984) The extraction of drainage networks from digital elevation data. Comput Vision Graph Image Process 28:323–344
32. Pham HT, Marshall L, Johnson F, Sharma A (2018) A method for com- bining SRTM DEM and ASTER GDEM2 to improve topography estimation in regions without reference data. Remote Sens Environ 210:229–241
33. Rai PK, Mohan K, Mishra S, Ahmad A, Mishra VN (2017) A GIS-based approach in drainage morphometric analysis of Kanhar River Basin. India Appl Water Sci 7:217–232.https://doi.org/10.1007/s13201-014-0238-y
34. Rana1 VK, Suryanarayana TMV (2019) Visual and statistical comparison of ASTER, SRTM, and Cartosat digital elevation models for watershed. J Geovisualization Spatial Anal 3:12
35. Reddy OGP, Maji AK, Gajbhiye SK (2004) Drainage morphometry and its influence on landform characteristics in a basaltic terrain, Central India—a remote sensing and GIS approach. Int J Appl Earth Obs Geoinformatics 6:1–16
36. Reddy GPO, Maji AK, Gajbhiye KS (2002) GIS for morhophometric analysis of river basins. GIS India 11(9):9–14
37. Reddy GPO, Maji AK, Chary GR, Srinivas CV, Tiwary P, Gajbhiye KS (2004) A GIS and remote sensing applications in prioritization of river sub basins using morphometric and USLE parameters—a case study. Asian J Geoinform 4(4):35–49
38. Schumm SA (1954) The relation of drainage basin relief to sediment loss. Int Assoc Sci Hydrol 36:216–219
39. Schumm SA (1956) Evolution of drainage systems & slopes in Badlands at Perth, New Jersey. Bullet Geol Soc Am 67:597–646
40. Schumm SA (1963) Sinuosity of Alluvial Rivers on the great plains. Bullet Geol Soc Am 74:1089–1100
41. Sethupathi AS, Lakshmi Narasimhan C, Vasanthamohan V, Mohan SP (2011) Prioritization of mini watersheds based on morphometric analysis using remote sensing and GIS in a drought prone Bargur Mathur sub watersheds, Ponnaiyar River basin, India. Int J Geomat Geosci 2(2):403–414
42. Singh P, Gupta A, Singh M (2014) Hydrological inferences from watershed analysis for water resource management using remote sensing and GIS techniques. Egypt J Remote Sens Space Sci 17:111–121
43. Singh S, Singh MB (1997) Morphometric analysis of Kanhar river basin. Natl Geogr J India 43(1):31–43
44. Strahler AN (1952) Hypsometric analysis of erosional topography. Bullet Geol Soc Am 63:1117–42
45. Strahler AN (1956) Quantitative slope analysis. Bullet Geol Soc Am 67:571–596
46. Strahler AN (1996) Quantitative geomorphology of drainage basin and channel network. Handbook of applied hydrology, pp 39–76
47. Strahler A (1964) Quantitative Geomorphology of Drainage Basins and Channel Networks. In: Chow, V., Ed., Handbook of Applied Hydrology, McGraw Hill, New York, 439–476
48. Tarboton DG, Bras RL, Rodriguez-Iturbe I (1989) The analysis of river basins and channel networks using digital terrain data", TR 326. Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Boston
49. Venkatachalam P, Krishna Mohan B, Kotwal A, Mishra V, Muthyramakrishnan V, Pandya M (2001) Automatic delineation of watersheds for hydrological applications. Paper presented at the 22nd Asian Conference on Remote Sensing", 5–9 November 2001, Singapore
50. Vittala SS, Govindiah S, Honne Gowda H (2004) Morphometric analysis of sub-watersheds in the pawagada area of Tumkur district, South India, using remote sensing and GIS techniques. J Indian Soc Remote Sens 32(4):351–362

Chapter 11
Internet of Medical Things (IoMT) & Secured Using Steganography for Development of Smart Society 5.0

Sachin Dhawan, Rashmi Gupta, Arun Kumar Rana, and Sharad Sharma

Abstract Nowadays, Internet of Things (IoT) is important part in the present world, which fills in as a feature of our standard life exercises. Despite the fact that IoT gives benefits in many ways. Like through IoT the communication between the doctors and patients were possible only through texts, visits, and telecommunications. However, the advancement in technology and the **evolution of IoMT in health care** has significantly changed the scenario. In this chapter we will see the role of IOT in health sector in every aspect like COVID-19 pandemic. Then we will see IOT healthcare networks along with its architecture. Along these lines, the requirement of a secure climate is indispensable to get the communicating information through IoT. Therefore, in this chapter, a secure plan is also recommended on utilizing steganography as an elective security system related to a healthcare sector to get the communicated information from IP camera as the IoT gadget to different gadgets.

Keywords Internet of Medical Things (IoMT) · Security · Image steganography

11.1 Introduction

Our health is one of the most essential part of our life. Unfortunately, the rise in various types of illness and lack of awareness regarding fitness is placing significant strain on modern healthcare systems [1]. But fortunately, today's modern technology is also growing its roots in health sector [2]. Advanced technology like (IoT, ML, AI) [3], has been widely identified and accepted as the potential solutions to lower the pressure on healthcare system [4]. Though these technologies are very much advanced, but still there is a lot to achieve. Here, we will see how these technologies

S. Dhawan (✉) · A. K. Rana
Panipat Institute of Engineering and Technology, Samalkha, India
e-mail: ersachind@gmail.com

R. Gupta
Netaji Subhas University of Technology, East Campus, Geeta Colony, Delhi, India

S. Sharma
Maharishi Markandeshwar University, Mullana, India

V. Bali et al. (eds.), *Decision Analytics for Sustainable Development in Smart Society 5.0*, Asset Analytics, https://doi.org/10.1007/978-981-19-1689-2_11

are currently playing there part in health sector how health sector is getting benefited from these technologies. A major problem in which all patients, especially those living in remote areas, were identified was the lack of access to medical care and treatment in severe cases [5]. This has had a devastating effect on people's minds about hospitals and their services [6, 7]. Nowadays, with the introduction of new technologies using IoT devices for healthcare monitoring systems, these problems are organized in a certain way but there is still much to be done. IOT has potential in not only keeping the patients safe and healthy but also to decrease the risks involving in critical surgeries.

Machine Learning (ML) already plays an important role in multiple situations in health care. It also helps in research of thousands of various data prongs and advise outcomes, gives time to time risk scores, exact resource allocation. It gives different methods and tools which will help in solving multiple trouble in healthcare sector. It is also used to manage or reduce complications and to help the patient care during the active phase of his treatment moreover to determine the ideal treatment for a disease. Smart urban areas address a calculated model of metropolitan improvement that depends on human, aggregate, and innovative capital exploitation [8]. The idea of shrewd urban communities [9] can be reached out to keen society; indeed, Society 5.0, alongside AI, can utilize normalized cycles to assess human requirements. Subsequently, a carefully empowered and information-based society should run after friendly, natural, and financial maintainability. Individual and community capital is at the core of advancements in shrewd urban areas and savvy people and requires imaginative strategies for prescient and versatile cycles. The point is to plan an information with computerized foundation that can cooperate and can empower dynamic constant associations between different brilliant city subsystems [10]. As of now, the writing contains many portrayals of what a shrewd or smart city is, in spite of the fact that there is no general understanding [11–14]. At any rate, it very well may be expressed that brilliant urban areas are human-driven social orders, in which AI chooses measures with the objective of getting the best conditions for quality of life utilizing savvy plants furnished with keen gadgets (sensors/actuators), programmable rationale regulators (PLCs), measure the executives and assembling execution frameworks, undertaking asset arranging SW, and HCPS. The thought is that a request can be made anyplace in the organization, and the assembling measures can be distantly controlled [15] and revamped with plug-and-produce innovation, in view of savvy framework and ontologies. This will permit the combination, trade, or substitution of creation hardware without particular HR support for framework reconfiguration. Indeed, fitting-and-produce innovation permits the shrewd reconfiguration and connection of keen gadgets associated with the PLC, which empowers brilliant participation with different gadgets [16].

To handle the issue of protection in the IoT association, we recommend a medical security contrive that incorporates relationship with the Web.

11.2 Literature Survey

There are various important elements of smart city. These are: Traffic, Learning, Smart homes, Transport, Government, nearby living areas.

The term "Smart urban communities" by and large alludes to the utilization of improved residential society with the association of government for the personal satisfaction [17–19].

A city can be depicted as smart where everything like natural, household equipments, public equipments, banking sector, health sector, and traffic sector are adjusted and connected more efficiently [20]. IoT is the invention due to which smart urban communities are planned and also empowered sensor innovation to make connectivity of everything in the world [21, 22].

Splendid city organizers utilize current developments, for instance, adaptable appropriated processing, electronic things, associations, sensors, and AI headways to engage the different portions of sharp metropolitan networks to work together and partner with the association plan [23]. The common complexity and new strategies for inhabitant collaboration required for the change to existing establishment, includes the immense political, managerial, and particular challenges for governments and regional trained professionals. One of the imperative challenges in the improvement of adroit metropolitan networks is the planning and the leading body of data. This relates to data adequately introduce inside city informational indexes yet moreover the interfacing of data with new structures and sensors present inside the splendid city, that influence security and insurance [24]. The risks are coming from data security, data assurance and includes the criticality of keeping an eye on these issues early inside the arrangement and improvement period of sharp metropolitan regions [25]. The composing has developed different assessments that have presented reviews of the adroit city composing [26–28]. In any case, experts seem to have blocked to bring to the table a huge assessment of the colossal risks to data security and the trademark complexities including assurance inside sharp metropolitan regions. This investigation means to defeat this issue in the composition by driving an extensive examination of the various issues and key complexities relating to assurance, security, and peril issues inside smart metropolitan networks (Table 11.1).

The current examination intends to address the accompanying inquiries: What is the present status of information identifying with security, protection, and hazard inside smart soceity? and What are the brilliant urban communities' difficulties in territories identifying with protection, security, and hazard from various partner viewpoints?

11.3 Steganography in Medical Image

In development of smart society, secured steganography plays the major role. In today's era, security of confidential data is very important and this task can be achieved with the help of steganography. If we observe the characteristics of smart

Table 11.1 Survey of steganography techniques used in medical images

Sr. no	Technique	References	Work
1	LSB substitution algorithm	Chang et al. [29]	Highly secured
2	Data hiding by fuzzy	Wang et al. [30]	Less distortion
3	Reversible data hiding algorithm	Chang et al. [31]	Improved image quality and less image distortion
4	Image steganography using fuzzy	Sajasi and Moghadam [32]	Improved stego image quality and less image distortion
5	Medical picture verification method using steganography algorithm	Martiri et al. [33]	High security and quality
6	Novel content-based information hiding scheme	Kong et al. [34]	Extremely secure transmission with more secrecy as well as privacy
7	Region-based steganography technique	Hamid et al. [35]	Extreme visual quality and less distortion
8	Steganography method based on sign embedding and fuzzy classification	Li et al. [36]	Increased embedding capacity and Extreme stegoimage
9	Gray block embedding method	Janakiraman et al. [37]	Extreme payload capacity and secure due to complex structure
10	LSB Technique for data hiding	Kavitha et al. [38]	Less error
11	Steganography scheme with the combination of the fuzzy logic	Goodarzi et al. [39]	Improved payload capacity
12	Data hiding technique using shared method	Pandey et al. [40]	High robustness

society, medical infrastructure will be on top priority. Without enhanced medical facility, we can not achieve our target of smart society5.0. In COVID-19's era, doctor is also communicating with patient through online mode. Therefore, steganography will also play major role for communicating confidential data from patient to doctor. An examination by [41] used ECG sign to have a confidential record of patient; the investigation chose the proper situation for implanting the secret clinical report. LPF channel was utilized in encoding and unraveling the ECG signal. Another investigation by [42] presented a clinical picture recovery framework to get highlights that will be utilized for steganography measures after implant [42]. An alteration of the Blowfish calculation was introduced by [43] for ensuring the security of clinical pictures and text record; they recommended the utilization of force work for privileged data encryption. They additionally presented the control of the UACI with (NPCR) to improve a security of the strategy versus assaults. According to [44], coefficients of DCT are useful in accomplishing amazing results when parceling the cover pictures

into 8 × 8 pixel [45, 46]; with no covering among them; they additionally presented that change space is more useful with watermark strategies. In this examination, the significant point is to build up another strategy that will address the difficulties of the current strategies and improve the outcomes.

11.3.1 Smart Medical Technology

In smart society, medical health care will be on top priority. It incorporates smart clinical hardware and units that are presently being conveyed and utilized by paramedics to give prompt assistance to patients who are in pressing need of clinical consideration and help. One model is the utilization of clinical robots to perform such an assignment [47]. Clinical robots were initially acquainted with react to crises identified with patients experiencing heart failures [48], since these robots are the quickest to show up at the crisis scene. The robots would be coordinated to travel to explicit objections, which saves time and accordingly, saves lives since paramedics may wind up stranded in rush hour gridlock, and will be unable to react as fast depending on the situation. This empowers the dependence on keen clinical robots [49] to perform careful tasks inside a clinic setting. Virtual/augmented reality and Artificial Intelligence (AI)-based clinical innovations were likewise utilized for different clinical purposes. This incorporates virtual reality to perform different sensible activities, for example, reenacted preparing, crisis preparing, and Cardiopulmonary Resuscitation (CPR) preparing. Figure 11.1 shows the smart healthcare system. Simulated intelligence based clinical innovations are likewise being utilized to guarantee a higher precision rate.

11.4 Proposed Healthcare System Using Steganography and IoT

Deep learning algorithms are utilized for high-quality healthcare steganography and IoT-based smart medical system. As shown in Fig. 11.2, block diagram of smart medical system is shown which depicts secured medical care system using steganography.

Typically, the advanced picture incorporates unique picture components named pixels. In this work, a grayscale picture is utilized as a cover picture. In this way, a picture is addressed by a cluster that incorporates numerous bytes. The proposed framework fundamentally incorporates four segments: first is encryption, second phase is hiding stage, third phase is feature improvement, and the last is extraction stage. As displayed in Fig. 11.1, initially, the clinical information is considered as cover, and the information identified with patient is considered as a privileged information. Here, the size of the restricted information ought to be not exactly the

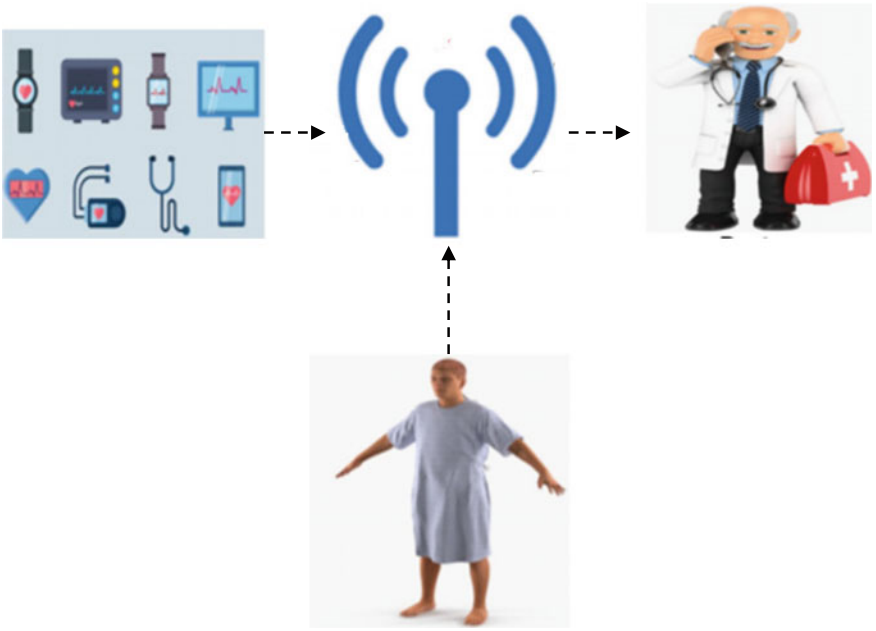


Fig. 11.1 System for smart healthcare

size of the cover information. In our proposed framework, for far places correspondence of specialist and patients, we are making our framework safer by utilizing IoT convention.

11.5 IOT in Health Care

- **Provides end-to-end connectivity and affordability**

IOT provides end-to-end connectivity which creates a strong link in between patients and doctors. This technology is pretty much affordable when seen from patient perspective [50]. IOT also provides connectivity between doctors and even different hospitals no matter how far they are. This ensures that the patients should get best possible treatment.

- **Improving the health of patients**

Since there is end-to-end connectivity, that means patients and doctors are in touch every time. This leads to the improvement in health of patients.

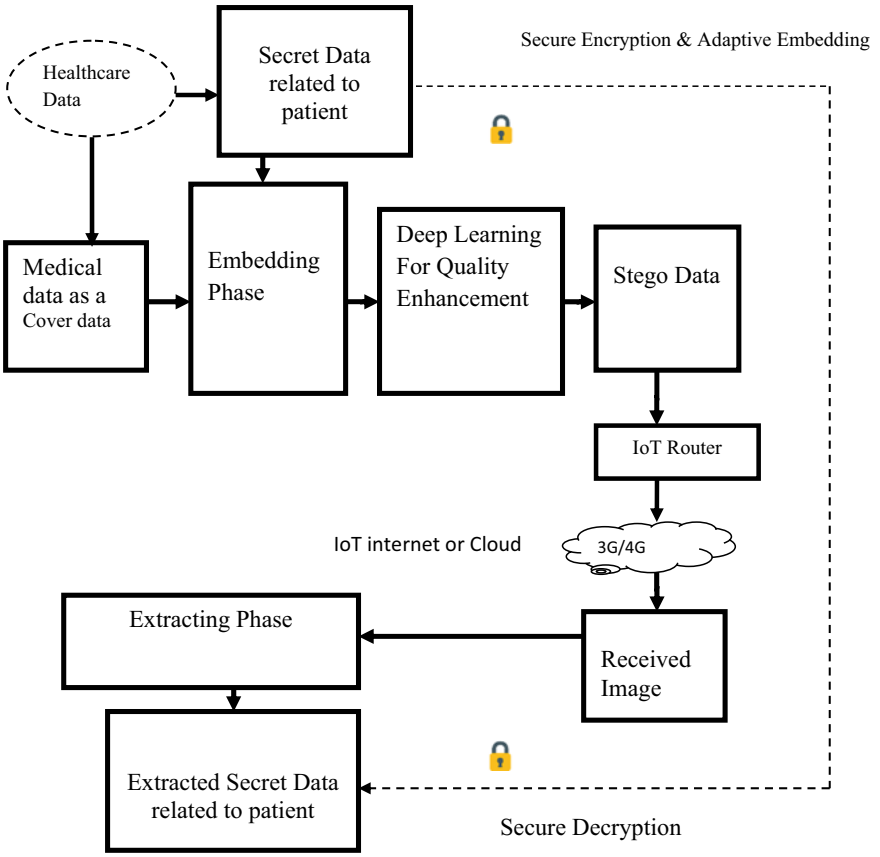


Fig. 11.2 Proposed smart medical healthcare system

- **Simultaneous monitoring and support**

Nowadays, we see various types of machine in a hospital. These machines ensure continuous monitoring on patient health and if something wrong is detected the same is conveyed to the required medical staff in no time and full support is provided.

- **Tracking and alerts**

Nowadays IOT is present in various forms [51]. One of them is in the form of wearable technology (Fig. 11.3). This includes gadgets like smart bands or smart watches. These gadgets provide continuous tracing of heart beat, blood pressure, and many more. If something wrong is detected, instant alert is sent to the owner.

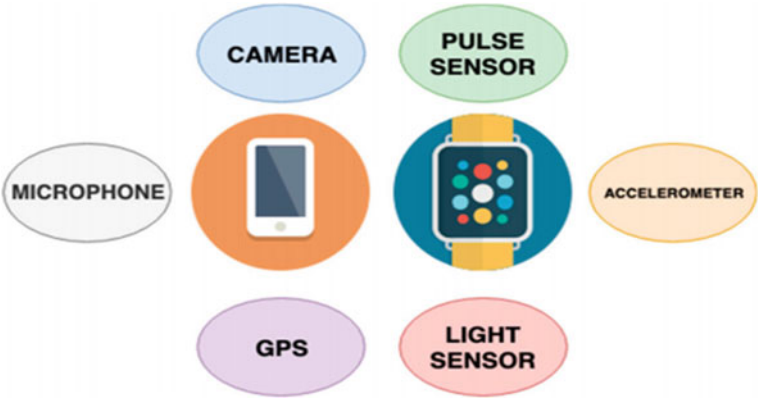


Fig. 11.3 Smart wearables

- **Research and data analysis**

Currently, IOT is playing a big role in research and data analysis. This is clearly evident from the fact that the global IOT in health to reach CAGR of 13.20% from 2020 to 2027 [52].

11.6 Challenges in IOT

- **Data security and privacy**

Quite possibly the main difficulties IoT has confronted is Data Security and Privacy (Fig. 11.4). In spite of the fact that IoT-empowered cell phones catch information progressively, the majority of them don't observe information arrangements and guidelines.

There is a critical vagueness regarding information proprietorship and information the executives [53].

- **Integration: Multiple protocol and devices**

The integration of many types of devices poses a barrier to IoT use in the healthcare sector. The reason for this obstacle is that device manufacturers have not yet reached agreement on compliance with standards and standards. This results in a situation where every manufacturer makes its own unique environment for IoT devices not working with devices and the use of competing manufacturers. In such a case, no sync protocol may be followed for data integration. These disparities reach delays in the process and reduce the extent of IoT intensity in health care [54].

Fig. 11.4 Data security cycle

- **Excessive data and accuracy**

IoT, ML, and AI provides accuracy even in case of excessive database. IoT helps to connect many things and makes all the task easy and provides high accuracy.

- **Existing software infrastructure is no longer operational**

IT infrastructure in many hospitals is no longer viable. They will not allow proper integration of IoT devices. Therefore, healthcare facilities will need to update their IT processes and use new, state-of-the-art software. They will also need to take advantage of virtualization (technologies such as SDN and NFV), as well as fast wireless networks and mobile phones such as Advanced LTE or 5G.

11.7 Artificial Intelligence in Health Care

- **Early warning and alerts**

PC-based knowledge can quickly take apart inconsistent result and other "admonitions" and, in this way, ready the patients and the human administrations specialists [55]. It helps with giving snappier dynamic, which is monetarily canny. It helps with developing another end and the chiefs structure for the COVID-19 cases, through important figurings. Man-made insight is valuable in the finish of the spoiled cases

with the help of clinical imaging progressions like Computed tomography (CT), Magnetic resonation imaging (MRI), etc., yield of human body parts.

- **Tracking and predictions**

PC-based insight can help take apart the level of pollution by this contamination recognizing the gatherings and "pain points" and can viably do the contact following individuals and moreover to screen them [56]. It can anticipate the future course of this disease and likely return.

- **Data dashboards**

Recreated insight is valuable for the future disease and sicknesses contravention, with the help of past guided data over data inescapable at different times [57]. It recognizes traits, causes, and clarifications behind the spread of illness. In future, this will wind up being a critical development to fight against various epidemics and pandemics. It can give a preventive measure and fight against various ailments. In future, AI will expect a key occupation in giving progressively perceptive and preventive social protection.

- **Reducing the workload oh health workers**

During COVID-19 pandemic, AI and ML is used to lessen the excess main job of restorative administrations workers [58]. This can be helpful in finding and giving treatment at a starting stage using modernized methodologies.

- **Treatment and cures**

With the help of progressing data assessment, AI can give revived information which is valuable in the expectation of this ailment. It will in general be used to expect the conceivable objections of pollution, the union of the contamination, necessity for beds and social protection specialists during this crisis.

11.8 Challenges in Healthcare System of AI

Artificial intelligence (AI) holds incredible guarantee as far as advantageous, precise, and successful preventive and corrective mediations. Simultaneously, there is likewise mindful of the possible dangers and perils that might be brought about by the unregulated improvement of AI [59].

SDG 3 structure a criticalness to advance wellbeing and decreasing wellbeing disparities, while nations measure how they are getting along offers another layer of inspiration. However, ready to face more challenges to accomplish objectives that solid qualities joined to making an auxiliary quandary in clinical morals and man-made consciousness.

At the point when individuals are needing medical services, worries about a dubious, possibly risky, long haul effect of getting help from AI-based frameworks

may not be their main concern. At the political level, consider the anticipation of an exceptionally unsure effect will come to be disregarded for a viable arrangement, which assists with settling the difficulties of public wellbeing. Moreover, capacity to bear unintended results of the capability of applying man-made reasoning will probably rely upon how the present status of undertakings unfortunate considered [60].

11.9 Machine Learning in Health Care

- **Identifying diseases and diagnosis**

One of the main ML applications in medical care is to identify diseases and diagnosis those diseases with the help of various ML techniques shown in Fig. 11.5 [61].

- **Drug discovery and manufacturing**

One of the fundamental clinical employments of AI lies in starting stage drug revelation measure. This in like manner joins R&D developments, for instance, forefront sequencing and precision drug which can help in finding elective ways for treatment of multifactorial diseases. As of now, the AI techniques incorporate independent acknowledging which can perceive plans in data without giving any assumptions [62].

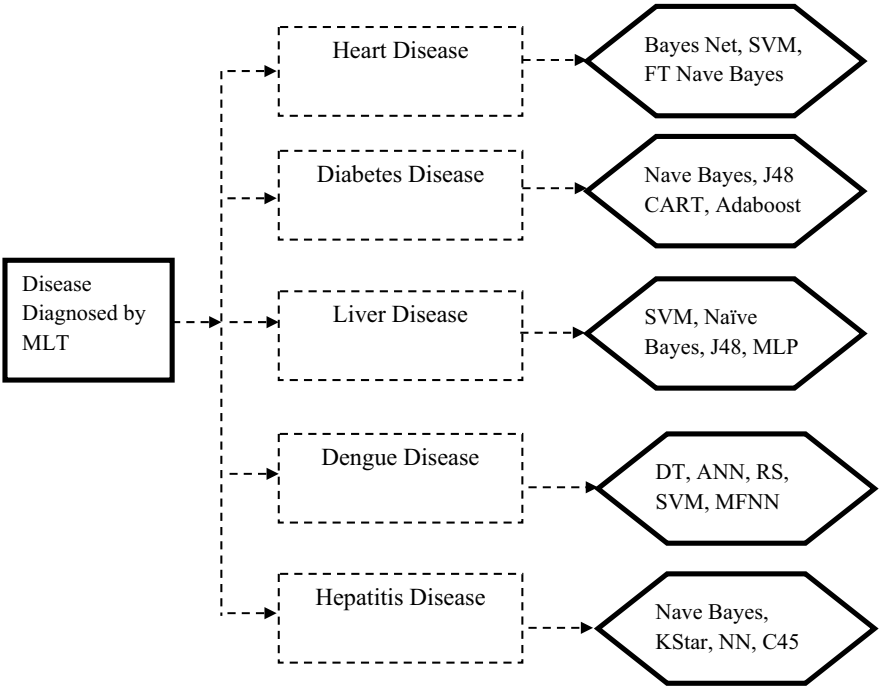


Fig. 11.5 Diseases diagnosed By MLT

- **Medical imaging diagnosis**

Man-made intelligence and significant learning are both liable for the headway development called Computer Vision. This has found affirmation in the inner eye movement made by Microsoft which manages picture logical gadgets for picture investigation [63].

- **Personalized medicine**

Tweaked meds can notwithstanding the way that more be fruitful by coordinating with singular prosperity with judicious assessment is similarly prepared are for extra investigation and better disease appraisal [64]. As of now, specialists are confined to investigating a specific course of action of examinations or check the threat to the patient reliant on his characteristic history and available genetic data [65].

- **Machine learning-based behavioral modification**

Direct change is a huge piece of preventive prescription, and since the time the augmentation of AI in clinical benefits, incalculable new organizations are jumping up in the fields of threat aversion and ID, open minded treatment, etc.

11.10 Challenges in Healthcare System of ML

In handling medical care errands, there are factors that ought to be considered cautiously in the plan and assessment of AI ventures: causality, missingness, and result definition [66]. These contemplations are significant across both displaying structures (e.g., directed versus solo) and learning targets (e.g., grouping versus relapse).

- **Tailored data acquisition**

At the point when we began Spry Health, there was no extraordinary information base of physiological signs we could use to make our models [67]. All together for any preparation to work, we needed to have both an adequate measure of information and adequate data in our sign. We thought of a particular sensor setup and held our own clinical preliminaries with several patients to accumulate the fundamental edge of information for our models.

- **Simple is better**

It may sound self-evident, however, when information is pre-prepared appropriately, straightforward models yield superb outcomes [68]. Basic models turned out better for us when we made AI calculations to foresee nonstop crucial signs just as the state vector of the patient's physiology.

- **Utilize expertise, not only data**

Another motivation behind why we evade profound learning is on the grounds that a specialist can perform far superior to any profound learning model. The presentation of any learning model, human or machine, relies upon the measure of information used to prepare it. Freely accessible datasets for ML are divided and generally little.

- **Need for traceable conclusions**

At the point when somebody's life is on the line, clinicians need to see how information is controlled to get to a specific outcome. Clinicians we talked with were not happy with an answer that solitary offers a last response—they needed to know the hidden elements that prompted a particular end.

- **Health data requires high level of customization**

Any execution of ML estimations as a general rule needs most likely some level of customization. Because of prosperity data, the level of required customization is amazingly high for three reasons: the natural capriciousness of the human body, the transparency and significance of data sources, and blend into the current clinical consideration structure. Associations like Spry Health are working determinedly dealing with these issues: as of now, we are seeing the promising finish to current conditions. As more data opens up, the hindrances to using ML and AI will evaporate. Right now, we really need to continue circumspectly [69].

11.11 Conclusion

The reputation of IoT, AI, and ML in a variety of areas such as automotive, education, industrialized, etc., is growing rapidly. Now comes the time when all these leading technologies also transform the medical industry by solving many issues. With the advent of associated equipments, better predictions and recommendations healthcare providers will not see any missed appointment, a better understanding of improved conditions and outcomes, and everything in less time. Once fully adopted, these technologies can considerably assist to get better services and experiences, reduce costs, and promote better medical services, but meanwhile, we should not neglect the problems that we encounter in the practical world and should focus on finding out the solutions to eradicate all these problems and make the best use of these modern technologies.

References

1. Nikfarjam A, Sarker A, O'connor K, Ginn R, Gonzalez G (2015) Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. J Am Med Inform Assoc, 671–681
2. Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X (2020) Artificial intelligence and machine learning to fight COVID-19
3. Manrai AK, Bhatia G, Strymish J, Kohane IS, Jain SH (2014) Medicine's uncomfortable relationship with math: calculating positive predictive value. JAMA Int Med, 991–993
4. Akhtar M, Kraemer M, Gardner L (2019) A dynamic neural network model for predicting risk of Zika in real time. BMC Med 17(171)
5. Babar S, Stango A, Prasad N, Sen J, Prasad R (2011) Proposed embedded security framework for Internet of Things (IoT). Proposed embedded security framework for Internet of Things (IoT). In Proceedings of the 2011 2nd international conference on wireless communication, vehicular technology, information theory and aerospace & electronic systems technology (Wireless VITAE), pp 1–5
6. Bruijne D, Marleen (2016) Machine learning approaches in medical image analysis: from detection to diagnosis, pp 94–97
7. Bullock JL, Luengo-Oroz M (2020) Mapping the landscape of artificial intelligence applications against COVID-19. ArXiv
8. Angelidou M (2015) Smart cities: a conjuncture of four forces. Cities 47:95–106
9. Yigitcanlar T, Kamruzzaman M, Foth M, Sabatini-Marques J, Da Costa E, Ioppolo G (2019) Can cities become smart without being sustainable? A systematic review of the literature. Sustain Cities Soc 45:348–365
10. Mehmood R, Bhaduri B, Katib I, Chlamtac I (eds) (2018) Smart societies, infrastructure, technologies and applications. Springer, New York
11. Hollands RG (2008) Will the real smart city please stand up? Intelligent, progressive or entrepreneurial? City 12(3):303–320
12. Komninos N (2011) Intelligent cities: variable geometries of spatial intelligence. Intell Build Int 3(3):172–188
13. Nam T, Pardo TA (2011) Petri nets for systems and synthetic biology. Lect Notes Comput Sci 5016:215–264
14. Wolfram M (2012) Deconstructing smart cities: an intertextual reading of concepts and practices for integrated urban and ICT development. In: Proceedings of the 2012 International conference on urban planning and regional development in the information society, May 14–16, Schwechat, Austria
15. Zuehlke D (2010) SmartFactory—towards a factory-of-things. Annu Rev Contr 34(1):129–138
16. Foresti R (2013) Plug and produce 2013 [Internet]. [cited 2019 Oct 2]. https://it.wikipedia.org/wiki/Plug_and_Produce
17. Chourabi H, Nam T, Walker S, Gil-Garcia JR, Mellouli S, Nahon K, Scholl HJ (2012) Understanding smart cities: an integrative framework. In 2012 45th Hawaii international conference on system sciences (pp 2289–2297). IEEE
18. Yahia NB, Eljaoued W, Saoud NBB, Colomo-Palacios R (2019) Towards sustainable collaborative networks for smart cities co-governance. Int J Inf Manage. https://doi.org/10.1016/j.ijinfomgt.2019.11.005
19. Yu W, Xu C (2018) Developing smart cities in china: an empirical analysis. Int J Public Administ Digital Age (IJPADA) 5(3):76–91
20. Komninos N (2013) What makes cities intelligent?. In Deakin M (ed). Smart cities: governing, modelling and analysing the transition. Taylor and Francis, p 77. ISBN 978–1135124144
21. Albino V, Berardi U, Dangelico RM (2015) Smart cities: definitions, dimensions, performance, and initiatives. J Urban Technol 22(1):3–21
22. Praharaj S, Han JH, Hawken S (2018) Urban innovation through policy integration: critical perspectives from 100 smart cities mission in India. City Cult Soc 12:35–43

23. Ismagilova E, Hughes L, Dwivedi YK, Raman KR (2019) Smart cities: advances in research—An information systems perspective. Int J Inf Manage 47:88–100
24. Van Zoonen L (2016) Privacy concerns in smart cities. Gov Inf Q 33(3):472–480. https://doi.org/10.1016/j.giq.2016.06.004
25. Chatterjee S, Kar AK, Gupta MP (2017) Critical success factors to establish 5G network in smart cities: inputs for security and privacy. J Glob Inf Manag 25(2):15–37. https://doi.org/10.4018/JGIM.2017040102
26. Albino V, Berardi U, Dangelico RM (2015) Smart cities: Definitions, dimensions, performance, and initiatives. J Urban Technol 22(1):3–21
27. Antonopoulos K, Petropoulos C, Antonopoulos CP, Voros NS (2017) Security data management process and its impact on smart cities' wireless sensor networks. Paper presented at the South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference, SEEDA-CECNSM 2017, https://doi.org/10.23919/SEEDA-CECNSM.2017.8088238.
28. Bibri SE, Krogstie J (2017) Smart sustainable cities of the future: an extensive interdisciplinary literature review. Sustain Cities Soc 31:183–212
29. Chang CC, Lin MH, Hu YC (2002) A fast and secure image hiding scheme based on LSB substitution. Intern J Pattern Recognit Artif Intell 16(04):399–416
30. Wang SS, Fan SJ, Li CS (2012) A new reversible data hiding based on fuzzy predictor. In: 2012 International conference on fuzzy theory and its applications (iFUZZY2012). IEEE, pp 258–262
31. Chung DJ, Jin HL, Choe YS (2011) Reversible data hiding algorithm for high quality stego images. In: TENCON 2011–2011 IEEE region 10 conference. IEEE, pp 251–254
32. Sajasi S, Moghadam AME (2013) A high quality image steganography scheme based on fuzzy inference system. In: 13th Iranian conference on fuzzy systems. IEEE, Qazvin
33. Martiri E, Baxhaku A, Barolli E (2011) Steganographic algorithm injection in image information systems used in healthcare organizations. In: Intelligent networking and collaborative systems (INCoS), 2011 third international conference on. IEEE, pp 408–411
34. Kong J, Jia H, Li X, Qi Z (2009) A novel content-based information hiding scheme. In: International conference on computer engineering and technology, 2009. ICCET'09, vol 1. IEEE, pp 436–440
35. Hamid N, Yahya A, Ahmad RB, Al-Qershi O (2012) Characteristic region based image steganography using speededup Robust features technique. In: Future communication networks (ICFCN), 2012 international conference on. IEEE, pp 141–146
36. Li Q, Yu C, Chu D (2006) A robust image hiding method based on sign embedding and fuzzy classification. In: 2006 6th World congress on intelligent control and automation, vol 2. IEEE, pp 10050–10053
37. Janakiraman S, Suriya N, Nithiya V, Radhakrishnan B, Ramanathan J, Amirtharajan R (2012) Reflective code for gray block embedding. In: Pattern recognition, informatics and medical engineering (PRIME), 2012 international conference on. IEEE, pp 215–220
38. Kavitha KK, Koshti A, Dunghav P (2012) Steganography using least significant bit algorithm. Int J Eng Res Appl 2(3):338–341
39. Goodarzi MH, Zaeim A, Shahabi AS (2011) Convergence between fuzzy logic and steganography for high payload data embedding and more security. In: Telecommunication systems, services and applications (TSSA), 2011 6th international conference on. IEEE, pp 130–138
40. Pandey V, Singh A, Shrivastava M (2012) Medical image protection using cryptography, data hiding and steganography. Int J Emerg Technol Adv Eng 2(1):106–109
41. Pandey A et al (2019) Bernoulli's chaotic map-based 2D ECG image steganography a medical data security approach. Med Data Secur Bioeng 2019:208–241
42. Eze P et al (2019) Integrity verification in medical image retrieval systems using spread spectrum steganography. In: Proceedings of the 2019 on international conference on multimedia retrieval. ACM
43. Akande NO et al (2019) Electronic medical information encryption using modified blowfish algorithm. In: International conference on computational science and its applications. Springer, Cham

44. Parah SA et al (2019) Realization of an adaptive data hiding system for electronic patient record, embedding in medical images. Security in smart cities: models, applications, and challenges. Springer, Cham, pp 47–70
45. Dhawan S, Gupta R (2020) Analysis of various data security techniques of steganography: A survey. Inf Secur J A Global Perspect, 1–25
46. Dhawan S, Gupta R (2019) Comparative analysis of domains of technical steganographic techniques. In: 6th International conference on computing for sustainable global
47. Haidari LA, Brown ST, Ferguson M, Bancroft E, Spiker M, Wilcox A, Ambikapathi R, Sampath V, Connor DL, Lee BY (2016) The economic and operational value of using drones to transport vaccines. Vaccine 34(34):4062–4067
48. Pulver A, Wei R, Mann C (2016) Locating AED enabled medical drones to enhance cardiac arrest response times. Prehospital Emerg Care 20(3):378–389
49. Ming Li, Russell H. Taylor (2004) Spatial motion constraints in medical robot using virtual fixtures generated by anatomy. In: Robotics and automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on, vol 2, IEEE, pp 1270–1275
50. Doukas C, Maglogiannis I (2012) Bringing IoT and cloud computing towards pervasive healthcare. Proc Int Conf Innov Mobile Internet Services Ubiquitous Comput (IMIS), pp 922–926
51. Ghassemi M (2020) A review of challenges and opportunities in machine learning for health. AMIA Summit Trans Sci Proceed 191
52. https://22century.ru/popular-science-publications/artificial-intelligence-in-medicine. (n.d.)
53. Iliashenko OZ, Dubgorn A (2019) Opportunities and challenges of artificial intelligence in healthcare. E3S Web Conf 110:20–28
54. Imadali S, Karanasiou A, Petrescu A, Sifniadis I, Veque V, Angelidis P (2012) eHealth service support in IPv6 vehicular networks. Proc IEEE Int Conf Wireless Mobile Comput Netw Commun (WiMob), 579–585
55. Islam SM, Kwak D, Kabir MH, Hossain M, Kwak K (2015) The internet of things for health care: a comprehensive survey. IEEE Access 3:678–708
56. Iswanto IE, Maseleno A (2019) Identifying diseases and diagnosis using machine learning
57. Pearl J (2009) Causality: models, reasoning, and inference. Ambridge University Press
58. Siripongdee K, Pimdee P, Tuntiwongwanich S (2020) A blended learning model with IoT-based technology: effectively used when the COVID-19 pandemic. J Educ Gifted Young Sci 8:905–917
59. Hernán MA, Robins JM (2010) Causal inference. CRC Boca Raton
60. Magoulas G, Prentza A (1999) Machine learning in medical applications. Adv Course Artif Intell, 300–307
61. Marques G, Pitarma R, Garcia NM, Pombo N (2019) Internet of things architectures, technologies, applications, challenges, and future directions for enhanced living environments and healthcare systems: a review. Electronics 8(10):1081
62. Riberio MR, Singh S, Guestrin C (2016) Explaining the predictions of any classifier. ACM Conferences on knowledge discovery and data mining, San Francisco, CA, USA, pp 1135–1144
63. Naudé W (2022) Artificial Intelligence against COVID-19: an early review
64. Schulam P, Saria S (2017) Reliable decision support using counterfactual models. Neural Information Processing Systems (NIPS)
65. Vaishya R, Javaid M, Khan IH, Haleem A (2020) Artificial Intelligence (AI) applications for COVID-19 pandemic. Diabetes & metabolic syndrome. Clinical research & reviews [vaws]
66. Rghioui A, Oumnad A (2018) Challenges and opportunities of internet of things in healthcare. Int J Electr Comput Eng 2088–8708:8
67. Shahamabadi MS, Ali BB, Varahram P, Jara AJ (2013) A network mobility solution based on 6LoWPAN hospital wireless sensor network (NEMO-HWSN). Proc. 7th Int. Conf. Innov. Mobile Internet Services Ubiquitous Comput. (IMIS), pp 433–438
68. Lei T, Barzilay R, Jaakkola T (2016) Rationalizing neural predictions. http://arxiv.org/abs/1606.04155

69. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E et al (2019) Applications of machine learning in drug discovery and development. Nat Rev Drug Discovery 18(6):463–477
70. Vafea MT, Shehadeh F, Mylona EK, Kalligeros M, Mylonakis E (2020) Emerging technologies for use in the study, diagnosis, and treatment of patients with COVID-19. Cell Mol Bioeng 13(4):249=257
71. Obermeyer Z, Lee TH (2017) The limits of the human mind and the future of medicine. New England J Med, 1209–1211
72. Sivaprakash A, Rajan SNE, Selvaperumal S (2019) A novel robust medical image watermarking employing firefly optimization for secured telemedicine. J Med Imag Health Inf 9(7):1373–1381

Chapter 12
An Enhanced Multi-channel MAC Protocol with Intelligent Sleep Scheduling Capabilities for High Lifetime Smart City Networks

Ashutosh Lanjewar, Sunil Kumar, Latesh Malik, and Supriya Sawwashere

Abstract Medium access control (MAC) protocols are responsible for handling the interactions between different communication mediums. They perform tasks like frame delimitation, addressing of nodes, providing transparency to data communication, protection against errors, etc. In this paper an improved MAC protocol design is proposed, which hybridizes MAC protocols like Time division multiple access (TDMA), bit-map-assisted (BMA) and sensor MAC (SMAC), and then adds multi-channel capabilities to it in order to improve the network Quality of Service (QoS) parameters. The proposed protocol also incorporates sleep scheduling capabilities to the hybridized multi-channel MAC protocol in order to further optimize the network lifetime, and reduce routing computational complexity. The proposed protocol is compared with non-hybridized implementations of standard MAC protocols, and it is observed that a delay reduction of more than 20% is achieved, along with an improvement of 10% in the overall network lifetime, which can be used for sustainable development in smart cities that require low energy.

Keywords MAC · Hybrid · Multi-channel · Sleep scheduling · Hybridized multi-channel MAC · Delay reduction · Energy consumption · Packet delivery ratio

12.1 Introduction

MAC protocols have paved the way for better communication control in all kinds of networks. These protocols assist in every phase of network communication right from address assignment to final packet routing. The main aim of these protocols is to improve the overall network communication quality with the help

A. Lanjewar (✉) · S. Kumar
Electronics & Communication Engineering, Kalinga University, Raipur, India
e-mail: ashutoshlanjewar39@gmail.com

L. Malik
Computer Science Engineering, Government Engineering College, Nagpur, India

S. Sawwashere
Computer Science Engineering, Kalinga University, Raipur, India

V. Bali et al. (eds.), *Decision Analytics for Sustainable Development in Smart Society 5.0*, Asset Analytics, https://doi.org/10.1007/978-981-19-1689-2_12

of efficient medium access control. Initially simple protocols like Channel Sense Multiple Access (CSMA) with collision detection (CD) or collision avoidance (CA) were used for medium access control. But as technology matured, protocols like BMAC (Berkley MAC), XMAC (improvement to BMAC), SCP-WUSTL (Ultra-Low Duty Cycle MAC with Scheduled Channel Polling), TDMA (Time division multiple access), SS-TDMA (Sleep scheduling TDMA), BMA (bit-map-assisted MAC), SMAC (Sensor assisted MAC), and various other protocols were designed [1]. A brief classification of these protocols can be observed in Fig. 12.1, wherein MAC protocols are classified into schedule based and contention based.

Schedule-based protocols are usually routing aware and work on the principle of channel scheduling and node scheduling in order to improve the medium access. While contention-based protocols are usually defined in terms of handling scheduling data on the channel such that there is limited contention in the network. Usually a combination of scheduling-based and contention-based protocols are used in order to improve the overall quality of service (QoS) parameters for the network [2]. This paper introduces a hybridization of SMAC, BMA, and TDMA protocols along with multi-channel and sleeps scheduling capabilities in order to improve the overall network performance.

In order to compare the performance of the proposed protocol, the next section describes the recent research done in this area, and the performance of the protocols recently proposed. This is followed by the details about our proposed protocol, and performance analysis with comparison of the proposed protocol with the reviewed MAC algorithms. Finally, this paper is concluded with some interesting observations and future work directions that can further improve the efficiency of the proposed MAC protocol.

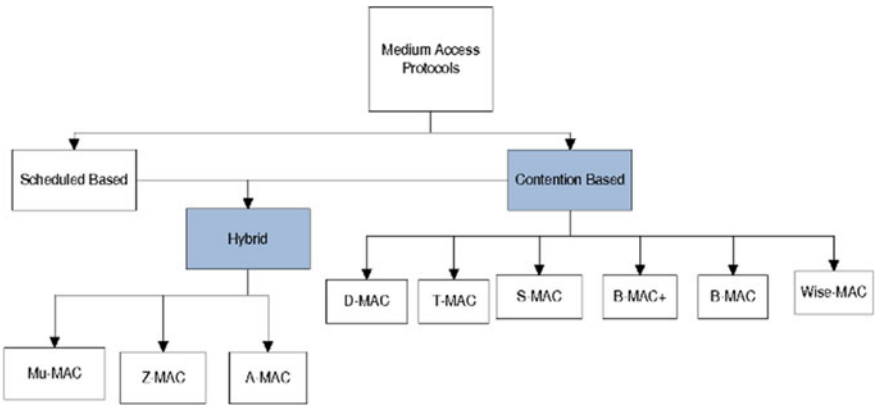


Fig. 12.1 A brief classification of MAC protocols

12.2 Literature Review

Communication networks are everywhere, right from aerospace, land communication, satellite communication, underwater communication and underground communication. A Medium Access Control (MAC) protocol is designed for each of these networks and is implemented such that these networks have high communication throughput with minimum loss of packets. The work in [3] proposes a Receiver Preambling with Channel Polling or RPCP MAC. The MAC protocol starts by preambling the receiver side data, which minimizes listening to packets in idle mode. Post that a novel mechanism for polling the available channels is adopted, which will determine any missed data frames during the sleep period and also reduce energy wastage of the node by keeping them active only for required cycles. Finally, the algorithm adopts a back-off mechanism for resolving collisions if multiple packets are received simultaneously. A diagrammatic description of this can be observed in Fig. 12.2, wherein all the MAC phases, namely sleep, wake-up, and active are described along with receiver preambling.

The RPCP-MAC is designed for underwater communications but can be extended to any kind of network due to its generic nature. It showcases a 10% improvement in energy efficiency when compared with low-power listening (LPL) and receiver-initiated packet train protocol (RIPT) protocols. But the delay of this protocol is high due to the back-off mechanisms and preambling. In order to reduce the delay, the work in [4] can be used. This work proposes a fast-mobility adaptive FMA-MAC. It enhances both mobility and communication management in wireless sensor networks (WSNs). It proposes a novel protocol for detection of vehicle mobility, followed by a novel inter-cluster handoff mechanism for these nodes, along with a priority-based contention mechanism which helps mobile nodes to access the medium and finally a multi-channel strategy is designed which enhances the throughput and assists in managing communication aspects in the network.

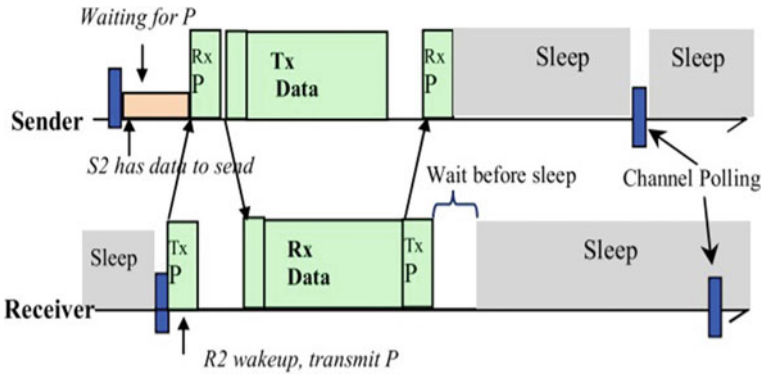


Fig. 12.2 The RPCP-MAC phases

The performance of the FMA-MAC is compared with CSMA, AM-MAC, EMS-MAC, MT-MAC, and MS-SMAC. Most of these protocols do not posses multi-channel capabilities. Due to which the energy, delay and throughput performance of these protocols is not very good when compared with FMA-MAC. Therefore, the underlying research also uses multi-channel communication for improved performance.

The FMA-MAC protocol can be further improved by adding interference awareness to the system. This interference awareness can be observed in [5], wherein an interference-aware multi-channel IAM-MAC protocol is described. The protocol relies on the concept of backup cluster head (BCH) nodes, and therefore in case of any contention in the network, the backup cluster head is selected for further communication. This reduces the dependency of the network on a single communication channel-path combination, thereby improving the overall reliability of successful packet communication. Due to this, the delay of communication is reduced by almost 15% when compared with MMIMO and DMMA protocols, while the average packet delivery ratio increases by almost 10%, by keeping constant energy consumption across different communications. Another multi-channel MAC suited for aerospace networks is defined in [6], wherein load awareness is taken into account while defining medium access control protocols. Moreover, a similar back-off mechanism as described in [4] is implemented to reduce contentions. Here a multi-priority queuing strategy is applied, wherein each node is clustered in groups and sub-groups depending upon the packet priority of the data being transmitted by the node. The multi-priority structure can be observed in Fig. 12.3, wherein a central server takes care of all the queues.

Due to the multi-channel and multi-priority schemes the delay of the communication is reduced by 15% when compared with MILD and BEB mechanisms [6], which results in increasing the throughput by almost 10% when compared with the same protocols.

The work in [3] is extended by [7], wherein energy efficiency is improved with the help of multi-channel communications along with collision avoidance. It uses a 2-hop routing mechanism for route discovery, in order to optimize the path selection

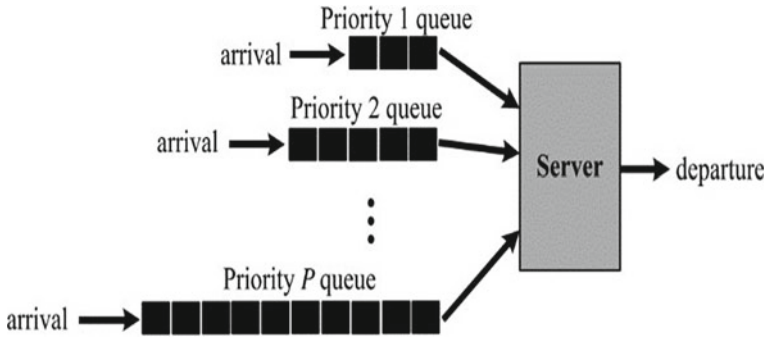


Fig. 12.3 Multi-priority queues

process. The data is transmitted between different pairs of nodes, so that even if one path is under contention then the other path can be used for effective communication. Due to this, the overall delay is reduced by almost 10% when compared with 802.11 protocols, and energy consumption is reduced by 30% for the same network structure.

A similar protocol is applied to vehicle-to-vehicle communication and is demonstrated in [8], it confirms the results of [7] but under a different kind of application. The performance evaluation of these MAC protocols can be further investigated using the work in [2], wherein different parameters like Collision probability, Throughput, Delay, Packet reception rate, Duration of each kind of intervals, etc. have been proposed. Protocols like VCI, Q-VCI, DCI, CAMAC, MP-MAC, FCM-MAC, and APDM have been studied. All these protocols aim at maximizing throughput by minimizing the delay of communication in the network. While throughput optimization is a primary metric for any MAC-based communication interface, the capacity utilization of these networks must also be studied. The work in [9] suggests the use of contention window management for capacity-based contention avoidance in wireless networks. The work proposes a receiver collision sensing scheme RCSS MAC protocol, which allows the receiver to check for collisions and inform the transmission antennas whether to communicate on a channel or not. Due to receiver-based sensing the energy consumption of the transmission equipment is reduced, therefore the overall lifetime of the network is improved. The contention window is managed using an additional victory field in the communication packet, which can be calculated from Eq. 12.1, as follows:

$$V_r = \frac{P_s}{T_p} \ldots \tag{12.1}$$

where V_r is the victory rate, P_s is the number of successfully received packets, and T_p is the total received packets. Due to the inclusion of victory rate, the overall throughput of the network is improved by 5%, while improving energy efficiency by 10% when compared with AOMR, AOMDV, and ACW protocols. This protocol is further extended in [10], wherein multi-channel capabilities are added to the system. Due to this, the throughput is further improved and energy consumption is further reduced when compared with the existing MAC protocol.

Specific MAC protocols are also defined for cognitive radio environments. For instance, the work in [11] proposes cognitive multi-channel opportunistic reservation-based MAC protocol (CogMOR-MAC), wherein a multi-channel MAC protocol is proposed which prioritizes primary users over secondary users, and uses opportunistic reservation for them. The protocol has been tested on different types of networks, and it is observed that the proposed protocol has higher throughput when compared with SHCS-based MAC protocols. This protocol's performance can be further improved with the help of game theoretical approach as mentioned in [12], wherein TDMA slot sharing (TSS) mechanism is used in order to schedule any kind of nodes in the network. This concept of TSS improves the node efficiency for channel access, thereby reducing the mean-waiting time for the node to access

the channels. This also improves the utility of each channel, thereby enhancing the overall QoS of the network. This network can be further enhanced by adding multi-channel capabilities as observed in [6]. Each of these networks has showcased enhanced performance when multi-channel capabilities are added to them, this motivated the underlying research to include multi-channel communications when designing the hybrid MAC protocol. This can again be confirmed in [13], wherein a multi-channel preamble sampling-based MAC protocol is defined. The protocol is defined for 802.15.4 physical layer but can be extended to any kind of physical layer.

The work in [11] is extended in [13], wherein a combination of back-off strategy, random slotting, and collision detection has been made in order to improve the QoS of the network. This improves the throughput and improves the packet delivery ratio of the network. A control channel interval (CCI) based strategy is defined in [14], wherein different CCIs for start and termination of communication are defined. It uses a virtual TDMA beaconing mechanism for control of these channel intervals, which improves the overall QoS of the network. The network is also based on multi-channel communications and solves the hidden terminal problem via cluster member request and cluster head requests. The packet delivery ratio and throughput are observed to be higher when compared with CAVI, TCM, and 802.11p MAC protocols. These MAC protocols can be further improved with the help of self-organizing schema (SoS). The SoSes help in achieving high throughput by speeding up the handover process in networks. The work in [15] proposes a Self-organizing Multi-channel MAC (SOM MAC) protocol which assists in improving the handover capabilities of the MAC protocol. The working on this SOM MAC protocol can be observed in Fig. 12.4, wherein coverage area of each node is defined, and based on this coverage area the overall handover process is controlled.

This protocol assists in further improving the packet delivery ratio by performing intelligent packet scheduling, which reduces the contention in the network.

Control channels are designed for reducing the network traffic on data channels, due to which there are minimum contentions in the network. The work in [16] proposes the design of a common control channel for any MAC protocol, using which the overall network performance can be improved. Protocols like ILP-MAC, DOSS, HC-MAC, SYN-MAC, C-MAC, Su. MAC, HD-MAC, Cog-Mesh, and DUB-MAC have been evaluated in order to evaluate the performance of these networks. It is observed that each of these protocols showcases a performance enhancement when a

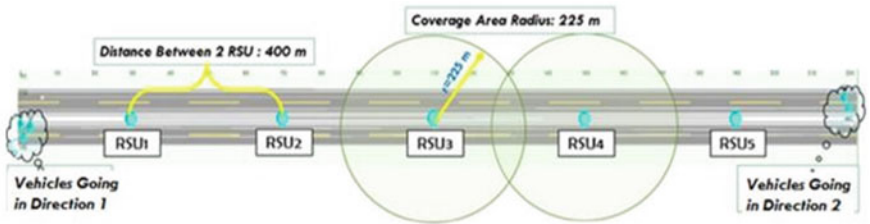


Fig. 12.4 The SOM MAC protocol

common CSMA-based control channel is established between the nodes. The CSMA protocol for this control channel is inspired by [17], wherein an energy efficient collision avoidance protocol for any kind of wireless communication is defined. The work in [18] further extends the concept of cognitive radio networks to IoT networks and proposes a MAC protocol for cognitive radio enabled IoT networks. This protocol is a token-based multi-channel protocol, and thus avoids contention in the network, because tokens are granted to the nodes based on their access patterns. This protocol is observed to have a 20% improvement in network throughput and a 15% reduction in delay per hop when compared with standard MAC protocols. The IoT networks can be further improved with the help of beacon-based MAC protocols. The work in [19] proposes the uses of a beacon-based multi-channel MAC protocol that utilizes receiver-initiated communications.

Due to receiver initiation, there is limited energy loss, and most of the nodes are in sleep mode before communication. This concept when combined with multi-channel communication improves the energy efficiency two-fold, and reduces the delay needed for communication as well. Another cooperative underwater multi-channel MAC protocol is described in [20], wherein the triple hidden terminal problem is resolved using propagation delay estimation, and channel allocation prediction. These problems can be easily resolved using token-based mechanisms, and thereby the work in [20] must only be used for very specific networks where token-based allocation is not possible.

The work in [21] proposes a dynamic back-off mechanism for contention removal in distributed networks. Design of a multi-channel medium access control (MAC) protocol with dynamic back-off contention (MMAC-DB) is proposed in [21]. This protocol analyzes the back-off periods and suggests variations in the same in order to improve the channel utilization and reduce idle channel periods. The algorithm is implemented in the contention phase, wherein a trade-off is made between the data transmission phase and the contention phase in order to reduce the contentions and increase the data transmissions in the network. Due to this, the delay reduces and the network throughput increases by 10% when compared with SMC MAC protocols. This network can further be improved with the help of Packet Reservation Multiple Access Protocols as defined in [22], wherein based on the type of packets, reservations are done, and based on these reservations network communications are performed. The protocol is a hybrid between slotted ALOHA and CSMA/CA protocols. This hybrid combination is able to improve the overall QoS of the network, when individually compared with slotted ALOHA and CSMA/CA protocols. This inspired the underlying research to combine different MAC protocols into a single hybrid MAC protocol for improvement of network QoS. Moreover, addition of multi-channel communication has always been a performance boost for any MAC protocol, as it can be observed from [23, 24], thus the underlying protocol proposes a hybrid multi-channel MAC protocol with sleep scheduling capabilities. The sleep scheduling capabilities can be observed from [25, 26], wherein receiver initiation is combined with multi-channel MACs for QoS improvement. The next section describes the proposed hybrid MAC protocol, followed by its performance comparison with some of the state-of-the-art methods.

12.3 Proposed Hybrid Multi-channel MAC with Sleep Scheduling Capabilities (HMC-MAC-SS)

From the review it is observed that performance of the current MAC protocols can be improved via incorporation of sleep scheduling capabilities. These capabilities can be incorporated using machine learning models for improved sleep scheduling performance. In order to do this, the proposed hybrid multi-channel MAC protocol with sleep scheduling capabilities is defined and can be observed in Fig. 12.5, wherein the node parameters are analyzed in order to select between TDMA, SMAC, and BMA protocols. Once the protocol selection is done, then the data is transmitted using multiple channels, each of these channels is selected such that there is no hidden channel problem. Finally, the data is transmitted only when receiver initiates the communication, otherwise all the nodes are placed into sleep mode. Routing nodes wake-up by receiver initiation, due to which the final path is traced from destination to source. This helps in reducing unnecessary energy loss, and thereby improves the energy efficiency of the network. Moreover, due to the initiation of communication requests from receiver side the final communication route is optimized,

The proposed HMC-MAC-SS protocol works in two phases, each of these phases is described in their respective sub-sections. A combination of these phases results in the final MAC protocol.

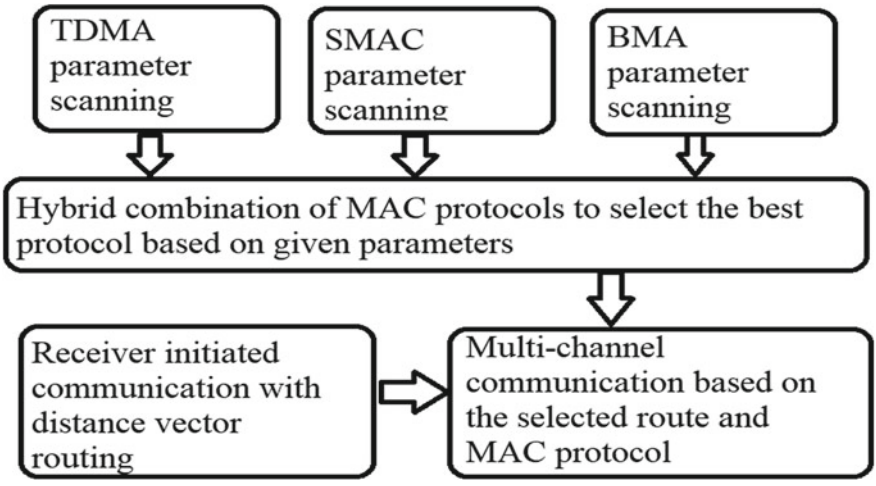


Fig. 12.5 Proposed HMC-MAC-SS protocol

12.3.1 The Hybrid MAC Protocol

The hybrid MAC protocol combines advantages from TDMA, SMAC, and BMA protocols. Each of these protocols is applied to the network, and parameters from these protocols are evaluated. This evaluation is done using a Genetic Algorithm (GA) which works using the following steps:

- Initialize GA parameters,
- Number of iterations (Ni),
- Number of solutions (Ns),
- Learning factor (Lf),
- For each iteration, and for each solution which needs to be changed perform the following steps:

  a. Select a random MAC protocol (TDMA, SMAC, BMA) = Ms.
  b. Perform "N" random communications for this protocol using different protocol specific parameters.

     - Change the time-slots for TDMA.
     - Change the number of sensors for SMAC.
     - Change the cluster numbers for BMA.

  c. At the end of 'N' communications, evaluate the average energy consumption, average delay, average throughput, and average packet delivery ratio for the configuration.
  d. Find out the fitness function for this solution using Eq. 12.2,

$$F_i = \frac{\frac{D_{avg}}{D_{max}} + \frac{E_{avg}}{E_{max}}}{\frac{T_{avg}}{T_{max}} + \frac{P_{avg}}{P_{max}}} \ldots \tag{12.2}$$

  where D is the delay, E is the energy, T is the throughput, P is the PDR, avg is the average value, and max is the maximum value for the parameter,

     - Evaluate the value of fitness for all solutions, and finally evaluate the fitness threshold from Eq. 12.3,

$$F_{th} = \frac{\sum F_i}{N_s} * L_f \ldots \tag{12.3}$$

- Now evaluate the solution change array, wherein if the value of the array is 1, then the solution is changed, else it is kept unchanged. This solution array value is evaluated using Eq. 12.4,

$$S_{\text{change}_i} = 1 \ldots \text{when } F_i \geq F_{th}, \text{ else } 0 \ldots \tag{12.4}$$

Table 12.1 Sample entries for MAC selection

Protocol	Configuration	Fitness
TDMA	Slots = 5	0.557
SMAC	S = 3	0.124
BMA	C = 4	0.115
TDMA	Slots = 10	0.328

- Repeat the computations for Ni iterations, to obtain the configuration table as shown in Table 12.1,

Select the MAC protocol and its configuration with minimum value of fitness. Each of these values is constantly updated after every communication. If the value from the current configuration is higher than the one used initially, then Table 12.1 is updated, and the new lower value is selected for network channel access. Moreover, if more than 50% of the table values gets changed, then the entire process is run again for a more effective network. The system is implemented using network simulator 2 (NS2), and its details are given in Sect. 12.4 of this text. Once the selection of a particular MAC protocol is complete, then receiver-initiated routing protocol is activated as discussed in the next section.

12.3.2 The Receiver-Initiated Sleep Scheduling-Based Routing Protocol

The routing protocol used by our proposed HMC-MAC-SS protocol uses receiver initiation for communication. The protocol works using the following steps:

- All nodes are initially in sleep mode.
- Select a pair of source (S) and destination (D) nodes.
- The destination node finds out the routing path by sending ping packets to one-hop neighbors by waking them up.
- The neighbors which respond the quickest are selected for the routing path.
- For a neighbor to be selected, the following Eq. 12.5 must be true:

$$d_{ref} > d_{sx} + d_{xd} \ldots \tag{12.5}$$

where d is the Euclidean distance between nodes. d_{ref} is the reference distance between source and destination, d_{sx} is the distance between current node and source, d_{xd} is the distance between current node and destination node,

- This equation makes sure that the selected node is in the path between the source and destination.
- Once the equation is satisfied, then the node which sends the fastest response is selected for communication.

- Once the nodes are selected, then all the nodes (except the source node) go into sleep mode again (including the destination node), thereby saving energy.
- The source node wakes up the nearest neighbor selected by the routing protocol and sends data to it via multiple channels.
- Due to multi-channel communication the speed of data transmission is higher.
- Once the data is sent from source to intermediate node, then source goes back into sleep mode, and the next hop node is woken up by the intermediate node.
- This process continues till the data reaches the destination node, upon which the destination node also goes to sleep.

Due to this protocol, the overall energy consumption of the network reduces, and due to multi-channel communication, the delay of packet transmission also reduces. This also increases the packet delivery ratio and throughput of the network. The system is implemented using network simulator 2 (NS2), and its details are given in Sect. 12.4 of this text. A statistical comparison of the proposed protocol with standard protocols is done in the next section.

12.4 Performance Comparison and Evaluation

In order to implement the system and evaluate its performance for the given MAC protocol, network simulation is done using standard parameters for all the MAC protocols. The following network parameters are decided while implementation for simulating the network.

Based on these parameters, the values for delay, energy consumption, packet delivery ratio (PDR), and throughput are evaluated. The value of delay is considered by considering the complexity of genetic algorithm and is evaluated using the following equation,

$$\text{Delay} = D_{\text{comm}} + D_{\text{GA}}.$$

where D_{comm} is the communication end to end delay, and D_{GA} is the delay needed to evaluate the Genetic Algorithm. These values are tabulated in the following tables, wherein the respective labels indicate the performance of the respective algorithms.

From the results in Table 12.2, it can be observed that the delay performance has been improved by more than 40 % when compared to other systems. Similar comparisons are made for energy, throughput and PDR. These values can be observed in Tables 12.3, 12.4, and 12.5 as follows:

The energy performance is improved by 35 % when compared to other implementations.

From the results it can be observed that the proposed MAC implementation is able to optimize the delay and energy in the network with 40% and 35% respectively while keeping packet delivery ratio and throughput performance to a respectable value when compared with other implementations (see Table 12.6)

Table 12.2 Network parameters

1	Channel type	Wireless channel
2	Propagation model	Two ray ground
3	Network interface	Wireless physical
4	MAC protocol	Mac/802.11
5	Interface queue type	Drop tail priority queue
6	Antenna type	Omni directional antenna
7	Number of nodes	30
8	Routing protocol	AODV
9	Network X size	300
10	Network Y size	300
11	Packet size	1000 bytes per packet
12	Packet interval	0.01s per packet

Table 12.3 Delay performance

Num. nodes	Delay (ms) TDMA [1]	Delay (ms) BMA [1]	Delay (ms) SMAC [4]	Delay (ms) HMC-MAC-SS
20	0.26	0.30	0.37	0.15
30	0.27	0.31	0.39	0.16
40	0.35	0.40	0.50	0.21
50	0.49	0.56	0.70	0.29
60	0.58	0.67	0.84	0.35
70	0.66	0.76	0.95	0.40
80	0.78	0.90	1.12	0.47
90	0.89	1.02	1.27	0.53
100	0.92	1.05	1.31	0.55

Table 12.4 Energy performance

Num. nodes	Energy (mJ) TDMA [1]	Energy (mJ) BMA [1]	Enrgy (mJ) SMAC [4]	Energy (mJ) HMC-MAC-SS
20	5.30	6.10	7.60	3.17
30	5.90	6.79	8.46	3.52
40	6.20	7.13	8.89	3.70
50	6.50	7.48	9.32	3.88
60	6.70	7.71	9.60	4.00
70	6.90	7.94	9.89	4.12
80	7.50	8.63	10.75	4.48
90	7.90	9.09	11.32	4.72
100	8.20	9.43	11.75	4.90

Table 12.5 Throughput performance

Num. nodes	Thr (kbps) TDMA [1]	Thr (kbps) BMA [1]	Thr (kbps) SMAC [4]	Thr (kbps) HMC-MAC-SS
20	321.00	279.13	400.09	500.11
30	335.00	291.30	417.54	521.92
40	337.00	293.04	420.03	525.04
50	339.00	294.78	422.52	528.15
60	342.00	297.39	426.26	532.83
70	345.00	300.00	430.00	537.50
80	347.00	301.74	432.49	540.62
90	352.00	306.09	438.72	548.41
100	355.00	308.70	442.46	553.08

Table 12.6 PDR performance

Num. Nodes	PDR (%) TDMA [1]	PDR (%) BMA [1]	PDR (%) SMAC [4]	PDR (%) HMC-MAC-SS
20	98.50	98.60	98.55	98.55
30	98.70	98.80	98.75	98.75
40	98.90	99.00	98.95	98.95
50	99.10	99.20	99.15	99.15
60	99.10	99.20	99.15	99.15
70	99.20	99.30	99.25	99.25
80	99.30	99.40	99.35	99.35
90	99.40	99.50	99.45	99.45
100	99.50	99.60	99.55	99.55

12.5 Research Implications

The following findings are observed from the proposed model,

- Smart city lifetime can be extended by up to 15% which reduces cost of running.
- Delay needed for communication can be reduced by 20%, which can assist in high speed and real-time updates in smart city development.
- Due to increase in throughput, there is an improvement in bandwidth offered by the network, which can assist in increasing number of users handled per node for smart city.

12.6 Limitations

The following limitations are observed from the proposed model:

- Security level of the network is not high enough and must be worked upon.
- Handoff strategies must be incorporated for better performance evaluation.
- Addition of better network handling protocols based on transfer learning can be used.

12.7 Conclusion

From the obtained results, it can be observed that a combination of TDMA, SMAC, and BMA with multi-channel and receiver-initiated sleep scheduling improves the QoS parameters of the network. The network performance is improved due to the inclusion of Genetic Algorithm for MAC selection along with receiver-initiated communications, which reduces the energy consumption by almost 40% when compared to individual TDMA, SMAC, and BMA protocols. The delay performance is improved due to multi-channel communications along with receiver-based communications.

12.8 Future Scope

In future, the system performance can be further improved by addition of machine learning and deep learning algorithms, because the continuous learning Genetic Algorithm requires the network to keep on learning from previous iterations which increases the computational complexity of the network. Moreover, the routing algorithm can be modified to add blockchain-based routing, which will further optimize the security and QoS performance of the network. Further improvements can be done by adding trust establishment and privacy preservation capabilities to the system for better security performance, which might reduce QoS of the network. Once these techniques are added, then the existing machine learning model can be modified to improve the system's QoS performance.

References

1. Fernandes R, de Almeida M, Brandão D (2020) Performance evaluation of asynchronous multi-channel MAC protocol for WSN. Wireless Pers Commun 113(2):1115–1133
2. Nguyen V, Khanh T, Pham X, Lee G, Huh E (2019) Performance analysis of adaptive MACprotocol in VANETs considering the potential impact on throughput and transmission delays. Int J Commun Syst 33(1):e4172

3. Roy A, Sarma N (2020) RPCP-MAC: receiver preambling with channel polling MAC protocol forunderwater wireless sensor networks. Int J Commun Syst 33(9):e4383
4. Hamidi-Alaoui Z, El Belrhiti El Alaoui A (2019) FM-MAC: a fast-mobility adaptive MAC protocol for wireless sensor networks. Trans Emerg Telecommun Technol 31(6)
5. Kakarla J, Majhi B, Battula R (2015) IAMMAC: an interference-aware multichannel MAC protocol for wireless sensor-actor networks. Int J Commun Syst 29(4):801–822
6. Zheng B, Li Y, Cheng W, Wu H, Liu W (2020) A multi-channel load awareness-based MAC protocol for flying ad hoc networks. EURASIP J Wirel Commun Netw 1:2020
7. Bouabdallah F, Zidi C, Boutaba R, Mehaoua A (2019) Collision avoidance energy efficient multi-channel MAC protocol for under water acoustic sensor networks. IEEE Trans Mob Comput 18(10):2298–2314
8. Yang F, Zou S, Tang Y, Du X (2016) A multi-channel cooperative clustering-based MAC protocol for V2V communications. Wirel Commun Mob Comput 16(18):3295–3306
9. Gafur AM, Ben Saleh A (2019) Capacity utilization based on contention window management in mobile ad hoc network. Int J Numer Model Electron Netw Devices Fields 32(6)
10. Incel O, van Hoesel L, Jansen P, Havinga P (2011) MC-LMAC: a multi-channel MAC protocol for wireless sensor networks. Ad Hoc Netw 9(1):73–94
11. Feng P, Bai Y, Huang J, Wang W, Gu Y, Liu S (2019) CogMOR-MAC: a cognitive multi-channel opportunistic reservation MAC for multi-UAVs ad hoc networks. Comput Commun 136:30–42
12. Jouhari M, Ibrahimi K, Tembine H, Benattou M, Othman JB (2019) Signaling game approach to improve the MAC protocol in the underwater wireless sensor networks. Int J Commun Syst 32(13)
13. Alahmadi H, Bouabdallah F (2019) Multichannel preamble sampling MAC protocol for wireless sensor networks. Int J Distrib Sens Netw 15(5):155014771985095
14. Pal R, Prakash A, Tripathi R (2018) Triggered CCHI multichannel MAC protocol for vehicular ad hoc networks. Veh Commun 12:14–22
15. Ağgün F, Çıbuk M, Ur-Rehman S (2020) A new self-organizing multi-channel MAC schemafor RSU-centric VANETs. Phys Stat Mech Appl 551:124098
16. Chuanqing L, Zhiming W, Hongmei S (2010) Design on common control channel of MACprotocol of cognitive radio networks. In: 2010 international conference on electrical and control engineering
17. Bouabdallah F, Zidi C, Boutaba R, Mehaoua A (2019) Collision avoidance energy efficient multi-channel MAC protocol for underwater acoustic sensor networks. IEEE Trans Mob Comput 18(10):2298–2314
18. Wu C-M, Kao Y-C, Chang K-F (2020) A multichannel MAC protocol for IoT-enabled cognitive radio ad hoc networks. Adv Technol Innov 5(1):45–55
19. Gomes RD, Benavente-Peces C, Fonseca IE, Alencar MS (2019) Adaptive and beacon- based multi-channel protocol for industrial wireless sensor networks. J Netw Comput Appl 132:22–39
20. Rahman P, Karmaker A, Alam MS, Hoque MA, Lambert WL (2019) CUMAC-CAM: a channel allocation aware MAC protocol for addressing triple hidden terminal problems in multi- channel UWSNs. SN Appl Sci 1(7)
21. Wei Z-H, Hu B-J, Lin Z (2019) Multichannel MAC protocol with dynamic backoff contention for distributed cognitive radio networks. EURASIP J Wirel Commun Netw 1:2019
22. Singh RP, Sangwan A, Godara K (2019) Modified-PRMA MAC protocol for cognitive radio networks. Wireless Pers Commun 107(2):869–885
23. Kannan K, Devaraju M (2018) QoS supported adaptive and multichannel MAC protocol in vehicular ad-hoc network. Clust Comput 22(S2):3325–3337
24. Yousefi MHN, Kavian YS, Mahmoudi A (2018) RTMCH: real-time multichannel MAC for wireless video sensor networks. Multimed Tools Appl 78(6):7803–7818
25. Gupta N, Prakash A, Tripathi R (Eds.) (2020) Internet of vehicles and its applications in autonomous driving. Springer International Publishing AG
26. Zhao H, Zhang M, Gao K, Mao T, Zhu H (2018) A Multi-channel cooperative demand-aware media access control scheme in vehicular ad-hoc network. Wireless Pers Commun 104(1):325–337

Chapter 13
An Analytical Approach Towards Data Stream Processing on Smart Society for Sustainable Development

Devesh Kumar Lal and Ugrasen Suman

Abstract In real-time processing, stream has to be processed as soon as it's generated. Data streams generated from IOT sensors are processed into a finite-size window. A sheer window is considered for processing of data streams in a particular time stamp. In these sheer windows, compare reduce aggregate (CRA) algorithm is applied for determining linear relation for multiple feature vectors. A real-time inference pattern is determine using hash-based classification. In this chapter, sheer window hash-based classification using binarised window analytic (SHCUBA) approach is proposed. This approach is beneficial for calculating linear relationship between sheer windows. SHCUBA approach is comprised of transformation and virtualisation. Here, time and space complexity for this approach is $O(n)$ and $O(1)$, respectively. This reduces latency and space requirements for various real-time use cases such as smart applications, sentiment analysis, IOT-based solutions, fraud detection and prevention, stock market prediction, etc. Data generated in smart societies can be correlated using SHCUBA approach for inferring useful decisions.

13.1 Introduction

A foundation of sustainable development of smart society is laid over Internet of Things (IoT) technology. Smart society eases citizen's living by providing automated utility services. Smart society solution helps in managing institutions in efficient manner. The overall development raises economic growth of a society. Here, IoT technologies plays significant role for achieving sustainable development [1]. IoT platform generates and processed real-time data from various sensors. These unbounded data streams can be used in wide varieties of application such as traffic management, smart waste management, smart healthcare system, smart transportations, etc. [2–6].

These applications are data-driven, where data stream are processed in real time by using special analytical algorithms. Some of the real-time algorithms are lossy counting, count-min sketches, bloom filters, etc. [7–9]. In real-time data processing

D. K. Lal (✉) · U. Suman
School of Computer Science & IT, Devi Ahilya University, Indore, India
e-mail: devesh2222@gmail.com

V. Bali et al. (eds.), *Decision Analytics for Sustainable Development in Smart Society 5.0*, Asset Analytics, https://doi.org/10.1007/978-981-19-1689-2_13

use cases, latency and space requirements should be minimum. Different approaches can be applied over data streams to determine the frequency count, aggregation, joins, etc. This approach is applied over a fixed size of data streams termed as 'window'. Real-time processing of data stream depends upon the overall size of a window. An appropriate sized window is useful for processing data streams in minimum latency. In this chapter, a sheer window approach is used for performing analytics in real time.

Sheer window is a term obtained from SBASH architecture describe as a self-sufficient unipartite sized window [10]. Sheer window is an adequate size of streams which consists data streams for a particular time stamp. This time stamp is calculated using various parameters such as, rate of flow, resource availability, latency, etc. A consideration of aforementioned parameter may provide efficient utilisation of processing capabilities. These continuously generated sheer window has to be processed in real time. A particular time interval may consist a chunks of data streams after occurrence of an event. Unbounded data stream flow into a channel is passes through multiple transformations. Transformation is a techniques applied over data streams, where it transformed into aggregated representation. A suitable virtualisation technique is beneficial for gaining useful insights from data streams. In this chapter, we have proposed sheer window hash-based classification using binarised window analytical (SHCUBA) approach. SHCUBA approach is comprised of two stages; namely, transformation stage and visualisation stage.

Streaming data is transform into a meaningful insights by applying a suitable transformation techniques. Different type of transformations are map, flatmap, filter, groupByKey, etc. SHCUBA approach uses real-time correlation transformation over data stream. A correlation technique is used for determining linear relation between two sheer windows. In SHCUBA approach, a real-time correlation transformation is performed by using CRA (Compare Reduce Aggregate) algorithm [11]. CRA algorithm is applied over continuously generated sheer windows. This transformation is applied over multiple feature vectors with column-oriented manner. The processed correlated results pass on to next stage of SHCUBA approach for virtualisation. Visualisation stage of SHCUBA approach is help in gaining inference from hash table. Binarised window and hash-based classification are two-step processes used for visualising real-time correlation of sheer window.

A correlation between sheer windows in real time may provide trends for a particular time interval. This trends are beneficial to infer about the directional growth of data streams in terms of associativity such as positive, negative or zero. Here, sheer window is transformed into fixed-sized binarised windows. After binarisation, a binarised window categorise into specific groups using hashmap technique. Sheer window is transformed into equivalent binarised window by passing through CRA technique. This representation includes concept of classification where multiple groups are included. This approach may provide a summarised view of entire data stream of particular time stamp. These proliferated data streams can be represented in a finite memory space.

An ideal analytical approach of determination of linear relation in real time is achieved by using following arrangement as shown in Fig. 13.1. This solutions

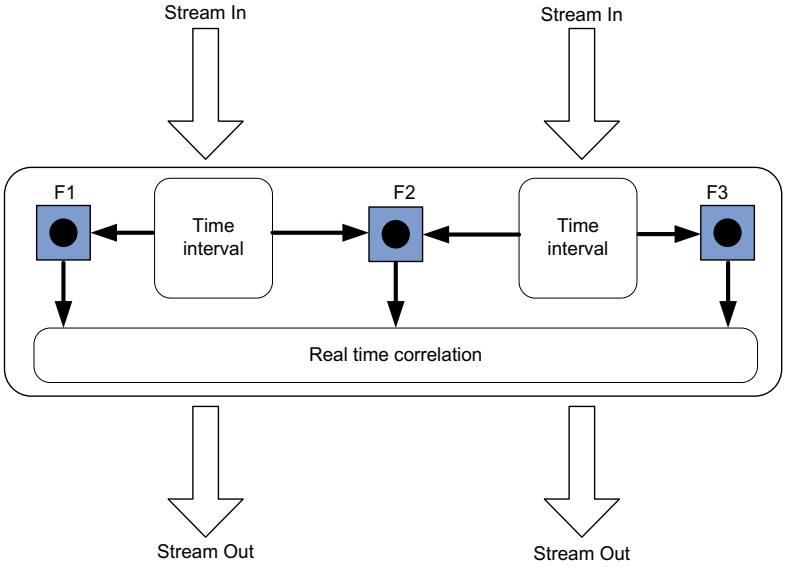


Fig. 13.1 Real-time correlation mechanism

includes requirements of various components such as stream IN, real-time correlation approach (processing elements) and stream OUT. Here, correlation is determined in single pass, which is beneficial for real-time applications. In Fig. 13.1, processing elements receive data streams with a different time slots, which is responsible to ingest one to many tuples in any period of time. Three function are used to aggregate similarity between streams in a specific time allowing to evict older data from other end. This composition is adapted by SHCUBA approach

In this chapter, we have adapted a different mechanism which comprised together to form SHCUBA approach. Here, multiple correlation is applied over different sets of feature vector explained in Sect. 3.1. Here, a fine-sized window is used for data processing at a time known as sheer window. Suppose a finite size sheer-sized window is allocated for processing at a time ranging from t_range_0 to t_range_n. Sheer window is unipartite sized window generated over regular time interval. These sheer window is formed by sheer window manager (SWM) as mention in SBASH architecture [10]. Here, data stream contains tuples, which is a combination of multiple features vectors denoted by $f0$ to fn. Arbitrary tuple consists information for any span of length $l0$ to ln and time t_range_0 to t_range_n is $A[m, n] \forall : \{m \in (f0 \rightarrow fn) | n \in (l0 \rightarrow ln)\}$, where "$m$" and "$n$" indicates row and column of an array, respectively. These unbounded and unprocessed raw data streams from different feature vectors may not be suitable for determining inference. These feature vectors is correlated by column oriented for producing a relation. A fn number of feature vectors may produce $fn - 1$ number of correlation in column manner. These correlated result among all feature vectors are essential for

determining a relation between two sheer windows. A linear relation between sheer windows may provide an extracted essence from volume of unbounded data stream. The degree of dependency in term of correlation technique is applied over sheer windows.

A stream flows on unbounded rate into a specific channel which requires a deterministic time and space to process. These summarised window is necessary for reducing space complexity. SHCUBA approach provides $O(1)$ space complexity. This individual window shows some relation among each other. In this approach, we may perform correlation among columns of feature vectors between sheer windows. These sheer window is transformed into a summarised window using binarised approach, which has a time complexity of $O(n)$. In this chapter, we have introduced a mechanism for determining classification between two sheer windows using hashing and hash table storage. The space and time complexity of hash-based classification is $O(1)$. Lastly, we have shown a technique for reducing a variability into a stream by lossless XOR operation.

The chapter is organised as follows. Section 13.2 explains about related work on the field of real-time stream processing correlation. Section 13.3 describes about SHCUBA approach. Section 13.4 describes the experimental set-up and implementation. The performed results and comparison is presented in Sect. 13.5. Section 13.6 concludes the chapter with future work.

13.2 Related Work

Data streams are classified using linear discriminant analysis which is applied on sequential as well as fraction of data [12]. Real-time clustering is performed over continuous generated parallel data streams [13, 14]. Ensemble tree methods is used for indexing data streams [15]. Data streams are classified into novel classes in concept drifting scenarios [16]. A real-time data stream of time series are represented into Boolean manner and each streams is compared with its mean value [17, 18]. The Boolean streams combination are useful for determining correlation between data streams. A summarisation technique is proposed for data streams which helps in reducing overall memory storage [19–21]. Distributed correlation is applied over streaming time series data [22–24] in sliding window manner. Here, Pearson's correlation coefficient is applied over two consecutive sliding window. Correlation is used in pre-processing phase of data mining for feature selection [25]. In multidimensional data analysis, Pearson's correlation coefficient is applied on static data [26, 27].

The major challenges associated with existing analytical stream processing approaches is the selection of appropriate window-sized data stream. Another aspect is selection of stream processing analytical approach, which is applied to generate inference in real time. The calculation of linear relationship for multiple feature vectors of data streams in real time is also a research challenge. The classification of unbounded data streams into different clusters also possess a challenge. Another

challenge is associated with a storage capabilities of processed data. A fixed memory size is required to map enormous volume of unbounded data streams into it.

After considering above challenges, we have proposed SHCUBA approach. This work is based on independent stream and techniques applies over continuously generated streams. Here, sheer window approach is used by applying CRA correlation algorithm over multiple features vectors. A summarised representation is proposed, which provides a fixed-sized binarised window. A hash-based mapping in used for classification of sheer window. In this techniques, a function mapping key is used as decimal representation of binarised window.

13.3 SHCUBA Approach

Sheer window hash-based classification using binarised window analytical (SHCUBA) approach is used to obtain linear relationship in real time. This linear relation between sheer windows is obtained by sequential execution of three different phases. The phase are, namely, multiple CRA phase, binarised formation window phase and hash-based classification phase. The combination of these phases are termed as SHCUBA approach. We have proposed three different algorithm for these phases, which is incorporated together and executed into sequential manner. These combined algorithms are considered as a steps by which it's executed into a particular order.

In Fig. 13.2, multiple feature vectors is represented by $f1, f2, \ldots..fn$, where tuple $t1, t2, \ldots..tn$ consists of aggregated value of a particular time stamp. SHCUBA approach is applied for multiple feature vectors with data streams of any given time stamp. SHCUBA approach is shown in Fig. 13.2 by applying algorithms over data streams. Stream constituted of multiple variables from which most suited feature vectors are selected. CRA correlation is performed over combination of feature vectors. Data streams are passes through CRA algorithm which provides correlated values ranges from '-1' to '1', where '-1', '0' and '1' are negative, zero or positive correlations, respectively. These correlated values are computed for multiple feature vectors. These values passes through second stage for binarised window formation.

Binarised algorithm is applied over this compared values, which provides windows in form of zeros and ones. This binarised window of length 'l' acquired combination of zeros and ones. This binarised window can be converted into an equivalent decimal representation. Decimal representation of binarised window is known as by term 'True Value'. The minimum and maximum possible True values ranges from '1' to '2^l', respectively. These sequence of binarised window forms an abstract view for finite time frame. A comparison between a consecutive binarised windows, provide similar values. Single binary changes in higher order window may result increase in overall True value, whereas lower order changes in binarised window reduces the increase of True values.

A difference between two consecutive True values generated from binarised window will provides Relative values. In SHCUBA approach, it become flexible

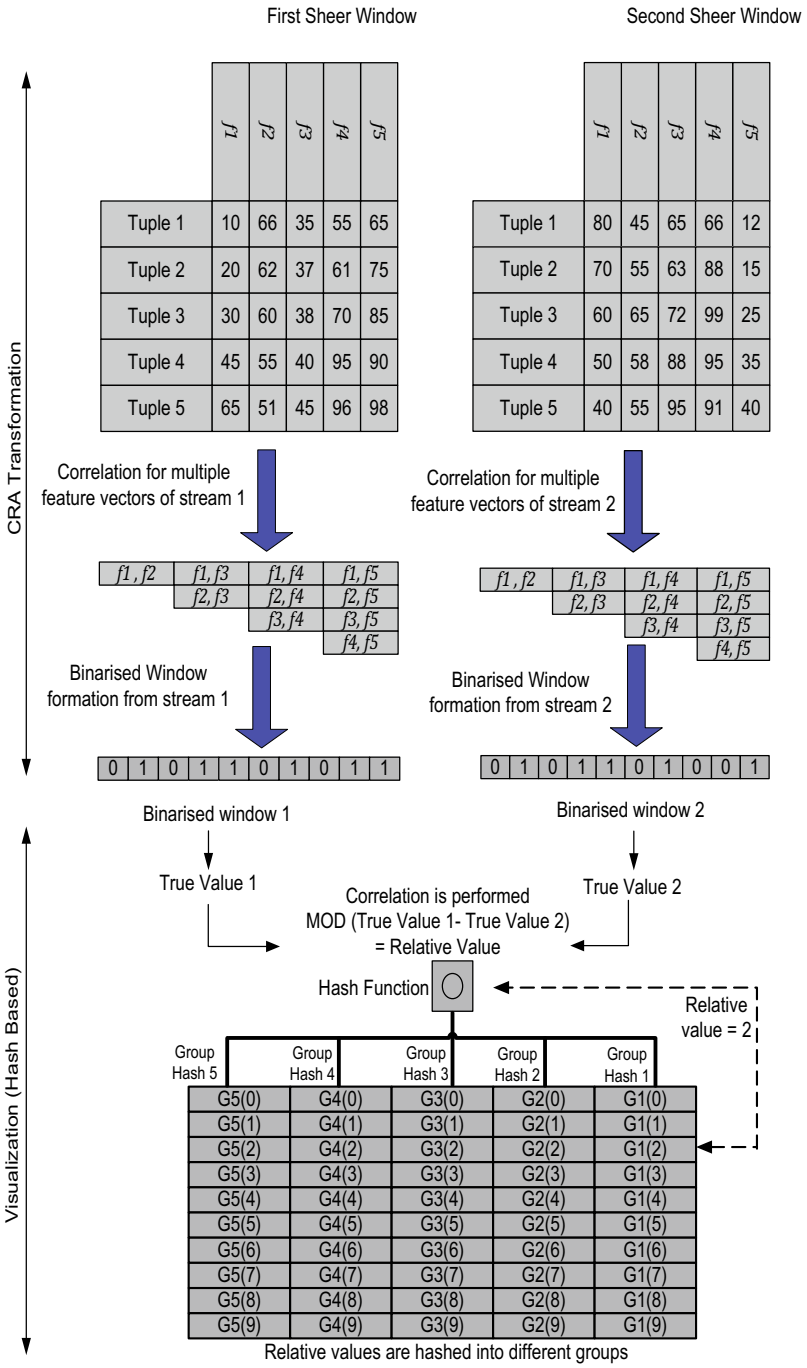


Fig. 13.2 SHCUBA approach with different phases

to apply CRA algorithm over relative values to find correlation. Moreover, a relative values are hash mapped into a respective hash table. Relative values are beneficial for identifying referential groups based on their hashmap with individual length. The possible number of group depends on length of relative value. For example, 'l' length relative value require 'l' number of groups. These values are mapped based on hash function into a sets of groups or termed as hash table. This window maps to suitable categories using hash-based classification approach. In classification approach, relative values with least variability (length equal to one) is higher probable to maps into group 1. The mapping of other relative values into group 2, group 3 and so on according to their rate of variability (length more than one). In short, qualitative or categorical data stream are transformed into quantitative data streams, where we may identify a linear relation between two sets of data. For example, it became easier to find relation between two numbers such as 245 and 242, which is formed after applying binarised algorithm.

An individual group is useful to infer about nature of data streams. The aggregated values on this groups are useful to infer about linear relation between sheer windows. Adaptive linear relation between sheer windows provides real-time trends in term of positive, negative and zero. Difference between two binaries window hashed into a hash table. A calculated multiple correlated values are binarised and form a window termed as 'binarised window'. Arrangement of feature vectors in a particular order plays significant role in determining associativity. A least variable feature vectors arrange in higher order of binarised window, whereas lower order consists higher variable feature vectors. This arrangements may provide affected binary bits in lower order binarised window. Decimal value obtained from binarised window is least tangible different from previous windows. A detailed explanation about complete algorithm is covered in next subsequent section.

13.3.1 Multiple CRA Phase

Here, CRA algorithm is used to perform correlation between multiple feature vectors. Data streams propagate from respective feature vectors and constituted in the form of sheer window. These sheer window assigned into CRA block and produces binarised window as shown in Fig. 13.3. Sheer window passes through CRA algorithm which provides multiple values ranges from '−1' to '1'. CRA approach is alternate approach of rank-based correlation such as Pearson's rank based coefficient of correlation, Kendal's correlation, etc. In CRA approach, two streams from different sources are compared and assigned values accordingly. This values are replaced with the actual parameters in reduce phase, whereas, in last phase in aggregation, the reduce values are combined together. These phases of CRA approach may provide six different parameters. CRA correlation obtain results using Eq. 13.1, where weight is indicated by 'W', whereas other parameters such as $h1, l1, h2, l2, s$ and d represented as aggregated values in stream.

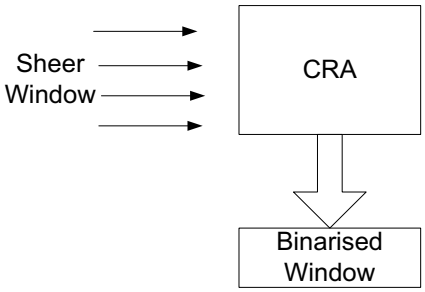


Fig. 13.3 Multiple SW passes through CRA and provide BW

$$\text{Correlation} = (W) * \frac{\max(s,d) - |(|h1 - l1| - |h2 - l2|)|}{s + d} \tag{13.1}$$

In real-time processing, streams are generated in the form of features vectors. Initially, at the time of stream accumulation, feature vectors are undefined. Suppose, some features where selected from a streams as $s_y = \{f_1, f_2, f_3, f_4\}$. These feature vector may produce relation of $s_y - \varnothing = \{\{f1, f2\}, \{f1, f3\}, \{f1, f4\}, \{f2, f3\}, \{f2, f4\}, \{f3, f4\}\}$. These provides total combination of $\frac{f_n(f_n-1)}{2}$, where f_n indicates total number of feature vectors. Time complexity of $O(n^2)$ requires to perform one to one correlation among all feature vectors. In time stamp of $\{t_1, t_2, t_3, t_4, \ldots .t_n, \}$, a number of sheer window (sh) as $\{\{sh_1, sh_2, sh_3, \ldots\ldots..sh_n\}, \{sh_{11}, sh_{22}, sh_{33}, \ldots\ldots..sh_n\}, \{sh_a sh_b, sh_c, \ldots.sh_n\}\}$. The overall complexity associated with correlation which is performed among sheer window in constant f_n with arbitrary t_1 is $= \sum_{sh1}^{shn}\left\{\prod_1^p\left(\prod_{p-1}^q(f_{p*q})\right)\right\}$. A matrix representation of multiple feature vectors of data stream are shown in Eq. 13.2. In this matrix, feature vectors of column $f_{(i)(j)}$, $f_{(i+1)(j+1)}$ is selected for any particular time stamp.

$$\text{Streams}(f_{ij}, f_{(i+1)(j+n)}) = \begin{bmatrix} f_{x1x1} & f_{12} & f_{13} & f_{14} & & f_{1j} \\ f_{21} & f_{22} & f_{23} & f_{24} & & f_{2j} \\ f_{31} & f_{32} & f_{33} & f_{34} & & f_{3j} \\ f_{41} & f_{42} & f_{43} & f_{44} & \cdots & f_{4j} \\ f_{51} & f_{52} & f_{53} & f_{54} & & f_{5j} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ f_{i1} & f_{i2} & f_{i3} & f_{i4} & & f_{ij} \end{bmatrix} \tag{13.2}$$

A multiple correlation is performed over entire feature vectors and total computation requires $\frac{f(f-1)}{2}$. Therefore, time complexity for multiple correlation is $O(\sqrt{n})$. In multiple correlation, it requires one to one storage which is similar to $\frac{f(f-1)}{2}$, thus space complexity is $O(n^2)$.

In this algorithm, (nth)-dimensional feature vectors streams are fetched with two for loops. Binarised correlation is performed within the scope. The pseudo-code for this approach is shown below.

Pseudo Code : Multiple CRA phase

In this algorithm, multiple correlation is performed for which two different *for loops* is considered on line 3 and 4. Line 5 executed for number of feature vectors presents in a data streams and results are stored into a mutable list $Correlated\_result[val]$. A CRA algorithm is performed for each feature vectors $\boldsymbol{f_{ij}}$ varies from $\boldsymbol{f_{i\to 0\ to\ n\ and\ j\to 0\ to\ m}}$. Line 6 return the correlated results whereas line 7 call the procedure $MCP(arg1, arg2)$.

Input parameter : Stream with multiple $f_0 \rightarrow f_n$.
Output parameter: Multiple correlated results.

1. $\boldsymbol{procedure}\ MCP(arg1, arg2)$
2. $Correlated\_result[val] = null$
3. $for(i = arg1; i \leq n; i++)$
4. $\quad for(j = arg2; j \leq n; j++)$
5. $\quad\quad Correlated\_result[val] = \boldsymbol{f_{ij}}\ (\boldsymbol{CRA})\ \boldsymbol{f_{ij}}$
6. $return\ Correlated\_results$
7. $MCP(f_0, f_n)$

13.3.2 Binarised Window Formation Phase

A binarised algorithm applied over entire feature vectors and form correlation matrix. A total number of $\frac{n(n-1)}{2}$ comparison is required, which is equals to size of binarised window. A total size of binarised window is calculated using Eq. (13.3).

$$\text{Size of binarised window} = \frac{n(n-1)}{2} \tag{13.3}$$

Multivariate correlated values are mapped into binary window in form of 0 and 1 as shown in Eq. 13.4. Positive correlated values are mapped with 1, whereas negative correlated values are mapped with 0.

$$\text{Correlation}\left(\boldsymbol{f_{ij}}, \boldsymbol{f_{(i+1)(j+n)}}\right) = \begin{cases} 0, x < 0 \\ 1, x \geq 0 \end{cases} \tag{13.4}$$

This binary window constitute a decimal number equivalent is less than $2^{\frac{n(n-1)}{2}}\{n \in$ number of feature vectors$\}$. This window may assign into different group based on binary patterns of window as mentioned in hash-based classification algorithm.

This algorithm requires one comparison to decide whether a value stored in binarised window is 0 or 1. Time complexity for n numbers of correlated values are $O(n)$, whereas space complexity to store binary values in window is $O(1)$. Here, size of binarised window is constant for n numbers of inputs. Pseudo-code for this algorithm is shown below, where a list b_w is append the value 0 or 1.

Pseudo Code : Binarised Window Formation phase

In this algorithm, it receives multiple correlated results as an input and produces binarized window. Line 2 consists $list\ b_w$ for binarised window. In line 3, $Correlated\_result$ is iterated upto its total length. Its instantaneous value are compared and append using zero and one as mention in line 5 and 7.

Input parameter: Multiple correlated results.
Output parameter: Binarised window b_w.

```
1.  procedure binarised_window(Correlated_result)
2.  list  b_w
3.  for x in correlated_result
4.          if (x ≤ 0)
5.                  append b_w(0)
6.          else
7.                  append b_w(1)
8.  end procedure
```

13.3.3 Hash-Based Classification Phase

A different labelled groups are mapped with multivariate correlated binarised window. A binarised window form a decimal equivalent value useful for determining respective classification. The categorisation of window may be performed by keeping aggregate frequency record in hash table. A hashing technique is used to map window into a particular hash table. Group is assigned to particular window by using Eq. 13.5. Every times group is maps into a locations an aggregated frequency stored. This equation provides result which is most suited for limited numbers of feature vectors.

$$\text{group} + := (b_{wi} - b_{w(i+1)}) \bmod (10) \tag{13.5}$$

If increased difference of $(b_{wi} - b_{w(i+1)})$ size satisfy the Eq. (13.6), where multiple hash table and hashing fuction is used. In this equation, the binarised window differences are converted into string and a length of string is determined. This length of strings y may be greater than or equals to two.

$$\text{length}\big(str\big(b_{wi} - b_{w(i+1)}\big)\big) == y : \{y \leq 2\} \tag{13.6}$$

$$x_i = \text{length}\left(str\left(b_{wi} - b_{w(i+1)}\right)\right)/10$$

$$\text{group}_{\text{hash}(i)} + := (x_i)\text{mod}(10) \tag{13.7}$$

In Eq. 13.7, hash table is decided for a group with a specific x_i. This equation is executed for total length of y. Hash-based classification algorithm is shown in next section.

In this algorithm, time complexity for n inputs requires same number of hashing, thus the complexity is $O(n)$, whereas constant hashmap table provides $O(1)$ space complexity.

Algorithm 2: Hash Based Classification phase

Here, it receives binarised window b_w as an input and transform into equivalent decimal representation by *proceduce digitize() as mention in line 1.* This procedure returns decimal values by function *digit(b_w).* Line 3 defines *proceduce grouping()*, where difference between two decimal values are determine in variable y as calculated in line 5. One *for loop* is used in line 6, which able to form possible numbers of groups as define in line number 7. If y length is greater than two then the difference is hash map into respective tables as pseudo code shown from Line 8 to 12.

Input parameter: Binarised Window b_w.
Output parameter: classified group structure g_w.

1. $\boldsymbol{procedure}\ digitize(b_w)$
2. $\quad return\ digit(b_w)$
3. $\boldsymbol{procedure}\ grouping()$
4. $\quad x_i = digitize(b_w)$
5. $\quad y = length(str(x_i - x_{i+1}))$
6. $\quad \boldsymbol{for}\ (i = 0 \rightarrow n\ \boldsymbol{do})$
7. $\quad\quad group := (x_i - x_{i+1})\ mod\ 10$
8. $if\ y \leq 2$
9. $while\ (y!\,NULL)$
10. $x_i = length\left(str\left(b_{w_i} - b_{w_{(i+1)}}\right)\right)/10$
11. $group_{hash(i)} + := (x_i)\ mod\ (10)$
12. $y - -$
13. $\boldsymbol{end\ procedure}$

SHCUBA approach can further modified to provide better results by ingesting optimised data stream into CRA algorithm. This mechanism is discussed in next section with different graph plots.

13.3.4 Optimising Feature Vector Instances

In ubiquitous systems, controlling hardware devices in terms of their security features may be the correlated results where maximise with low variability in a stream. It is explained with example where feature vectors f_1, f_2 consists of integer such as $f_1 = \{1, 2, 3, 4, \ldots n\}$ and $f_2 = \{n, n-1, n-2, \ldots .3, 2, 1\}$. These example possess minimum variability, where f_1 has increasing number of streams and f_2 is opposite of f_1. These feature vectors are negatively correlated. Therefore, to transform our data stream into this order it requires a metric to universally decrease or increase from a set. It also become essential to maintain and regenerate the original stream. To solve this problem, our approach perform better results for reducing into a particular order as explained below.

In this approach, bit-wise XOR operation is performed as shown in Fig. 13.4. XOR operation is performed between variable X and threshold T, which provides result as X'. This values consist less variability as compared to X. The value X may be regain by performing XOR operation with a same threshold parameter. Here, T value has to be ranges closer to X for gaining smaller values of X' as observed from Figs. 13.5 and 13.6.

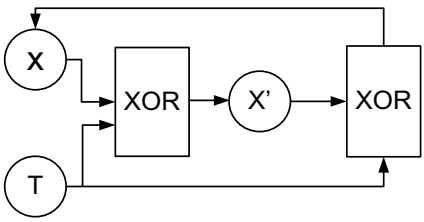


Fig. 13.4 Data reduction for less variability

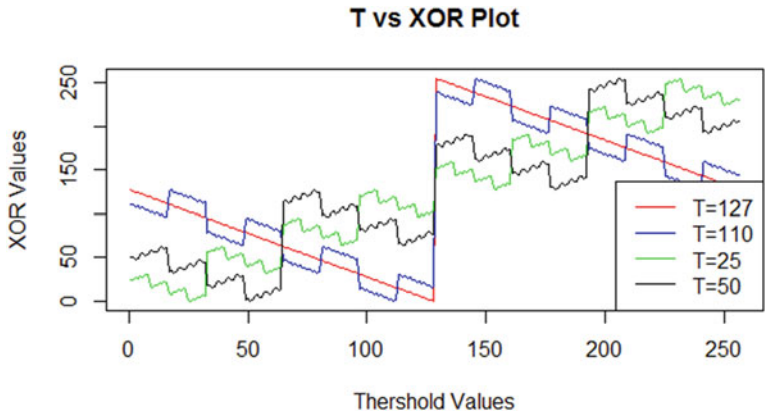


Fig. 13.5 Threshold values $(T) \leq 127$ versus XOR plot

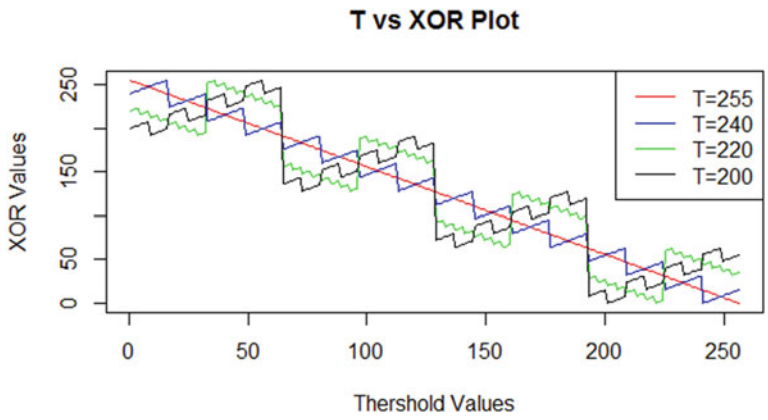


Fig. 13.6 Threshold values (T) ≤ 255 versus XOR plot

XOR experimentation is performed of data variability ranges from 0 to 250. A different T values is considered such as $T1 = \{25, 50, 110, 127\}$ as shown in Fig. 13.5. A graph is plot between and T and XOR values. Here, $T1 = \{127\}$ may produce least XOR values which lies to a mid-point of a data stream. Therefore, it become necessary to select a suitable T values.

In Fig. 13.6, different threshold values such as $T2 = \{200, 220, 240, 255\}$ is selected. The graph shows decrease in XOR values with the increase in T values. Here, the maximum values of T with respect to the highest attributes of stream may provide better results. These results infer about a suitable selection of T values may leads to smaller X'.

13.4 Experimental Set-Up and Implementation

All the experiment is performed on Intel core i3 CPU 4010U @1.7GHz. The machine has 4 GB of DDR3 RAM with a disk size of 1TB, running ubuntu16.04 LTS x86_64. Datasets generated by python 2.7.12 with gcc version 5.4.0, execution is performed on R programming language. We have used real and synthetic dataset.

The real-time streams is generated from tweeter API, a time-based top ten hashtag correlation is determined. An aggregated hashtag for a particular time stamp is shown in Table 13.1. Equation 13.8 helps for data pre-processing. These aggregated values are passes from XOR operation.

$$\text{Values} = \left\lceil \frac{\text{Local Count} \times \text{Weight}}{\text{Total Count}} \times 100 \right\rceil \tag{13.8}$$

Data stream from Twitter API is collected as shown in Table 13.1 after applying pre-processing operations. A flow of stream processing stages of this datasets is

Table 13.1 Hashtag versus time stamp values

Time stamp by hash tag counts comparison										
	Hash tag 1	Hash tag 2	Hash tag 3	Hash tag 4	Hash tag 5	Hash tag 6	Hash tag 7	Hash tag 8	Hash tag 9	Hash tag 10
T1	15	35	200	85	120	75	65	30	180	25
T2	50	37	250	90	145	85	85	55	210	45
T3	85	45	270	120	186	100	95	60	250	90
T4	90	55	290	140	210	105	120	85	290	135

depicted in Fig. 13.2. The experimentation is divided into two part transformation and visualisation. This tweets are bundled in a sheer window with more than 100 tuples. This tuples are correlated with each feature vectors and form binarised window. For visualisation, we used heatmaps and bar graph for different groups. The groups is calculated using decimal difference between two binarised windows. A correlation can be determined between various hashtags. These may provide similarities among different hashtag.

Multiple compared window is formed which provides values ranges from 0 to 65536. The most frequent count value is zero, which inference maximum associativity between two sheer windows. It is mapped on groups based on their total length. These type of visualisation is essential to infer about frequency count of a differenced decimal values. The visualisation results are depicted and discussed in Sect. 13.5.

13.5 Results and Limitations

The result is performed using 500 comparison between sheer windows. These sheer window passes through CRA algorithm using multiple correlation approach. This step provide a binarised windows. Here, we have used a binarised window of size 16 bit. This window provides maximum decimal number of 65536 to minimum of zero. These comparison provides hash mapped into categories belongs to zero to nine. These groups are numbered based on their face-value among its respective place-value on numeric size of binarised window. Final results are comprised with multiple hashmap table plots in form of bar graphs. An individual bar graph is plotted for different number of groups. These graphs indicates frequency from hash-mapped values to a particular group. These five different bar plot graph are shown in Fig. 13.7.

Here, group 1 consists large number of frequencies as its higher probable for similar consecutive binarised window selection. The remaining group 2–4 maps accordingly as group 1 as shown in Fig. 13.7 section (b), (c) and (d), whereas, in group 5, only consists hashmap from one to six. Some hashmap values may not be obtained in group 5 such as zero and more than six values. This bar plot is occurs due to the highest possible range for 16 bit is 65536, and subtraction from this may provide values which is always lower than 65536. Therefore, this group may provide inference about similarities between sheer windows. These results are used to draw a heatmap for further understanding as shown in Fig. 13.8.

In Fig. 13.8, heatmap is drawn with results obtained from hashmap of different groups, which is essential to identify number of possible clusters. Group 1 indicates least variability in unbounded stream of sheer windows. In Group 1, relative values is highly probable to be mapped in possible hash table location as 0, 5, 8, 7 and 3. This location will reflects the feature vectors of low significant position is affected. Here, group 2 are shows similar behaviour as group 3. Therefore, a hierarchical representation is drawn alongside with all the groups. This may be beneficial for identify the similarity and dissimilarity between sheer window. Also, the specific feature vectors

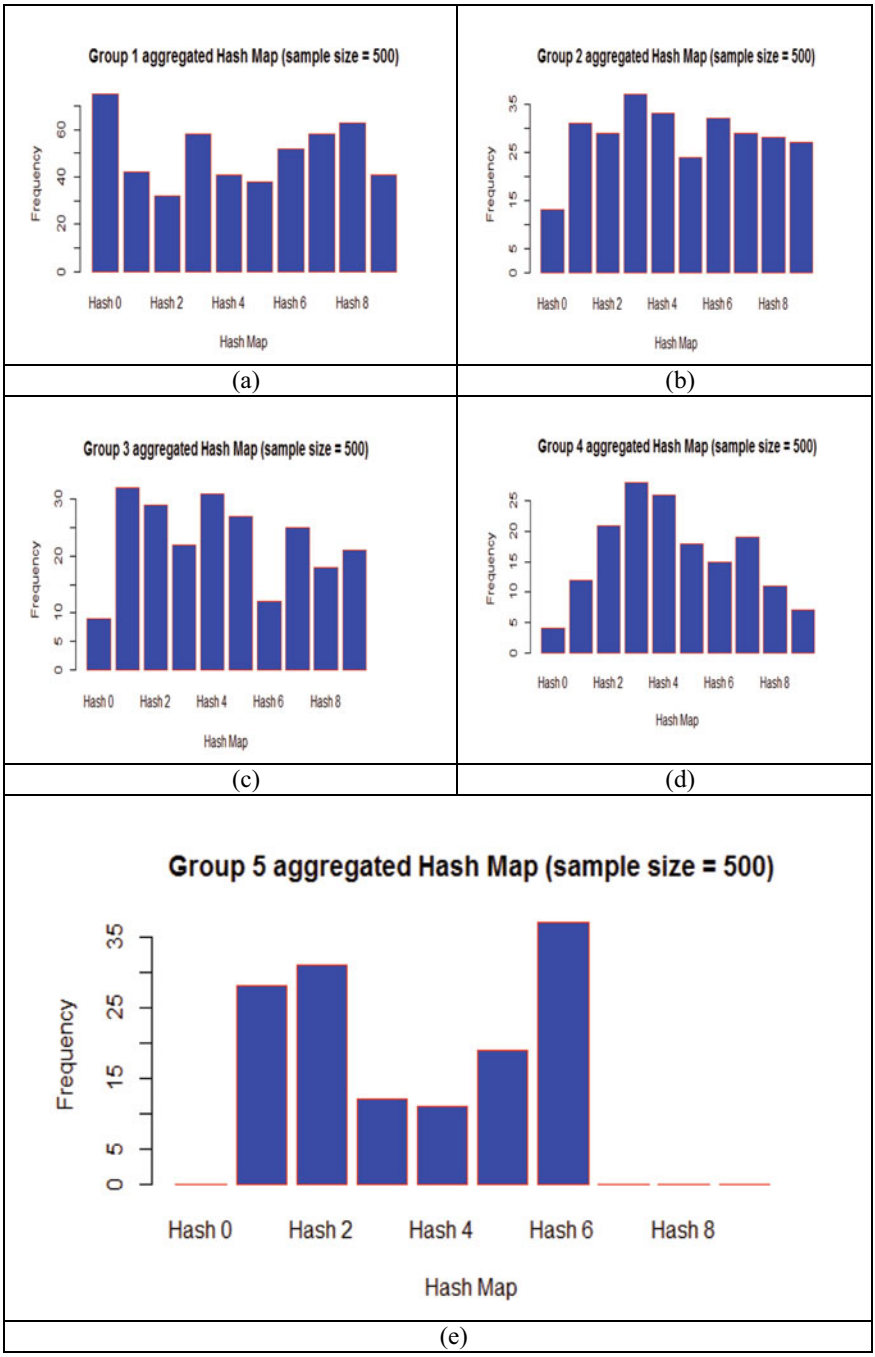


Fig. 13.7 Aggregated hashmap **a** Group 1, **b** Group 2, **c** Group 3, **d** Group 4, **e** Group 5

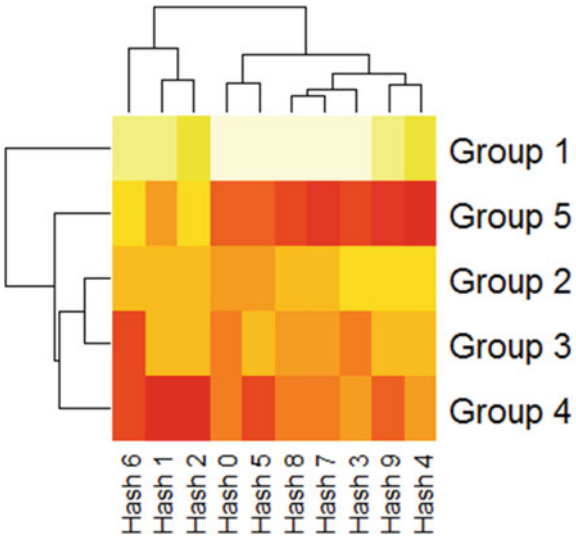


Fig. 13.8 Heatmap between Hashmap and different group categories

can be highlighted for contribution of dissimilarities. These observation highlights the level grouping frequency counts.

In this approach, some limitation are associated which has to be considered before applying in real-time application scenarios. Binarised approach uses CRA algorithm, which is an ordinal-based association where for any values ($a < b$) is consider as (0 or 1) and its scalar values may not effect in correlation.

13.6 Conclusion and Future Work

In this chapter, SHCUBA approach is proposed to determine online correlation between sheer windows. In addition to this, a CRA-based multiple correlation techniques is used for more than two feature vectors. A binarised algorithm is proposed for generating reduced sized window from an unbounded data streams. A hash-based classification techniques is proposed for forming different groups or clusters. The different groups are used to infer about nature of data stream for a particular time stamp. This approach provides time and space complexity as $O(n)$ and $O(1)$, respectively. In future work, this approach can be modified in a manner where it becomes compatible for different real-time data processing use cases.

Acknowledgements I offer my most sincere gratitude towards Council of Scientific and Industrial Research (CSIR), Government of India, for providing me research grant in the form of Senior Research Fellowship.

References

1. Maheswar R, Kanagachidambaresan GR (2020) Sustainable development through Internet of Things. Wireless Netw 26(4):2305–2306
2. Artikis A et al (2014) Heterogeneous stream processing and crowdsourcing for urban traffic management. In: EDBT, vol 14
3. Carvalho O, Roloff E, Navaux POA (2017) A distributed stream processing based architecture for IoT smart grids monitoring. In: Companion proceedings of the10th international conference on utility and cloud computing
4. Abdullah N, Alwesabi OA, Abdullah R (2018) Iot-based smart waste management system in a smart city. In: International conference of reliable information and communication technology. Springer, Cham
5. Cortés R et al (2015) Stream processing of healthcare sensor data: studying user traces to identify challenges from a big data perspective. Procedia Comput Sci 52:1004–1009
6. Biem A et al (2010) Ibm infosphere streams for scalable, real-time, intelligent transportation services. In: Proceedings of the 2010 ACM SIGMOD international conference on management of data
7. Dimitropoulos X, Hurley P, Kind A (2008) Probabilistic lossy counting: an efficient algorithm for finding heavy hitters. ACM SIGCOMM Comput Commun Rev 38(1):5–5
8. Cormode G, Muthukrishnan S (2005) An improved data stream summary: the count-min sketch and its applications. J Algorithms 55(1):58–75
9. Kumar A, Xu J, Wang J (2006) Space-code bloom filter for efficient per-flow traffic measurement. IEEE J Sel Areas Commun 24.12:2327–2339
10. Lal DK, Suman U (2020) SBASH stack based allocation of sheer window architecture for real time stream data processing. Int J Data Anal (IJDA) 1.1:1–21
11. Lal DK, Suman U (2021) An online approach to determine correlation between data streams. In: Companion of the ACM/SPEC international conference on performance engineering
12. Pang S, Ozawa S, Kasabov N (2005) Incremental linear discriminant analysis for classification of data streams. IEEE Trans Syst Man Cybern Part B (Cybernetics) 35.5: 905–914
13. Beringer J, Hüllermeier E (2006) Online clustering of parallel data streams. Data Knowl Eng 58(2):180–204
14. Mencagli G et al (2017) Harnessing sliding-window execution semantics for parallel stream processing. J Parallel Distrib Comput
15. Zhang P et al (2014) E-tree: an efficient indexing structure for ensemble models on data streams. IEEE Trans Knowl Data Eng 27.2:461–474
16. Masud M et al (2010) Classification and novel class detection in concept-drifting data streams under time constraints. IEEE Trans Knowl Data Eng 23.6:859–874
17. Mueen A, Nath S, Liu J (2010) Fast approximate correlation for massive time-series data. In: Proceedings of the 2010 ACM SIGMOD international conference on management of data
18. Zhang T et al (2009) Adaptive correlation analysis in stream time series with sliding windows. Comput Math Appl 57.6:937–948
19. Cormode G, Muthukrishnan S (2005) Summarizing and mining skewed data streams. In: Proceedings of the 2005 SIAM international conference on data mining. Society for industrial and applied mathematics
20. Nelson J (2012) Sketching and streaming algorithms for processing massive data. XRDS Crossroads ACM Mag Stud 19.1:14–19
21. Yang T et al (2017) SF-sketch: a two-stage sketch for data streams. arXiv:1701.04148
22. Palma-Mendoza R-J et al (2019) Distributed correlation-based feature selection in spark. Inf Sci 496:287–299
23. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3:1157–1182
24. Amen B, Grigoris A (2018) A theoretical study of anomaly detection in big data distributed static and stream analytics. In: 2018 IEEE 20th international conference on high performance computing and communications; IEEE 16th international conference on smart city; IEEE 4th international conference on data science and systems (HPCC/SmartCity/DSS). IEEE

25. Zhang Y et al (2017) Improved visual correlation analysis for multidimensional data. J Vis Lang Comput 41:121–132
26. Hauke J, Kossowski T (2011) Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Quaest Geograph 30(2):87–93
27. Zar JH (2005) Spearman rank correlation. Encycl Biostat 7

Chapter 14
Decision Analytics Using Predictive and Prescriptive Analyses of Student's Satisfaction Towards Quality of Education for Sustainable Society in Oman

S. L. Gupta, Niraj Mishra, Pooja Kansra, and Mohammad Yawar

Abstract Quality of education is essential for the teachers and the learners, and it is also crucial for the analyst. The process of maintaining high education quality is critical and challenging too. Education is a very compound and complicated system which has a cause and effect relationship that is nonlinear. Thereby, present chapter explores the satisfaction of the students towards the higher educational institutions in Oman for sustainable society. The study was based on a primary data and a structured questionnaire has been designed to collect the data. The measurement model has been tested to measure the internal reliability. All the constructs have indicated high internal reliability as Cronbach's alpha, and composite reliability values have been higher than 0.70. When all the conditions for the measurement model have been met, structural equation modeling is conducted. The structural model has been analyzed by way of bootstrapping methodology. The consistent PLS bootstrapping found that certification most significantly impacts satisfaction followed by website satisfaction. The analysis of the bootstrapping exhibited that certification is the most significant variable which influence the satisfaction of the students followed by the website design. Along with the website and certification, it was suggested that the educational institutions need to concentrate on the placement policies which can allow the students to be placed at international companies. Educational institutions can look forward to international standards to examine themselves and improve their standards. The role of accreditation agencies is vital in ensuring that the quality

S. L. Gupta
Birla Institute of Technology, Noida, India

N. Mishra
Birla Institute of Technology, Ranchi, India

P. Kansra (✉)
Mittal School of Business, Lovely Professional University, Punjab, India
e-mail: pooja.kansra@lpu.co.in

M. Yawar
Waljat College of Applied Sciences, Muscat, Sultanate of Oman

V. Bali et al. (eds.), *Decision Analytics for Sustainable Development in Smart Society 5.0*, Asset Analytics, https://doi.org/10.1007/978-981-19-1689-2_14

is fulfilled and they maintain the same quality with regular interval checks on the accredited educational institutions.

Keywords Education · Quality · Satisfaction · Students · Institutions · Certification · Website · Bootstrapping

14.1 Introduction

Education is increasingly being practiced in an online learning environment as the field of education extends the use of IT technology in online learning [3]. As a result, researchers and developers have started to use decision analytics using predictive and prescriptive analysis of student's satisfaction towards quality of education in Oman and, ultimately, enhance educational services. The goal of decision analytics is to use predictive and prescriptive analysis and give a novel prediction to facilitate the teachers and institutions to customize educational opportunities [2].

Students' satisfaction is one of higher education's main performance indicators (KPI) since overall satisfaction appears to lead to higher loyalty, which in turn leads to higher student performance from a variety of perspectives [22]. Nowadays, the growing variety of data is regarded as a contentious issue for research [12]. Student satisfaction (SS) is a significant factor in assessing the originality and accuracy of an educational curriculum in a higher learning institution. A higher satisfaction of students will lead to greater ability to improve their skill growth, course knowledge, and attitude [17].

Reforms in education of Oman were introduced to make Oman an economically established country that can use its resources to develop Oman. Education contributes significantly to the country's economic growth [46]. The educational reforms were made mainly to focus on making the education sector strong and independent and make the students get higher education in Oman itself. It also aimed to inculcate vocational training among the students. The educational reforms' application helps in the development of the level of education and the geographical aspect of the growth of education. The Sultanate of Oman is full of human and financial resources, and these reforms were made to utilize both these resources [21]. But these resources are underutilized. Therefore, the Government of Oman has introduced such reforms that can help and utilize the available resources. They provide full support to education development and provide financial and moral support [7]. The effect of these reforms can be measured when the students who complete the basic education will gain higher education. After that, when they enter the labor market, it will help get the appropriate result of various practices followed by the department of education in Oman. Some definitions of interaction focus on the social intent and processes of interaction, particularly when it comes to student–student (SS) and student–teacher (ST) interactions [14].

According to many researchers, the real impact can only be judged when the practical implementation is determined mainly by the teachers' quality [37]. The criteria

of opting for higher education depend upon the student's academic performance and decision analytics using predictive and prescriptive analysis for their satisfaction towards quality education in secondary education and the availability of seats in the higher education institutions in Oman. Various governing bodies are formed to settle down all these issues, such as the Council of Higher Education and the individual institutions, which regularly revise the system and upgrade it [21]. The institutional structure of the education system in Oman is mainly divided into two levels. The first level is represented by the Council of Higher Education, which focuses on policy-making. The second level is the implementation of the policies governed by different ministries and institutions [9].

Following significant shifts in the world economy's operating rules and market competition trends in the 1990s, businesses recognized the need for economic globalization [39]. In different areas, vast swaths of data are readily accessible and education is no exception. The impetus to extract concrete perspectives from knowledge in tandem gains urgency. A flurry of research has resulted from recent advances in deep learning, particularly in the field of voice and image recognition and so-called complete information games such as chess, go, and StarCraft. We investigate the usefulness and applicability of deep learning for educational data mining and learning analytics using two educational datasets [19].

The variables which influence the retention in undergraduate engineering includes satisfaction of students with the majors, intentions to persist, subject-taking habits, decisions to pursue jobs in the given field and real persistence to degree completion (elitzler uw edu Litzler & asu edu [20]). When it comes to their research plans, students in higher education must make a lot of choices [18]. Oman is highly aligned towards expanding the education system and maintains the quality of education in its higher education institutions. The expansion brings many pros and cons with it. If a large number of institutions are opened, it becomes equally important to maintain education quality. It is seen that a lot of malpractices started to happen in higher education as well, due to competition in these institutions [5].

According to a report on Human Development in 2015, Oman holds 52nd rank in the world. Oman is at the cutting edge in developing its human resources to develop in economic terms. To fulfill these needs, the government is investing a lot in the education sector. This aims at providing education to the students as resources are scarce in the Gulf region. Education in Oman started flourishing in the 1990s, but the whole process was initialized only a few years ago. Oman is a highly dependent country upon its oil resources, which is a source of energy, but it is nonrenewable. Therefore, the government is aiming to invest in the country's youth's education, not to face any kind of economic crisis. The educated youth will stabilize the country's resources, and they will also help make the country more independent and economically stronger. The government has also opened doors for private institutions [6].

Learning analytics is a new field that looks at the assessment, compilation, interpretation, and recording of information about the students' performance and their surroundings [40]. Learning analytics, helps to analyze the socio-technical data and widely applied in the educational contexts to improve the student's performance in the higher education institutions by collecting and analyzing data from students,

learning processes, and learning environments in order to obtain meaningful input and scaffolds when appropriate [26]. A broader policy is framed to provide higher education to the students and make it easily accessible to the country's citizens. The private institutions also help provide a world-class education to the students as they are affiliated with foreign institutions. The students are made ready to meet the competition in the global market. The private institutions were working in collaboration with foreign partners. Hence, the quality, which was desired by the higher education institution in the Sultanate of Oman, was maintained. There are 27 higher education institutions in the country. The private higher education institutions were needed to be affiliated with international universities [23]. This is deemed to be the first step towards making the higher education institutes of Oman internationalized and privatized. These institutes were monitored by the various departments like the Ministry of Higher Education, the Ministry of Defense, the Ministry of Commerce and Industry, and the Ministry of Manpower, apart from the private higher education institutions themselves. We propose a more meaningful descriptive approach to LAD that represents pedagogical intent, as opposed to predictive dashboards that build organizational black boxes [47].

Due to their early exposure to new technology, the students are often believed to be technologically fluent as "digital natives" [29]. Many emphases were laid on the institutional accreditation in order to maintain the quality of education in the higher institutions in the Sultanate of Oman and hence improve on decision analytics using predictive and prescriptive analysis of student's satisfaction towards quality of education in Oman. The development of the accreditation process and other licensing procedures were monitored by the Ministry of Higher Education (MoHE) in Oman. To assist the Omani higher education, a non-profit organization was set up in 2006 for sharing good practices and information in various aspects of quality assurance. The Education Council (EC) was set up to ensure that the citizens of Oman get a world-class quality of higher education and, at the same time, maintain the international standards of education. The public and private institutions' working was monitored by the Oman Academic Accreditation Authority (OAAA). This is the chief accreditation authority in the country. A quality audit is done to maintain the quality of education in the higher education Institution. In this audit, the external auditors and reviewers' recommendation is considered to meet the quality standards and maintain the same. A lot of education is delivered through offshore programs, which are planned and executed on the guidelines made by the foreign institutions. Therefore, it was very important to take measures to ensure that the standard of education quality is maintained in off-campus institutions so that the students are benefitted [10].

Learning behaviors are influenced by motivation, which is a dynamic and powerful factor. Many educational and instructional design studies have been conducted with the goal of identifying successful approaches that can boost and sustain learning motivation [25]. As online higher education programs grew in popularity, they created a complex tension that sparked ambivalence in some higher education sectors [36]. High quality must be insured in carrying out the academic programs. The international standards must be met right from deciding the curriculum to teach the students

to learn and utilize their study plans and assignments [1]. The Oman Academic Accreditation Authority (OAAA) has mentioned that the private higher education institution that operates in Oman for the quality of education and student's satisfaction in predictive and prescriptive analysis for education. The university of its affiliation must meet the standards in imparting education to its students. In some instances, the Oman Academic Accreditation Authority (OAAA) found a discrepancy in meeting the education standards. For quoting an example, an institution offering a joint degree, the university lacked collaboration to provide joint degree by two different universities. They were not able to standardize their academic programs. Another report of the Oman Academic Accreditation Authority (OAAA) stated that the foreign universities' teachers must also teach in Oman's institutions to benefit by getting education through the teachers who teach in foreign institutions. It was also reported that the employees' roles and responsibilities must be classified and clearly delineated so that there is no difference between the set standards and the delivered standard [11].

Both the private university and the local private institute are partners. Their joint responsibility is to look after the matters related to imparting education's standard pattern [8]. They must also provide online learning materials, and they must also incorporate online libraries so that the students can consult the same as and when required. This is also essential for the teachers who teach in the local private institutions. By consulting the foreign materials, they can understand the level of education which is imparted in the foreign institutions. This helps them not only to teach the students but also in grooming themselves and their capacities. The foreign private institution's role is equally important in providing a research platform to the students to meet the global learning standards and make them equally competitive enough to crack the competition and hold higher job positions.

Only by improving product quality, manufacturing processes, material requirements, marketing quality policy, quality system, and human resource management can quality be improved [35]. It can be said that the local institution and the foreign private institution must work together in maintaining the institutional infrastructure. Both these institutions must work together to promote research work as these are in high demand for delivering a qualified youth for the country. The role and significance of various factors in predicting service quality is a highly subjective phenomenon [45]. A research-oriented curriculum is essential for the teaching staff and the students as this helps generate skilled and trained individuals. The students and the teachers also benefit from various workshops, seminars, and conferences conducted by foreign institutions. They help upgrade the quality standards and increase the institution's professional capacity and development. Therefore, it is very important that the affiliated university incorporates the recommendations of the Oman Academic Accreditation Authority (OAAA) to give quality education to its students and maintain the standards of the education system. This will lead to the improvement of the quality assurance and education system. Many organizations have adopted total quality management (TQM) practices as a result of intense domestic and foreign competition, heightened quality consciousness, and the need to reduce operating costs [15].

Individual college achievement has been meticulously associated with the dropout and graduation rates, and these completion indices have long been regarded as the most imperative indicators of the achievement of the undergraduate [13]. Scholars from different backgrounds have always been interested in how people and organizations make decisions. Individuals and organizations can only deliberately influence the results that are important to them by making choices [43]. Education and research are critical components of a country's overall growth. "Quality education" is recognized as the pillar for the sustainable, stable, and prosperous future in many countries in the globe [42]. Quality assurance plays a very important role in maintaining the quality of education in higher education institutions. The institutions in the higher education needs to play a very active in order to maintain and raise the quality of education so that the students and the teachers can benefit from grooming themselves (Nasser 2019). To check the quality of education in the HEIs, the Omani institutions have introduced institutional accreditation [49]. They also help the higher education institutions to formulate policies and procedures so that the quality of education can be standardized. This also helps in meeting out the set criteria for the higher education institutes in the country. The quality audit also checks the higher education institution's quality assurance based on its set goals and the determined objectives against the Oman Academic Accreditation Authority quality standards.

Teaching Analytics (TA) is a modern theoretical methodology that incorporates teaching experience, visual analytics, and design-based analysis to help teachers develop their diagnostic pedagogical ability to use data and evidence to improve the quality of their instruction [38]. Such participation necessitates a long-term dedication to lifelong learning that goes beyond a basic level of proficiency [27].

14.1.1 Decision Analytics

Decision analysis (DA) is a type of decision-making that entails identifying and evaluating all aspects of a decision, as well as taking actions based on the decision that yields the best result [16]. The goal of DA is to ensure that all relevant information and options are considered when making decisions. The foundations of DA, as a form of decision-making, can be used to solve a wide range of problems, from complicated business issues to basic daily problems [4].

14.1.2 Descriptive Modeling

Descriptive modeling can be applied to find and analyze trends in raw data principally through the data aggregation and data mining techniques, in order to provide historical insights. It poses the question, "What happened?" by accurately presenting relationships between input data and target variables [41]. It can be applied to trace out the behavior pattern of the high academic achievement students. It can be used

to analyze the clustering of SNS-usage and class attendance patterns of the drop out students in order to generate descriptive modeling applications in data-driven student characterization [44, 48].

14.1.3 Predictive Modeling

Predictive modeling allows assumptions about the probability of the future outcomes of the given target variables based on present and past input dataset [30]. It attempts to address the question as: What might actually occur? Predictive models may use trends discovered in previous data to predict potential threats and opportunities. Repeated descriptive modeling by applying the input data and target variables can improve the descriptive model, allowing it to predict target variables solely on the basis of input variables. Predictive modeling, in general, necessitates a large number of input and goal variables. Predicting modeling can be used to estimate the probability of the dropout students by using the latter unstructured and structured datasets of input data are two examples of predictive modeling [24]. Predictive analytics analyzes existing data and makes predictions about the future using a variety of methods which includes data mining, statistics, modeling, machine learning, and artificial intelligence [28].

14.1.4 Prescriptive Modeling

Prescriptive modeling can be applied to forecasting the future by recommending decisions that can be taken to benefit from predictions and displaying the consequences of various policy choices to decision makers [33]. Prescriptive analytics not only helps to predicts what can happen, but it can go beyond it as how those occurrences can be brought on. Formulation of a set of choices for effectively reducing dropout probabilities, and commitment of constructive steps for enhanced academic success are two examples of prescriptive modeling. Since predictive modeling necessitates a large set of data, it should be focused on highly accurate descriptive and predictive models. As a result, prescriptive modeling will be beyond the reach of the current study [31, 34].

14.2 Objectives of the Study

The higher education institutions have to be very active in maintaining and raising the quality of education. The students must gain a high standard of academics. The quality standard can be maintained through regular quality audits of the accreditation documentation and assessment of education quality through students' interviews. Keeping the same in mind, the present chapter examines decision analytics

using predictive and prescriptive analysis of student's satisfaction towards quality of education in Oman. The method applied to evaluate, analyze, and interpret the large amounts of data into a meaningful form is known as analytics [32].

14.3 Materials and Methods

14.3.1 Sampling

The study was based on descriptive research. In the present study, the population consists of the students of higher education institutions of Oman. A sample of 157 students has been selected based on the simple random sampling, where the sampling frame was fixed with the list of students.

14.3.2 Data Collection

The data has been obtained from the students with the help of a structured and pre-tested questionnaire. The views of the students play a vital role to understand the educational institutions' existing position, which made the questionnaire structure into two parts: internal and external factors. The questionnaire was subjected to Cronbach's alpha testing for the reliability of the data collection instrument and it was found to be satisfactory for carrying the data collection process with the built questionnaire. The measurement model has been tested to measure internal reliability. All the constructs have indicated high internal reliability as Cronbach's alpha, and composite reliability values have been higher than 0.70. The variables chosen for the present study has been described in appendix.

14.3.3 Statistical Analysis

The analysis of data has been made with the help of the structural equation modeling. When all the conditions for the measurement model have been met, then structural equation modeling is conducted. The structural model has been analyzed by way of bootstrapping methodology. Along with structural modeling descriptive statistics were also used when needed. All the data has been checked for normality, and possible outliers have been removed. The structural equation modeling has been done with the help of the R-Program.

14.4 Results and Discussion

This section exhibited the opinions of the students by applying structural equation modeling. The structural equation model was built for examining the quality of education. The measurement model has been tested to measure the internal reliability. The results of the reliability and validity of the constructs have been presented in Table 14.1. All the construct items where loadings were greater than the threshold value of 0.708 as recommended by Hair, Hult, Ringle, & Sarstedt (2013) were included. All the constructs have indicated high internal reliability as Cronbach's alpha, and Composite Reliability (CR) values have been higher than 0.70 (Sarstedt, Ringle, & Hair, 2017). When all the conditions for the measurement model have been met then structural equation modeling can be applied to attain the desired results. A good measurement model helps in removing the bias and ensures the correctness of the results.

The structural model has been analyzed by way of bootstrapping methodology. The results of the bootstrapping with 1000 samples shown in Table 14.2 of the following section titled results of consistent PLS reveal which of the structural model relationships have been significant. The model shows the impact of all the constructs on student satisfaction. The consistent PLS bootstrapping found that certification most significantly impacts satisfaction, followed by the website. The R^2 focuses on in-sample predictive power. All the independent variable has been able to explain 73.04% variation in the dependent variable, i.e., satisfaction. The representation has been done by of model diagram in Fig. 14.1.

It has been observed that only two variables have been significant; however, the rest of the variables have been insignificant. It signifies that students have been more concerned about the certification options provided and usage of the institute's website. Hence, policymakers should focus more on these two indicators. However, overall, the model prediction power has been moderate; thus, it can be concluded that other factors also contributes a critical role in the students' satisfaction in general.

14.5 Conclusion

In the present study, an attempt has been made to identify the factors influencing satisfaction of students towards educational institutions. The analysis of the bootstrapping exhibited that certification is the most significant variable which influence the satisfaction of the students followed by the website design. It has been observed that all the independent variable has been able to explain 73.04% of the total variation in the dependent variable i.e., satisfaction. The study gives insights to the policymakers and the educational institutions to work on these parameters in order to meet the expectations of the students using predictive and prescriptive analyses

Table 14.1 Reliability, Validity, and Factor loadings

Construct	Items	Loadings	Cronbach's alpha (α)	Composite reliability (CR)	Average variance extracted (AVE)
Availability of faculty	Q14.1	0.779	0.911	0.931	0.694
	Q14.2	0.821			
	Q14.3	0.878			
	Q14.4	0.796			
	Q14.5	0.894			
	Q14.6	0.824			
Career growth	Q17.1	0.709	0.711	0.710	0.545
	Q17.2	0.872			
Certification	Q19.1	0.849	0.721	0.868	0.767
	Q19.2	0.901			
Classroom management	Q13.1	0.785	0.756	0.832	0.622
	Q13.2	0.763			
	Q13.3	0.817			
College infrastructure	Q9.1	0.744	0.887	0.909	0.559
	Q9.2	0.799			
	Q9.3	0.742			
	Q9.4	0.710			
	Q9.5	0.753			
	Q9.6	0.786			
	Q9.7	0.755			
	Q9.8	0.741			
Exhibitions	Q15.1	0.937	0.852	0.931	0.871
	Q15.2	0.929			
Extracurricular activities	Q3.1	0.837	0.824	0.879	0.646
	Q3.2	0.758			
	Q3.3	0.781			
	Q3.4	0.834			
Hostel facilities	Q11.1	0.849	0.857	0.897	0.637
	Q11.2	0.815			
	Q11.3	0.809			
	Q11.4	0.825			
	Q11.5	0.682			
Location	Q1.1	0.709	0.769	0.772	0.421
	Q1.2	0.726			
	Q1.3	0.718			

(continued)

Table 14.1 (continued)

Construct	Items	Loadings	Cronbach's alpha (α)	Composite reliability (CR)	Average variance extracted (AVE)
	Q1.4	0.742			
	Q1.5	0.817			
Practical orientation	Q4.1	0.891	0.739	0.731	0.584
	Q4.2	0.720			
Satisfaction	Q22.1	0.818	0.765	0.864	0.680
	Q22.2	0.809			
	Q22.3	0.845			
Website	Q16.1	0.881	0.787	0.867	0.766
	Q16.2	0.869			

Source Primary Data

Table 14.2 Results of consistent PLS bootstrapping on satisfaction

Independent variables	Path coefficients	t-statistics
Availability of faculty	0.119	0.799257
Career growth	0.034	0.373208
Certification	*0.371*	*3.641192**
Classroom management	–0.060	0.398977
College infrastructure	−0.039	0.306966
Exhibitions	0.146	1.001508
Extracurricular activities	−0.043	0.469905
Hostel facilities	0.110	0.887285
Location	−0.105	1.267907
Practical orientation	0.158	1.796064
Website	*0.321*	*2.795908**

Source Primary Data
Note * indicate significance at 5% level of significance

for student's satisfaction in quality of education in Oman. The students are restricted from the certificate options and usage of the institute's website, which is proving to be a bottleneck for the students in developing their careers. The website's usage has to be made unrestricted to make them have access to the required study-oriented aspects that help to improve the students' career perspectives. The educational institutions have to modify their policy towards the students' certificate options and improve the level of completion using predictive and prescriptive analyses. The research study has committed itself to bring out a real state of existing affairs of the quality of education in Oman. The stakeholders have portrayed a real picture of the educational institutions operating in the country. The perspectives of the stakeholders reveal that the educational institutions operating in the country are having a significant role in

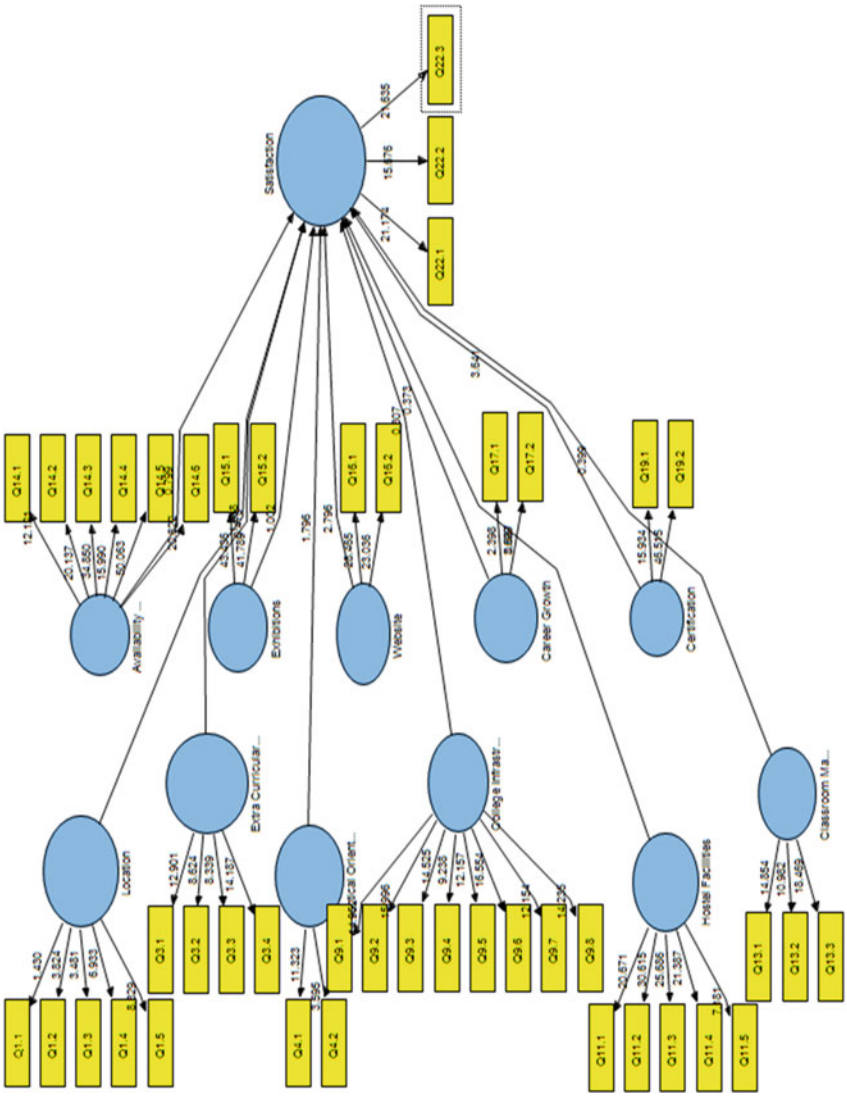


Fig. 14.1 Bootstrapping Model depiction

maintaining and enhancing the quality of education imparted. Educational institutions can look forward to international standards to examine themselves and improve their standards. The accreditation agencies of OAAA are having the mentoring and quality control role in examining the quality of education from the stakeholders' perspective to have insights into the real picture. The learning techniques and probabilistic models have highlighted the various areas that prove to be challenges for educational institutions in providing quality education. The educational institutions and accreditation agencies' policies have to give special importance to these areas suggested by the probabilistic model to uphold the existing standards and build an integrated quality education system for the students of Oman.

14.6 Future Research

The present study provides the following insights into the future research:

i. Gradual progress from descriptive to predictive to prescriptive modeling, i.e., we start with persuasive and accurate descriptive models, then focus on improving predictive modeling and analyzing the quality of education in Oman. We will move on to prescriptive modeling once we have solid results from descriptive and predictive modeling.
ii. Use of recurring data acquisition to the fullest extent possible. The same student datasets can be taken into account, and data collection can be simplified. The extensive use of repeated datasets for the variables would aid in making the best possible use of the data for future analysis.
iii. Different level of segmentation, which is the research's other strategy for completely exploiting different levels of segmentation, and we assume that a certain level of segmentation will improve the modeling efficiency.

14.7 Policy Implications

Despite of the above-mentioned parameters, it has been observed that the educational institutions have to play a very active role for the sustainability. The following exhibits some of the major recommendations from the stakeholders' perspectives, which might help revamp the educational standards and focus on improving the key areas that upgrade the country's quality of education:

(i) The complexity in the program offered creates a sense of disbelief among the students to complete the education. These have to be minimized by creating a systematic plan by the teachers, enabling the students to understand those complexities paving the way for completing the program.

(ii) Educational institutions can look forward to international standards to examine themselves and improve their standards. The accreditation agency of OAAA plays the mentoring and quality control role in examining the quality of education from the stakeholder's perspective to have insights into the real picture. The role of accreditation agencies is vital in ensuring that the quality is fulfilled, and they maintain the same quality with regular interval checks on the accredited educational institutions.
(iii) The educational institutions' strategies and quality accreditation strategies have to focus on improving the international collaboration of educational institutions in the country, which brightens the prospects for students.
(iv) The international placements of the educational institutions have to be strengthened by creating viable understanding among the international companies looking out for talents from the country. The educational institutions have to concentrate on placement policies that allow the students to be placed at international companies.
(v) The international placements of the educational institutions have to be strengthened by creating viable understanding among the international companies looking out for talents from the country. The educational institutions have to concentrate on placement policies that allow the students to be placed at international companies.

14.8 Limitations of the Study

(i) The study was confined to some selected variables and inclusion of more variables can provide more appropriate results.
(ii) An increase in the sample size can give better insights into the researchers and the policymakers.
(iii) An increased sample size can lead to improved generalizations for effective policymaking.

Appendix I: Description of Variables

Construct	Items
Availability of faculty	14.1 = How much satisfied you are with faculty's help in developing your own values and ethical standards?
	14.2 = How much satisfied you are with faculty's help in developing your abilities, interests, and personality?
	14.3 = How much satisfied you are with faculty's listening and caring abilities?
	14.4 = How much satisfied you are with time allocated by faculties for students?
	14.5 = How much satisfied you are with faculties' efforts in encouraging students to express their feelings?
	14.6 = How much satisfied you are with guidance of an academic advisor in career development and personal development?
Career growth	Q17.1 = How much satisfied you are with management's provision of career plans?
	Q17.2 = How satisfied are you with the adequate background and specialization that you have acquired for further education in professional, scientific, or scholarly field?
Certification	Q19.1 = How satisfied are you with your college degree being internationally valid?
	Q19.2 = How satisfied are with your knowledge regarding your field with respect to international standards?
Classroom management	Q13.1 = How often does your institution provide the extra practice hours for the students who are weak in studies?
	Q13.2 = How are often are the teachers strict in handling the cellphone-related problems in their classrooms?
	Q13.3 = How often does a teacher's fitness level contribute to effective classroom management?
College infrastructure	9.1 = Your institution uses electronic medium to discuss or complete the assignment
	9.2 = Your institution uses advanced lab equipment's for practical training
	9.3 = Your institution provides necessary resources in college library

(continued)

(continued)

Construct	Items
	9.4 = Your institution provides recreational facilities(pools, fitness equipment's, courts)
	9.5 = The cleanliness of campus is acknowledged
	9.6 = The use of various campus facilities
	9.7 = How often does your institution undergo the fire safety test?
	9.8 = How often does your institution have any committee in order to tackle any anxious moments among students/staff?
Exhibitions	Q15.1 = How much satisfied you are with the exhibitions conducted by your institution?
	Q15.2 = How much satisfied you are with the anticipating funds to encourage the participation of students in the exhibition?
Extracurricular activities	Q3.1 = How often does the management give importance to physical activities?
	Q3.2 = How often does your institution conduct social or cultural events?
	Q3.3 = How often does your management help in implementing your innovative ideas?
	Q3.4 = How often do you visit your library?
Hostel facilities?	Q11.1 = How much satisfied are you with the hostel facilities of the institution?
	Q11.2 = How satisfied are you with the quality of food provided in your institution?
	Q11.3 = How satisfied are you with the hostel environment?

(continued)

(continued)

Construct	Items
	Q11.4 = How satisfied are you with the overall rules and regulations in hostels?
	Q11.5 = How satisfied are you with the security measures in hostels
Location	Q1.1 = Easy access to transportation from institute
	Q1.2 = Institute's comfortable surroundings
	Q1.3 = Institute's transportation facility
	Q1.4 = Institute's parking area
	Q1.5 = Locality of your institute to promote your learning ability
Practical orientation	Q4.1 = How often does your department indulge in conducting seminars to motivate the idea of your course?
	Q4.2 = How often do your college instructors use computer technologies in your course?
Satisfaction	Q22.1 = How satisfied are you with your institution as a place to learn and survive?
	Q22.2 = How satisfied are you with your parent's support in your achievement?
	Q22.3 = How satisfied are you with your degree of success in implementing the institutions curriculum?
Website	Q16.1 = How often do you access your institute's official website that gives full accessibility to all internal and external aspects of the institution?
	Q16.2 = How often is the official website updated?

References

1. Al Abri R, Al Marzouqi F, Kuvarega AT, Meetani MA, Al Kindy SMZ, Karthikeyan S, Kim Y, Selvaraj R (2019) Nanostructured cerium-doped ZnO for photocatalytic degradation of pharmaceuticals in aqueous solution. J Photochem Photobiol Chem 384:112065
2. Abu-Naser SS (2012) Predicting learners performance using artificial neural networks in linear programming intelligent tutoring system
3. Aggarwal CC (2018) Neural networks and deep learning. Springer
4. Agrawal RS, Pandya MH (2015) Survey of papers for data mining with neural networks to predict the student's academic achievements. Int J Comput Sci Trends Technol (IJCST) 3(I5)
5. Akareem HS, Hossain SS (2016) Determinants of education quality: what makes students' perception different? Open Rev Edu Res 3(1):52–67
6. Al-Barwani T, Chapman DW, Ameen H (2009) Strategic brain drain: implications for higher education in Oman. High Educ Pol 22(4):415–432. https://doi.org/10.1057/hep.2009.1
7. Al-Hajri S (2008) The adoption of e-banking: the case of Omani banks. Int Rev Bus Res Papers 4(5):120–128
8. Al-Qaysi N, Mohamad-Nordin N, Al-Emran M (2020) What leads to social learning? Students' attitudes towards using social media applications in Omani higher education. Edu Inf Technol 25(3):2157–74. https://doi.org/10.1007/s10639-019-10074-6. Accessed August 10, 2021
9. Al-Amri, AS, Mathew P, Zubairi YZ, Jani R (2020) Optimal standards to measure the quality of higher education institutions in Oman: stakeholders' perception. SAGE Open 10(3). https://doi.org/10.1177/2158244020947440. Accessed August 10, 2021
10. Al-Maamari SN ((2020)) Developing global citizenship through geography education in Oman: exploring the perceptions of in-service teachers. Int Res Geograph Environ Edu 1–16
11. Al-Nabhani HZA (1996) Factors influencing the academic performance of undergraduate students of Sultan Qaboos University (Oman) learning in English, with special reference to attitude. Mot Acad Self Concept. http://hydra.hull.ac.uk/resources/hull:3938. Accessed August 10, 2021
12. Banumathi S, Aloysius A (2017) Predictive analytics concepts in big data-a survey. Int J Adv Res Comput Sci 8(8):27–30
13. Berea A, Tsvetovat M, Daun-Barnett N, Greenwald M, Cox E (2015) A new multi-dimensional conceptualization of individual achievement in college. Decis Anal 2(1). https://doi.org/10.1186/s40165-015-0012-8. Accessed August 10, 2020
14. Bernard RM, Abrami PC, Borokhovski E, Wade CA, Tamim RM, Surkes MA, Bethel EC (2009) A meta-analysis of three types of interaction treatments in distance education. Rev Edu Res 79(3):1243–1289. https://doi.org/10.3102/0034654309333844. Accessed August 10, 2020
15. Bouranta N, Psomas E, Suárez-Barraza MF, Jaca C (2019) The key factors of total quality management in the service sector: a cross-cultural study. Benchmarking 26(3):893–921. https://doi.org/10.1108/BIJ-09-2017-0240. Accessed August 12, 2020
16. Carlsson C (2018) Decision analytics-key to digitalisation. Inf Sci 460:424–438
17. Chin Wei C, Sri Ramalu S (2011) Students Satisfaction towards the University: does service quality matters?. Int J Edu 3(2):1–15. https://doi.org/10.5296/ije.v3i2.1065. Accessed August 12, 2020
18. De Laet T, Millecamp M, Ortiz-Rojas M, Jimenez A, Maya R, Verbert K (2020) Adoption and impact of a learning analytics dashboard supporting the advisor—student dialogue in a higher education institute in Latin America. B J Edu Technol 51(4):1002–1018. https://doi.org/10.1111/bjet.12962. Accessed August 12, 2020
19. Doleck T, Lemay DJ, Basnet RB, Bazelais P (2020) Predictive analytics in education: a comparison of deep learning frameworks. Edu Inf Technol 25(3):1951–63. https://doi.org/10.1007/s10639-019-10068-4. Accessed August 12, 2020

20. Elitzler uw edu Litzler E, Asu edu Young JJY (2012) Understanding the risk of attrition in undergraduate engineering: results from the project to assess climate in engineering. J Eng Edu 101(2):319–345. http://www.library.umaine.edu/auth/EZProxy/test/authej.asp?url=http://search.ebscohost.com/login.aspx?direct=true&db=eft&AN=75042214&site=ehost-live. Accessed August 11, 2020
21. Emam MM, Hendawy Al-Mahdy YF (2020) Building school capacity for inclusive education in the sultanate of Oman: a construct validation of the inclusive school climate scale. Leadersh Policy Schools 1–16. https://doi.org/10.1080/15700763.2020.1770803. Accessed August 11, 2020
22. Heo J, Lim H, Yun SB, Ju S, Park S, Lee R (2019) Descriptive and predictive modeling of student achievement, satisfaction, and mental health for data-driven smart connected campus life service. In: ACM international conference proceeding series, pp 531–538. https://doi.org/10.1145/3303772.3303792. Accessed August 11, 2020
23. Heyneman SP (2004) International education quality. Econ Educ Rev 23(4):441–452
24. Hung JL, Shelton BE, Yang J, Du X (2019) Improving predictive modeling for at-risk student identification: a multistage approach. IEEE Trans Learn Technol 12(2):148–157
25. Huang W, Huang W, Diefes-Dux H, Imbrie PK (2006) A preliminary validation of attention, relevance confidence and satisfaction model-based instructional material motivational survey in a computer-based tutorial setting. Br J Edu Technol 37(2):243–259
26. Ifenthaler D, Yau JYK (2020) Utilising learning analytics to support study success in higher education: a systematic review. Edu Tech Res Dev 68(4):1961–1990
27. Jeffreys MR, Dogan E (2012) Evaluating the influence of cultural competence education on students' transcultural self-efficacy perceptions. J Transcult Nurs 23(2):188–197
28. Kawchale N, Satao R (2016) Predictive analytics in education context. Int J Mod Trends Sci Technol 02(11):77–82
29. Kim HJ, Hong AJ, Song HD (2018) The relationships of family, perceived digital competence and attitude, and learning agility in sustainable student engagement in higher education. Sustainability (Switzerland) 10:12. https://doi.org/10.3390/su10124635
30. Lepenioti K, Bousdekis A, Apostolou D, Mentzas G (2020) Prescriptive analytics: literature review and research challenges. Int J Inf Manag 50:57–70
31. Luan H, Huang Q (2016) Prescriptive modeling and compensation of in-plane shape deformation for 3-D printed freeform products. IEEE Trans Autom Sci Eng 14(1):73–82
32. Manohar A, Gupta P, Priyanka V, Uddin MF (2016) Utilizing big data analytics to improve education. In: ASEE northeast section conference
33. Martin AJ (2011) Prescriptive statements and educational practice: what can structural equation modeling (SEM) offer? Educ Psychol Rev 23(2):235–244
34. Marley SC, Levin JR (2011) When are prescriptive statements in educational research justified? Educ Psychol Rev 23(2):197–206
35. Mishra P, Satapathy S, Patel SK (2013) A methodology for evaluation of e-electricity service quality using neural networks. Int J Indian Cult Bus Manag 7(2):152
36. Moskal P, Dziuban C, Hartman J (2013) Blended learning: a dangerous idea?. Internet High Educ 18:15–23
37. Al Musawi A, Amer T (2017) Stakeholders' perceptions of quality and potential improvements in the learning resources centers at Omani basic education schools. Int Educ Stud 10(9):124
38. Ndukwe IG, Daniel BK (2020) Teaching analytics, value and tools for teacher data literacy: a systematic and tripartite approach. Int J Educ Technol High Educ 17(1)
39. Anitha P, Patil MM (2018) A review on data analytics for supply chain management: a case study. Int J Inf Eng Electron Bus 10(5):30–39
40. Park Y, Jo IH (2015) Development of the learning analytics dashboard to support students' learning performance. J Univ Comput Sci 21(1):110–133
41. Pretz CR, Kozlowski AJ, Dams-O'Connor K, Kreider S, Cuthbert JP, Corrigan JD, Whiteneck G et al (2013) Descriptive modeling of longitudinal outcome measures in traumatic brain injury: a National Institute on Disability and Rehabilitation Research Traumatic Brain Injury Model Systems study. Arch Phys Med Rehabil 94(3):579–588

42. Rajni J, Malaya DB (2015) Predictive analytics in a higher education context. IT Prof 17:24–33
43. Siebert JU, Kunz RE, Rolf P (2020) Effects of proactive decision making on life satisfaction. Eur J Oper Res 280(3):1171–1187
44. Shayakhmetova BK, Ayazhanov SS, Ten TL, Drozd VG, Medeubaeva AM (2021) Descriptive big data analytics in the field of education. In: 2021 IEEE international conference on smart information systems and technologies (SIST), pp 1–5
45. Sureshchandar GS, Rajendran C, Anantharaman RN (2021) Determinants of customer-perceived service quality: a confirmatory factor analysis approach. J Serv Mark 16(1):9–34
46. Tawafak RM, Romli A, Malik SI, Shakir M, Al Farsi G (2019) A systematic review of personalized learning: comparison between e-learning and learning by coursework program in Oman. Int J Emerg Technol Learn 14(9):93–104
47. Winer A, Geri N (2019) Learning analytics performance improvement design (LAPID) in higher education: framework and concerns. Online J Appl Knowl Manag 7(2):41–55
48. Wise JA, Mockovak WP (1973) Descriptive modeling of subjective probabilities. Organ Behav Hum Perform 9(2):292–306
49. Yarahmadi F (2019) The evolution of higher education in Oman under the gravity of globalization and innovation. Contrib Econ 351–66

Chapter 15
Why Big Data and Data Analytics

Nagendra Singh Yadav and Vishal Goar

15.1 Big Data

Big Data describes the form of data whose size is massive in volume, which turns out to be both structured and unstructured, which aims to drive operations of a business [13]. When we speak of Big Data, it is not important that how much amount of data a business is concerned about. The most important thing matters with the huge amount are data is, how the organization is going to use the data or what they are going to do with that data [4]. This huge amount of Big Data is analyzed for insights that lead to efficient decision-making and help to plan the strategic business moves. Big Data is defined by its vastness, and it can comprise a variety of data types, such as urban data [26].

15.1.1 Big Data—An Introduction

Data consist of several sets of raw information that are processed to find meaningful purpose out of the information pool by performing computer operations and which can be stored and transmitted to others in order to facilitate the information sharing.

Big Data consists of forms of data that could be structured, unstructured, and semi-structured, which is gathered by an organization for the purpose of data mining, which acts an input for project related to machine learning and data analytics [5].

Big Data comes with a huge size of data. Big Data describes a huge repository of data with larger volume and which can grow dynamically over time. These types of

N. S. Yadav (✉) · V. Goar (✉)
Government Engineering College Bikaner, Bikaner, Rajasthan, India
e-mail: nksyadav100@gmail.com

V. Goar
e-mail: dr.vishalgoar@gmail.com

V. Bali et al. (eds.), *Decision Analytics for Sustainable Development in Smart Society 5.0*, Asset Analytics, https://doi.org/10.1007/978-981-19-1689-2_15

data are too large to handle and store using legacy Data Management tools because they are not ready-made tools to handle larger volumes of data. The efficiency comes as promised with Big Data [3].

The systems that can process and store Big Data are equipped and preloaded with a common piece of data management architecture. One cannot measure the Big Data in smaller units of data volume, as the Big Data Deployments occur in TB (Terabytes), PB (Petabytes) and EB (Exabytes) in few cases for the data gathered over a span of time.

However, the Data has its natural value and that is of no use if the value is not located or found. The most important question arises here: how authentic is the data and how much one can depend on it?

Today, Big Data has become a valuable asset for the giant tech companies in the field of information technology. These giants relay on the value of the data using which they can analyze and build new applications or products with promised efficiency.

Recent innovations in technology perspective have reduced the cost associated with storage of data, and it has made computation easier and cost-friendly, allowing to store larger amounts of data. The organizations can make the business decisions easily when it comes to equipping Big Data, as an increased amount of Big Data has turned cost-friendly.

15.1.2 The Background of Big Data

Big Data is an emerging technology, but it initially originated in the 1960s and 70 s when the data gathering for a huge set of data and storing the data started in the first data mart. It started along with the evolution and development in relational databases.

By the year 2005, everyone in the digital community started realizing that how much data is being generated with several social media and video streaming platforms, i.e. Facebook, YouTube. An open-source framework was developed in 2005 to store this huge amount of data and analyze the resultant data sets was called Hadoop. This contributed to the popularity of NoSQL at this point of time [21].

The revolutionary development of the open-source framework, especially Hadoop was the reason, because of which the Big Data has grown so much. Big Data allows you to operate on data with ease along with data storage at a lower cost. Several users around the globe are generating a huge amount of data, and it applies to all the industries as well [17].

Side by side, rapid evolution and advances in IoT (Internet of Things) allowed the larger no of devices to be connected with the help of the internet. This aims to gather customers' data based on several usage patterns to improve the performance of the product.

This is the beginning of the usefulness of the Big Data, as it has traveled so far. Several data possibilities are expended by cloud computing as well. A cloud is a place where a developer can test data subsets.

With the availability of more information, Big Data allows users or developers to gain more insights into the data. This contributes to gaining more trust and confidence in the data, which leads to a new approach to overcome and handle the problems.

15.1.3 Foundation Pillars of Big Data—Characteristics

Systems that are capable to store the Big Data after its processing are the usual component when it comes to the management of data in organization architecture. Below are the listed V pillars in Big Data (Fig. 15.1).

15.1.3.1 Volume

Big Data represents "Amount of data whose size is infinite or unknown, as it has to measure (in fewer cases)". If one wants to find out value of data, we must know the size of data as the size of data plays a vital part. If someone wants to consider whether a data is a Big Data, they must consider the volume of data in the first place. Volume is an important aspect that should be considered as one of the prime characteristics if we speak of handling Big Data. The amount of data matters a lot to carry out day to business and transactions. Using Big Data, we can process massive amounts of unstructured data [9]. The value of data could be unknown in a few cases, for example, data feeds on Facebook or Facebook posts, a YouTube stream, or hardware equipment that are equipped with sensors. The amount of data to be captured is

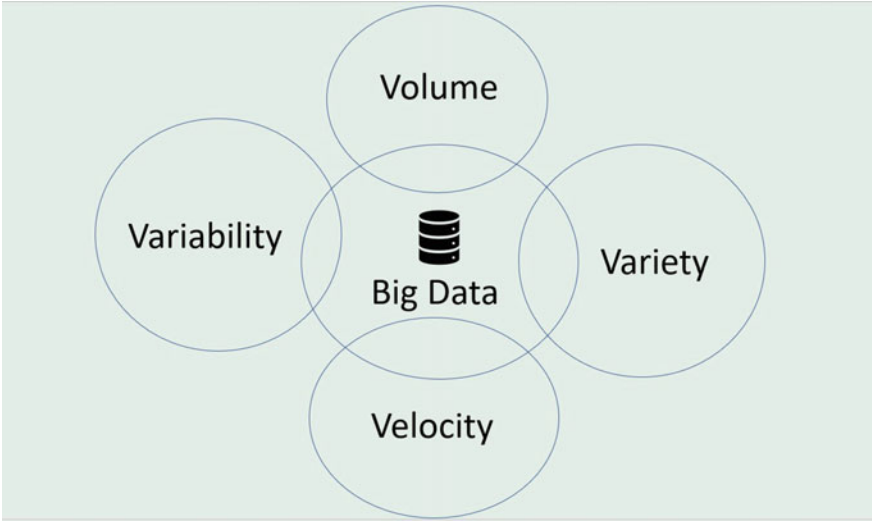


Fig. 15.1 Foundation pillars of big data

based on organization usages as for a few of them the data could be in terabytes or petabytes. It is not mandatory that all the Big Data environments should contain a larger amount of data but in exceptional cases, most of them do contain a larger amount of data in terms of volume of data. This completely depends upon the data that is captured and stored inside Big Data.

15.1.3.2 Variety

Variety refers to diversified nature and source of data, which could be either structured or unstructured or both at the same time [14]. In the past, most of the applications rely on a few data sources like spreadsheets and databases when it comes to the diversified form of data [6]. These days the data is found in the form of multimedia files (audio and video), PDF, etc. This form of data is used for analysis purposes to predict and determine the behavior and meaning out of data using analysis applications. However, various forms of unstructured data are the biggest challenge for data mining, data storage and analysis of data.

15.1.3.3 Velocity

The speed of generating the data determines velocity. It focuses on, how quickly data can be generated and processed further so that demands can be satisfied. Velocity controls and drives the true potential of data. Velocity in terms of Big Data describes speed in which from different sources data is flown, i.e. application logs, social media platforms [15]. The data flow is continuing and its size turns out to be massive. DataStream that of higher velocity can be directly written to memory instead of writing it to disks. Some of the smart products that are internet-enabled operate in real time and they do require time for evaluation.

15.1.3.4 Variability

Variability refers to the behavior of data, which results in the inconsistency of data several times. This could be due to the process of handling and managing the data is hindered, which results in reduced efficiency in terms of managing and handling the data.

15.1.4 Examples Associated to Big Data

The Big Data is gathered and collected using quantities from different sources, which could be a transaction system of a business. Databases store customer details, medical

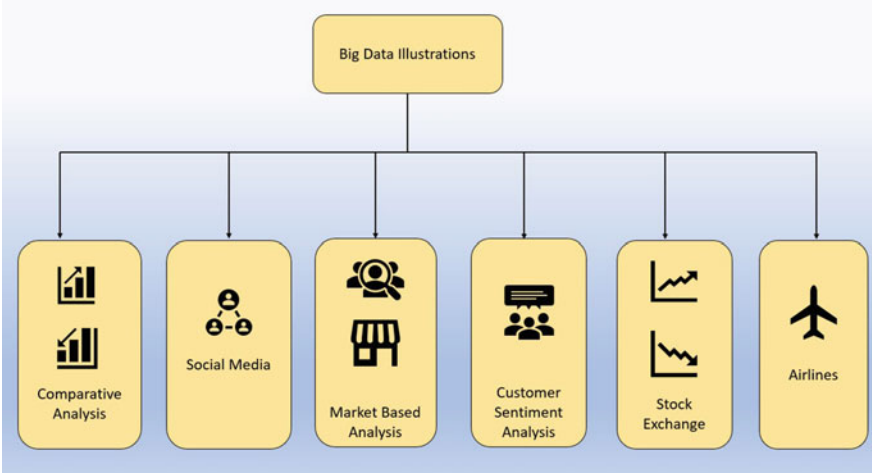


Fig. 15.2 Big Data Illustration

records of the patients. The number of several web pages visited by a user and several other social media platforms (Fig. 15.2).

Different set of analysis can be performed on the above set of data using.

15.1.4.1 Comparative Analysis

The comparative analysis focuses on user behavior observation and metrics that are customer engagement based used to compare with company services, brands and products to find out the competitors.

15.1.4.2 Social Media

This contains all the information related to a brand in terms of customer opinions about the product on social media platforms. This can be gathered using feedback surveys or polling. This can help the businesses to improve and meet the expectation of the customer in the long run. The data gathered using social media platforms helps the business to drive the organic traffic towards the product and help them engage with their audience. Example—Facebook, which generates terabytes of data just in the forms of photos and videos.

15.1.4.3 Market-Based Analysis

The market-based analysis consists of information that includes promoting the new products to the audience, building up brand image, and promoting the product by keeping in mind the target audience with creating a new audience for the product.

15.1.4.4 Customer Sentiment Analysis

The gathered information can be used to find out how a customer feels about a brand or a company when they consume certain commodities. Finding out, the customer wants to go for the same products as being loyal to the brand and how customer experience can be improved in terms of customer assistance.

15.1.4.5 Stock Exchanges

The stock exchanges generate a huge amount of data certainly around Terabytes per day. The example of this is The New York Stock Exchange, which produces around 1 TB (Terabyte) of new trading data on daily basis.

15.1.4.6 Airlines

Did you ever thought of the aviation sector, how much data they generate during the whole flight time. A single jet engine of an airplane is equipped to generate more than 10 terabytes of data during the flight of 30 min. Just imagine thousands of flights per day, the generated data can reach more than petabytes.

15.1.5 Classification of Big Data

The Big Data can be classified into (Fig. 15.3).

15.1.5.1 Structured Data

This contains few characteristics in which a data can be stored in some table-like structure using specific formats. When a data possesses all these characteristics it is said to be structured data. Over time, the field of computer science has seen some innovation when speak of data and kinds of data. This is a behavior that predicts where and when what type of format has to be used for Data categorization or storage or to organize it. Implementation of such kinds of data can be seen in the Relational Database in which firstly Database is created, then tables are created and at last

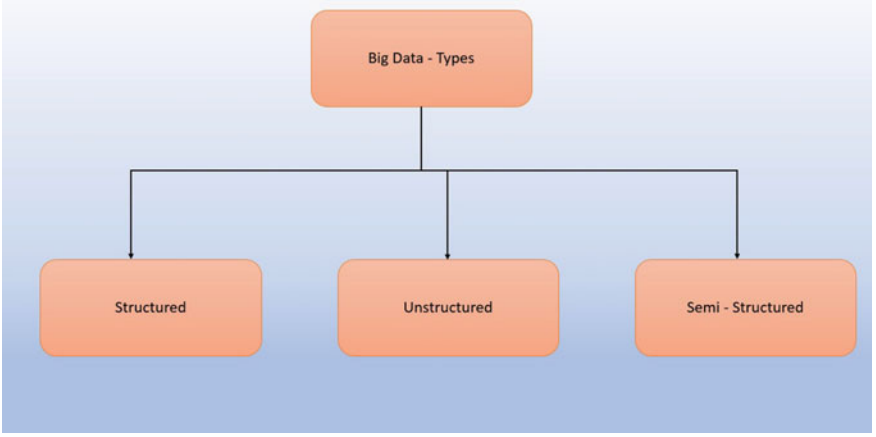


Fig. 15.3 Classification of big data

EMP_ID	EMP_NAME	GENDER	DEPARTMENT	EMP_SALARY
1101	James	Male	Finace	70000
1102	Richard	Male	Admin	80000
1103	Donald	Male	Marketing	88000
1104	Vince	Male	Finance	90000
1105	Kate	Female	Human Resource	100000

Fig. 15.4 Example of structured data

data is inserted in the table, which is a combination of rows and columns. However, the Relational Database management system is the effective implementation of a structured database.

Example of Structured Data—Consider the below-mentioned Employee Table (Fig. 15.4).

15.1.5.2 Unstructured Data

A data of which form is unknown is categorized as Unstructured Data. Though the Unstructured Data is bigger in size, it does create multiple challenges when it comes to the process of that data in order to find a value of it. Unstructured data can be seen in the form of heterogeneous data sources that are formed with the combination of, i.e., videos, Images, Text Files. These days, the organizations do have a huge repository of unstructured data but they are not aware of a process that can be used to extract the meaning from that data.

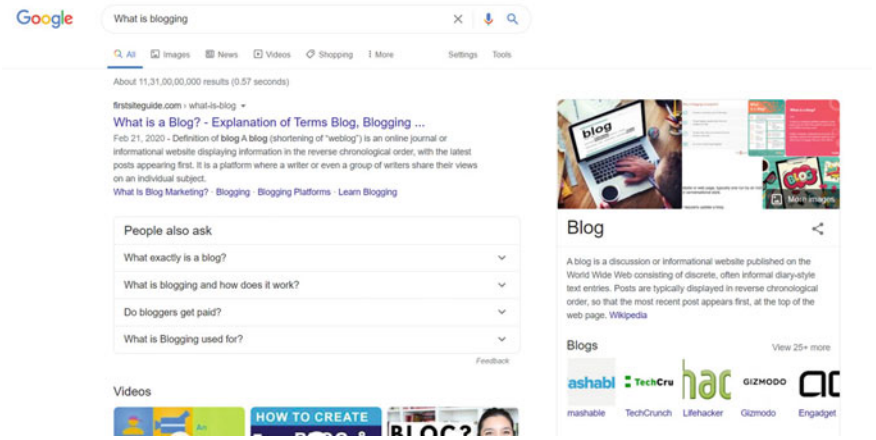


Fig. 15.5 Unstructured data. *Source* David Taylor [2]

The Implementation of Unstructured Data can be seen in search engines and Google. Searching is the best-known example of it (Fig. 15.5).

15.1.5.3 Semi-Structured

This kind of data comes in combination with data that is structured and unstructured.10] When one looks at the Semi-Structured Data that looks like the form of a particular structure but it is not defined as we see in the Database Tables. Below is an example that reflects semi-structured data in a XML File. This is a representation of data that is inserted in an XML File (Fig. 15.6).

15.1.6 Big Data—Applications

Big Data helps a business to address several business activities, which range from analytics to customer experience and orientation. Below are few applications offered by Big Data (Fig. 15.7).

```
<rec><name>Ron Von</name><sex>Male</sex><age>25</age></rec>
<rec><name>April S</name><sex>Female</sex><age>31</age></rec>
<rec><name>Ranjeet Singh</name><sex>Male</sex><age>39</age></rec>
<rec><name>Vishwajeet Singh</name><sex>Male</sex><age>26</age></rec>
<rec><name>James S</name><sex>Male</sex><age>25</age></rec>
```

Fig. 15.6 Example of semi-structured data

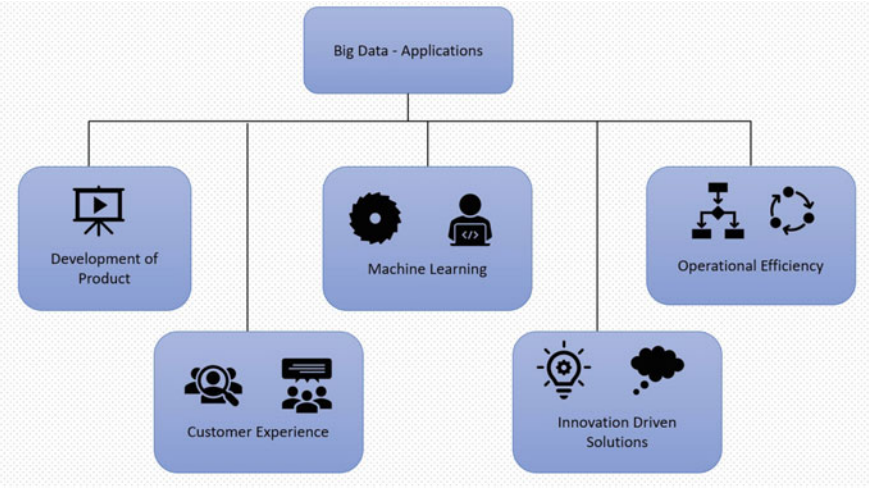


Fig. 15.7 Application of big data

15.1.6.1 Development of Product

Companies in the field of entertainment field, i.e. Netflix use Big Data to predict customer demand. They transform predictive models specific to new products.

15.1.6.2 Customer Experience

All the giants are chasing and competing with each other to influence the larger audience towards their own platforms. We can relate it to the example of the entertainment industry. Big Data facilitates companies to capture the data related to customer choices such as social media, website visits, etc.

15.1.6.3 Machine Learning

Machine learning is an ongoing trend in the field of Information Technology. Using Big Data, we can now teach the machines instead of writing programs for them. The availability of Big Data has helped a lot to train the machine learning models.

15.1.6.4 Innovation Driven Solutions

Big Data allows us to use the data insights in order to improve on decisions related to finance and planning. It also helps us to understand what a customer wants from a new product in terms of services offered by them.

15.1.6.5 Operational Efficiency

Big Data has created a huge impact when it comes to improving operational efficiency. This helps to analyze and assess the production, customer experiences, and helps to predict future demands. This improves the decision-making process so that it aligns with the current market demands.

15.1.7 How Big Data is Processed?

The computation power is required to process the massive volume of data rapidly. The various sets of data can be overwhelming for a single server cluster or server. The organizations are required to equip appropriate processing capacity for Big Data assignments if they want to obtain the needed velocity. A public cloud provider is capable to store data in any size and can scale up the server so that the targeted Big Data projects can be completed within plenty of time [16]. The business has to bear the cost for storage and time used for the computation. However, cloud instances can be turned off if they are not required. In order to improve the services level, the public cloud providers purpose capabilities of Big Data with the help of managed services that include—

- Microsoft Azure HDInsight
- Google Cloud Dataproc
- Amazon EMR (previously known as Elastic MapReduce)

Big Data is stored in a cloud environment using:

- HDFS (Hadoop Distributed File System)
- NoSQL Database
- Relational Database
- Amazon Simple Storage Service (which is a low-cost cloud object storage method) (Fig. 15.8)

Big Data provides insights so that a business can explore new opportunities and business models. There are three actions involved when it comes to the processing of data.

15.1.7.1 Integration

The data in Big Data arrives from various sources irrespective of a specific type. Legacy Data Integration tools such as Extract, Transform and Load (ETL) are not up to the task. They require the formation of new strategies along with technology so that they can analyze the Big Data at various storage levels in sizes, i.e. terabyte, petabyte. During the Integration process, it is required that a business must ensure that the data is processed along with the formatting. This makes the job of Business Analyst easier.

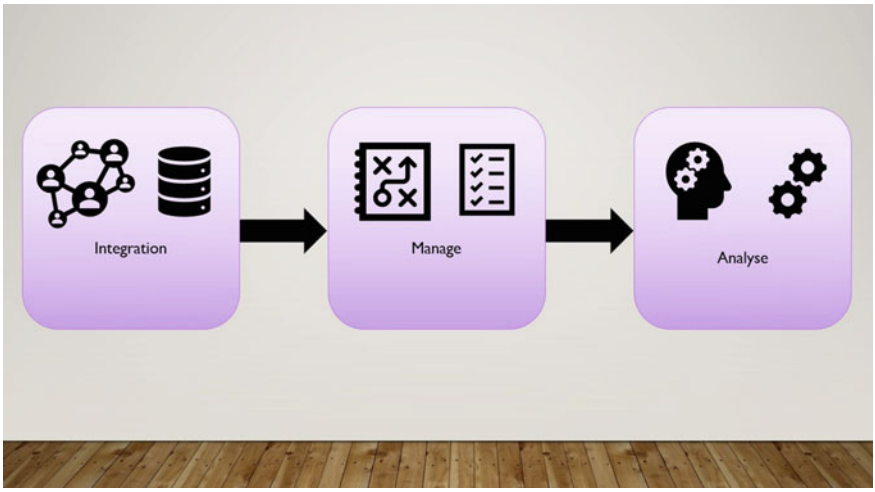


Fig. 15.8 Working of big data

15.1.7.2 Manage

Big Data requires larger storage as the name "Big Data" itself reflects the massive storage space required to store the data. One alternative could be to store the data on the cloud, and to store that data which requires frequent access in day-to-day operations. This enables a business to make the decision-making process smoother and faster as the data is available on-demand on the cloud. A business makes its decision of data storage based upon where their current data resides. The cloud storage has gained a lot of buzz and popularity around information technology as it provides support for current computation requirements and gets the resources whenever required.

15.1.7.3 Analyze

The return of investment on Big Data comes in handy and greater when a business knows how they should act on the data by analyzing it. This enables a business to:

- Create new data models with the help of artificial intelligence and machine learning.
- More clarity along with visual analysis across several data sets.
- To explore data furthermore to discover the new findings.
- Get the data to work.

15.1.8 Importance of Big Data

Using Big Data, organizations can rely on the end results obtains after analysis. Below is the importance of Big Data:

- Improves operations
- Provides improved customer service
- Based upon the customer preferences, personalized market campaigns are created
- Increases profitability
- Helps in predicting the future scope of business
- Provides a competitive advantage
- Faster and more informed business decisions.
- Effective use of data
- Offers valuable insights to the companies
- Increases customer engagement
- Predictive modeling
- Helps to business the sustain in a dynamic environment
- Helps to identify the risk factors

15.1.9 Advantages Offered by Big Data

Using Big Data, a staging area is created for the new data that arrives and this is performed to determine if the new data has to be stored in a data warehouse or not. The combination of technology that consists of Big Data and Data warehouse can help organizations to get rid of the occasionally used data. Big Data brings a lot of advantages alongside the adoption of it. Few of them are listed as mentioned below .

15.1.9.1 Builds Trust in Business

Big Data helps the business to trust the intelligence while making crucial decisions. With easy access to social media, businesses can retransform their strategy based upon the ongoing dynamic environment changes to sustain in the long run.

15.1.9.2 Enhanced Customer Solution

Newly designed systems powered by Big Data technology have replaced the traditional customer feedback systems. These new systems are combined with natural language processing technology to evaluate and read customer responses.

15.1.9.3 Early Identified Risk to Product and Services

Big Data helps businesses to identify the risk when it comes to products or services. Predictive Modeling helps to eliminate the risk at the earlier stage to ensure the cutting-edge growth of the business.

15.1.10 Factors That Challenge Big Data

Big Data is preloaded with a lot of features, it does have some challenges to address. The design, development and techniques of deploying apps for smart cities face a broader variety of problems [24]. Big Data is said to be spreading across all facets of information technology [28]. Below are the challenges faced by Big Data.

15.1.10.1 Big Data Is Huge

Big Data comes with massive storage. The new technology was developed for massive data storage but data volumes seem to double up every 2 years of time [18]. This creates problems for organizations, as they need to find efficient ways to store the data, which is growing over time.

15.1.10.2 It's Not All About Storing the Data

If an organization has adopted Big Data, it is not about the Data Storage. The stored data should be used for valuation purposes instead of keeping the data [18]. The Data has to be filtered target to a specific purpose and it needs to be organized for analysis purposes which are again a huge job.

15.1.10.3 Dynamic Technology

Big Data is changing constantly at a quick pace. Few years back, Apache Hadoop use to be the technology that was used to Handle Big Data [22]. Later on in the year 2014, Apache Spark was introduced, which was a combination of two frameworks, proves to be the leading approach. The ongoing challenge with Big Data is keeping up with the changing technology.

15.2 Impact of Big Data on Smart Cities

15.2.1 Introduction

In business and government, "smart cities" have gotten a lot of interest since they stress constant monitoring through various computer methods. Cloud computing or ubiquitous computing, for example [25]. Big Data enabled solutions provide admin control over the data in the repository, as well as storage, backup, and visualization. Big Data has enabled technologies to improve the efficiency of complicated data structures. Smart cities benefit from the possibilities of such technologies. This covers the Internet of Things (IoT), as well as smart gadgets (sensors).

15.2.2 Smart Cities

Sensors and related connected devices are used in smart cities to collect and evaluate data. This information may be used for a variety of purposes, including resource optimization and management. Smart cities may make the most of technological advancements to enhance a variety of issues, i.e. traffic management, water and electricity supply, and so forth. The Smart City strives to improve the residents' quality of life by improving the availability of utilities [27].

15.2.3 Layers in Smart Cities

This may be broken down into three levels, as seen below:

15.2.3.1 Technology Layer

Several sensors and linked devices are included in this, which assist to support a larger range of services.

15.2.3.2 Application Layer

This comprises IS (Information Systems) to help cities to make better use of their services and improve citizen operations.

15.2.3.3 Usage Layer

Comprises completed solutions/applications that can be used by a specific group of people.

Key elements in technology layer of Smart City

- **IoT (Internet of Things)—**The linked gadgets and sensors may communicate with one another and send data to management systems. In this scenario, the Internet of Things is being utilized to collect the same data and fix the problems.
- **ICT (Information and Communication Technology)—**This serves as a platform for residents to communicate with one another. It facilitates transportation by enabling smart and efficient transport which ends up reducing the cost [24]. Smart cities can employ ICT to make the necessary modifications to current services based on studied data and citizen/user feedback. Considering the several dimension of smart cities, ICT becomes the key aspect when it comes to the development of smart city operations [23].
- **Sensors—**Hardware having a wide range of applications. For example, it records temperature and vehicle count, among other things.

15.2.4 How the Big Data Can Impact Smart Cities

There are a variety of effects on several parts of city life, including transportation, citizen safety, and many more.

15.2.4.1 Safety of Citizens

This may be achieved by implementing predictive Big Data analytics, which can identify important places that are likely to be crime hotspots and attempt to forecast where the actual crime will occur. The use of location and historical data will aid in the transformation of a safer environment for residents.

15.2.4.2 Traffic Resolutions

Traffic congestion is an issue in many places, and it costs a lot of money. The data may be analyzed by traffic authorities, who can then attempt to manage transportation. This data can reveal previously unknown trends, allowing authorities to design data-driven solutions and lowering traffic congestion.

15.2.4.3 Cost Cutting

Smart cities may be transformed for a large sum of money. Such investments are made for either rehabilitation or new construction. The studied data from Big Data may assist in determining which regions require renovation or transformation, as well as the type of transformation necessary. This will make it easier to make the best investing decisions.

15.2.4.4 Continuous Growth

The officials can keep up to speed on the needed modifications thanks to the daily study of the city's expansion. Daily updates are critical for ensuring sustainability since they aid in gaining a comprehensive understanding of the needed new developments. The growth of smart cities is driven by data.

15.2.5 Where the Data Should Be Stored?

In order for smart cities to work effectively, a bigger amount of archival and real-time data is required:

- **Data related to smart cars—**In today's world, all smart automobiles can be connected to mobile devices. Now, using the same principle, automobiles may talk with one another, reducing traffic congestion, improving navigation systems, and lowering the chance of road accidents.
- **Data related to capture devices—**Surveillance using a camera may be a useful tool for controlling and managing a variety of factors. They might include traffic surveillance and the development of intelligent technologies that can subsequently be used as crime detectors.

The data from smart cities can primarily be kept in three locations, which are described below.

15.2.5.1 Cloud Storage

When it comes to smart cities, a bigger volume of data is necessary for analytics. Cloud systems include a solid-state driven in their data center, which may eliminate superfluous data while also encrypting the data transferred. There are a variety of cloud solutions available, with various pricing choices and on-site or off-site data centers.

15.2.5.2 Edge Computing

Edge computing allows data to be processed close to its source. Instead of sending data to a specified remote site and then routing the same data back to the appropriate authorities, edge computing is supplied at a lower cost. AI-based traffic management is one of the Edge computing capabilities that is already under development. AI-based traffic management uses automation to detect traffic obstacles and accidents, allowing us to respond quickly in a variety of scenarios.

15.2.5.3 Hybrid Data Storage

Hybrid data storage systems combine cloud and edge storage capabilities. Such systems enable fresh decision-making based on warnings generated by real-time data as well as previously stored data.

15.3 Big Data Analytics

For advanced analytics techniques, Big Data is used that applies to a diversified larger set of data that could be either structured data, semi-structured data and unstructured data [12]. Such data is captured or obtained using several sources that differ from their total sizes from terabytes to zettabytes.

15.3.1 Introduction

The data volumes have reached unimaginable levels during the past decade alongside reduced prices for data storage. Several private firms and research institutes capture a huge amount of data, which could be related to customer interactions, social media and they do capture data from sensor-enabled devices such as a smartphone [20]. Though gathering of data is quite easy but finding the meaning out of data is the one difficult task. This is where need of Big Data is realized.

Big Data Analytics plays a vital role in data collection from various data sources, merges them in such that the handy data can be ready-made to consume by analysts and the analyst finds the meaning out of data to power up business decisions [7].

Big Data Analytics is a process in which a huge amount of data is captured from different data sources for the desired product is always useful for organizations and at a later stage this becomes the foundation [17].

Big Data Analytics allows the researchers and businesses to improve in decision-making based upon the data, which was previously not accessed and usable [21]. The Big Data can be used by several advanced analytics techniques, i.e. prediction-based

analysis, machine learning so that new insights can be gained from data that was captured previously but never used later then.

However, the use of Big Data is limitless as we can use several advanced analytics techniques to find the meaning out of data, which can impact the business decision and essential for a business to grow [23]. It is all about picking up the desired and correct analytics tools or techniques.

15.3.2 Why Big Data Analytics

Big Data analysis has formed a revolution in information technology. The usages of data analytics by businesses have increased in recent years. The main focus of a business is over the customer. The more they can influence the customer, the more the business grows. Data analytics is required for the business and is required to determine the future scope and expected outcomes in the near future [23]. Big Data analytics does help a lot of businesses in determining the best data fit based upon which the business decisions are made [8]. It facilitates business to sustain in the dynamic environments where predicting the future is the key to survival. New insights can be gained using Big Data analytics to determine the growth areas of the business [19]. However, the scope of Big Data has to be realized as its outcomes depend on the selection of data.

15.3.3 Data Life Cycle in Big Data Analytics

To provide the desired framework in order to organize the job required by the organization and to provide crystal clear insights on Big Data, one has to think of a process, which is equipped with several stage stores underneath it. All the stages are related to each other.

15.3.3.1 The CRISP—DM Life Cycle

The CRISP—DM Methodology or "Cross Industry Standard Process for Data Mining" is a process that outlines several approaches in which experts of Data Mining use in order to address and solve problems in classical BI Data Mining. It's still used in classical BI Data mining Teams (Fig. 15.9).

Below are the stages in CRISP—DM life cycle.

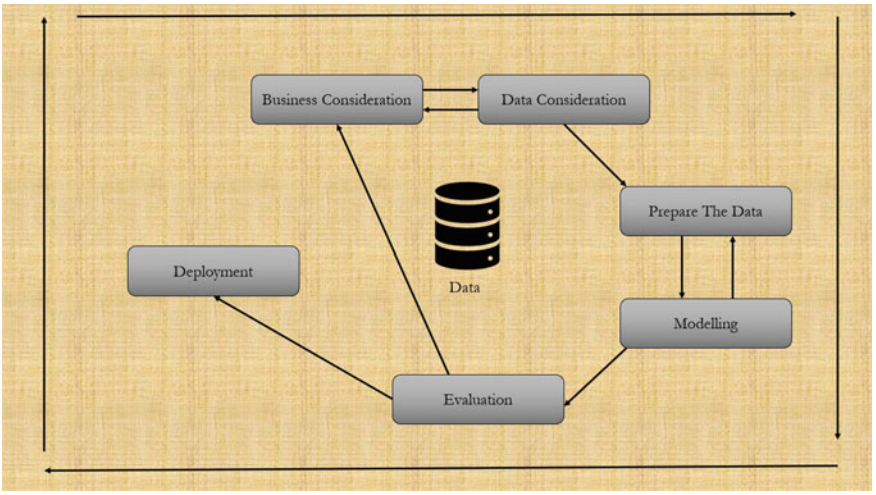


Fig. 15.9 The CRISP—DM life cycle

Business Consideration

The focus is set on understanding the requirement and objectives of the project from a business perspective. Later on, the gathered knowledge is transformed into the problem definition of data mining to obtain the objective an introductory plan is designed.

Data Consideration

The phase begins when the data collection starts and it goes further which involves activities so that we could get familiarize with the data. This helps to recognize Data Quality problems and to find out very first insight on data or to detect the hidden information.

Prepare the Data

This phase consists of all the activities in order to build the final dataset, which acts as an input for tools and to is fed into modeling tools. This task can be performed more than once and it doesn't require an advised order. This can include the selection of records, attributes, and tables as well as data cleaning for modeling tools.

Modeling

This phase involves the adoption of modeling technique and selected techniques are applied along with parameters that are measured to best values. In order to solve the data mining problems, there are multiple techniques available. A few techniques have particular requirements in the form of data. This frequently requires to go one step backward towards Prepare the data stage.

Evaluation

When we have reached this phase, it implies that we have successfully built a high-quality model that is used for the Analysis of Data. Before moving on to the deployment, it is recommended that the model should be evaluated and verified to its core alongside steps review, which were performed while construction of the model. This is required to verify that we have achieved the business objectives.

Deployment

After the creation of the model, the job is not finished until the created model is deployed. Based on the requirements, this phase could be either generation of report or segment allocation or mining of data.

There are several cases in which the customer carries out the deployment steps instead of a Data Analyst.

15.3.3.2 The SEMMA Methodology

Sample, Explore, Modify, Model, Assess is popularly known as SEEMA that was developed by SAS specific to Data Mining Modeling. Below are the stages in this methodology (Fig. 15.10).

Sample

The process begins with the sampling of data. The Dataset should be so larger, which can contain enough information to retrieve and should be so small to be used efficiently.

Explore

In this phase, we discover anticipated and unanticipated relationship amongst variables and abnormalities by using Data Visualization.

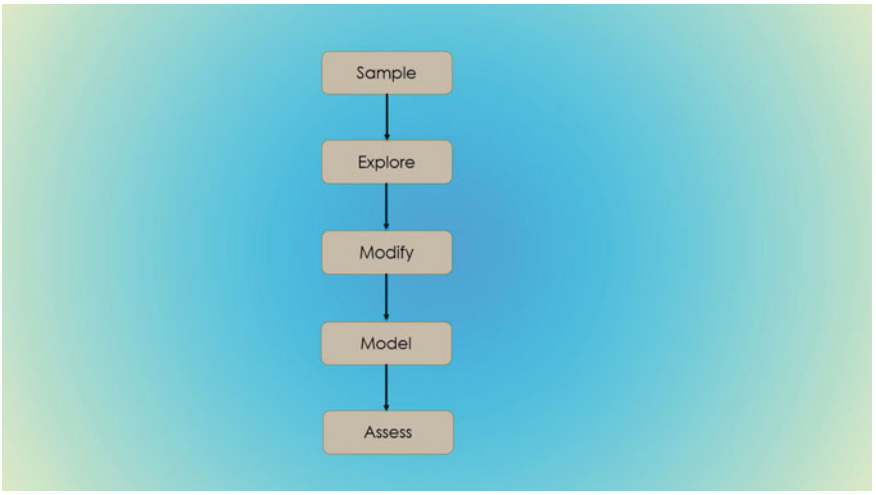


Fig. 15.10 The SEMMA methodology

Modify

This phase consists of methods, which are used to select, transform and create required variables for Data Modeling preparations.

Model

In this phase, the focus is shifted on applying several modeling techniques to prepared variables so that a model is created, which provides expected results.

Assess

The Assess Phase involves evaluations of modeling results, which reflects usefulness and reliability of created models.

15.3.3.3 Life Cycle of Big Data

Below are the stages in the big data analytics cycle (Fig. 15.11).

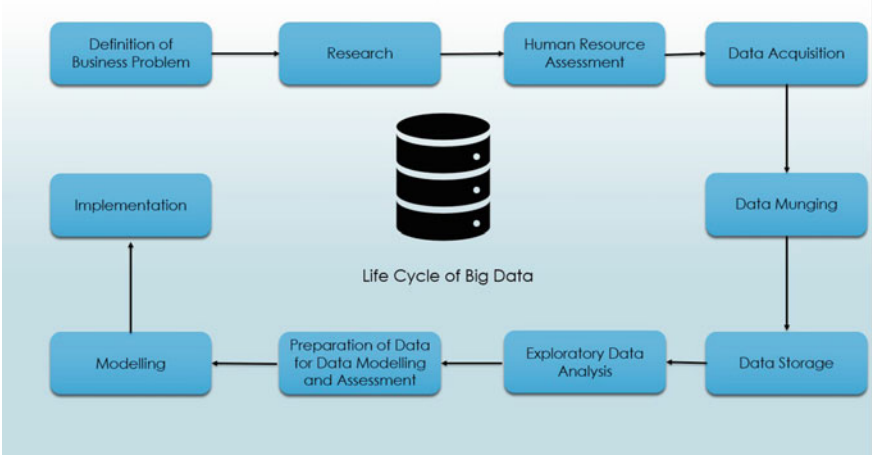


Fig. 15.11 Life cycle of big data

Definition of Business Problem

In this phase, the problems are defined and evaluated correctly to determine the potential growth of an organization. This has to be evaluated because as an end result of this expected gains and the project cost is determined.

Research

This requires analysis to be performed against the other companies in terms of what they followed in the current situation. This includes finding out solutions, which can be used for business. The methodology for future stages has to be defined in this phase.

Human Resource Assessment

After defining the problem, we have to determine that if we can complete the project successfully with the help of existing staff. Outsourcing has to be performed if with the required staff we are unable to meet the timelines.

Data Acquisition

This defines what are the types of profiles to be considered in order to deliver the resultant data product. This involves gathering of data from several sources, in which

the data is found as unstructured. Data could be in text, a different language that requires several hours of time to get completed.

Data Munging

Once the data is captured from different sources it is required to be stored in Easy-to-Use format. In order to make a decision, the data sources have to be combined. This process requires the conversion of the first data source into another form. The processes require a lot of time but they must be result-oriented with the quality delivered.

Data Storage

When the data is processed successfully, the processed data is required to be inserted into databases for storage. One can equip the Hadoop framework, which facilitates a user to use SQL to inset and query with database, describes "HIVE Query Language". This facilitates carrying out all the Analytics jobs on the Data. There are several other options available other than Hadoop, which are MongoDB, Redis, SPARK.

Exploratory Data Analysis

Insights on the data can be retrieved after the cleaning of data and its storage. This phase is a prerequisite, which has to be performed. This phase ensures that a common understanding with the data is developed with the use of statistical tools that are used to represent and plot the data.

Preparation of Data for Modeling and Assessment

In this, cleaned data is rephrased, which is retrieved with the help of statistical preprocessing of missing value imputation.

Modeling

The modeling stage involves trying out several models with an aim to solve the business problem. As a prerequisite, it is expected that the model must provide some insights into the business. At the end of this phase, the best model gets selected to determine its performance on a left-out dataset.

Implementation

As we adopt this phase, using which the newly developed data can be applied to business. This includes the setup of validation scheme when the data product is executed with a sole aim to monitor its performance.

15.3.4 Methodologies in Big Data Analytics

The analytics part of Big Data is completely different from the commonly used statistical methods, which are used for design purposes. As the data arrives, the analytics part comes into effect. We developed the model such that it can describe the response. This is performed to keep an eye on the response and behavior of the model or to develop an understanding of the input variable how it drives the response of the model. This could be matching the actual result with the expected result same as we aim to find out while testing an application. When it comes to statistical design experiments, data is retrieved as an end result of what we have developed in an experiment. This results in the generation of data, which is useful to statistical model.

Big Analytics is all about the representation of data. It is not feasible to design an experiment that has the sole purpose to satisfy the statistical model. A wider amount of data clean-up is required so that this could be used in modeling of machine learning in case of large-scale applications of analytics. At first place, when a business problem is defined, it is required to perform a research stage in order to determine the design of methodology. Statistical modeling is the biggest task in Big Data analytics. When data is cleaned and ready-made available for modeling, extra precautions have to follow in order to evaluate several models. Once a model is implemented successfully, it is recommended that reports related to evaluating and results should be published. The biggest drawback of predictive modeling is that it just focuses on the implementation of model instead of measuring performance.

15.3.5 Core Deliverables of Big Data Analytics

Below are the core deliverables of Big Data Analytics [11].

15.3.5.1 Implementation of Machine Learning

Implementation of machine learning could consist of classification of the algorithm, segmentation or regression model.

15.3.5.2 Recommendation-Based System

The main objective is to design a method that can recommend choices on the basis of user's behavior of choices picked in the past. For example, YouTube is a video streaming platform, where videos are suggested to the user on the basis of user's interest.

15.3.5.3 Dashboard

The business requires tools to represent aggregated data using graphical methods. A dashboard can be an ideal choice to visualize the statistical data in a more user-friendly form.

15.3.5.4 Ad-Hoc Analysis

To answer several business area queries, hypotheses, this could be used to provide a solution using Ad hoc analysis of data.

15.3.6 Finding Out Stakeholders in Big Data Analytics

In the case of large-scale organizations to succeed in the development of Big Data projects, the management should always show faith and put their trust in the project. This includes discovering a way around to showcase what are the advantages offered by the project. We always don't need a unique solution for a specific problem for searching for a sponsor when it comes to the project.

In order to find out a sponsor for the project, the key parameter is to understand a problem and build a model after implementation. This will help in convincing the management by highlighting the Big Data project importance to the management.

15.3.7 Data Analyst in Big Data Analytics

The profile is specific to report-oriented, which reflects relevant experience in extracting and analysis of data that is extracted from data warehouses using a query language such as SQL. The tasks of a data analyst could be either specific to storage of data or to develop reporting of business results. Data warehousing is not quite an easy job but it is far different from a data scientist Profile.

There are several companies who find it difficult to select a prospective data analyst and to train them so that they can become data scientists. Below are the skills that a data analyst must have

- Understanding of business
- Programming in a query language such as SQL
- Implementation and report designing
- Development of dashboards to gain daily insights on business.

15.3.8 Data Scientist in Big data Analytics

The key skills of a data scientist are specific to tasks that are related to the development of segmentation algorithm, System that recommends suggestions to the user based on user behaviors, predictive modeling, A/B Testing Frameworks and operates on unstructured data. This type of works requires a depth understanding of mathematics, computer programming in a specific programming language and applied statistics. Both data scientist and data analyst have the ability to query with the database, they analyze data, but what a decision has been made by the data scientist, has a huge impact on the organization.

The preliminary set of skills a data scientist is equipped with:

- Knowledge and hands-on experience on statistical packages, i.e. Python, SAS, R.
- Ability to clean, extract and explore the data gathered using several data sources.
- Implementation, research and design of the statistical model.
- A depth understanding and knowledge of deep statistical, computer science and mathematics.

15.4 Future Scope

Since the pandemic COVID-19 has pushed for more remote techniques, resulting in more devices and therefore more data, the data will continue to rise in the near future. Big data will be useful in the near future to store the rising amount of data. However, the following characteristics will continue to be important in future growth.

15.4.1 Data Will Continue to Grow Along with Migration to Cloud

The way large data is stored and handled is changing thanks to Amazon Web Services, Microsoft Azure, and Google cloud platforms. To operate their own data-intensive applications, firms formerly had to establish their own physical data centers. Everything is pay-as-you-go with the new cloud services, making data more agile and scalable.

Environments

- **Hybrid—**When organizations can't store sensitive data in the cloud, they keep certain data on premises and shift the rest to the cloud.
- **Multi-cloud—**Several businesses want to meet their business needs by storing data in multiple clouds, i.e. both public and private.

15.4.2 Growth of Machine Learning with Big Data Adoption

Due to a variety of open-source platforms, businesses have never had access to machine learning and AI. These platforms were created to overcome technological gaps, but businesses and people sometimes fail to customize them according to their own needs.

Given the advancements in algorithms, computers' capacity to learn from data is expected to increase. As a consequence, we will have more dependable systems. capable of forecasting and making the greatest use of data.

15.4.3 Privacy Will Be on Fire

Data privacy and security is a recurring problem that has to be addressed. The increasing volume of data necessitates the use of many tiers of data security to safeguard it against assaults.

Skill gap—Skill gaps emerge as a result of a lack of training opportunities, and they continue to widen over time.

Complex cyberattacks—Hacker risks are growing all the time, and resolving them is becoming increasingly difficult.

15.5 Conclusion

In smart cities, Big Data is critical. Smart cities have the ability to interpret data collected by sensors and IoT-enabled devices to find patterns and requirements. As a result of this research, road accidents and obstacles can be reduced, allowing cars to find the closest parking spots. The data can also aid in the reduction of crime and the improvement of water and management systems. IoT and Big Data combined have the potential to improve transportation systems in smart cities. This will take use of the basic features of IoT, such as ease of access to the cloud and worldwide connection.

At the end of the day, IoT allows services and businesses to connect, and that is what counts most. The Internet of Things (IoT) is in charge of leading the digital transformation, which provides limitless opportunities for business and service solutions. This brings up the topic of improving one's quality of life.

Big Data and Smart Cities are significant concepts, and many people have begun to build smart city apps that enable increased resiliency and governance. Each city has more data than the next, and repurposing or repurposing existing data across domains is the key to generating value for the city and its residents. The data is then saved and made available through the platform, where it may be combined with other data and utilized for prediction.

The utilization of Big Data and data analytics allows for all of the above. We can store any data with Big Data, and analytics can help us determine the value of that data.

Following the analysis, the data aids in effective decision-making and may contribute to the ROI (return on investment).

However, because of the associated costs, adopting new technology is not always viable. When making a decision, one should constantly examine the facts as well as the dangers.

References

1. https://unsplash.com/photos/z4H9MYmWIMA
2. https://www.guru99.com/what-is-big-data.html
3. https://searchdatamanagement.techtarget.com/definition/big-data
4. https://www.oracle.com/big-data/what-is-big-data.html
5. https://www.sas.com/en_in/insights/big-data/what-is-big-data.html
6. https://www.ibm.com/in-en/analytics/hadoop/big-data-analytics
7. https://www.tutorialspoint.com/big_data_analytics/index.htm
8. Dasoriya R (2017) A review of big data analytics over cloud. In: 2017 IEEE international conference on consumer electronics-Asia (ICCE-Asia), 2017, pp 1–6. https://doi.org/10.1109/ICCE-ASIA.2017.8307833
9. Wei S, Da-xin L (2005) Mining schemas in semi-structured data using fuzzy decision trees. In: Gervasi O et al. (eds) Computational science and its applications – ICCSA 2005. ICCSA 2005. Lecture notes in computer science, vol 3483. Springer, Berlin. https://doi.org/10.1007/11424925_79
10. Vanani IR, Mohammadipour F (2019) Big data analytics: a necessary roadmap for enterprises. In: Öner S, Yüregir O (ed), Optimizing big data management and industrial systems with intelligent techniques. IGI Global, pp. 24–52. http://doi:https://doi.org/10.4018/978-1-5225-5137-9.ch002
11. Suneetha K (2017) Big data analytics in health care. In: Trivedi S, Dey S, Kumar A, Panda T (eds) Handbook of research on advanced data mining techniques and applications for business intelligence. IGI Global, pp 240–249. http://doi:https://doi.org/10.4018/978-1-5225-2031-3.ch014
12. https://www.webopedia.com/TERM/B/big_data.html
13. https://hbr.org/2012/10/big-data-the-management-revolution
14. https://www.i-scoop.eu/big-data-action-value-context/
15. https://www.bernardmarr.com/default.asp?contentID=766
16. https://www.edureka.co/blog/what-is-big-data/
17. https://www.sap.com/trends/big-data.html
18. https://searchbusinessanalytics.techtarget.com/definition/big-data-analytics
19. https://en.wikipedia.org/wiki/Big_data
20. https://www.qubole.com/big-data-analytics/

21. https://www.ngdata.com/what-is-big-data-analytics/
22. https://www.sisense.com/glossary/big-data-basics/
23. Lim C, Kim K-J, Maglio PP (2018) Smart cities with big data: reference models, challenges, and considerations, cities, vol 82, pp 86–99. ISSN 0264–2751. https://doi.org/10.1016/j.cities.2018.04.011
24. Al Nuaimi E, Al Neyadi H, Mohamed N et al (2015) Applications of big data to smart cities. J Internet Serv Appl 6:25. https://doi.org/10.1186/s13174-015-0041-5
25. Kitchin R (2014) The real-time city? Big data and smart urbanism. GeoJ 79:1–14. https://doi.org/10.1007/s10708-013-9516-8
26. Batty M (2013) Big data, smart cities and city planning. Dialogues Human Geogr 3(3):274–279. https://doi.org/10.1177/2043820613513390
27. Neirotti P, De Marco A, Cagliano AC, Mangano G, Scorrano F (2014) Current trends in Smart City initiatives: Some stylised facts, Cities, vol 38, pp 25–36. ISSN 0264–2751. https://doi.org/10.1016/j.cities.2013.12.010
28. Wu X, Zhu X, Wu G-Q, Ding W (2014) Data mining with big data. IEEE Trans Knowl Data Eng 26(1):97–107. https://doi.org/10.1109/TKDE.2013.109

Chapter 16
Review on Opportunities and Challenges of Blockchain Technology for Tourism Industry in Future Smart Society

B. C. Girish Kumar, Parma Nand, and Vikram Bali

Abstract Tourism is a trending industry in the current era of the economic growth of the country. Tourism can be explained as traveling to a place which is different from your home city or country for numerous reasons. But the tourism industry has been facing problems like Identity theft, Intermediaries, Payment failures, fake reviews, transparency, settlement, overbooking, policy, and most importantly safety concerns due to the COVID-19 pandemic. To overcome the above drawbacks, after referring to some of the research papers we have come up with a solution called Blockchain. Blockchain is a chain of blocks where each block contains a set of transactions and the blocks are connected using cryptography. Mainly Blockchain and tourism have the potential to show into a really advantageous combination as this technology will offer additional security, Safety, eradication of middleman, data integrity, decentralized business models, immutability, transparency, and accurate contact tracing app for Tourism Industry in the Future Smart Society. This paper gives a clear idea of how BLOBIM helps to overcome difficulties/challenges in tourism industries. Our future proposed architecture Blockchain-based integrated model removes the third party intervention and improves security, safety, trust, transparency, etc.

Keywords Blockchain · COVID-19 · Tourism · BCT · Contact tracing

16.1 Introduction

Tourism is one of the trending industries in which it takes a major role in improving the economic growth of the country [1]. Tourism can be defined as the situation when a

B. C. Girish Kumar (✉) · P. Nand
Sharda University, Greater Noida, India
e-mail: 2020487380.girishkumar@dr.sharda.ac.in

P. Nand
e-mail: parma.nand@sharda.ac.in

V. Bali
IMS college, Ghaziabad, India
e-mail: vikramgcet@gmail.com

V. Bali et al. (eds.), *Decision Analytics for Sustainable Development in Smart Society 5.0*, Asset Analytics, https://doi.org/10.1007/978-981-19-1689-2_16

person is traveling to a place that is different from the parent city or country for various reasons or business dealings and staying there for some significant period of time [2]. We can also define it as a motion for a few purposes and to keep that isn't terribly long or permanent in nature. In the last decades, travelers need to visit a booking office for booking travel tickets and also travelers are not satisfied with the service provided by the travel agencies [3]. Since the 1990s, the Internet has had a strong impact on the tourism industry with a direct interaction among different stakeholders. Then the concept of e-tourism came into existence with the use of the Internet; e-tourism was stated as the digitalization of all the processes like tourism, travel, hospitality, and food industries that help organizations to improve their efficiency and effectiveness [1]. Tourism becomes an activity that builds the base for wide-scale consumption bringing far-reaching benefits. Tourism gains importance as an activity that has so much reaching positive impacts on industries, economy, society, and therefore for the whole growth, progress, and development of a country [4].

The booking process in tourism has improved after the internet allowed tourists to plan their trips without needing to go to a travel agent's office. With the help of the Internet, travelers can book hotel and other tourism services online [5]. The companies like Uber and Airbnb have provided the services by using traditional centralized architecture. This traditional architecture uses a heterogeneous payment gateway which leads to malicious activities such as identity theft, hacking of wallet accounts, and others. But the travelers are not satisfied with the provided services because of many issues and most of the travelers need good, safe service architecture which will improve the performance and security while making online transactions by using personal identification. The major challenges in the travel industry with the traditional architecture are transparency, settlement, loyalty, fraud, identity, over-booking, and IoT; [4]. Figure 16.1 highlights the traditional architecture of the typical

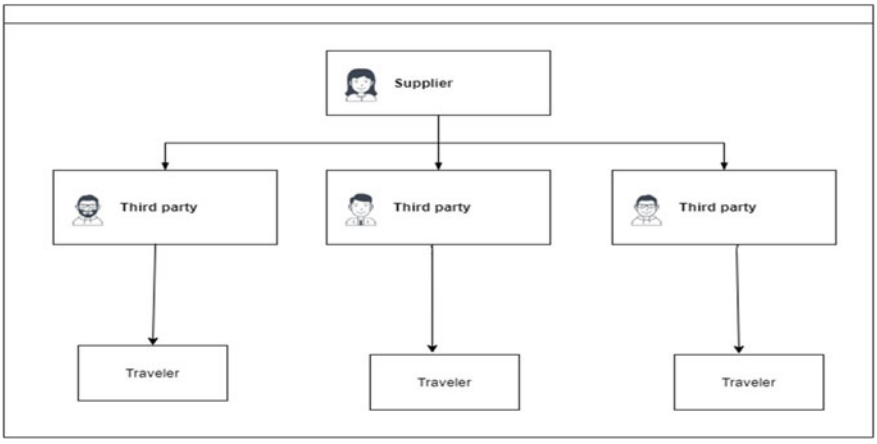


Fig. 16.1 Traditional architecture of tourism services [1, 4]

business process involved in tourism. This will help us understand BCT's impact on this industry and its intermediaries [4].

In 2018, the total tourism industry's contribution to the world economy was 8.8 trillion USD [2], which represents 10.4% of the global Gross Domestic Product (GDP), 6.5% of the total global exports, 27.2% of the total services exports, and 10% of the total world employment [2]. On this basis, the authors can define tourism industry as the consumption or use of all services and products by travelers throughout their tourism-oriented travel trip. This is the demand aspect definition of the tourism industry. The supply aspect business definitions are like: business could be a cluster of industries providing services and merchandise for varied wants of tourists and travelers on leisure, business, and alternative connected business characterized travel visits. What the tourists pay for the consumption of those services and merchandise becomes the revenues of business [2]. The aforementioned issues can be addressed by the decentralized mechanism called Blockchain technology. Tourism industry wants to merge technology, knowledge, and money to develop an innovative and new platform to fulfill the needs of customers [1].

16.1.1 Blockchain

Blockchain is a chain of blocks where each block contains a set of transactions and transactions are protected by a technology called cryptography. The dynamic features of Blockchain help us to build this operation safer and additionally clear since the accountability gets unfold throughout the entire network. The same can happen with foreign transactions, increasing the level of trust among all the involved parties [7].

The Blockchain is a principled computerized record of exchanges that can be robotized to store monetary dealings as well as all that which holds esteem. It stores data across a progression of PCs on a P2P organization. The information could be gotten progressively which can be seen all through organizations and make it conceivable to see the historical backdrop of the products just as its parts [1].

The above-mentioned Characteristics of Blockchain [1] support to overcome the drawbacks of traditional architecture used in the tourism industry (shown in Table 16.1). Firstly, Blockchain supports secure transaction with a public key encryption algorithm and hashing techniques [4]. Next is the removal of the intermediaries from the traditional architecture by applying Blockchain technology [4]. Third, data integrity will be achieved using the one-way hash function added with the Blockchain architecture. Fourth, the decentralized technology [2, 8] of Blockchain solves the problem of data lost. Fifthly, users' data stored in a Blockchain network are unchangeable until a specified portion of the network agrees to do so. Sixth, Blockchain is famous for its immutable nature. If any hacker is going to hack the data, it's not possible to extract information. In this way, we can easily identify if data had been tampered (Table 16.2).

Trust will be achieved with all users by making a direct communication without a third party and also trust in the security of payment processing [4]. Although

Table 16.1 Areas of impact for the Tourism industry

Source	Areas	Description
Rashideh [4]	Transparency	There is no transparency in the existing traditional architecture
Rashideh [4]	Settlement	Stakeholders facing the financial transactions settlement problems
Bodkhe et al. [2]	Loyalty	No genuine reviews and ratings
Rashideh [4]	Fraud	Not a secured architecture with a centralized server
Bodkhe et al. [2]	Identity	Identity theft problem
Treiblmaier [1]	IoT	Even though the existing architecture uses the IoT but not a complete model
Bodkhe et al. [2]	Overbooking	Due to the centralized nature problem with the traffic flow
Treiblmaier [1]	Policy	No trust with the cost and control
Xu et al. [6]	No accurate contact tracing app	During this COVID-19 pandemic, there is an urgent need for an accurate contact tracing app to stop the further spreading of the virus

Table 16.2 Blockchain Characteristics and benefits (cf. Rashideh [4], Treiblmaier [1], Xu et al. [6])

Areas	Description
Secure Transaction	Algorithm used in Blockchain supports secure transactions using different encryption algorithms to overcome the risk of data deception
Disintermediaries	Blockchain platforms reduce the involvement of intermediaries in the industries
Data integrity	Data modification will not happen
Decentralization	Blockchain is built with no centralized server hence no need for each other identity
Immutability	Data stored in a Blockchain is visible to all users connected to a chain and the same view will be shared on data
Transparency	Blockchain network allows the users to share similar master records and it helps to provide high accuracy and consistency among users
Trust	Blockchain platform supports reliability and security of payment dispensation leading to trust
Accurate contact tracing App	Blockchain-enabled contact tracing app will help to reduce the further spreading of the COVID-19 virus leading to an increase in the visitors to the travel industry

Blockchain is visible to a specified group of users with the same view of data hence it's possible to achieve transparency. Loyalty was also a major issue in tourism industries because travelers' experience is considered for future travelers so the review system should be genuine and trustfulness should also be achieved by Blockchain technology [2]. Finally, Contact tracing is a mechanism to avoid the further spreading of Coronavirus. But currently, contact tracing applications exist with some limitations or issues like (a) less downloads of the tracing application and (b) Denied of location access due to privacy concerns and cyber attacks. The study of incorrect data in a current application has lost trust. The above-stated issues are to be solved by Blockchain-based contact tracing app.

16.1.2 Working of Blockchain Technology

Figure 16.2 shows the working functionalities of Blockchain technology with Transaction operations. As discussed earlier, Blockchain is a node-to-node type of network, the meaning of a Blockchain is clear-cut [8]. A user needs to begin transactions, and on completion, a block is assigned to the said transactions. These blocks are also transmitted to all the nodes in the network and get the said information. After that, the block is mined and evaluated. It is also added to the chain, followed by a successful completed transaction. *Blockchain ledger* is a technique consisting of information about the transaction [9].

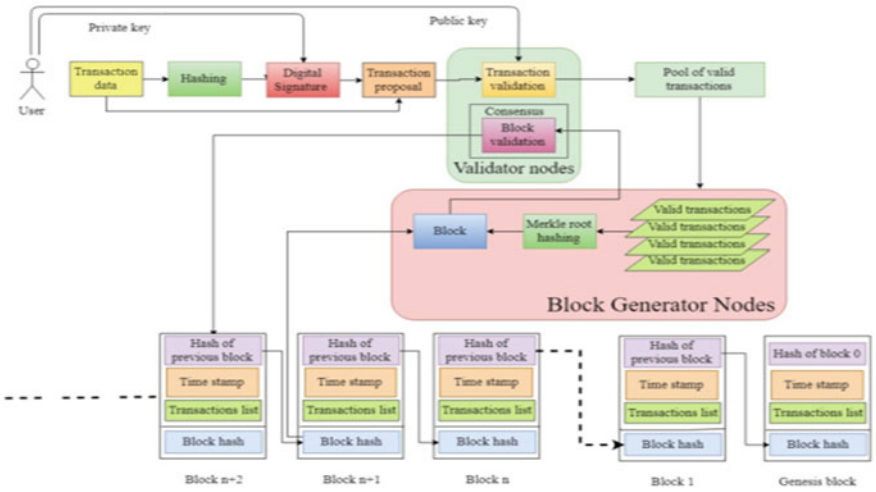


Fig. 16.2 The transaction operation on Blockchain technology [1, 4]

16.1.3 Steps Followed During the Working of Blockchain Technology

Step 1: *Transaction* is a process that changes the status of the Blockchain ledger. As part of the Smart Contract [11], the transaction can be the transfer of a monetary value depending on the application [12].

Step 2: Block header and Block data are part of block in Blockchain which contains a set of transactions. The block header comprises block meta-data information like Merkle tree root hash, the previous block hash, the timestamp, and the Version of block, and finally, the set of valid transactions are performed as data.

Step 3: *Merkle tree root hash*: Hashing Algorithm is used by each block for hashing individually. On combining the hash values for pair hash values are hashed again and again until a single hash value is obtained. This value is called the Merkle tree root hash value [13].

Step 4: *Block hash*: Unique identifier is obtained for a particular block by hashing the block header twice.

Step 5: *Previous block hash*: The current block is the parent of the preceding block. It is the hash of the block preceding the current block in the chain. To ensure the immutability of the Blockchain ledger, it uses the previous block hash value in a block header [14].

Step 6: *Timestamp*: It is the time that specifies the creation of the block.

Step 7: *Block version*: Specifies the Block version of the Blockchain protocols used.

Step 8: *Mining*:

On completion of the transaction, it is assigned a block and informed to all nodes in the network.

Step 9: *Genesis Block*: The first block in the ledger is the Genesis block and all other blocks are linked to the genesis block in the chain.

In general, it comprises the access-control policies, the consent protocol to be used, the hash function, the block creation period, the block size, and the relationship for the network characteristics [13] (Table 16.3).

16.1.4 Blockchain and Tourism

Blockchain and tourism industry have the truly profitable mix as this innovation will offer extra security and straightforwardness to significant touch focuses [15]. In the case of workplace booking flights and hotels for a shopper, it's to send the info to the varied companies. Blockchain technology comes with a bounty of idiosyncratic characteristics and traits which promise benefits for the tourism industry [9].

Table 16.3 Comparison between Traditional architecture and Blockchain-Based architecture in Tourism Treiblmaier [1]

Areas	Traditional architecture	Blockchain-based architecture
Intermediaries	Involvement of third party	Peer-Peer communication
Transparency	No transparency	Transparency
Architecture	Centralized	Decentralized
Trust	No trust	Trust will be achieved
Security	Less security	More security
Loyalty	Fake reviews	Trusted reviews

Table 16.4 Existing blockchain platforms in the tourism industry (*Source* Rashideh [4])

Lock chain	It mainly uses a rent-out property. No intermediaries with no commission
BeeToken	Gives Home distribution Platform
Beenest	It helps to keep users' payment secure and no commission fees
Winding tree	Baggage tracking
ShoCard and SITA	Decentralized character data set administration
Trippki	Guests' loyalty rewards system

The researcher has limited these characteristics which directly or indirectly help in achieving the research objective useful to develop a framework that will lead to a secure decentralized model with a direct communication between various stakeholders [4]. These characteristics are the base of the proposed blockchain-based framework in the next paper. Figure 16.1 highlights the traditional architecture used in the tourism industry. This architecture will help to show the impact of blockchain in the tourism sector and also the disadvantages of the involvement of third party architecture [4] (Table 16.4).

16.1.5 Tourism Demand

Vacationers' own encounters, just as openness to others' very own encounters intensified by web-based media, can affect their movement perspectives, goals, and future conduct. As per research, hazard discernments vary contingent upon the traveler's beginning country, extreme objective, age, sex, and travel type [5].

COVID-19 (proclaimed as a pandemic by WHO, 12 March 2020) altogether impacts the worldwide monetary, political, and socio-social frameworks. The travel industry is utilized to and has become tough in skipping back from emergencies and episodes (e.g. terrorism, earthquakes, Ebola, SARS, and Zika) [5].

16.2 Literature Survey

Bodkhe et al. [2] anticipated a framework called BloHosT for the registration of the tourist user and tourism stakeholders through a single unified cryptocurrency-enabled application. Also, stakeholders can execute smart contracts for interoperability. The TeDL design framework also proposes to generate rating scores for future travelers across the globe by running long short-term memory over the previously saved traveled itinerary. Using case studies, the authors have also predicted the suitability of their framework in the tourism industry. The framework provides end-to-end security. However, it is not a complete blockchain-based framework [2].

Sigala [5] said 'ripe' new setting whereby the travel industry experts can coordinate examination with significant end-customer. However, COVID-19 the movement business investigation should attempt to avoid the 'disperse or pass on' old assessment designs and develop the responsibility and backcountry of the movement business assessment and industry. It's the commitment analysts to guarantee that COVID-19 the movement business assessment can guarantee the last referenced. Covid the movement business research shouldn't simply uncover such The assessment didn't moreover fuse other huge the movement business accomplices like the movement business laborers, neighborhood organizations, the movement business visionaries and the movement business tutoring (scientists, understudies working costs of the movement business firms), COVID-19 the movement business examination ought to frantically investigate issues The COVID-19 consequences for the movement business make further squeezing factors on the movement business preparing that movement industry social undertakings having the chance to make social worth, tackle social issues made by the COVID-19 and supply help to people up the creek without a paddle [5].

Kansakar et al. [16]. The relaxation and neighborliness industry become the main thrust of the worldwide financial system. The far-reaching appropriation of novel innovations in this business over late years has generally reshaped the manner by which administrations are also given. In this paper, we investigate a portion of the condition of the art advancements presently utilized in the accommodation business and furthermore, how they are further developing visitor encounters and evolving the neighborliness administration stage. We likewise imagine some latent future friendliness administrations we can expect as the Internet of things (IoT) innovation continues to develop. We perceive that the innovative spine of numerous friendliness foundations needs to be upgraded to work with the changing scene of innovation in the advanced world. We talk about some essential provokes that should be defeated to establish an enduring future proof answer for the friendliness business. We additionally address the issues these difficulties present for visitors and friendliness administration suppliers (HSP) [16].

Rashideh [4] proposed a Framework for an expanded degree of disintermediation. Blockchain is an amazing innovation that offers benefits in expanding the degree of eliminating delegate layers inside ventures like the travel industry. This innovation quickly works on the few requirements of the movement business, for

example, its effective plan of action, Payment frameworks, security, execution, and trust, and so forth this evaluation of blockchain innovation followed the subjective technique to accomplish both the examination destinations and to respond to the inquiries presented at the start of this examination. Blockchain's potential positive effect offers a decent chance for different organizations to bring about comparative advantages by embracing this innovation. This exploration will help specialists in the movement business come out as comfortable with blockchain attributes and advantages. Scientists will perceive its handiness and how it's anything but a fundamental piece of the innovation they use. For a various reasons relying on mediators become a major source of concern for some people [4].

Xu et al. [6] said this arrangement gives an in time system to governments, specialists, organizations, and exploration foundations over the world to foster a confided in stage helpful for battle with the COVID-9 plague by following data sharing. The future work in regards to the extensible module security highlight for Beep Trace structure is being scrutinized. In the interim, the likely arrangement of removing the latent message work from Beep Trace and making the system into an overall Blockchain-empowered Privacy-saving Messaging administration (Beep Mess) is under arrangement. Notwithstanding the specialized part of Beep Trace, we are anticipating to carry out the Beep Trace in the current contact following plans with business off-the-rack (COTS) cryptography components. BeepTrace is being worked on and it has additionally gotten a greeting of preliminaries in the teamed up colleges and nearby networks, to rejuvenate Beep Trace soon. Also, we are in dynamic pursuit of possible answers for bringing down the escalated computational necessity of blockchain hubs with downsize practice endeavours, and the impact of downsizing the organization can diminish the opposition and correspondence prerequisite among peer hubs, however, yet to confirm the security and security assurance execution [6].

Erceg et al. [3] Tourism has improved essentially since the web permitted clients to look for and make their outings without the need to go to an office. The travel industry should incorporate information, cash, and innovation for the point of making new plans of action. One of the quickly developing advances which is likewise one among the state-of-the-art innovations entering the travel industry is 'blockchain'. The execution of this innovation can assist with expanding upper hand, further develop consumer loyalty, and upgrade execution, as per the creators of the paper. The goal of this study is to look into the travel industry's ability to implement blockchain innovation in the Republic of Croatia and the Republic of Macedonia tourism industry [3].

Treiblmaier [1] shows how Blockchain innovation can possibly significantly change the travel industry. Its highlights like unchanging nature, straightforwardness, programmability, and decentralization leave imaginative approaches to style client connections. It can likewise empower novel authoritative constructions and measures and work with new types of inter-organizational joint effort. It might actually cause the disintermediation of officeholders and the development of most recent parts in the travel industry area [1].

Detotto [17] directed the examination study to evaluate the job of administration quality in the age of the travel industry income. Creator said that discoveries depict that higher saw administration quality, from an expansive perspective, altogether affects the travel industry income factor. All in all, administration adequacy and administrative quality essentially affect the capacity of a nation to create the travel industry income. Second, the apparent ability of an administration to carry out an administrative structure that advances private area action emphatically affects the travel industry inbounds consumption. The presentation and productivity comprise the travel industry administration [17].

Ozdemir [18] said, we contend that the most essential yield of this investigation is the proposed blockchain fundamentals measures set that empowers specialists and leaders to evaluate blockchain applications. In light of our examinations of the flow research, that set contains rules, for example, blockchain administration model, blockchain stages, kind of agreement, utilization of digital currency, utilization of keen agreement, and utilization of tokens. For instance, it's extremely mind-boggling interaction to distinguish blockchain administration frameworks. It likewise assists with discovering its qualities including namelessness, permanent, productivity, straightforwardness, and information trustworthiness. Partners in the travel industry will get the advantage. We contend that having the option to perform such assessments is priceless to fathom and further develop current blockchain applications [18].

16.3 Challenges

After studying the above research papers, we have come up with the challenges in tourism and hospitality management; they are (1) Removing of the third party, (2) Security, (3) Payment system, (4) Review ratings, (5) high cost, and 6) Accurate Contact Tracing app [19]. Disintermediaries is one of the major issues faced by the customer in terms of cost while making the hotel reservation and other services used in tourism [18]. Also there are no such security measures taken during sharing of personal data and online transactions in the current traditional architecture [8]. The payment settlement is also one of the major issues in existing architecture involvement of third parties leading to payment failure or delay in payment. Customer review is one of the major tools to attract new travelers but in the existing architecture, there is no trust in the review system [20]. Finally, there is no accurate contact tracing application available to trace the COVID-19 infected patients while traveling because of so many reasons/vulnerabilities exist in current applications like identity theft, no accurate data, etc. [19].

16.4 Discussion

To overcome the above challenges, we have come up with a solution called Blockchain-based tourism and hospitality management system [21]. It helps the Visitors to check in safely and securely meaning whenever the visitors visit the tourist place they have to install the Blockchain integrated app by entering the pin or fingerprint. Then he/she needs to scan the QR code which is available in the tourist place without sharing any personal information [6]. By using this information, the application can monitor the visitors, if any symptoms of the virus are found, inform the nearby health workers for further process to trace the contacts of the infected person (tourist) [6]. In this way, our proposed system will work in the combat of COVID-19 disease in tourism management [5]. Our proposed architecture was built up with Blockchain technology with a distributed system using a cryptographic hashing algorithm; hence, the security is high. By using this proposed architecture, visitors can directly access the services with a secure and convenient payment gateway. Our application BLOBIA also helps customers to interact with stakeholders directly without the intervention of the third party (disintermediaries) [4] and the security, financial settlement problems that exist with the intermediaries are also resolved by using blockchain technology.

Improved security [14], eradication of middleman [10], data integrity, decentralized business models, Immutability [20], transparency, trust [21], cost, and traceability are some of its characteristics. As we know, the travel industry should be founded on interests into applications that will give customers a starting point when deciding whether to visit on objective or not and make it easier for vacationers to stay at their locations. In light of this, we have studied numerous research papers that provide information on how the tourism sectors will resist improving the countries economic expansion. Not only must country grow, but it must also meet customer demands in the tourism and hospitality sectors in terms of safety, security, high cost, trust, disintermediation, transparency, and other factors [4].

The real-time example which I was faced by the third party provider (**Flipkart)** is as follows: A few months back I was planned to travel from Bangalore to Delhi through Airline by booking the air ticket through third party provider (intermediaries) like Flipkart App and also got the ticket confirmation from Flipkart. But on the day of travel, I got the message regarding flight cancelation due to some problem with airways and with a guaranteed refund of flight ticket fare. After that I contacted Air India airlines for a refund. But Airlines refused to refund the amount by stating that I had booked a flight ticket through a third party provider (Flipkart) and they will refund the amount by communicating with the airlines. Once I got the clarification from Air India I contacted Flipkart customer care for refunding of the amount till now I haven't received any refund from both the parties. These were some of the problems faced by many customers by Intermediaries with a financial settlement. After experiencing this type of problem, I had planned to propose a framework that overcomes the difficulties of traditional architecture [4].

16.4.1 Blockchain E-Passports for Tourism Industry

Blockchain Company ShareRing has dispatched a mysterious e-identification application with contact tracing [22]. The application permits clients to transfer significant documentation, for example, e-visa on appearance (eVOA), identification data, travel protection, flight and convenience appointments, and a negative COVID-19 test outcome. At the point when a client tests positive for the infection, the application will namelessly send the important data to the public authority, permitting others to be informed on the off chance that they have been in touch [17, 23].

"When someone signs up for a ShareRing ID, we take their photo, video selfie, name, DOB, address, etc. and store it in an encrypted file that never leaves the user's device. We also take a 'fingerprint' of the data and documentation and store that on the Blockchain." Covid-19 is a new passport app developed by ShareRing, which aims to make it easier for governments and businesses to integrate the passport app on existing solutions without using the ShareRing brand. The company's co-founder says the app is low cost and appealing to consumers. "It encourages adoption among the population by building it in a way that probably safeguards their privacy," he said [24].

16.4.2 Robots Service as a Tool for Physical Distancing in Tourism

The COVID-19 pandemic is adversely influencing the travel industry and friendliness industry. Administration robots can make a mechanical safeguard among vacationers and representatives that builds the physical and enthusiastic distance between them. The travel industry organizations need to supplement robots with different advancements to give social connectedness and offset the adverse results of physical distancing [16, 25].

16.4.3 Contact Tracing

Contact following permits us to disconnect the infection when it shows up and keep it separated, says Dr. Watson. "Contact following is the best device we need to oversee COVID-19 in a continuous manner," she says. The John Hopkins Institute has been utilizing contact following for a very long time.

The World Health Organization suggests fast analysis and confinement of cases, thorough following, and preparatory self-seclusion of as of late uncovered close contacts. To forestall ahead transmission, the individual who tests positive for COVID-19 is either isolated or holed up. From that point onward, contact following is completed to speed the revelation of an influenced individual's contacts to forestall

further transmission. This cycle is called contact following, and its motivation is to forestall conceivable transmission chains [22, 25].

As of now, contact following exercises depend on the memory of the tainted individual to give a rundown of contact-people to the medical care experts or COVID-19 reaction group. In any case, there have been occurrences when the contamination tainted individual can't recall his/her nearby contact and doesn't have contact subtleties. Innovation-based contact following applications can work with and computerize the cycle. This can be empowered by utilizing a worldwide situating framework (GPS), remote innovation, Bluetooth innovation, social diagram, network-based API, versatile following information, card exchange information, and framework actual location [6, 22].

16.4.4 Possibilities for Integrating Contact Tracing with Arising Technologies

With fast expansion in COVID-19 cases and deaths [26], the reconciliation of arising advancements into contact following exercises is inescapable. Germany carried out a contact following application that executes on savvy watches and screens the client utilizing heartbeat, temperature, and rest design. Singapore executed a Bluetooth-based cell phone application, Trace Together, to help contact following efforts [22]. More instances of contact following applications are portrayed in Table 16.6 [6, 22] (Table 16.5).

Difficulties of coordinating arising advancements into COVID-19 contact following exercises [22]. COVID-19 contact following is a versatile contact following application created by the Department of Defense DoD for use in crisis circumstances (Table 16.6).

Benefits of Blockchain

Greater Transparency, Better Securities, Immutability, Increased Efficiency, and Improved Traceability [1, 3].

16.5 Conclusion

Tourism is a trending industry in the current era of the economic growth of the country but the current traditional architecture used in the tourism sector has some limitations like Identity theft [27], Intermediaries [4], Payment failures [9], fake reviews [2], transparency [9], settlement [9], overbooking, policy, and most importantly safety concerns due to COVID-19 pandemic [25]. The literature review paper will make a comparison with all existing architecture which will lead us to decide our objective. After taking the input from the literature review, we have identified the Gaps in

Table 16.5 Openings for coordinating contact following with arising advancements for Tourism Industry in Future Smart Society (*Source* Mbunge [22])

Arising innovation	Potential exercises of COVID-19 contact following applications
Artificial intelligence	The Irish government has announced plans to introduce a new system for monitoring the body temperature of people who are suspected of being at risk of heatstroke in the Republic of Ireland COVID-19. It aims to make it possible for people to check their body temperature using Smartphone, smartwatches, smart thermometers, and other devices
Internet of Things-IoT	Permit specialists to have continuous admittance to COVID-19 information for contact-people. Remote checking of COVID-19 patients in isolate focuses as well as self-seclusion
Geographical information systems(GIS)	The GIS will provide current time tracking of patients and contact-persons' movements, as well as providing information on the impact of the epidemic on the health of the local population
Big data	Be associated with public wellbeing data frameworks (HIS) and store information (portability designs) on an ongoing reason for additional investigation. This aids wellbeing policymakers to viable screen COVID-19 and allotment of assets. Raise mindfulness, training, and correspondence. Support video conferencing exercises
Blockchain technology	Confirm and approve COVID-19 patients before the arrangement of guiding administrations and managing patients' drug in isolate facilities. Encrypt wellbeing information as it moves from one hub to the next in a shared engineering
5G technology	Increase the size of the transmission capacity and the rate at which information is moved to ensure continuous access to information. Teleconferences between medical exports, COVID-19 patients, and contact persons are possible.

the tourism industry. In the future we will clearly define the solution to fulfill the Gaps using Blockchain technology. After finding the solution, analysis will propose a novel framework using Blockchain technology called '**BLOBIM** (Blockchain-Based Integrated Model)' for the tourism industry. This will provide a trustworthy rating system with a decentralized model and accurate contact tracing application for Tourism Industry in the Future Smart Society.

Table 16.6 Difficulties of coordinating arising innovations into COVID-19 contact following exercises (*Source* Mbunge [22], Xu et al. [6])

Source	Challenges	Description
Mbunge [22]	Asymptomatic individuals	Asymptomatic people communicate COVID-19, however, they don't have indications of the infection. This postures contamination and control difficulties yet their commitment isn't notable. This implies contact following stressing more on close contacts is probably not going to be adequate except if clinical testing is finished
Mbunge [22]	Surfaced significant technical limitations	The Blue Trace software for Bitcoin is a contact tracking system that allows users to track down their phone numbers, emails, and other information on the Bitcoin network. To build, install, configure, and manage the system, highly trained specialists are required. Many regions of the world have a scarcity of such specialists, making it difficult to locate them
Xu et al. [6], Mbunge 22]	Absence of supporting data and correspondence	Combination of arising innovations into wellbeing frameworks is as yet incipient in certain nations. There are no rules or strategies on electronic information assurance in the COVID-19 pandemic approach. Nations like Chad and the Central African Republic with poor ICT foundation think that it is difficult to send restricted assets toward mechanical advancement

(continued)

Table 16.6 (continued)

Source	Challenges	Description
Mbunge [22]	Socio-economic inequalities	Interfacing individuals with high-hazard bunches requires web association and figuring gadgets which can be costly. Organization specialist co-ops along with the influenced government can zero rate COVID-19 contact following applications or decrease the expense of web information, report says
Mbunge [22]	Deactivation of cell phones' WIFI, GPS administrations	The current contact following applications depends on Bluetooth conventions, WIFI, and GPS innovation to screen developments of contaminated individuals. These advancements can be hindered by the vicinity, heterogeneity of conventions, and satirizing of GPS signals. Individuals have the rights to design their cell phones 'association settings'
Velliangiri and Karthikeyan Karunya [13]	Inter-operability and standardization issues	Contact following applications work and embrace the Internet of Things. Information from patients and contact-people is moved promptly continuously to the public wellbeing data frameworks. These contact following applications are uniquely crafted and not normalized around the world, which implies every nation has its own information designs

(continued)

Table 16.6 (continued)

Source	Challenges	Description
Xu et al. [6, Velliangiri and Karthikeyan Karunya 13]	Security risks	Contact following applications abuse contact-people's information security, secrecy, respectability, and accessibility of information and now and again connect moral and protection of patients. There is a need to ensure against wellbeing information misfortune and unapproved access when carrying out arising advancements in handling COVID-19. Most COVID-19 contact following applications support simultaneous access of wellbeing information
Xu et al. [6, Velliangiri and Karthikeyan Karunya 13]	Privacy issues	Wellbeing frameworks have not considered a singular's security in contriving contact following arrangements. Some versatile applications encroach on individuals' protection as they gather, break down, and approach individual wellbeing information like wellbeing conduct, status, voyaging history, family facilitates, and area
Velliangiri and Karthikeyan Karunya [13]	Political and structural responses	Absence of political will and contribution of state and neighborhood governments seriously influence the reception of electronic wellbeing intercessions. Notwithstanding the convergence of ICTs, further developing medical care administration conveyance isn't focused on in certain nations. This could be credited to cost related with setting-up, reconciliation, and upkeep of arising advancements

(continued)

Table 16.6 (continued)

Source	Challenges	Description
Velliangiri and Karthikeyan Karunya [13]	Ethical and legal risks	There are four kinds of COVID-19 computerized applications. They were produced for pandemic administration, isolate control, following and contact following, indications testing, and stream demonstrating. These computerized applications compromise standards of general wellbeing morals and information morals by giving little protection to tainted people
Velliangiri and Karthikeyan Karunya [13]	Symptoms checkers	Keen and computerized thermometers are being utilized as indicative reconnaissance apparatuses in Spain. Keen thermometers matched with portable applications are empowering medical services experts to gather temperature information from individuals. Manifestation checkers are savvy for screening COVID-19 suspects during emergency and scatter populaces
Velliangiri and Karthikeyan Karunya [13]	Consent and voluntariness	Contact following applications ought to permit individuals to rehearse assent withdrawal. Information sharing dependent on assent mitigates protection chances, however, there are obstructions. Language obstructions, absence of contact following applications, and absence of decision to deny assent are featured. There must be a harmony between assent withdrawal and COVID-19 regulation

(continued)

Table 16.6 (continued)

Source	Challenges	Description
Velliangiri and Karthikeyan Karunya [13]	Abuse of contact following applications	Most contact following applications use GPS and Bluetooth, which can be utilized by pernicious clients for following. Contact following applications should fuse step acknowledgment and other action acknowledgment techniques that use biometric confirmation to keep bogus data from being shipped off wellbeing specialists
Velliangiri and Karthikeyan Karunya [13]	Discrimination	Contacts following applications represent a high danger of segregation particularly among the influenced people. Information from contact following applications can impact the assignment and dissemination of COVID-19 assets. This information can prompt segregation and crazy conduct in certain spaces. The application is accessible for Android and iOS gadgets
Mbunge [22]	Digital divide	In certain main lands including Africa, the advanced separation stays a. challenge in numerous wellbeing frameworks. The reconciliation of contact following applications accepts that the entire populace approaches. Current advancements yet the hole actually exist between ICT access and usages among the populace

Acknowledgements Our aspiration is to express our heartfelt gratitude to His Holiness "**Jagadguru Sri Shivarathri Deshikendra Mahaswamiji"** Pontiff of Sri Suttur Math.

I extend my sincere thankfulness for continuous remarkable motivation and support from my Guide Prof.Dr.Parma Nand and Co-Guide Prof.Dr.Vikram Bali and my family members for completing this research paper.

Funding No funding.

Conflict of Interest Authors declare that we don't have a difference in interest.

Ethical approval In this paper, we the authors are not focusing our study on an individual participant or animal.

References

1. Treiblmaier H (2020) Blockchain and tourism. In: Handb. e-Tourism, pp 1–21. https://doi.org/10.1007/978-3-030-05324-6_28-2
2. Bodkhe U, Bhattacharya P, Tanwar S, Tyagi S, Kumar N, Obaidat MS (2019) BloHosT: blockchain enabled smart tourism and hospitality management. In: CITS 2019-Proceeding 2019 international conference on computer, information, and telecommunication systems, pp 1–5. https://doi.org/10.1109/CITS.2019.8862001
3. Erceg A, Sekuloska JD, Kelic I (2020) Blockchain in the tourism industry-a review of the situation in Croatia and Macedonia. Informatics 7(1):1–16. https://doi.org/10.3390/informatics7010005
4. Rashideh W (2020) Blockchain technology framework: current and future perspectives for the tourism industry. Tour Manag 80(April):104125. https://doi.org/10.1016/j.tourman.2020.104125
5. Sigala M (2020) Tourism and COVID-19: impacts and implications for advancing and resetting industry and research. J Bus Res 117:312–321. https://doi.org/10.1016/j.jbusres.2020.06.015
6. Xu H, Zhang L, Onireti O, Fang Y, Buchanan WB, Imran MA (2020) BeepTrace: blockchain-enabled privacy-preserving contact tracing for COVID-19 pandemic and beyond, vol 4662, no c, pp 1–14. https://doi.org/10.1109/jiot.2020.3025953
7. Tasatanattakool P, Techapanupreeda C (2018) Blockchain: challenges and applications. In: International conference information network, vol 2018-Janua, pp 473–475. https://doi.org/10.1109/ICOIN.2018.8343163
8. Zheng Z, Xie S, Dai H, Chen X, Wang H (2017) An overview of blockchain technology: architecture, consensus, and future trends. In: Proceeding-2017 IEEE 6th international congress on big data, BigData Congress, pp 557–564. https://doi.org/10.1109/BigDataCongress.2017.85
9. Dutta P, Choi TM, Somani S, Butala R (2020) Blockchain technology in supply chain operations: applications, challenges and research opportunities. Transp Res Part E Logist Transp Rev 142(July):102067. https://doi.org/10.1016/j.tre.2020.102067
10. Sayadi S, Ben Rejeb S, Choukair Z (2019) Blockchain challenges and security schemes: a survey. In: Comnet 2018-7th international conference on communications and networking, 2019. https://doi.org/10.1109/COMNET.2018.8621944
11. Panchamia S, Byrappa DK (2017) Passport, VISA and immigration management using blockchain. In: Proceeding-23rd annual international conference in advanced computing and communications ADCOM 2017, pp 8–17. https://doi.org/10.1109/ADCOM.2017.00009
12. Khare A, Khare A (2010) Travel and tourism industry yet to exploit the Internet fully in India. J. Database Mark Cust Strateg Manag 17(2):106–119. https://doi.org/10.1057/dbm.2010.9

13. Velliangiri S, Karthikeyan Karunya P (2020) Blockchain technology: challenges and security issues in consensus algorithm. In: 2020 International conference on computer communication and informatics, ICCCI 2020. https://doi.org/10.1109/ICCCI48352.2020.9104132
14. Draper A, Familrouhani A, Cao D, Heng T, Han W (2019) Security applications and challenges in blockchain. In: 2019 IEEE International conference on consumer electronics. ICCE 2019, pp 1–4. https://doi.org/10.1109/ICCE.2019.8661914
15. Önder I, Treiblmaier H (2018) Blockchain and tourism: three research propositions. Ann Tour Res 72(March):180–182. https://doi.org/10.1016/j.annals.2018.03.005
16. Kansakar P, Munir A, Shabani N (2019) Technology in the hospitality industry: prospects and challenges. IEEE Consum Electron Mag 8(3):60–65. https://doi.org/10.1109/MCE.2019.2892245
17. Detotto C, Giannoni S, Goavec C (2021) Does good governance attract tourists? Tour Manag 82(March):104155. https://doi.org/10.1016/j.tourman.2020.104155
18. Ozdemir AI, Ar IM, Erol I (2020) Assessment of blockchain applications in travel and tourism industry. Qual Quant 54(5–6):1549–1563. https://doi.org/10.1007/s11135-019-00901-w
19. Işın A (2019) Current issues in tourism and hospitality management. Curr Issues Tour Hosp Manag (November):193–206
20. Bodkhe U et al (2020) Blockchain for Industry 4.0: a comprehensive review. IEEE Access 8:79764–79800. https://doi.org/10.1109/ACCESS.2020.2988579
21. Calvaresi D, Leis M, Dubovitskaya A, Schegg R, Schumacher M (2019) Trust in tourism via blockchain technology: results from a systematic review. Inf Commun Technol Tour 2019(January):304–317. https://doi.org/10.1007/978-3-030-05940-8_24
22. Mbunge E (2020) Integrating emerging technologies into COVID-19 contact tracing: Opportunities, challenges and pitfalls. Diabetes Metab Syndr Clin Res Rev 14(6):1631–1636. https://doi.org/10.1016/j.dsx.2020.08.029
23. Iyengar K, Upadhyaya GK, Vaishya R, Jain V (2020) COVID-19 and applications of smartphone technology in the current pandemic. Diabetes Metab Syndr Clin Res Rev 14(5):733–737. https://doi.org/10.1016/j.dsx.2020.05.033
24. Wang C, Luo J, Lin S, Zhang J, Wang Z, Luo A (2018) Cloud platform for the management of tourism resources. In: International Conference Geoinformatics, vol 2018-June, pp 1–9. https://doi.org/10.1109/GEOINFORMATICS.2018.8557190
25. Seyitoğlu F, Ivanov S (2021) Service robots as a tool for physical distancing in tourism. Curr Issues Tour 24(12):1631–1634. https://doi.org/10.1080/13683500.2020.1774518
26. Bhatt T, Kumar V, Pande S, Malik R, Khamparia A, Gupta D (2021) A review on COVID-19. Stud Comput Intell 924(April):25–42. https://doi.org/10.1007/978-3-030-60188-1_2
27. Oksiiuk O, Dmyrieva I (2020) Security and privacy issues of blockchain technology. In: Proceeding-15th International conference on advanced trends in radioelectronics, telecommunications and computer engineering. TCSET 2020, pp 531–535. https://doi.org/10.1109/TCSET49122.2020.235489

Chapter 17
Smart Society 5.0 for Social and Technological Sustainability

Reena Thakur, Pradnya S. Borkar, and Mohit Agarwal

Abstract Based on recent publications, we hope to gain a better understanding of the framework of Society 5.0 in commercial, as well as its latent for proactivity in the Sustainable Development Achievements, as well as possible consequences on sustainability and accounting reporting. Given the breadth of the deviations forced by the 4.0 industry, it is very critical to comprehend the current effects and forthcoming opportunities of this progress with Cross-industry 5.0. All residents are expected to become dynamically involved in Society 5.0, incorporating digital technology into a range of systems and speeding up their implementation. As a result, it is projected to expand on the notion and possibility of the specific technology interaction in the advancement of their own and society's quality of life, with an emphasis on sustainable development, sustainability, and reporting. As a result, it is a contribution that aims to contribute to this extremely important and vital conversation by offering its progression. It is determined that Industry 4.0 and, more newly, Industry 5.0 are here to remain, putting artificial intelligence (AI), big data, robotics, and other technologies by the provision of man in a world where everything will be connected and society will have to adapt. This position paper aims to create a theoretically focused perspective on sustainable digital inventions in the framework of Society 5.0, as well as the underlying empirical potentials of this association.

17.1 Introduction

The Government of Japan released information on the "Fifth Basic Plan for Science and Technology (2016–2020)" in January 2016. The effort, dubbed "Society 5.0," aims to build a supportable society that also adds to people's comfort and safety.

R. Thakur (✉) · P. S. Borkar
Jhulelal Institute of Technology, Nagpur, India
e-mail: rina151174@gmail.com

M. Agarwal
School of Engineering & Technology, Sharda University, Greater Noida, India

V. Bali et al. (eds.), *Decision Analytics for Sustainable Development in Smart Society 5.0*, Asset Analytics, https://doi.org/10.1007/978-981-19-1689-2_17

The term "Society 5.0" refers to a group of people who have come together as individuals who are dependent on a particular cyber-physical system [1, 2]. Numerous systems are included in it. For the moderation of both global and local threats, systems are connected via the Internet.

Society 5.0 is a concept that tries to solve societal challenges from a fresh perspective. Different aspects of society would be connected in this new era, and technology would link a super-intelligent society with full integration of artificial intelligence (AI), the Internet of Things (IoT), big data, and people facilities to ease physical and digital infrastructures for humans. The goal of this project is to build societal foundations that allow anyone to produce value at any period and in any location, in a harmless atmosphere, and in accordance with natural settings, deprived of the constraints that already exist [3].

The chapter is organized such as Sect. 1.1 gives the introduction; in Sect. 1.2, we focus on what is Society 5.0. In Sects. 1.2.1, 1.2.2, and 1.2.3, we described the revolution of Society 5.0 in detail followed by Societal effects of 4IR, evolution, respectively. Section 1.3 highlights the Schema of Society 5.0. Section 1.4 gives new solutions for new problems with detailed applications. Section 1.5 shows how we enter into Society 5.0 with examples. In Sect. 1.6, Cybersecurity 5.0 is explained. Then in Sect. 1.7, we explained the difference between real space and cyberspace followed by Sect. 1.8 describing what is the use of cyberspace andg Sects. 1.8.1 and 1.8.2 wherein working of cyberspace and the difference between the Internet and cyberspace are described, respectively. In Sect. 1.9, we listed 5.0 society. Section 1.10 gives Sustainability and environmental harmony and Sect. 1.11 Equality and Sustainability 5.0 with some important key terminologies. Section 1.12 describes Sustainable Digital Innovations in Society 5.0.

17.2 What is Society 5.0?

Society 5.0 is a scenario of a world, in which people and machines work together to "co-create" answers to societal problems by combining the Internet and real space. And, contrary to popular belief, most of the technology we will require is already available.

Society 5.0 refers to a new era marked by transformations fueled by scientific and technical innovation in order to address societal difficulties such as declining birth rates, elderly populations, and environmental and energy concerns. Advanced technologies such as the Virtual Reality (VR/AR), Internet of Things (IoT), Big Data, Artificial Intelligence, and Robotics have a significant impact on people's quality of life.

Human-centred society is a Society 5.0 which stabilizes financial growth with the determination of communal issues through a scheme that tightly integrates physical space and cyberspace; goods and services will be made available to people irrespective of their language, location, gender, age, or other limitations.

The idea of Society 5.0 has been established primarily in Japan, defining a perfect situation in the direction of which each country should advance to fully influence incessant technological revolutions, thus benefiting all their peoples [4].

Society 5.0 can be understood by building a unified common service model that standardizes and combines data formats, low code development, microservices, API, open innovation, information processing technologies, databases, advanced systems, projects, systematic services, visualization, and so on.

The innovations related to the transport, handling, and gathering of data on the Internet are crucial advancements in structuring our world-leading great smart society and extracting value from vast amounts of data. As a result, Japan will expedite the formation of the central invention that comes with it.

- Cybersecurity: invention that supports secure communication and data, taking into account the characteristics of the Internet of Things, like lengthy life cycles from formation to transfer.
- The system architecture technology of the Internet of Things enables the presentation of programming and equipment as segments, as well as the construction and operation of large-scale structures [5].
- Big data (BD) analytics is a type of innovation that involves extracting knowledge and value from a large amount of data, particularly unstructured data.
- AI is a technology that underpins IoT, massive data analysis, and enhanced communication.
- Device technology is an advancement that allows for the fast, actual preparation of huge quantities of data while overwhelming minimal power.
- Network innovation is a type of technology that allows enormous amounts of data to be accessed at a high rate.
- Edge computing is a technology that permits quicker and more precise real-time giving out at the system level, which is critical for encompassing the IoT's utility.

17.2.1 The Revolution: Society 4.0

The incorporation and dissemination of calculating power utilized for automation in information technology and production ushered in the Industry 3.0 revolution, which saw a move from electronic and mechanical devices to digital technology. Nowadays, the Industry 4.0 Revolution is based on the third digital revolution, which is being shaped by a confluence of technologies distorting the boundaries between the digital, biological, and physical realms.

Because of the speed, scope, and severity of this change, we must reconsider in what way countries should grow, how businesses should generate worth, and even what it means to be communal. Everybody benefits from harnessing technologies to create an inclusive, citizen-centred future, with leaders, lawmakers, and people.

Industry 4.0 is a term that refers to the digital revolution in transforming economies, allowing for more inexpensive digital breakthroughs, products, and services to better people's lives all around the world. Taking advantage of rapid

breakthroughs in AI and IoT technology, billions of individuals are associated via hand-held devices that have unrivalled computing power, storage space, and access to vast knowledge, revolutionizing companies around the world.

Remodelling machines gain the skill to communicate and share important data with nominal social interaction, thanks to a combination of operational technology and Industrial Internet of Things (IIoT).

Every day, new breakthroughs in disciplines like artificial intelligence, robots, the Internet of things, autonomous cars, quantum computing, energy storage, material science, biotechnology, nanotechnology, and 3-D printing expand the possibilities. Computational design, synthetic biology, material engineering, and additive manufacturing are all being combined with biological science to orchestrate an association between human bodies, commodities, microorganisms, and even our houses.

We are in the early phases of the Fourth Industrial Revolution (4IR), a period of tremendous technical improvement in "cyber-physical systems" that has resulted in dramatic changes in society and a re-imagining of manufacturing through digitization.

To increase output, the First Industrial Revolution used steam and water power. The second did the same thing with electricity. Computers and automation were utilized to speed up production during the Third Industrial Revolution.

The Fourth Industrial Revolution (4IR) relates to our present economic world's fundamental change towards a new approach based on the merger of digital and physical worlds in cyber-physical systems and the increasing use of emerging technologies like AI, Robotics, IoT, and cloud.

A subtype of 4IR centred on the use of cyber-physical systems and artificial intelligence to alter manufacturing which is the most well-known example of 4IR transformation is Industry 4.0. Industry 4.0 is already transforming the way we produce goods. By introducing machine and system autonomy, it expands on the capability of computerized automation. Huge volumes of industrial data are collected and processed by computers, machine learning, big data, and, increasingly, artificial intelligence (AI), which is now of the "narrow" or "weak" variety, via wireless networks of receivers, sensors, and processors.

These self-contained configurations of digital and physical computing elements are interested in real learning. They make decisions based on super-fast analysis of actual and previous data collected from the manufacturing environment to continually enhance production operations.

The demand for human labour was lessened during the First Industrial Revolution. The second improved productivity by automating massive production lines. The third used computers to further automate these procedures, but production was still managed by people.

The 4IR goes much further, rendering human participation in industrial applications nearly obsolete. Smart factories, for example, are envisioned as self-contained cyber-physical systems (CPSs) in which humans are only required for specialized tasks, machine maintenance, high-level network management, and strategic leadership. In terms of the technologies used and the concept of uniting the biological worlds, physical, and merging of cyber, Society 5.0 is akin to 4IR. Society 5.0, on

the other hand, is a broader notion that encompasses more than just industry and trade and envisions a full overhaul of our way of life. Society 5.0 is a human-centred suggestion that aims to tackle societal problems by using the same linkages between cyberspace and physical space as 4IR.

17.2.2 Societal Effects of 4IR

What will happen to the nature of labour, communities, and social structures if AI and automation eliminate many human jobs?

What will occur to economies as a result of medical advances and an ageing population? What will become of the environment when human productivity and consumption expand?

These are wicked problems, even though they are the result of largely positive trends towards more widespread human wellbeing. And they would not be vexing us were it not for technology.

Of course, this does not make technology bad, or even good—it is agnostic—but it does raise the question: if we used technology to get ourselves into these dilemmas, can we use it to get ourselves out?

The idea of Society 5.0 is a resounding affirmative response to that question. It claims that we may establish a forward-thinking society in which everyone can live an active and joyful life by incorporating technology deeply as shown in Fig. 17.1.

17.2.3 The Evolution: Society 5.0

Data Overabundance to store, identification of real and applicable data to evaluate, and limited possibility of act because of physical ability and a deficiency of laws and guidelines are putting a stress on the nations' current social infrastructure, economic, and industrial, preventing them from taking satisfactory measures to determine and resolve any critical issues in a timely manner as shown in Fig. 17.2. Globalization and life expectancy are increasing, as are economies' progress, international competitions, and social and regional inequities. Sustainability in all industries, as well as Social Innovation, Climate Control, and Green Energy, are critical.

The immense possibility of the Industry 4.0 Revolution is paving the technique for countries to the encirclement of Civilization 5.0, a future authenticity that will serve as a milestone to a thriving data coordinated Super Smart human-centred society. Social Innovation is the process of using new technologies such as Big Data, Robotics, AI, IoT, and Advanced Analytics to create a thriving society that equalizes financial progress with the determination of social issues.

By gathering big data from many sources via gadgets and sensors, people, objects, and systems are all associated in kicking off the Society 5.0 evolution, integrating real space and cyberspace. Big Data is examined with the help of Artificial Intelligence

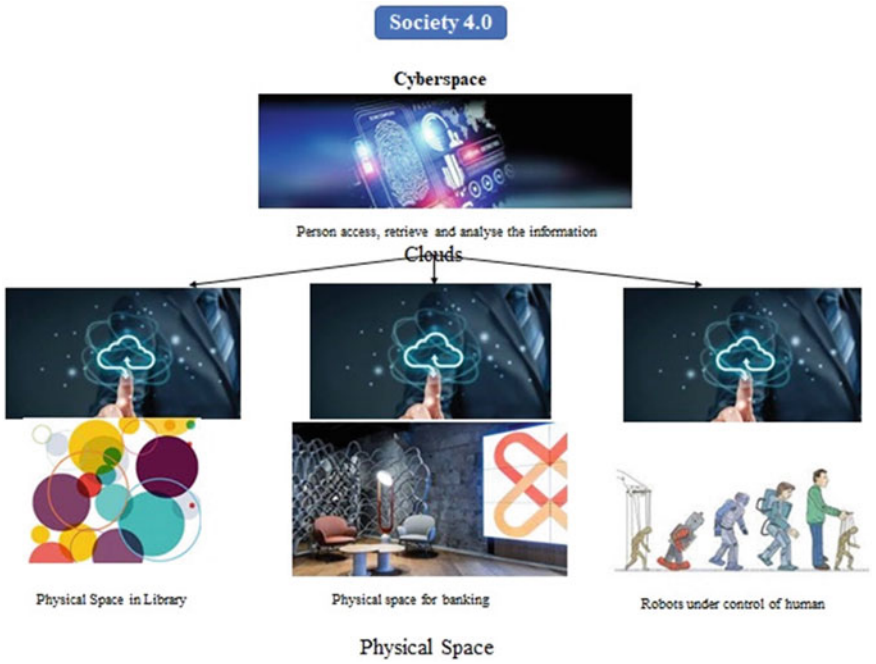


Fig. 17.1 Transition from Society 4.0–5.0

capabilities to integrate back into the physical space, by new values for individuals, industries, and businesses through various media and forms in order to accomplish both financial development and resolutions to social problems at the same time.

In Society 5.0, the innovative value created by Social Invention bridges geographic, gender, age, and language divides, allowing for customization of products as well as services to meet a wide range of individual demands. It demonstrates the ability to tackle a wide range of problems in industries such as energy, disaster management, manufacturing, food, agriculture, healthcare, and many more.

The vision of Hitachi—"Sustainable Society"—in which everybody may enjoy a safe as well as joyful life is fully integrated into the Society 5.0 idea. Hitachi is prepared to work with the government in attaining this genuineness by emerging a vigorous framework for a smooth transition to Society 5.0 and assisting in the resolution of various social challenges through new-age digital technologies. Hitachi's collaborative co-creation with the government in e-Government, Urban Development, Agriculture, Finance, and Railways, as a key partner to the government's "Digital India" project, is driving India to become a nation that is able to fulfil the expectations of Society 5.0 in the future.

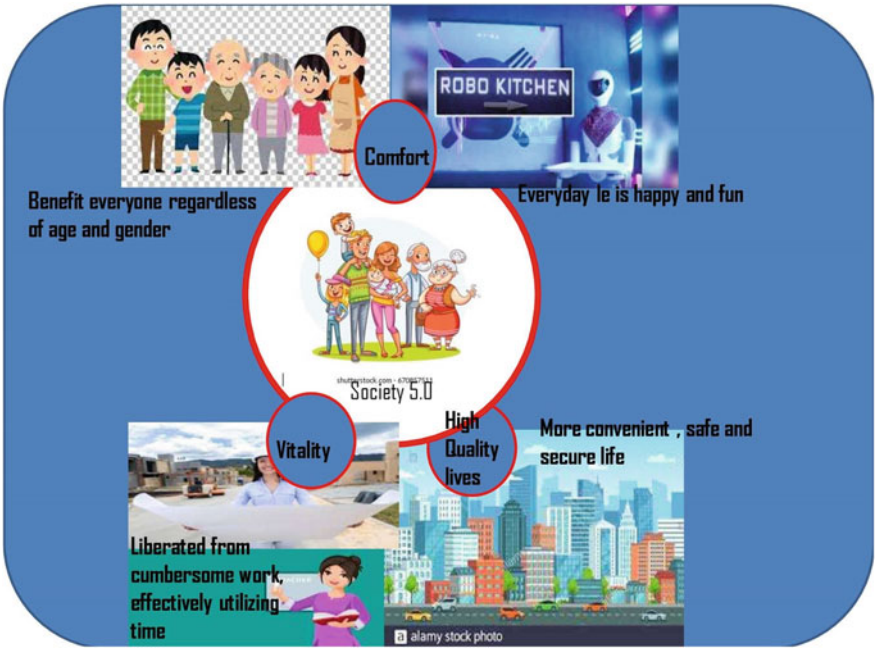


Fig. 17.2 Society 5.0

17.3 The Schema of Society 5.0

Data is taken from the "real world" and handled by computers, with the findings actually being applied in the real world, according to Society 5.0's core schema. Air-conditioning units, for example, preserve a room's environment at the temperature programmed into the appliance. The air conditioner notices the temperature of the room on a regular base, and an inbuilt microcomputer compares the reading to the recorded temperature set. The airflow is inevitably stopped or triggered based on the result, guaranteeing that the room preserves the proper temperature. This basic principle is used in many of the systems we rely on in society. It underpins the mechanisms that keep our homes well-supplied with energy and the trains operating on time. Computerized automatic controls are used in this system. Information society refers to a society in which each of these systems takes data, processes it, and then uses the results in a specific real-world setting. Society 5.0 will include systems that work together to keep society running smoothly. It is not enough to have pleasant room temperatures to ensure enjoyment and comfort. We need comfort in all elements of our lives, including transportation, energy, shopping, medical care, education, and employment.

In short, Society 5.0 will follow an iterative aspect cycle in which data is collected, processed, and then translated into useful information, which is then used in the actual world; moreover, this cycle operates at a society-wide level.

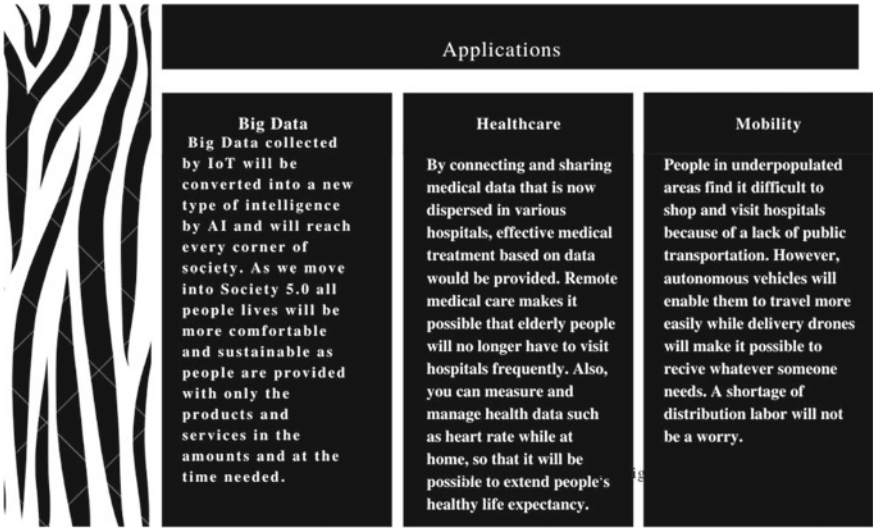


Fig. 17.3 Application of Society 5.0

17.4 New Solutions for New Problems

Because of the speed and scope of globalization, new difficulties have surfaced that were either not foreseen or not expected for some time. A more integrated world also means more integrated problems, which are more difficult to tackle.

Maintaining economic growth while reducing income inequality and environmental destruction; enhancing the welfare of an ageing population while ensuring opportunities for the youth; feeding more people with limited resources; and slowing, stopping, and then reversing the effects of climate change are all wicked problems, according to Watkins and Wilber.

Society 5.0 imagines technology and humans working together to approach these Gordian knots in a number of different areas (Fig. 17.3).

Society 5.0 Japan's idea signifies the fifth form of society in our human history, sequentially following information, industry, hunting, and farming. The Fourth Industrial Revolution is generating new values as well as services one after another, carrying a wealthier life to all.

17.5 Entering Society 5.0

See (Fig. 17.4).

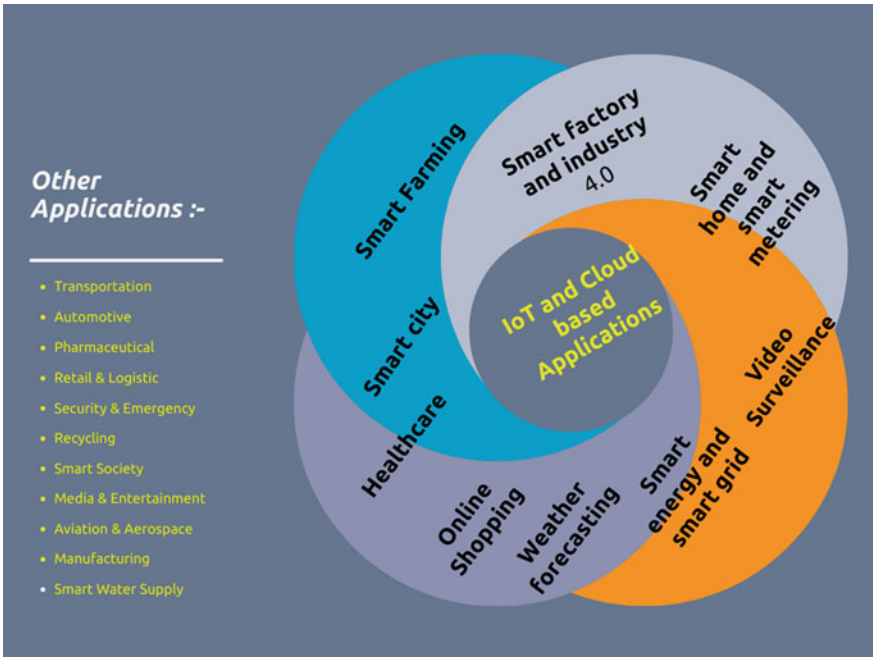


Fig. 17.4 IoT and cloud based applications

17.5.1 Healthcare

Japan is recognized for having a society that is highly weighted towards older citizens—roughly a third of the population is 60 years or older—which is one of the reasons why Society 5.0 places a major emphasis on greater health and wellbeing, particularly for the elderly.

However, as medical technology advances in quality and price around the world, all countries will face the issues of an ageing population. Increasing medical and social security costs, as well as the obligations of care for the elderly, are among them.

Wearable medical devices will permit healthcare and physiological data to be gathered, uploaded, and analysed remotely in Society 5.0, allowing for early (AI-driven) ailment identification and diagnosis. Drones and driverless cars will carry medication and healthcare services, offering older people in rural places equal access to high-quality care. Robots and artificial intelligence (AI) will assist in providing living assistance to the elderly, including providing them with the dialogue and companionship that is necessary for better mental health.

These findings will reduce the strain on public healthcare systems by reducing the number of hospital visits and improving the accuracy and efficiency of diagnoses and treatments.

17.5.2 Mobility

Traffic and transportation system overload are becoming more severe as the world's urban populations grow. Depopulated rural locations, on the other hand, have fewer, if any, public transportation options.

As we approach the fifth iteration of civilization, technology will play an increasingly important role in tackling these issues.

Traffic control systems in cities and congested metropolitan areas will be directed by ubiquitous sensors and cameras. These will generate massive amounts of data, which will be coupled with weather data and regional event data using artificial intelligence to improve traffic flows.

Different individuals have their own choices for travel, cuisine, and leisure which will be combined with universal transportation data to provide customized trip suggestions.

For public transportation in rural areas, autonomous taxis and buses will be promoted. The reach of distribution and shipping services will expand.

17.5.3 Infrastructure

In Society 5.0, social care for public infrastructure and services will become proactive, just as it is for individual health. In smart cities, this will form the backbone of civil management.

Streets, structures, tunnels, and reservoirs will all be monitored by sensors that provide a continuous stream of data. Preventive maintenance and efficient deployment of technicians with specialized skills will be possible with this information.

As a result, accidents will be decreased, as will the time and money have spent on building and maintenance work. The level of safety and productivity will rise.

17.5.4 Agriculture

A workforce shortage in agriculture is resulting from a global rural population decline, which comes at a time when the sector is under increasing pressure to increase production while dealing with the problems of more harsh climate patterns.

AI analysis of large data, such as weather data, crop growth data, economic conditions, and food patterns and demands, will lead to hyper-efficient agriculture management in Society 5.0. Autonomous farming vehicles and equipment will make these "intelligent" data-based decisions. Robots, drones, and self-driving farm equipment will take over many traditional farm labour jobs, from crop establishment to crop harvesting to planting (Fig. 17.5).

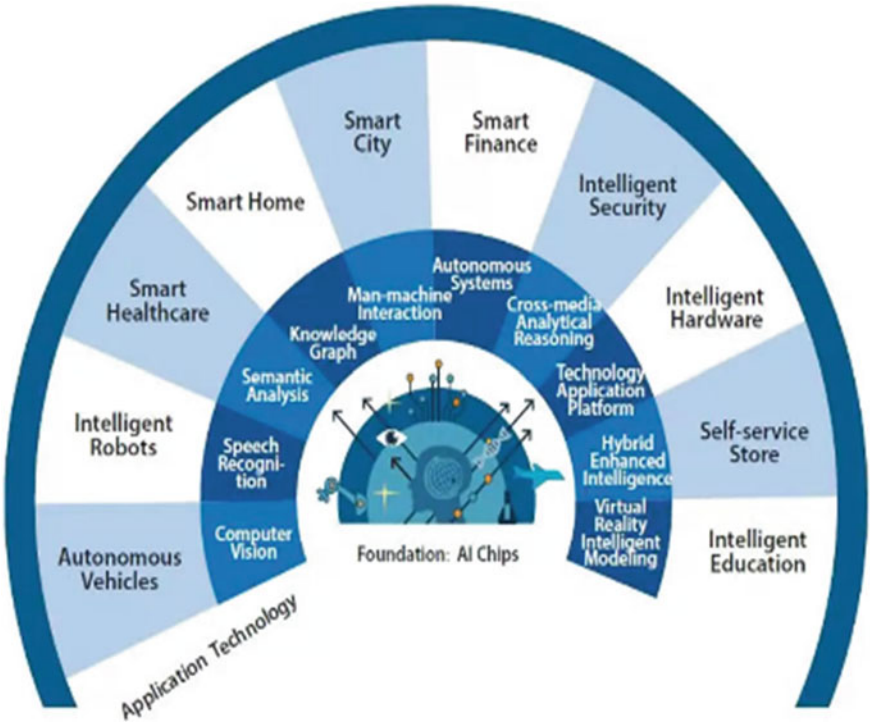


Fig. 17.5 Stacks

By 2050, the global population is predicted to surpass 9 billion people. We will only be able to feed so many people using AI and machine-optimized agricultural management.

17.5.5 Disaster Prevention and Response

The potential importance of predicted climatological and geological information is becoming increasingly obvious as we see more examples of severe weather around the world.

Data collected from terrestrial weather radar, observatories, drones, geological sensors, and public observation systems will become invaluable as Society 5.0 unfolds. This data will be processed in real time using AI to provide those crucial minutes or hours of warning of imminent tragedy that can save lives.

Widespread access to mobile networks will enable for direct distribution of safety and preventive broadcasts to end users, as well as the use of devices to geolocate individuals in need.

Drones can transport relief and rescue goods to persons caught by natural catastrophes, and they will be able to transmit back video footage of the victims' condition.

17.5.6 Energy

Much of the fight for resources in a world of 9 billion people will be a competition for energy. A peaceful society will require optimal energy development and maintenance.

Weather is becoming increasingly significant as energy generation shifts to green alternatives such as wind and solar. Weather data analysis and accurate forecasting will be critical components of dependable electricity production.

AI-powered big data processing will also optimize electricity flows throughout the system to meet demand and supply fluctuations. This will be especially crucial in smart cities, where energy will be managed down to the minute by responsive systems in buildings and public spaces, and most modes of transportation will be electric.

17.6 Cybersecurity 5.0

The ultimate strength of Society 5.0 will be found in its level of integration. The stronger the integration of the virtual and physical worlds, the larger the benefits we will reap.

Cyberrisks, on the other hand, are the same. The greater the risk to our personal and collective safety as technology is integrated into every aspect of our social being, including our physical being [7].

The foundation of Society 5.0 is a sophisticated network of sensors, devices, and systems—a massive Internet of everything. Each of these elements increases the attack surface while also raising the stakes in the event of a cyber-attack.

It's easy to understand the potential pitfalls when technology is woven into the fabric of everything we do. Autonomous vehicles, AI-assisted public transportation, drone fleets, and vital disaster-prevention systems can all be compromised. That is still true in today's society, but in Society 5.0, all relationships are cyber-kinetic. Virtual events have real-world consequences. People are injured. Or even worse.

17.6.1 Society 5G.0

Society 5.0's connectivity requirements are practically unfathomable. Devices and humans will be in constant real-time communication in all locations, from urban to rural. Without 5G, this will not be possible.

Only 5G has the ability to achieve the Society 5.0 vision, with its lightning speeds, high device connection capacity, and near-zero latency. It will make AI's full potential accessible, allowing it to evaluate oceans of data in real time and make critical decisions that affect millions of people.

Edge computing and network slicing will become even more significant than they are currently as a result of Society 5.0's vast geographical spread of services.

5G is already one of the most important Critical Infrastructures. However, in the fifth generation of society, it could be the key to a better existence for all of us.

17.6.2 Cyberspace

With the arrival of the Internet in the mid-1990s, the word "cyberspace" became popular, which was coined by Gibson as a science fiction idea and then extended to computer-based communications and virtual reality technology. For example, cyberspace has been variously called as follows: "Cyberspace is a globally networked, computer-sustained, computer-accessed, and computer-generated, multidimensional, artificial, or 'virtual' reality," according to Wikipedia. Interactivity Space: "Cyberspace is defined by interactivity between remote computers…cyberspace is not necessarily imagined space—it is real enough in that it is the space created by individuals who communicate using remote computers." Conceptual Space: "Rather than the technology itself, the conceptual space within ICTs (information and communication technologies)."

17.7 Difference Between Real Space and Cyberspace

The basis of any digital transaction is packet prioritization, which is the distinction between physical space and cyberspace. Ones and zeros do not carry any independent information by default; nevertheless, a real-space transaction does carry indivisible secondary data. Digital transmissions can simply transmit; unless explicitly stated, there is no secondary information embedded in the transmission. As a result, more information must be carried with cyberspace transactions for identity purposes for verification purposes. Identity theft is a concern when more information is provided through digital communication. Because there is no way to prohibit the transmission of fraudulent identity information or the duplicate of someone else's identity information. To avoid these issues, the sender's true identity should not be communicated with the message; instead, a verification mechanism should be employed to persuade the recipient that the message was indeed delivered by the sender. It is no longer necessary to send one's true identity. The principle of validating rather than exposing provides the sender with an additional degree of security. The digital certificates that were issued to validate these qualities are another source of risk. These certificates are intended to be used solely by their owner, but if they are obtained by a

third party, that third party can impersonate him and represent him as the person for whom he holds digital certificates. We must determine how to store by using these certificates in terms of architecture. The certificates can be kept on a smart card for usage on a computer terminal, or they can be stored in a "identity server" that is protected by a password or biometrics and accessible for transmission. In real life, selecting, verifying, or revealing elements of one's identity is difficult: most forms of identification contain far more information than is required for any transaction. Because of the unbundling capabilities of cyberspace, pieces of an individual's identity can be disassociated and validated by a third party. This not only creates the capacity to verify using the least revealing method possible, but it also creates the framework for anonymous transactions, as it is possible to simply verify the correct information without ever disseminating the same characteristic. Furthermore, users of cyberspace have control over the strength of the relationship between their physical and virtual selves.

17.8 What is the Use of Cyberspace?

Let us now discuss the benefits of cyberspace to us. We live in an Internet era, and the Internet's indispensability is something we cannot dispute. Cyberspace has evolved from the ever-expanding computer network, technology, and the Internet. It is a virtual environment in which computer networks communicate with one another.

Cyberspace brings in various purposes. It allows you to do everything via the Internet. Everything today is related to what is known as cyberspace, whether it be education, military, money, or even education. There isn't a single aspect of our lives that isn't influenced by social media.

The Internet has made data storage and management more efficient. It has helped to organize and systematize man's life. Cyberspace is omnipresent, whether for e-banking, ordering tickets, or even working online.

17.8.1 Working of Cyberspace

We all realize that life today would not be possible without online. So, how does the Internet work? Understand how the Internet allows you to send and receive data from anywhere on the planet, including space. Getting online appears to be rather simple. But there's a lot more going on behind the scenes.

Complex and huge cables, as well as networking satellites, are hidden beneath the sea level and above the earth's surface, allowing you to stream your favourite movie and use the maps to get to your preferred place. There are a variety of physical installations that allow you to connect to the Internet wirelessly.

The majority of cyberspace infrastructure is developed and maintained by private companies. We're all online, but there's no international or centralized authority in

charge of what happens online or how cyberspace is handled and organized. There are submarine cables that employ fibre optic technology to carry data. These submarine cables are the primary data carriers, and they transfer a large amount of data at a low cost.

17.8.2 Is Cyberspace the Same as the Internet?

Cyberspace and the Internet have the ability to create a virtual world for a variety of cultural and social behaviours. It is now possible to observe, communicate, and handle data using virtual cyberspace reality. The Internet, or cyberspace, is a virtual world of computers that allows people to communicate across the Internet. Information is conveyed through the Internet in today's environment. The Internet in cyberspace, on the other hand, is not the same as the Internet. The Internet is a worldwide network of computers that provides information and promotes communication via interconnected networks. This is accomplished through the use of established communication protocols.

On the other hand, cyberspace Internet is a virtual world of computers that exists within a virtual computer network environment.

To fully comprehend the concept of cyberspace and its differences, the Internet can be defined as a collection of computer networks that communicate via the Internet protocol. This is where you'll find the intranet. Cyberspace is a virtual environment connected to the Internet.

17.8.2.1 5.0 Society

The Japanese government suggested the notion of the "5.0 Society" to mitigate the potential negative consequences of technology. A vision of the future in which digital transformation enables people to realize their goals while also contributing to the construction of a more sustainable and resilient society.

A 5.0 civilization strives for digital technology innovation to be the pillar that leads us to a future with limitless potential for humanity. The following principles will characterize this future society.

17.8.3 Diversity

As a result of the standardization of products and procedures, we have established a civilization of patterns and uniformity. Having everyone think and act in the same way is frequently the goal.

Individual diversity would be a significant component of the human race's evolution in a 5.0 society. Different kinds of people with different abilities may be capable

of contributing to a society in which high levels of creativity are necessary to recognize and translate various types of demands and issues into business concepts.

Momlancers, for example, is a Mexican platform that connects businesses with freelancers with children who are looking for flexible hours and/or project-based employment. The idea is for them to be able to balance their personal and professional life at home.

This platform exemplifies how technology enables us to tap into a talent pool that would otherwise be unemployed.

The biggest challenge for momlancers is persuading more companies to trust and value what a per-project worker can bring to the table. If the platform succeeds in this, more mothers will be able to profit from its efforts.

17.8.4 Value Creation

The purpose of modern consumer culture is to create economic value above everything else. A 5.0 society is a concept in which enterprises, governments, and consumers work together to solve problems and create value.

Every day, more entrepreneurs choose to create company models that go beyond financial rewards in order to make a difference, where genuine value is made for either customers or the environment.

With its Vi-Sor service, the Chilean company Red Apis exemplifies this understanding of value generation. It's a customer service language interpretation system that works in real time. Vi-Sor is primarily designed to assist non-Spanish-speaking Haitian migrants in communicating.

Migrants benefit greatly from Red Apis' system since they can better integrate and complete legal procedures in businesses and government entities.

Because not every firm or government agency sees the advantage in investing in a system like Vi-Sor, Red Apis' impact is limited.

17.8.5 Decentralization

One of the key causes of social inequality today is the concentration of wealth and information in the hands of a few.

In a 5.0 society, technology is used to deliver possibilities to everyone, regardless of their location, education, social class, or other factors. Relationships become more horizontal in this manner. Everyone can play a part in creating a more equal society by contributing their unique talents.

The financial system in Latin America is a good example of a lack of access to opportunities. According to the IDB, half of the region's adult population does not have a bank account. As a result, they are unable to use the financial system.

Sumatoria, a digital crowdlending platform, was created in Argentina to address this problem. It enables anyone to invest in entrepreneurial enterprises that might otherwise be ineligible for a bank loan.

The platform provides investors with a proportion of a guaranteed financial return as well as traceability data on the social impact of their money.

Sumatoria has the problem of educating consumers so that they opt to deposit money on a loan for a network entrepreneur rather than keep it in a typical bank savings account by offering such an innovative solution.

17.8.6 Resilience

Externalities such as natural disasters and criminality are currently posing a significant threat to civilization or, for example, the deterioration of public infrastructure, such as health care.

Through technology, Society 5.0 aims to improve human and natural resilience. The objective is to reduce or eliminate the harmful effects of natural disasters or failing health systems.

AdApp designed a mobile application that involves people and farmers at the Gran Chaco to achieve this goal. Between Bolivia, Paraguay, and Argentina, there is an area known as the Andes.

Users acquire information from rainfall monitoring and data collecting, as well as river water levels, to make decisions such as harvesting or transporting items and animals from one location to another.

The residents of the Gran Chaco have been able to better adjust to the consequences of climate change as a result of this application, saving their products and even their lives.

17.8.7 Sustainability and Environmental Harmony

It is critical that we adapt our model of mass consumption of natural resources and limit the environmental impact of productive activities in order to ensure the existence of the human species and the planet.

Society 5.0 proposes to harness technology advancements such as Big Data to promote environmental harmony.

We can eliminate food waste, develop more sustainable cities by using alternative and efficient energy sources, and reduce production and mass consumption by using the collaborative economy.

Kingo is a firm that brings energy to houses in rural Guatemala using solar energy, the Internet of Things (IoT), and Big Data.

Kingo has developed a pre-paid energy sales model, in which clients purchase the energy consumption time they require, helping them to better manage their restricted budgets.

Despite the beneficial impact Kingo has on these communities, the high cost of developing and maintaining Kingo energy units prevents them from scaling at the desired rate, limiting the company's growth.

17.8.8 Equality and Sustainability 5.0

The notion of society 5.0 encourages us to use technology to co-create a more sustainable and inclusive future for all by bringing together all actors in society [8]. In the digital context, innovation, mobilizing the digital, is the key concept [9–11] (Table 17.1).

Table 17.1 Key terminologies

Key terminologies	
Terminology	Detail
Machine learning:	Machine learning is an important part of many business applications and research ventures, ranging from medical diagnosis and care to social networking with friends [6]
Internet of Things	Internet of Things communication methods and better results for a precise smart environment. Nowadays, Cloud computing as well as IoT are common to the people. Moreover, researchers are integrating Cloud Computing and Internet of Things [12]
Cyberspace	Cyberspace is a globally artificial, multidimensional, computer-generated, computer-accessed, computer-sustained, networked, or 'virtual' reality," according to Wikipedia
Deep learning	Deep Learning (DL) is a recent machine learning field which has been launched with the aim of getting machine learning (ML) and artificial intelligence (AI) nearer
Cloud computing	Cloud computing is the newest development in Information Technology, it carries desktop computing to the whole web, and the user does not need to concern about managing and adapting any of the devices
Big data analytics	Big data analytics is the use of advanced analytic techniques to very large, heterogeneous data sets, which can contain structured, semi-structured, and unstructured data, as well as data from many sources and sizes ranging from terabytes to zettabytes
Computer vision	Computer vision is a branch of artificial intelligence (AI) that allows computers and systems to extract useful information from digital photos, videos, and other visual inputs, as well as to conduct actions or make recommendations based on that data

17.8.9 Sustainable Digital Innovations in Society 5.0

Sustainable invention can be defined as a process in which financial, social, and environmental sustainability issues are integrated into organizational systems from idea development to Research and Development and following commercialization of the outcomes. This applies to new business and organizational structures as well as products, services, and technologies (Charter & Clark, 2007). The importance of the environmental, social, and economic factors essential in the innovation process is referred to when the concept of sustainability is mentioned as a vital ingredient gearing innovation. The profit component of sustainability is concerned with concerns such as economic growth, resource efficiency, and the financial viability of businesses. The environmental dimension is concerned with pollution control and the wise and effective use of natural resources. Equal chances, fair economic distribution, equity, justice, and ethical behaviour are all aspects of the social dimension. According to the European Commission (2010), advocating for a more competitive economy that promotes sustainable economic growth with more and better jobs and social cohesion comprises inclusive growth, sustainability, and intelligence. The triple bottom line can be used to examine the relationship between innovation and long-term development (Fig. 17.2). The sustainability triangle serves as the foundation for this viewpoint, which provides a systematic relationship between the innovative dimensions (Fig. 17.6).

Furthermore, the big data itself is not fully accessible and can act as a factor of economic and social differentiation in surveillance and control strategies [13] as shown in Fig. 17.7.

17.9 Summary

Due to the advancement in technology, Society 5.0 is the new terminology that rises with the high rate of involvement of all modern technologies such as Artificial.

Intelligence, big data, and IoT. The concept of Society 5.0 aims to provide a luxurious and good quality of life. The revolution from Industry 4.0 to Industry 5.0 also impacts the betterment of life which makes the individual to get everything on a single click. In this chapter, the various factors responsible for Society 5.0 and applications such as innovations in healthcare, Banking, and agriculture which make the society as Society 5.0 are discussed. The various key terminologies that are effective and impactful with respect to society 5.0 are also described. The Cyber-physical spaces that are the main source for inventing Society 5.0 are also discussed and compared.

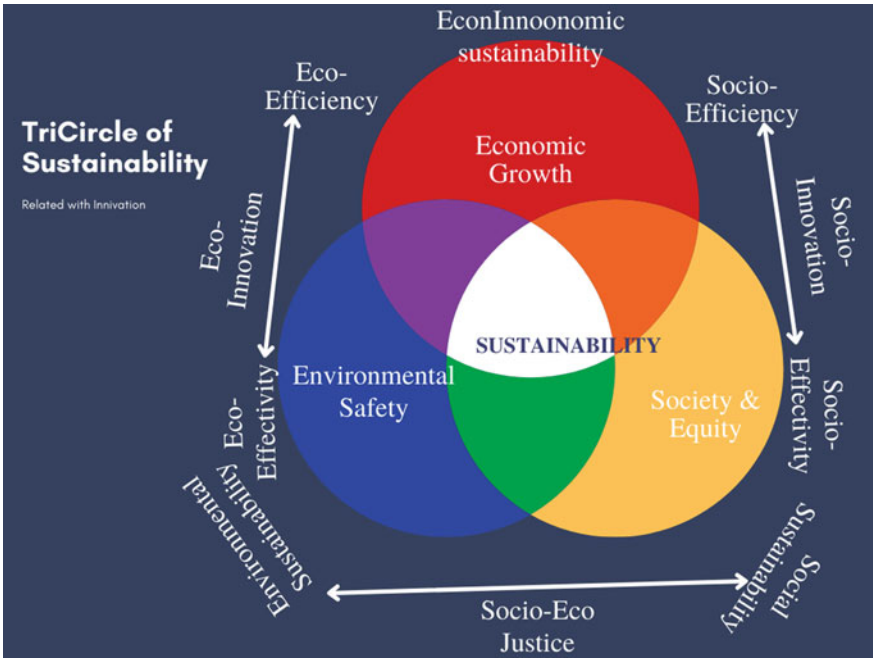


Fig. 17.6 TriCircle of sustainability

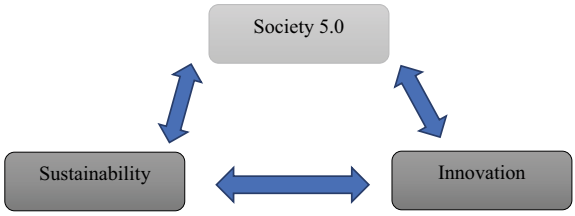


Fig. 17.7 Trimetric interconnection of innovation, Society 5.0, and sustainability

References

1. Fujii T, Guo T, Kamoshida A (2018) A consideration of service strategy of Japanese electric manufacturers to realize super smart society (SOCIETY 5.0). In: Proceedings of the Programmierenfüringenieure und Naturwissenschaftler. Springer Science and BusinessMedia LLC, Berlin, Germany, pp 634–645
2. Nagahara M (2019) A research project of society 5.0. In: Proceedings of the 2019 IEEE conference on control technology and applications (CCTA), Kitakyushu, Japan, 19–21 Aug 2019, pp 803–804
3. Cabinet Office (2015) Report on the 5th science and technology basic plan council for science, technology and innovation. https://www8.cao.go.jp/cstp/kihonkeikaku/5basicplan_en.pdf
4. Serpa C, Ferreira S (2018) Society 5.0 and social development: contributions to a discussion. J Manag Organ Stud 5:26–31
5. Mohammadian HD, Mohammadian FD, Assante D (2020) IoT-education policies on national

and international level regardingbest practices in German SMEs. In: Proceedings of the 2020 IEEE global engineering education conference (EDUCON), Porto, Portugal, 27–30 Apr 2020, pp 1848–1857
6. Thakur R, Rane D (2021) Book chapter: Smart environment monitoring models using cloud-based data analytics: a comprehenssive study in book machine learning approach for cloud data analytics in IoT. Published Scrivener Publishing Wiley
7. Nakajima A, Yano M, Chris D (2019) Blockchain and society 5.0—The creation of a new marketplace based on distributed consensus. https://www.rieti.go.jp/en/rieti_report/222.html. Accessed 27 Mar 2021
8. Pochenchuk G, Fedkovich Y, Babukh I, Baraniuk D (2018) Innovations of the digital era and economic choice. USV Ann Econ Public Adm 18(1):63–69
9. Lokuge S, Sedera D, Grover V, Dongming X (2019) Organizational readiness for digital innovation: development and empirical calibration of a construct. Inf Manag 56(3):445–461
10. Corrales-Garay D, Mora-Valentín EM, Ortiz-de-Urbina-Criado M (2019) Open data for open innovation: an analysis of literature characteristics. Future Internet 11(3):77
11. Fukuyama, M. (2018). Society 5.0: Aiming for a new human-centered society. Japan SPOTLIGHT, 47–50.
12. Thakur R, Borkar P (2021) Book chapter: Research aspects of machine learning: issues, challenges and future scope. Published, Taylor and Francis, CRC Press
13. Possamai-Inesedy A, Rowe D, Stevenson D (2017) Sociology in the 21st century: challenges old and new. J Sociol 53(4):723–729