Signals and Communication Technology

Nishu Gupta
Sumita Mishra *Editors*

# Internet of Everything for Smart City and Smart Healthcare Applications

Springer

# Signals and Communication Technology

This series is devoted to fundamentals and applications of modern methods of signal processing and cutting-edge communication technologies. The main topics are information and signal theory, acoustical signal processing, image processing and multimedia systems, mobile and wireless communications, and computer and communication networks. Volumes in the series address researchers in academia and industrial R&D departments. The series is application-oriented. The level of presentation of each individual volume, however, depends on the subject and can range from practical to scientific.

Indexing: All books in "Signals and Communication Technology" are indexed by Scopus and zbMATH

For general information about this book series, comments or suggestions, please contact Mary James at mary.james@springer.com or Ramesh Nath Premnath at ramesh.premnath@springer.com.

Nishu Gupta • Sumita Mishra
Editors

# Internet of Everything for Smart City and Smart Healthcare Applications

Springer

*Editors*
Nishu Gupta
Department of Electronic Systems
Faculty of Information Technology and Electrical Engineering
NTNU
Gjøvik, Norway

Sumita Mishra
Amity School of Engineering and Technology, Amity University, Lucknow
Uttar Pradesh, India

ISSN 1860-4862 ISSN 1860-4870 (electronic)
Signals and Communication Technology
ISBN 978-3-031-34600-2 ISBN 978-3-031-34601-9 (eBook)
https://doi.org/10.1007/978-3-031-34601-9



This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

*Dr. Nishu Gupta dedicates this book to his Parents*

*Dr. Sumita Mishra dedicates this book to her Husband Mr. Gaurav Mishra*

# Foreword

I am pleased to write this foreword as I find that the book deeply emphasizes the state-of-the-art technologies that comprise many research explorations in the field of Internet of Everything that widely includes 'things' as the major component and that it can offer in taking care of various applications through artificial intelligence, networking and automation.

I am confident that this book will provide an effective learning experience and a guided reference for researchers, professionals and students that are interested in the integration of artificial intelligence in IoT-based embedded technologies and its advances to the engineering field.

We encounter challenges in addressing technological problems. These challenges are both difficult and interesting. Researchers are working on them to develop new approaches and provide new solutions to keep up with the ever-changing potential threats. This book is a good initiative and a combined effort of a lot of authors across the globe. It provides a strong foundation to the existing and upcoming technologies, especially in the fields of smart cities and smart healthcare.

I highly recommend this book to a variety of audiences, including academicians, commercial engineers, researchers, students and scholars. It is my desire and expectation that this book will provide an effective learning experience, a contemporary update and a practical reference for all those who are interested in this versatile and trending field.

Wing Commander (Retd.), Deputy Pro Vice Chancellor, Anil Kumar
Amity University, Lucknow Campus,
Lucknow, India

# Foreword

I am delighted to write the foreword for this edited book *Internet of Everything for Smart City and Smart Healthcare Applications*. This book highlights the importance of the Internet Technology, particularly the applications of the Internet of Things (IoT) towards smart cities and healthcare. The book intends to demonstrate to its readers useful applications and use-cases that cater to diversified technological requirements.

This book provides a window to the research and development in the field of "Internet for everyone" in a comprehensive way and enumerates the evolutions of contributing tools and techniques. The range of topics covered in this book is quite extensive and every topic is discussed by experts in their own field. The advances and challenges are discussed with a focus on successes, failures and lessons learned, open issues, unmet challenges and future directions.

Additionally, this book is a good initiative in a direction that addresses numerous issues related to IoT applications. I am convinced that this book will provide a solid platform to various realms of the existing and upcoming technologies, especially in the field of Smart Cities, Smart Healthcare, Artificial Intelligence, Human-Machine Interaction, Autonomous Vehicles and Intelligent Transportation. The authors can be confident that there will be many grateful readers who will have gained a broader perspective of the disciplines of machine interaction and its applications as a result of their efforts. I hope that this book will serve as a primer for industry and academia, professional developers, upcoming researchers across the globe to learn, innovate and realize the multi-fold capabilities of Artificial Intelligence and IoT applications.

I wish very good luck to the editors and contributors of this book.

Professor, Engineering Department/IEETA, UTAD University, Vila Real, Portugal

Manuel J. C. S. Reis

# Preface

The Internet of Everything (IoE) refers to devices or physical objects embedded with sensors and other technologies that connect and exchange data with other devices and systems over the Internet. Such devices are used for different applications including smart cities, smart healthcare, etc. to enhance the lifestyle of the people in terms of safer, sustainable and comfortable environment around us.

IoE is believed to integrate the novel futuristic trends at the cutting edge of study and research by featuring enormous applications in a proficient, adaptable and manageable way. It covers the primary mainstays of the Internet of Things (IoT) world giving a thorough description of the present advancements, systems and structures.

Written by international experts, this book intends to present to its readers about day-to-day and upcoming trends as an insight on the importance that IoT and Information and Communication Technology (ICT) solutions can offer towards smart city applications as well as in taking care of people's health proactively. Key features of this book include elaboration of recent and emerging developments in various specialization of curing health problems and their solutions; smart transportation systems, traffic management for smart cities, energy management, deep learning and machine learning techniques for smart health and smart cities, and concepts that incorporate Internet of Everything (IoTs and ICTs). The benefits of IoE solutions are enormous and the range of applicability is also significant. This book attempts to cover some useful IoE applications and architectures that caters to improved sustainability requirements.

The book is a perfect blend of a text as well as reference and suits to almost all levels of technical education and research as well as to novices interested in applications that Industrial IoT technologies can offer in the coming times thereby making almost everything smart, intelligent and self-adaptive. The book is divided into different parts having multiple chapters. The contents of the book have been organized in a reader-friendly manner. It is targeted at professionals, including university professors, graduate and Ph.D. scholars, industry practitioners and researchers, particularly in the field of computer communication, wireless communication, cyber-physical systems, machine learning and sensor networks.

The book attracted contributions from all over the world, and we would like to thank all the authors for submitting their works. We extend our appreciation to the reviewers for their timely and focused review comments. We gratefully acknowledge all the authors and publishers of the books quoted in the references.

Gjøvik, Norway — Nishu Gupta
Lucknow, India — Sumita Mishra

# Acknowledgements

**Dr. Nishu Gupta**

I acknowledge the inspiration and blessings of my mother Smt. Rita Rani Gupta and father Prof. K.M. Gupta, who are the pillars of my strength and personality. Having parents like them is the greatest gift and biggest advantage anyone could ever have given me. It is because of the confidence and values that they have instilled in me that made me who I am today.

I am full of gratitude to my sister Mrs. Nidhi Gupta, brother-in law CA Ritesh Shankar Gupta, wife Smt. Anamika Gupta, son Ayaansh Gupta and other family members for the patience shown and encouragement given to complete this venture.

I deeply acknowledge the blessings of my Academic Advisor and Mentor Prof. Rajeev Tripathi, Motilal Nehru National Institute of Technology (MNNIT) Allahabad, Uttar Pradesh, India. I am highly grateful to my Ph.D. Supervisor Prof. Arun Prakash, Electronics and Communication Engineering Department, MNNIT Allahabad, India, whose guidance has always encouraged me to do my best. I wholeheartedly acknowledge the motivation given to me by my Scientific Advisor Prof. Mohammad Derawi, Norwegian University of Science and Technology (NTNU) Gjøvik, Norway; academic collaborators including Prof. Sara Paiva, Polytechnic Institute of Viana do Castelo, Portugal; Prof. Manuel J. C. S. Reis, UTAD, Portugal; Prof. Ahmad Hoirul Basori, King Abdulaziz University, Saudi Arabia; Prof. Ariel Soares Teles, Federal Institute of Maranhão, Brazil; Dr. Anil Gupta, Centre for Development of Advanced Computing (CDAC), Pune, Maharashtra, India; Dr. Krishan Kumar, National Institute of Technology Hamirpur, Himachal Pradesh, India, for always encouraging and guiding me towards delivering the best. Thank you for shaping me into a person I am proud to be.

I profoundly acknowledge my friends Ms. Manisha Srivastava, India; Mr. Gaurvendra Singh, IIFL Wealth, New Delhi, India; Er. Jalaj Kumar Singh, KPMG Global Services, India; Ms. Isha Bharti, Capgemini Inc., USA. Thank you for everything you have done for me. I love you, more than words can ever express.

I express my heartfelt thanks to my colleagues, friends and students for their support and motivation in several ways.

**Dr. Sumita Mishra**

First and foremost, I acknowledge my mother Smt. Sushila Shukla for her blessing, love and support throughout my life. I profoundly acknowledge my father Mr. V. K. Shukla and Father-in-Law Late. Sri R. S. Mishra, for their support and blessings.

I want to thank my husband Mr. Gaurav Mishra for the unwavering and unconditional support provided to me at all times. I sincerely acknowledge my brothers Mr. Ramendra Shukla, Mr. Vinay Shukla and Mr. Vineet Shukla, daughter Sanvi Mishra, son Aditya Mishra and other family members for the patience, love, and encouragement.

I am eternally grateful to Dr. Ashok K Chauhan, Founder President, Amity Group, and Dr. Aseem Chauhan, Chairman Amity University, Lucknow Campus, for always encouraging us and providing excellent research facilities.

I wholeheartedly acknowledge the motivation, guidance and support given to me by Wg. Cdr. Dr. Anil Kumar, Dy. Pro. VC, Amity University, Lucknow Campus for achieving greater heights.

I am thankful to Dr. O. P. Singh, Head, Amity School of Engineering and Technology, Amity University, Lucknow Campus for always being helpful and inspiring.

I deeply acknowledge the support and encouragement from my co-editor Dr. Nishu Gupta in right shaping this book. I express my heartfelt gratitude to my friends and colleagues for their constant support and inspiration.

Last but not the least, we, the Editors, express our heartfelt gratitude to the publisher and the team behind it for their continued support and cooperation in publishing this book.

# Contents

# Editors and Contributors

## About the Editors

**Nishu Gupta** is a Senior Member, IEEE. He is a Postdoctoral Fellow (ERCIM Alain Bensoussan Fellowship) in the Department of Electronic Systems, Faculty of Information Technology and Electrical Engineering, at Norwegian University of Science and Technology (NTNU) in Gjøvik, Norway. He is also a Visiting Researcher at the University of Oviedo, Gijón, Spain, under the research group on Systems for Multimedia and the Internet of Things (SMIOT). He served as a Member, Zero Trust Architecture working group of MeitY-C-DAC-STQC project under 'e-Governance Standards and Guidelines', Ministry of Electronics and Information Technology (MeitY), Government of India. Recently, he won the Electronics, MDPI 2022 Travel Award. He received his Ph.D. degree in the year 2016 from the Department of Electronics and Communication Engineering, MNNIT Allahabad, Prayagraj, India, which is an Institute of National Importance as declared by the Govt. of India, and M.Tech. degree from Delhi Technological University (Formerly Delhi College of Engineering), Delhi, India.

Dr. Nishu is the recipient of Best Paper Presentation Award at the 4$^{th}$ International Conference on Computer and Communication Systems held at Nanyang Technological University, Singapore. He has published 5 patents and more than 55 research articles in IEEE Transactions, SCI and Scopus Indexed Journals. Dr. Gupta has supervised numerous theses at Master level and projects at Bachelor level in his main line of work. He has authored and edited several books with international publishers like Taylor & Francis, Springer, Wiley and Scrivener. Dr. Nishu is on the Editorial board of various Internationally reputed journals and transactions. He serves as reviewer of more than 20 SCI indexed journals and transactions such as *IEEE Transactions on ITS*, *IEEE Access*, *IET Communications*, etc. He was awarded twice for Outstanding Contribution in Reviewing by Elsevier. His research interests include autonomous vehicles, edge computing, augmented intelligence, Internet of Things, Internet of Vehicles, deep learning, ad-hoc networks, vehicular communication, driving efficiency, cognitive computing, human-machine interaction, traffic pattern prediction, etc.

**Sumita Mishra** received the M.Tech. degree in optical communication from the Shri Govindram Seksaria Institute of Technology and Science, Indore, India. She received her Ph.D. degree from Dr. Ram Manohar Lohia Avadh University, Faizabad, India. She is currently affiliated to the Electronics and Communication Engineering Department, Amity School of Engineering and Technology, Amity University, Lucknow Campus, Uttar Pradesh, India. Dr. Mishra is a member of IET (UK), IAENG and served as Senior Member of IEEE. She has more than 50 publications in international journals and prestigious conferences. She has published two patents and serves as reviewer and editorial board member of many highly reputed journals. Dr. Mishra has had key responsibilities in organizing several international conferences and various international events. She has also delivered invited speech at national and international forums. Her current research interests include deep learning, visible light communication, and Internet of Things and is guiding several Ph.D. students in this area.

## Contributors

**Anuj Abraham** Technology Innovation Institute, Abu Dhabi, United Arab Emirates

**Saul Barraza-Aguirre** Autonomous University of Ciudad Juarez, Chihuahua, Mexico

**Juan Cota-Ruiz** Autonomous University of Ciudad Juarez, Chihuahua, Mexico

**Carlos R. Cunha** Applied Management Research Unit (UNIAG), Instituto Politécnico de Bragança, Campus de Santa Apolónia, Bragança, Portugal

**Arlindo Flavio da Conceição** Federal University of São Paulo (UNIFESP), São Paulo, Brazil

**Nitin Vamsi Dantu** North Eastern University, Boston, MA, USA

**Miguel Franklin de Castro** Federal University of Ceará (UFC), Fortaleza, Brazil
Group of Computer Networks, Software Engineering and Systems (GREaT), Fortaleza, Brazil

**Jose Diaz-Roman** Autonomous University of Ciudad Juarez, Chihuahua, Mexico

**Francisco Enriquez-Aguilera** Autonomous University of Ciudad Juarez, Chihuahua, Mexico

**Paula Odete Fernandes** Applied Management Research Unit (UNIAG), Instituto Politécnico de Bragança, Campus de Santa Apolónia, Bragança, Portugal

**P. K. Garg** Geomatics Engineering Group, Civil Engineering Department, Indian Institute of Technology Roorkee, Roorkee, India

**R. D. Garg** Geomatics Engineering Group, Civil Engineering Department, Indian Institute of Technology Roorkee, Roorkee, India

**Henrique Gil** Age.Comm – Polytechnic University of Castelo Branco, Castelo Branco, Portugal

**Abhijit Dnyaneshwar Jadhav** Department of Computer Engineering, Pimpri Chinchwad College of Engineering & Research, Pune, India

**Evangelia Konstantopoulou** SCYTALE Group, Computer Engineering and Informatics Department, University of Patras, Hellas

**Amit Kumar** Department of Electronics Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, India

**Abhilash Mandloi** Department of Electronics Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, India

**Chetan B. Math** Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore

**Boris Mederos-Madrazo** Autonomous University of Ciudad Juarez, Chihuahua, Mexico

**Kavach Mishra** Geomatics Engineering Group, Civil Engineering Department, Indian Institute of Technology Roorkee, Roorkee, India

**André Moreira** Applied Management Research Unit (UNIAG), Instituto Politécnico de Bragança, Campus de Santa Apolónia, Bragança, Portugal

**Sunandhini Muralidharan** L&T Technology Services, Bengaluru, India

**T. S. Murugesh** Department of Electronics and Communication Engineering, Government College of Engineering Srirangam, Tiruchirappalli, Tamil Nadu, India
On Deputation from Department of Electronics and Instrumentation Engineering, Faculty of Engineering and Technology, Annamalai University, Chidambaram, Tamil Nadu, India

**M. S. Narassima** Great Lakes Institute of Management, Chennai, India

**Muhammad Bello Nawaila** Aminu Saleh College of Education, Azare, Nigeria

**Carlos Ochoa-Zezzatti** Autonomous University of Ciudad Juarez, Chihuahua, Mexico

**Ivana Ognjanovic** University of Donja Gorica, Donja Gorica, Podgorica, Montenegro

**Victória Tomé Oliveira** Efrei Research Lab, Paris, France
Federal University of Ceará (UFC), Fortaleza, Brazil
Group of Computer Networks, Software Engineering and Systems (GREaT), Fortaleza, Brazil

**Fezile Ozdamli** Management Information Systems Department, Near East University (Cyprus) Computer Information Systems Research and Technology Center, Nicosia, Cyprus

**Raghavendra Pal** Department of Electronics Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, India

**Maria Raquel Patrício** Age.Comm – Polytechnic University of Castelo Branco, Castelo Branco, Portugal
Centro de Investigação em Educação Básica (CIEB), Instituto Politécnico de Bragança, Bragança, Portugal

**Marciel Barros Pereira** Efrei Research Lab, Paris, France
Federal University of Ceará (UFC), Fortaleza, Brazil
Group of Computer Networks, Software Engineering and Systems (GREaT), Fortaleza, Brazil

**Luís Pires** UNIAG, Instituto Politécnico de Bragança, Campus de Santa Apolónia, Bragança, Portugal

**Shitala Prasad** Indian Institute of Technology (IIT Goa), Ponda, Goa, India

**Siniraj Pulari** Bahrain Polytechnic, Isa Town, Bahrain

**Raumit Raj** Department of Electronics Technology, Guru Nanak Dev University, Amritsar, India

**Vladimir Rocha** Federal University of ABC (UFABC), Santo André, Brazil

**Christiano A. P. Rodrigues** Efrei Research Lab, Paris, France
Federal University of Ceará (UFC), Fortaleza, Brazil
Group of Computer Networks, Software Engineering and Systems (GREaT), Fortaleza, Brazil

**Mohit Sharma** Technology Innovation Institute, Abu Dhabi, United Arab Emirates

**Nicolas Sklavos** SCYTALE Group, Computer Engineering and Informatics Department, University of Patras, Hellas

**Rachna K. Somkunwar** Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, India

**Pankaj P. Tasgaonkar** Electronics and Telecommunication Engineering Department, COEP Technological University, Pune, India
Geomatics Engineering Group, Civil Engineering Department, Indian Institute of Technology Roorkee, Roorkee, India

**Shriram K Vasudevan** Intel India Pvt. Ltd., Bengaluru, India

**Dario Vieira** Efrei Research Lab, Paris, France

**Yi Zhang** Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore

# Part I
# Internet of Everything: A Perspective

# Standardization in the Transformation of Civic Systems Using Safe and Secure Internet of Things Systems

**Abhijit Dnyaneshwar Jadhav**

## 1 Introduction

### *1.1 IoT in Human Lives*

Our world is becoming more and more connected via smart technologies. Through the use of artificial intelligence, machine learning, and big data analytics, these "Internet of Things" (IoT) integrate devices with sensor capabilities and cloud connectivity, sometimes drastically enhancing their capabilities. Instead of only having one home computer, everyday users now have a variety of devices that must all be controlled separately, which can be challenging [1]. For instance, many baby monitors allow arbitrary strangers on the web to observe unknowing people's homes since they neglect to update the default password. But in order to make such systems safe, secure, efficient, and useable, new research and industry techniques are needed given the ubiquity, capabilities, and interconnection of smart devices. The issue is so serious that the government cybersecurity agencies have advised people to isolate IoT devices on their own protected networks and be aware of the capabilities of the devices, which are instructions that are very unlikely to be carried out in reality [2, 3]. In this chapter, we contend that networks of smart devices pose novel problems that call for a deeper comprehension of how people may utilize such systems productively and a larger investment in tools and policies that inspire user confidence. Solutions that address all four concurrently are essential, in particular since security, physical safety, privacy, and usability concerns are closely intertwined.

A. D. Jadhav (✉)
Department of Computer Engineering, Pimpri Chinchwad College of Engineering & Research, Pune, India



N. Gupta, S. Mishra (eds.), *Internet of Everything for Smart City and Smart Healthcare Applications*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-34601-9_1

## 1.2 Impact of IoT

Estimates of the Internet of Things' economic impact abound, with projections that the number of deployed devices will reach 50 billion by 2020 and that the overall impact might reach $10 trillion by 2025. We already live in a linked device world, with many businesses selling smart appliances like smart thermostats and doorbells. The so-called Industrial Internet of Things and smart cities programs will penetrate cities and factories with networked smart devices, with significant efficiency and reliability gains anticipated [4]. For instance, the development of networked sensor devices has benefited hospitals, leading to better health outcomes and cheaper expenses. Unfortunately, the difficulty of controlling these groups of devices rises exponentially as their quantity and connection rise [5]. Imagine the complexity of understanding and effectively maintaining a network of several interconnected devices if controlling a single home computer is challenging for a non-technical individual [6, 7]. Think about a fictitious scenario where an Xbox, an Amazon Echo, an Apple iPhone, and a Ring doorbell collaborate. They are configured using the iPhone, and voice instructions are carried out using the Echo. For instance, a user may instruct the Echo to utilize the Xbox to display the video stream from the Ring on the TV. Monitoring of homes and energy use might be the subject of another example. A single device's security and safety already provide a challenging situation. Figure 1 shows the challenges associated with IoT designs including security challenges. As IoT systems are utilized to physically manage electrical products like light bulbs and heating systems in both homes and businesses, safety concerns in particular are becoming more and more crucial. The difficulties in making devices secure even in the absence of malicious attacks are

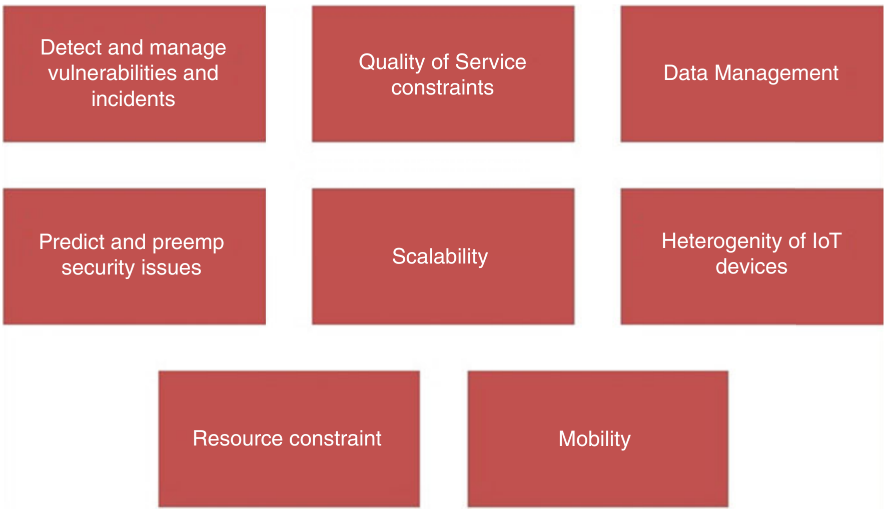



**Fig. 1** Challenges associated with IoT designs

demonstrated by the safety issues with Samsung Galaxy Note cellphones catching fire. It is extremely harder to make them safe when an attacker is present, and doing so necessitates reevaluating how such devices are created and put through safety testing [8]. When the Mirai virus was used to build a 380,000 IoT-based botnet used in a significant distributed denial-of-service (DDoS) assault, it brought to light the dangers of having numerous unsecure individual devices connected to the Internet. Just 1 month later, a significant cyberattack that involved tens of millions of machines, including a sizable number of IoT devices, was directed at the DNS infrastructure of the Internet. As a result, several important service providers, including Twitter, Netflix, Spotify, Airbnb, Reddit, Etsy, SoundCloud, and *The New York Times*, were disrupted. The more unsecured nodes that are attached, the greater leverage an attacker has over them.

### *1.3 Easiness in IoT Interaction*

Beyond the current difficulties in protecting individual devices, we need to make it easier for users to engage with a group of devices so they don't have to consider each one and how it could interact. What information is exchanged between an iPhone, Echo, Ring, and an Xbox, for instance, and what privacy regulations are in place regulating what information from a private house may be transferred to the various firms and how this information can be used? What security vulnerabilities, beyond those related to privacy, are produced by this specific grouping of devices, and who is in charge of alerting owners to such vulnerabilities? New technology is required to let people to more easily comprehend, set, and administer their collections of devices, just how operating systems have evolved to enable individual users to control them [9].

The linked gadgets in a smart home and the device collections in hospitals are the two situations we look at in this chapter where groupings of devices generate opportunities and obstacles. We may see parallels and contrasts in the needs for such systems by examining both a consumer-focused scenario and safety-critical business applications.

## 2 Smart Home Devices

Although there are numerous linked home solutions available, the field has grown so quickly that it has surpassed regulatory requirements, lifespan and safety considerations, security and privacy studies, and a general knowledge of how such systems represent human cognition and mental models. But the development of scalable smart home systems has the potential to have a significant influence on how we live our daily lives [10, 11]. As a result, we list a number of possibilities and difficulties for computer research in the context of smart home technologies. As linked products like Jarden's Mr. Coffee™ and Crock-Pot™ that can be controlled by software proliferate on the market, new physical safety risks also do as well. The safety risks of straightforward WiFi-enabled appliance modules and light bulbs have recently come to light. Smart home technologies require a comparable layer of security to the legally required safety features like electrical circuit breakers, GFCI switches, and fire-rated walls that shield customers from flaws in house infrastructure. The National Electrical Manufacturers Association (NEMA) and National Electrical Codes (NEC) both exist to offer safety requirements, and comparable safety enforcement procedures must develop for IoT products in the house. Building regulations will also need to change in order to accommodate new smart home technology. A collaborative effort involving the computing community, the Department of Housing, the Federal Communications Commission (FCC), Underwriters Laboratories (UL), and the National Institute of Standards and Technology will be necessary to address safety risks for residential IoT devices (NIST).

Smart home technologies and the IoT as a whole provide a new difficulty in the form of manufacturer abandonment, particularly for IoT startups that may launch a product into the market before swiftly going out of business or abandoning support entirely. These so-called "zombie" gadgets are still connected to a home network but will not receive any additional security or safety updates. These dangers are troublesome for technologies that are built into the structure of the home or for equipment that may stay in the house for a long time, posing both a technological and a legislative issue. Methods are required to efficiently find these abandoned systems and keep track of how they connect with other platforms [12]. The other extreme would be to mandate that when support ends, manufacturers remotely deactivate older equipment. Figure 2 shows the smart home using IoT.

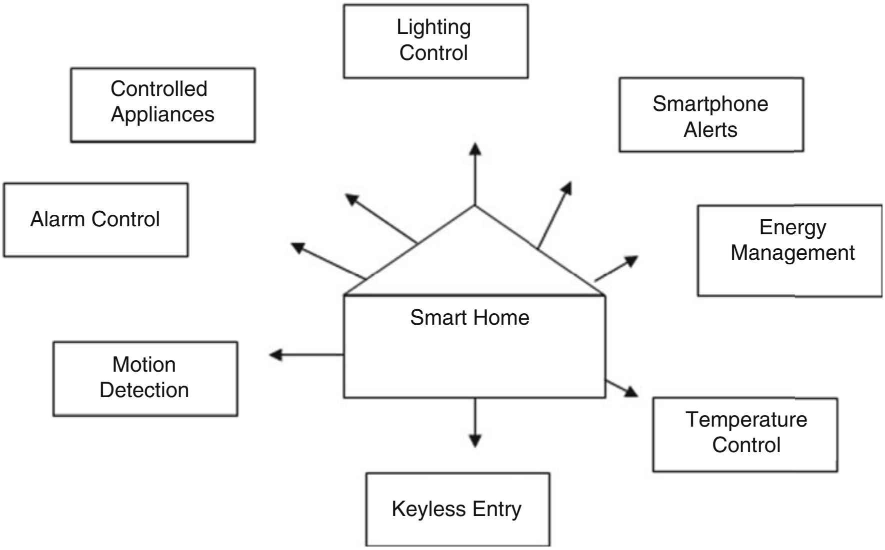



**Fig. 2** Smart home using IoT

## 3 Hospitals Using Smart Devices

The use of computing in hospitals and healthcare generally is quite advantageous. Through the use of electronic medical records, computation can provide more precise and knowledgeable patient care. Hospitals can operate more efficiently because of computation, which allows one nurse station to remotely monitor several patients at once. For instance, a nursing station might wirelessly and remotely monitor the medication dispensers giving out medications to all of the patients under their care. Even wireless implanted medical devices like pacemakers and implantable cardiac defibrillators use computation to operate within patients' bodies.

Sadly, it is well recognized that, along with the enhanced benefits of computing in hospitals, there is also a risk of patient injury if the software for the systems contains bugs [13]. Therac-25, a radiation treatment device from the 1980s, is a classic example. It was discovered to contain a software flaw that may result in patients receiving 100 times the recommended dose of radiation therapy. Patients suffered injury, and there were at least a few fatalities as a result of this software flaw, human factors, and poor project management. These injuries were accidental. In the field of cybersecurity, we need to consider what an ingenious, clever opponent could be able to do and how we might protect ourselves from one. Because the opponent can drive the systems into their worst-case settings, it is obvious that they can hurt systems at least as much as they might accidentally, if not more. Additionally, the potential attack surface for cyber adversaries is now significantly larger than it was in the 1980s due to the increased pervasiveness of computation inside the healthcare setting.

A thorough strategy for hospital cybersecurity must take into account each of the computational hospital equipment, as well as the things that those equipment rely on. For instance, cyberattacks on the hospital's electrical system might have a big impact on patient care. Water supply cyberattacks on the hospital might have a serious negative effect on patient care. There have been instances where ransomware has shut down hospital systems, forcing healthcare practitioners to switch back to paper-based records—something that many younger medical staff members may not be trained to use [14]. Imagine the effects of even more malevolent software, including medical malware that willfully changes patient electronic prescriptions or dosage data to potentially deadly medications or drug doses, building on the impact of the ransomware scenario. Similar effects could be expected if hospital equipment that directly affect patient care were compromised. Examples include computerized radiation therapy equipment and equipment that doctors use to wirelessly alter the settings on implantable medical devices like pacemakers and implantable drug pumps.

We emphasize that risk management is a key component of cybersecurity and that the potential damages are frequently more extensive than the harms that are really likely to materialize [15, 16]. Hospitals and the healthcare industry as a whole must be cautious in identifying the range of potential hazards which are prepared for unanticipated effects and realistic in their estimation of the danger of these damages. Best security procedures must be applied whenever practicable. Devices shouldn't, for instance, utilize default passwords. Additionally, if a device is recognized to have cyber vulnerability, a software update must be applied to it whenever practical.

## 4 Smart Health at Home

When one examines the rising usage of healthcare devices in the home, the preceding two scenarios come together in intriguing ways [17, 18]. Digital technologies are leaking out of traditional healthcare environments and finding their way to typical homes, whether this is due to older adults wanting to "age in place," the growing use of wearable sensors, the interest in accountable care, or the need to monitor patients "in the wild" to help ensure treatment success. Due to two systems' attempts to coexist in the same physical location and probably on the same wireless network (home and healthcare), we now have a perfect storm of security and safety flaws [19, 20]. The hospital gains access to the residence through a backdoor, and vice versa. Beyond security, what's at risk is the planned dependence on data produced at home to guide healthcare decision-making. This information may be crucial in assisting older people to avoid paying for institutional care, assisting patients undergoing treatment to avoid going to ERs unless absolutely necessary and reach them quickly in an emergency, and assisting patients whose illnesses have environmental triggers (such as asthma) in managing their treatment and behavior on a daily basis [21].

## 5 Consequences of the Scenarios

### *5.1 Safety and Physical Security*

The primary need for gadget collections is that they provide for both individual and physical security. Although there has been a lot of research and funding from businesses to stop cyberattacks, safeguarding collections of devices poses additional problems that have not yet been solved. Particularly, the capability of smart devices to control tangible aspects of the environment (like the temperature of a home or whether a door is locked) creates potential threats to a person's physical safety that call for even higher levels of assurance than current countermeasures for cyberattacks. Rethinking the fundamental ideas behind security and system management is also necessary given the dispersed and networked nature of the many systems present in device collections [22]. Without adopting a multi-system perspective, security solutions won't be able to foresee and mitigate vulnerabilities that result from improper settings or attacks that take advantage of flaws in how devices communicate with one another and with cloud computing.

The hospital scenario for managing collections of smart devices is better recognized since interactive gadgets have been used in hospitals for a while and because hospitals are subject to regulatory frameworks that demand greater degrees of compliance. Based on this experience, researchers have learned that (a) the lifecycle of the device, including how software is upgraded, must be taken into account; (b) the physical accessibility of devices, including the ability for an intrusive party to access interfaces like USB ports or WiFi networks, must be carefully controlled; and (c) the regulatory framework surrounding privacy makes it very difficult to reason about where data is collected, how it is shared, and where it is stored.

We get to numerous conclusions by contrasting two situations involving devices in homes vs devices in hospitals. First, each scenario requires a distinct level of security screening and research. Although there are existing legislative restrictions on medical devices, the complexity explosion and growing number of possible vulnerabilities necessitate a careful review of the certification levels necessary to offer the necessary levels of security and safety assurance for such applications [23]. The recent disclosure of security flaws in St. Jude pacemaker devices brings to light the difficulties in deciding on the appropriate level of cybersecurity assurance for both a given device and the entire collection of devices. Similar to the "ransomware" assaults now being carried out on hospital electronic health record systems, hospitals would be more desirable targets for coordinated attacks. Second, whereas hospitals hire IT specialists to handle large collections of devices, consumers lack such assistance yet are nonetheless confronted with the same complicated system challenges. Similar dangers are highlighted in the most recent report from the Commission on Enhancing National Cybersecurity for small enterprises who cannot afford an IT team. Both situations will benefit from advancements in people's ability to comprehend and handle such a collection of gadgets, but the consumer scenario calls for reconsidering how such systems may be communicated in words that are understandable to regular people.

### *5.2 Data Privacy*

In a world where an increasing number of smart gadgets collect data, distribute it, and monetize it, privacy is difficult to comprehend and ensure. At the device level, the paradigm of software monetization through advertising is being used. Numerous unpaid smartphone applications currently gather user data and sell it in methods that are neither clear nor explicit to the user [24]. Theoretically, algorithms like differential privacy might restrict the possible effects of data sharing, but they are rarely employed in practice, leaving us unsure of how more intrusive smart gadgets and sensors will affect our privacy. Individual users may become overwhelmed by the intricacy of knowing a single program's privacy policy, such as Facebook, and the burden of doing so for every device and application used takes focus and complexity that is beyond the range of most individuals.

Think about the difficulty in comprehending not just one gadget but numerous that interact in intricate ways. Users won't be able to appreciate the privacy consequences of their decisions without new means for describing what information is being gathered and shared, not by each individual device but rather in aggregate. Think of purchasing a smart fork, for instance. Beyond just counting the individual forklifts, how can a consumer determine what data the fork is gathering? Imagine if the customer then purchases a smart plate. Can the plate and fork communicate data? What can be learned from the combination of the data that cannot be learned from each data source, assuming that is the case? Think of a smart thermostat with an Internet TV service, for instance. The usage of cellphones to operate these devices generates information that may be used to identify residents of the house. The thermostat can thus determine who is home and when and where they are. A few IoT devices in the house can provide a fairly accurate chronology and map of household activity [25].

What is legal in terms of data gathering and sharing in the hospital context is determined by regulatory compliance with HIPAA and other rules. IT specialists' knowledge and skills are necessary to navigate the complexity of determining whether a certain device configuration is compliant. Even the most knowledgeable IT experts may struggle to fully comprehend the privacy implications of a specific setup as the complexity of the data being gathered rises and the uses for it grow more varied.

Beyond just comprehending how connected devices' intended behavior may have privacy ramifications, it becomes more difficult and dangerous to provide proper privacy protections when security flaws result in data breaches. Fortunately, improvements in the storage and use of encrypted data will probably lead to technical answers to some of the problems associated with avoiding data breaches [26]. However, the existence of nefarious state-sponsored actors that target the privacy of well-known people significantly raises the level of protection required to offer general confidence in such systems. The biggest obstacle to be overcome, in the end, may be social engineering assaults and attacks based on a lack of human knowledge of these systems.

### *5.3 Client Interactions and Usability*

We have already argued that there are serious security, safety, and privacy risks associated with the ability of experts or consumers to comprehend and operate complex systems. There are two ways to handle this issue: either simplify the systems to the point where they can be comprehended or provide better conceptual models for users and tools to lighten the load. Due to the extensive usage of open-source software, such as Linux, in the development of many smart devices, their setup is often complex and requires a high level of skill to comprehend and administer. There are certain restrictions on how much can be simplified by removing options and making the setup accessible as a "wizard." To expressly exclude devices from communicating with one another is another simplification. This plan lessens the user's administrative load, but it also drastically lowers the system's potential worth [27]. To close a garage door, for instance, a gadget that determines there is no one inside a home would seek to speak with a device that controls the garage door, but their contact would be blocked.

New methods for assisting people in gaining a broader perspective on their full device collection may be an alternative. A "device dashboard" in particular may show a view of all the devices and their configurations and relationships. A view like this may help users comprehend their whole network using principles they are already acquainted with from controlling individual PCs, such security and privacy settings. Tools that assist users in tracking the configuration, such as individual software updates, and ensuring the present configuration is secure can be created and sold with the aid of such an aggregate view.

Future study and policy on smart devices must focus heavily on how people perceive technology, their readiness to accept it, and their difficulties in sustaining it. If the person configuring the system doesn't give strong passwords or recognize that the system is setup incorrectly, no degree of software protection is sufficient [28]. In the past, the human aspect of design could be delegated to knowledgeable IT specialists, but increasingly, users must deal with these challenging usability issues themselves.

## 6 Security Measures in IoT for Safe Autonomous Civic Systems

We suggest the following strategy for advancing the research agenda and policy agenda based on developments in the Internet of Things and ad hoc collections of smart devices in light of this debate. Figure 3 as shown below shows the importance of IoT security.

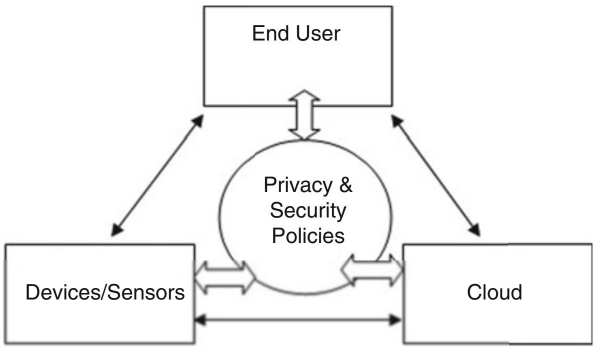



**Fig. 3** Importance of security in IoT

## 6.1 Importance of Security, Privacy, and Usability Together

These must be taken into account jointly, and federal expenditures should give preference to strategies that improve a person's capacity to comprehend and manage complex systems. It is necessary to specify and mandate minimal levels of cybersecurity assurance for wide-scale device deployment due to the possibility of threats to physical safety. For establishing the degree of analysis and testing necessary for smart device goods, milestones must be created (akin to targeted EPA emission requirements). Improvements are specially needed as follows:

- The openness of the software that the devices are using for analysis and inspection.
- The degree of examination and evaluation necessary for certification.
- The key components' degree of hardening (crypto, secure communication, secure update channels).

## 6.2 Security and Management of Individual IoT Devices

Existing initiatives like the Cybersecurity Assurance Program and the Report of the Commission on Enhancing National Cybersecurity offer standards and specifications to assist guarantee that individual devices are adequately protected. In addition to the present investments, we advise:

- Putting a focus on adversarial thinking when revising safety standards for Internet-connected electrical equipment, which will help to reduce the amount of harm that a remote attacker with malicious intentions might cause.
- Increasing the focus on creating hardware and software from tested parts. Modern verification techniques should be used to create increasingly sophisticated subsystems, including cryptographic implementations, as program verification technology is developing quickly.
- Software analysis and testing tools are increasingly needed to validate software installations in smart devices, with varying levels of analysis needed depending on how much a gadget could endanger physical safety.
- To enable software to be fixed as new vulnerabilities are found, software update requirements for deployed devices should be improved.
- Utilizing cutting-edge encryption to update exploit-resistant techniques.
- Establishing "cradle-to-grave" specifications that outline what happens to gadgets when they are no longer updated, for instance, because the firm that made them went out of business.
- Supporting studies to assist users in properly maintaining their hardware and software.

### *6.3 Various IoT Devices' Grouping Management*

Despite the fact that managing groups of devices is becoming more common, very little has been written about it. As a first step, we advise the development of:

- Explicit program that takes into account every device in a collection and gives a user an overview of them (device dashboard).
- Tools for configuration management that make it possible for the user to comprehend and modify the configuration over time to keep it secure.
- Reductions in the complexity of configuration management that shield users from making frequent mistakes that compromise privacy or security.
- A user interface that makes advantage of management principles people are already familiar with.

## 7 Summary

With sensing and intelligence beginning to be incorporated into every gadget, technology is fast advancing and having a larger influence on society than ever before. Although there are many advantages for individuals, businesses, and organizations due to the advancements, there are also hazards involved until the technology is well understood. We have discussed use-case scenarios and the aspects of safety, security, and privacy to highlight some of the repercussions of these developments. I think that because changes are occurring so quickly and because risk and uncertainty are at such a high level, it is important to priorities funding for research that might assist prevent possible issues. The potential benefits to human lives, our national interests, and the economy are great enough to justify significant research expenditures to maximize the technology's advantages. A user interface that makes advantage of management principles people are already familiar with.

## References


1. Fu, K., Kohno, T., Lopresti, D., Mynatt, E., Nahrstedt, K., Patel, S., Richardson, D., & Zorn, B. (2017). *Safety, security, and privacy threats posed by accelerating trends in the internet of things*. http://cra.org/ccc/resources/ccc-led-whitepapers/.
2. Mohammed, H., & Qayyum, M. (2017). Internet of things: A study on security and privacy threats. https://doi.org/10.1109/Anti-Cybercrime.2017.7905270
3. Kumar, J. S., & Patel, D. R. (2014). A survey on internet of things: Security and privacy issues. *International Journal of Computer Applications, 90*(11).
4. Abdur Razzaq, M., Sheikh, R. A., Baig, A., & Ahmad, A. (2017). Digital image security: Fusion of encryption, steganography and watermarking. *International Journal of Advanced Computer Science and Applications (IJACSA), 8*(5).
5. GeekPwn. (2017). *IoT devices have a large number of low-level loopholes*. [Online]. Available: http://www.sohu.com/a/129188339_198147

6. Lin, J., et al. (2017). A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications. *IEEE Internet of Things Journal, 99*, 1.
7. Roman, R., Zhou, J., & Lopez, J. (2013). On the features and challenges of security and privacy in distributed Internet of Things. *Computer Networks, 57*(10), 2266–2279.
8. Yang, Y., et al. (2017). A survey on security and privacy issues in Internet-of-Things. *IEEE Internet of Things Journal, 4*(5), 1250–1258.
9. Jia, Y. J., et al. (2017). ContexIoT: Towards providing contextual integrity to Appified IoT platforms. In *Network and distributed system security symposium* (pp. 1–15).
10. Mavrogiorgou, A., Kiourtis, A., Perakis, K., Pitsios, S., & Kyriazis, D. (2019). IoT in healthcare: Achieving interoperability of high-quality data acquired by IoT medical devices. *Sensors, 19*(9), 1978.
11. Lemayian, J. P., & Al-Turjman, F. (2019). Intelligent IoT communication in smart environments: An overview. In *Artificial intelligence in IoT* (pp. 207–221). Springer.
12. Mukhandi, M., David, P., Pereira, S., & Couceiro, M. S. (2019). A novel solution for securing robot communications based on the MQTT protocol and ROS. In *IEEE/SICE international symposium on system integration (SII)* (pp. 608–613).
13. Coman, F. L., Malarski, K. M., Petersen, M. N., & Ruepp, S. (2019). Security issues in internet of things: Vulnerability analysis of LoRaWAN, sigfox and NB-IoT. In *2019 global IoT summit (GIoTS)* (pp. 1–6). IEEE.
14. Alam, S., Siddiqui, S. T., Ahmad, A., Ahmad, R., & Shuaib, M. (2020). Internet of Things (IoT) enabling technologies, requirements, and security challenges. In *Advances in data and information sciences* (pp. 119–126). Springer.
15. Li, W., & Wu, D. (2019). Bridging the gap between security tools and SDN controllers. *ICST Transaction on Security and Safety, 5*(17), 156242.
16. Almiani, M., Abu Ghazleh, A., Al-Rahayfeh, A., Atiewi, S., & Razaque, A. (2020). Deep recurrent neural network for IoT intrusion detection system. *Simulation Modelling Practice and Theory, 101*, –102031.
17. Sarwar, M. I., Iqbal, M. W., Alyas, T., et al. (2021). Data vaults for blockchain-empowered accounting information systems. *IEEE Access, 9*(2021), 117306–117324.
18. Naqvi, M. R., Iqbal, M. W., Ashraf, M. U., et al. (2022). Ontology driven testing strategies for IoT applications. *Computers, Materials & Continua, 70*(3), 5855–5869.
19. Gondal, F. K., Shahzad, S. K., Iqbal, M. W., Aqeel, M., & Naqvi, M. R. (2021). Business process model for IoT based systems operations. *LGU, Research Journal for Computer Science and IT, 5*(4), 1–10.
20. Kabir, S. (2021). Internet of Things and safety assurance of cooperative cyber-physical systems: Opportunities and challenges. *IEEE Internet of Things Magazine, 4*(2), 74–78.
21. Gope, P., & Sikdar, B. (2021). A comparative study of design paradigms for PUF-based security protocols for IoT devices: Current progress challenges and future expectation. *Computer, 54*(11), 36–46.
22. Zhang, Y., He, D., Obaidat, M. S., Vijayakumar, P., & Hsiao, K.-F. (2021). Efficient identity-based distributed decryption scheme for electronic personal health record sharing system. *IEEE Journal on Selected Areas in Communications, 39*(2), 384–395.
23. Alazab, M., Reddy Maddikunta, S. P. R. M. P. M. P. K., Gadekallu, T. R., & Pham, Q.-V. (2022). Federated learning for cybersecurity: Concepts, challenges and future directions. *IEEE Transactions on Industrial Informatics, 18*(5), 3501–3509.
24. Eisele, M., Maugeri, M., Shriwas, R., Huth, C., & Bella, G. (2022). Embedded fuzzing: A review of challenges tools and solutions. *Cybersecurity, 5*(1), 18.
25. Aldahmani, A., Ouni, B., Lestable, T., & Debbah, M. Cyber-security of embedded IoTs in smart homes: Challenges, requirements, countermeasures, and trends. *IEEE Open Journal of Vehicular Technology*. https://doi.org/10.1109/OJVT.2023.3234069
26. Anagnostopoulos, T. (2023). Smart campus safety systems survey. In *IoT-enabled unobtrusive surveillance systems for smart campus safety* (pp. 39–95). IEEE. https://doi.org/10.1002/9781119903932.ch5

27. Padmanaban, S., Nasab, M. A., Shiri, M. E., Javadi, H. H. S., Nasab, M. A., Zand, M., & Samavat, T. (2023). The role of internet of things in smart homes. In *Artificial intelligence-based smart power systems* (pp. 259–271). IEEE. https://doi.org/10.1002/9781119893998.ch13
28. Jeribi, F., Amin, R., Alhameed, M., & Tahir, A. (2023). An efficient trust management technique using ID3 algorithm with blockchain in smart buildings IoT. *IEEE Access, 11*, 8136–8149. https://doi.org/10.1109/ACCESS.2022.3230944

# A Deep Learning Approach for the Sales Prediction in Retail Stores: An End-to-End Analysis and Implementation

**Shriram K Vasudevan, T. S. Murugesh, M. S. Narassima, Nitin Vamsi Dantu, Siniraj Pulari, and Sunandhini Muralidharan**

## Abbreviations

| | |
|---|---|
| ACF | Autocorrelation function |
| ANN | Artificial neural network |
| AR | Autoregressive |
| ARIMA | Autoregressive integrated moving average |

S. K. Vasudevan (✉)
Intel India Pvt. Ltd., Bengaluru, India

T. S. Murugesh
Department of Electronics and Communication Engineering, Government College of Engineering Srirangam, Tiruchirappalli, Tamil Nadu, India

On Deputation from Department of Electronics and Instrumentation Engineering, Faculty of Engineering and Technology, Annamalai University, Chidambaram, Tamil Nadu, India
e-mail: tsmurugesh@gces.edu.in

M. S. Narassima
Great Lakes Institute of Management, Chennai, India
e-mail: narassima.s@greatlakes.edu.in

N. V. Dantu
North Eastern University, Boston, MA, USA
e-mail: dantu.v@northeastern.edu

S. Pulari
Bahrain Polytechnic, Isa Town, Bahrain
e-mail: sini.raj@polytechnic.bh

S. Muralidharan
L&T Technology Services, Bengaluru, India
e-mail: sunandhini.muralidharan@ltts.com

 
N. Gupta, S. Mishra (eds.), *Internet of Everything for Smart City and Smart Healthcare Applications*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-34601-9_2

| | |
|---|---|
| BI | Business intelligence |
| CNN | Convolutional neural network |
| CPU | Central processing unit |
| DCNN | Deep convolutional neural network |
| FFNN | Feed-forward neural network |
| FMCG | Fast-moving consumer goods |
| FPGA | Field-programmable gate array |
| GPU | Graphics processing unit |
| LSTM | Long short-term memory |
| MA | Moving average |
| MAE | Mean absolute error |
| MAPE | Mean absolute percentage error |
| MLP | Multi-layer perceptron |
| MSE | Mean squared error |
| PACF | Partial autocorrelation function |
| ReLU | Rectified linear unit |
| RMSE | Root mean squared error |
| RMSLE | Root mean squared logarithmic error |
| RNN | Recurrent neural network |

## 1 Introduction

The manufacturing business world nowadays is highly growing and becoming more and more competitive. Any retail company vying to cope with the competitors to outperform them has to be well aware of their sales and their effect on the company's near future. The estimation of accurate future sales based on the pertinent current scenario stands decisive to achieving this. It helps the company understand the market better and predict future sales in a better way, often referred to as sales forecasting, a business intelligence (BI) technique. For any retail store under consideration, numerous products would be sold, and multiple customers would visit their stores to purchase the desired products. The demand for every product continuously fluctuates because of the ever-changing customer requirements, price changes, and promotion schemes. More often, the products in stores include vegetables, fruits, fast-moving consumer goods (FMCGs), and other grocery products, which have a decreased shelf life. Hence, it is crucial to place orders and purchase the products based on sales data, which can be made effective by forecasting the data history logged and maintained in the store. Accurate forecasts help minimize inventory and wastage to retain customers by fulfilling their demands. Apart from addressing the above issue, it also aids in appropriate planning, decision-making, and marketing of products. This work compares the performances of basic models mentioned in the related work on retail store data by using the standard metrics and also discusses an improved deep learning model which performs better when compared to other models, as detailed below.

## 2 State of the Art

Time series forecasting involves predicting future events, such as revenue evaluation, using available data [1]. Time series forecasting aims at predicting the future more precisely using historical data in any field, including business, stock market, sports, health care, etc. The most common forecasting method that helps retailers in marketing is sales forecasting, which allows them to make decisions on hiring and resource management. The sales forecast is a statistical job done in any business sector to predict future sales as accurately as possible. The following steps are to be remembered while performing time series forecasting on data [2]. Firstly, the nature of data required for forecasting and the structure in which the available data could be useful in forecasting are to be worked upon. The complete statistical information about each attribute of data should be known, and the best parameter for forecasting should be identified and selected wisely. The selection of parameters and forecasting methods would affect the accuracy of the forecast. The best model that fits the data more precisely is chosen, including basic statistical and deep learning models. The best model is found by evaluating using the metrics such as mean squared error (MSE) and mean absolute percentage error (MAPE) [2].

Sales forecasting is more of a regression problem than a time series problem. The main assumption for this problem is that the pattern of the past data would be reflected in the future. This might not hold well for products that are not consistent over a period or for products that are produced for single-time sales, such as jerseys manufactured for particular Olympic games or for a particular conference. Major components that are addressed during forecasting include level, trend, and seasonality. It is always essential that a product's seasonality should be identified properly as the data cycles around for every season with a change in the level and trend [3]. The effects on pattern include the level, which indicates the constant obtained from the regression performed on data; trend, which indicates the persistent long-term change in pattern; and seasonality, which defines the periodic data fluctuation and autocorrelation [3, 4]. Other major factors to be considered are removing noise or outliers in the data and shaping data. Stacking multiple models resulted in better accuracy than that of a single model. The model used has two layers: the base training layer has the XGBoost model, and the second layer has a basic neural network model for which features from the first layer are taken as input [4]. The advantages noticed while using XGBoost include consistent accuracy and faster learning, i.e., scalability.

The dataset that is to be used for forecasting has to be appropriately shaped for better results. The data are segregated into linear and nonlinear components using discrete wavelet transform. The linear components are fed as inputs to autoregressive integrated moving average (ARIMA) model, while the artificial neural network (ANN) handles the nonlinear components. This hybrid model utilizes the goodness of models astutely, and it has been tested on four real-world datasets. The proposed model has resulted in notable accuracy compared to single models like ARIMA and ANN separately [5]. Basic linear models have been used on different real-time datasets for forecasting. The basic average model predicts the future based on the average of historical observations. The drift model uses the recent past observations and the average change in observations as the key parameters to forecast future points [6]. The exponential smoothing method assumes that the recent observations would provide more relevant information than past observations, hence assigning decreasing weights (exponentially) to distant data points [6]. This method is effective if the time series changes shortly over time. The random forest also displays better results, which is an ensemble of decision trees whose results are finally combined with each other. The abovementioned models' performances are compared using metrics such as mean absolute percentage error (MAPE), mean squared error (MSE), and mean absolute error (MAE). Despite the simplicity of algorithms, exponential smoothing and ARIMA models worked better on a few datasets. Consistent improvement in accuracy is observed in the ensemble model, which is a combination of the abovementioned models [7].

Deep learning frameworks are also used for forecasting time series data, most popularly LSTM, a variant of recurrent neural network [8, 9]. Neural network is known for extracting deep features from the input data. A model, which is a combination of the wavelet transform, autoencoders, and LSTM, is used for stock price forecasting [7]. Wavelet transform works on denoising the data and is expedient while dealing with irregular time series data. The autoencoders are used for training the data, which have a simple three-layer architecture. The first and last layers are the input and output, respectively, whereas the middle layer is the hidden layer, which works on feature engineering and extracts the features suitable for forecasting. Then, a basic LSTM model is employed for forecasting using the extracted data from the previous layers. The results are then evaluated using basic metrics such as MAPE. The major drawback of using this technique is that the implementation of deep learning models is time-consuming as the data become complicated [7].

## 3 Data Description

The dataset used for analysis is approximately 2 GB and contains the following details:

(a) Customer details
(b) Store location/store information
(c) Purchase information
(d) Billing information
(e) Discount/loyalty details
(f) Transaction information

The dataset used for forecasting is the sales and transaction data of a big retail store from India that holds a memory of approximately 2 GB. As per the analysis, the dataset consists of not more than 7.3 million customers from seven different branches of the store. The data available is that of 3 years, from January 2015 to December 2017. The number of products sold in all 7 branches would sum up to 6000 distinct products from various categories, including grocery, cosmetic products, and so on. As far as sales prediction is concerned, proper feature selection plays a major role. Neglecting parameters that considerably influence sales might lead to inaccurate forecasting and hence steer the entire purchase and production process in the wrong direction. The input variables here include the lagged sales values for past time units, transaction date, product type, and promotions on a particular product. The output variable would be the sales-related features (daily).

## 4 Methodology

### *4.1 Workflow for Time Series Forecasting*

A high-level workflow representation is picturized in Fig. 1. Firstly, the data needs to be pre-processed. Then, the data should be fine-tuned to simplify further tasks and provide better prediction accuracy. Secondly, a model that fits the data more precisely is chosen by trying out various existing models or by improving and combining one or two models. It is not necessary that a model universally works well for all kinds of data. The model that best fits the data can be determined using standard metrics such as error rate and accuracy. The last step would be the prediction of future data. The model is made platform neutral as it is developed on top of Intel oneAPI. For the creation of high-performance, data-centric programs on a variety of architectures, Intel oneAPI is a key set of tools and libraries. The Intel oneAPI has streamlined the TensorFlow code so that it can run on hardware that supports running the DCNN model on CPU, GPU, FPGA, and other inference accelerators. The uppermost layer, which makes use of middleware and frameworks, receives the optimized programs, performs the inference, and delivers the result seamlessly.

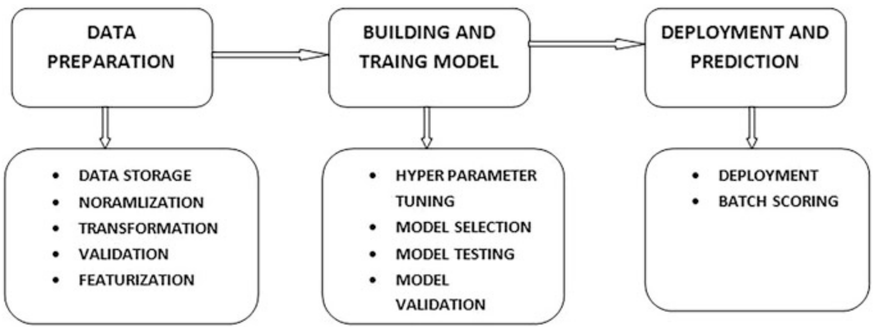



**Fig. 1** Workflow of time series forecasting

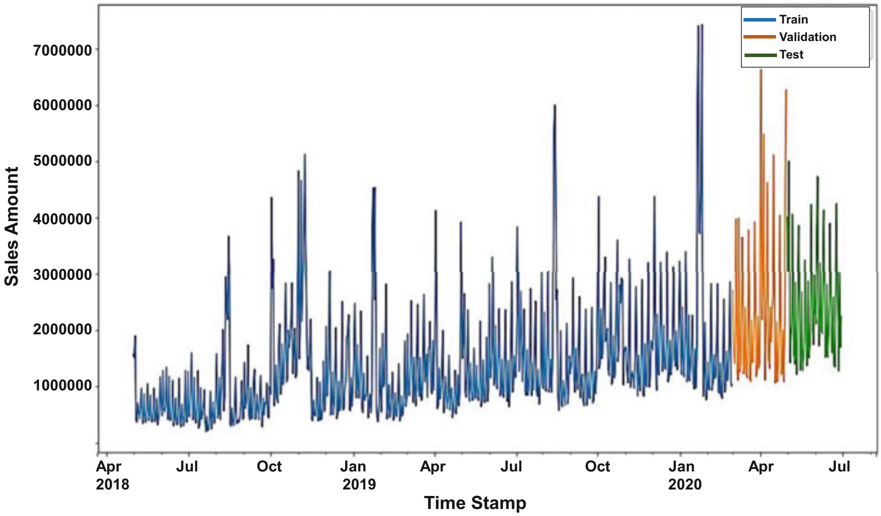



**Fig. 2** Representation of sales of a product in 3 years

### *4.2 Data Pre-processing*

The dataset that has been used contains the details about the customers, store, product, and promotion schemes, of which the feature of our interest for forecasting would be store ID, product type, promotion schemes, and the sales based on these features. Figure 2 displays sales data for a particular branch of a retail store from April 2018 to July 2020. This is a comprehensive set of data that include the complete sales detail. Through this representation, one can understand and visualize that peak in terms of the sales value.

Apart from this, the lagged values of the previous week's sales (Table 1) have been taken as a feature of interest for which the sales values have been normalized and shifted for the past 6 days.

## 5 Forecasting Models

### *5.1 ARIMA Model*

It is one of the most commonly used statistical methods, which works better than most complicated time series forecasting models. The data that could be tested by this method needs to be stationary, meaning that the data's trend and seasonality

**Table 1** Lagged values of sales

| Transaction date | Sale original | Y_t + 1 | Sale_t-5 | Sale_t-4 | Sale_t-3 | Sale_t-2 | Sale_t-1 | Sale_t |
|---|---|---|---|---|---|---|---|---|
| 2018.05.06 | 0.070640 | 0.050384 | 0.187910 | 0.179968 | 0.233610 | 0.027694 | 0.023498 | 0.070640 |
| 2018.05.07 | 0.050384 | 0.031079 | 0.179968 | 0.233610 | 0.027694 | 0.023498 | 0.070640 | 0.050384 |
| 2018.05.08 | 0.031079 | 0.058456 | 0.233610 | 0.027694 | 0.023498 | 0.070640 | 0.050384 | 0.031079 |
| 2018.05.09 | 0.058456 | 0.106248 | 0.027694 | 0.023498 | 0.070640 | 0.050384 | 0.031079 | 0.058456 |
| 2018.05.10 | 0.106248 | 0.025862 | 0.023498 | 0.070640 | 0.050384 | 0.031079 | 0.058456 | 0.106248 |

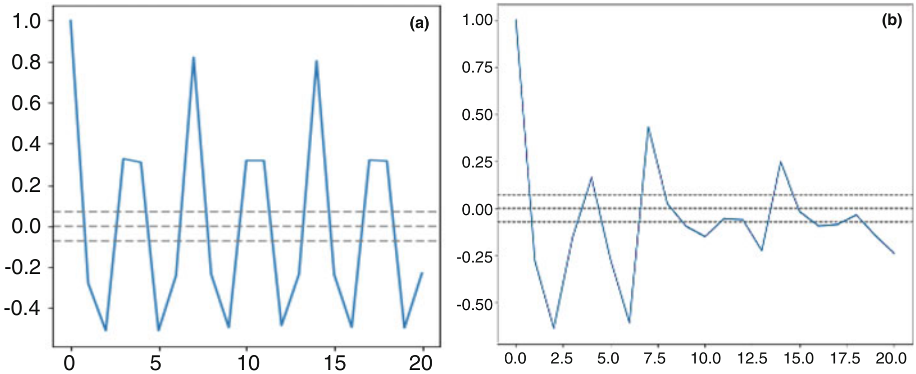



**Fig. 3** (**a**) Autocorrelation graph. (**b**) Partial autocorrelation graph

should be removed. It is done by computing the difference between consecutive data points, which stabilizes the variance of the series, and by taking the logarithm of the data which stabilizes the series' mean. The above-processed data is thus stationary and is suitable to be taken as input for our model. The model has two parts, of which the first part involves autoregression to predict the future values based on a combination of past values of the variable, i.e., with the lagged values of a variable.

$$y_t = a + b_1 y_{t-1} + \cdots + b_n y_{t-p} \tag{1}$$

The second part is the computation of the moving average to predict the values based on past forecast errors instead of lag values.

$$y_t = c + \varepsilon_t + d_1 \varepsilon_{t-1} + \cdots + d_n \varepsilon_{t-p} \tag{2}$$

The appropriate model is selected by deciding the number of lagged values to be included (AR), the number of times the data is differenced to make the series stationary, and the size of the MA window. These AR and MA terms are decided by analyzing the autocorrelation function (ACF) and partial autocorrelation function (PACF) graphs (Fig. 3). ACF defines if the series is stationary, whereas PACF gives the data correlation with its lagged values. The results of the ARIMA model are illustrated in Fig. 4.

The model works well when the data is univariate and the size of the data is less. As we include the lagged values or the dependent variables for prediction, the size of the data grows bigger, and hence, the time for calculating the order of AR and MA increases. As the order increases, prediction quality decreases, which is seen as the major challenge when ARIMA is chosen over other options.

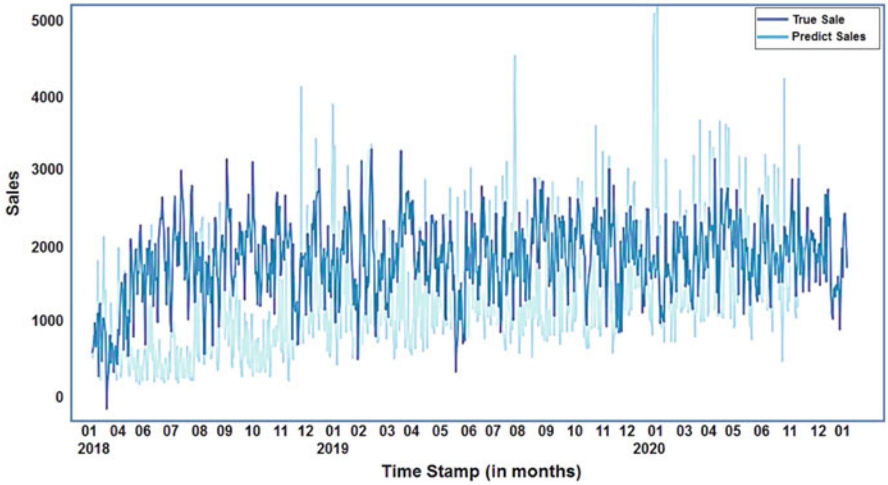



**Fig. 4** Results of the ARIMA model

## 6 Deep Learning Models

Deep learning models are considered because they outperform conventional machine learning and statistical models, especially when the data size is large. The amount of data being used for predictions in industries is large. But both these models have their pros and cons. Deep learning models consume time for training which in this case is at least 3.5 days, but the time for prediction is much lesser when compared with that of linear models. The hardware that is used for training data has 4 GB RAM, NVIDIA GeForce 830 M GPU, and a 1 TB hard drive.

### *6.1 Convolution Neural Network (CNN)*

CNN is developed to be used on images for classification and prediction. The basic layers of this model include the convolution layer and pooling layer linked to a fully connected layer (Fig. 5).

The work of the convolution layer is to read the input data, interpret it based on the filters, and extract the features from the input. The pooling layer provides attention to the important focused features, which is useful in the case of time series data. This model works better than linear models when the parameters are tuned properly. The model designed is simple enough which has 1 convolution layer with 20 kernels and 5 timestamps. This means that the convolution operation is done on the input data as 5 timestamps at a time for 20 times. A pooling layer reduces the

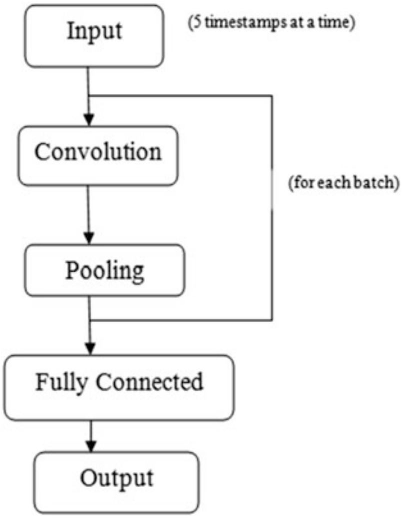



**Fig. 5** Simple CNN architecture

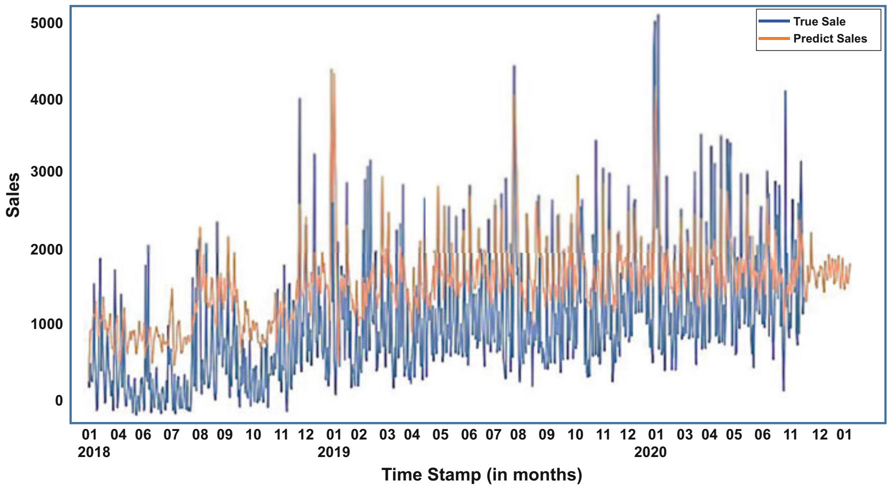



**Fig. 6** Prediction graph – CNN

size to less than half by extracting the features and concatenating them into one single vector by the fully connected layer (Fig. 6).

MSE is used to calculate the loss. The Adam stochastic gradient method is used for updating weights while training the model and for fitting the model for 500 epochs and a batch size of 128.

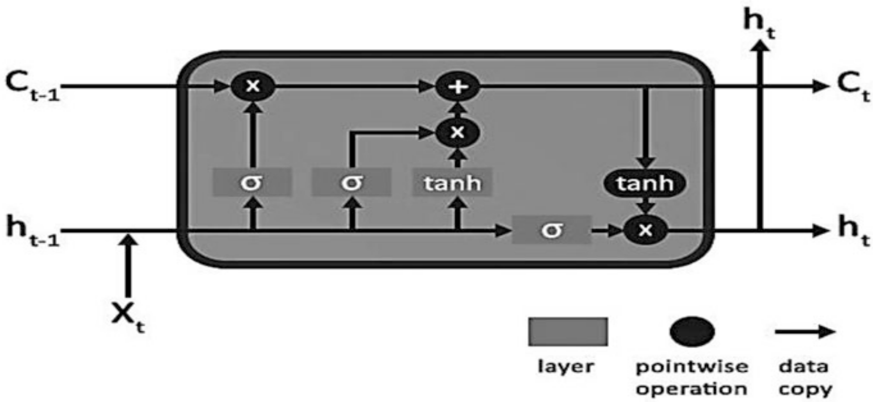



**Fig. 7** Basic LSTM cell

### *6.2 LSTM*

The LSTM model is a variant of the RNN model. This model is known for its performance in time series data. The RNN model has an input layer to take inputs, a hidden layer that performs operations on the input data, and an output layer to generate outputs, but the property change in this model is that it takes the output of the current step as the input for the next step. LSTM deals with data dependency, which is not accounted for in other models. RNN fails to keep track of the dependency as the series size becomes longer. LSTM is designed to solve this problem (Fig. 7), which can memorize the data and possess better backpropagation.

The basic LSTM cell gathers the data and stores it. The cells here connect the past module with the future module that takes over the past information to the present. This is achieved by using gates in the cell. Three gates act as sigmoid layers. The "forgot" gate, which gives 0 or 1 as output, signifies that if the value is near 0, the data is ignored; else, it is retained in the memory. The memory gate is a sigmoid layer followed by a tanh layer which decides on allowing the input data in the cell. The sigmoid layer decides on modifying the data, and the tanh layer maintains the vector of entering the data. The "output" gate decides on sending the filtered input data. It also uses supervised learning to update weights. As far as the model is considered, the input layer takes the input data at every time instant "t." A hidden stacked LSTM layer is shown in Fig. 8a. It calculates the value based on the input data at a particular instance, the previous hidden state, and an output layer. The model used is with 1 neuron fit for 500 epochs, and it is fit using ADAM optimizer (the name being derived from adaptive moment estimation) and MSE as loss function. A dropout layer is added, which excludes the data from the input layer to each block of LSTM from updating weights based on the probability. Instead of a single hidden layer, two hidden layers of LSTM are added. One layer takes the input

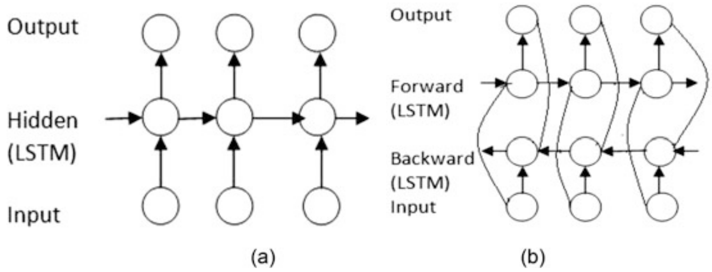



**Fig. 8** Architecture diagram. (**a**) Basic LSTM model. (**b**) Bidirectional LSTM model

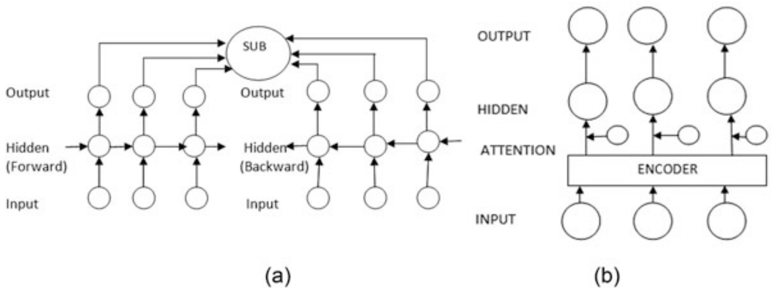



**Fig. 9** Architecture diagram. (**a**) Two-way LSTM. (**b**) Attention LSTM

in a forward direction and the other in a backward direction. Finally, the output of these two layers is concatenated, and the results are recorded as shown in Fig. 8b.

The predictions done by the second model are better than those of the first model, though the training time happens to be a bit more for the second model. The next model built was much more complicated, with two different LSTM models. One model takes input in increasing time series order, and the other model takes input in decreasing time series order, as shown in Fig. 9a. Both models' outputs are subtracted to get the final prediction values. This model shows adverse results compared to the above models including the CNN and consumes more training time. So, in the next LSTM model, an attention module is used in addition to the memory cells available in LSTM. This module involves a comparison of input and output for each timestamp. Based on the weights assigned to each instance, decisions are taken on whether or not the data would be allowed to the next layer, as shown in Fig. 9b. The next LSTM layer generates an output similar to the earlier ones. The predictions obtained seem better than some of the above models, which will be discussed in the succeeding sections.

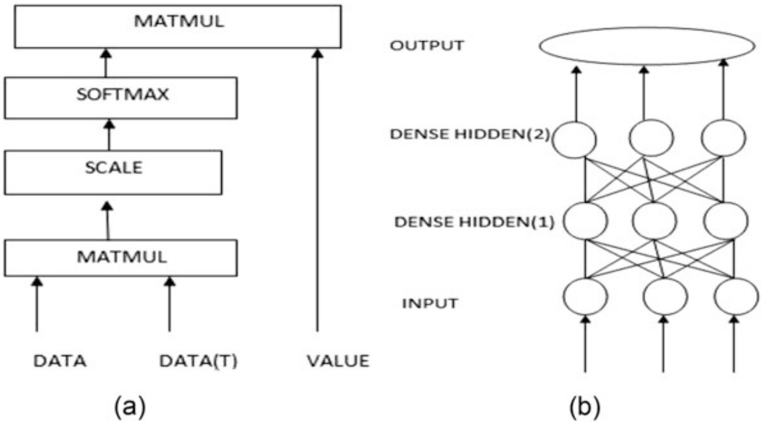



**Fig. 10** (**a**) Attention module. (**b**) Feed-forward neural network

### *6.3 Attention (Cognition)-Based Deep Learning Model*

The LSTM models perform well on sequential data, but the training time and the complexity of these algorithms are higher. To reduce the complexity, a simple model is proposed. The proposed model has two main layers, which include an encoder layer and a decoder layer which further has two layers that are two separate networks. The first layer is the attention mechanism, and the second is the simple feed-forward neural network or MLP. Initially, basic data regression is done; to avoid overfitting, the dropout layer is added with a probability rate of 0.75. Only the data with dropout values greater than the above rate are taken as inputs to the next layer. The attention module (Fig. 10a) requires the output of the encoder for each time step. These outputs are matched with the corresponding outputs of a decoder, and a score is calculated. The calculated score is normalized using the softmax function, which is assumed to be an instance's weight. Based on the weights, the data are sent to the next layer, a simple feed-forward neural network (Fig. 10b).

$$\text{FFNN (input)} = \left(\text{Relu}\left(W_1{}^{*}\text{input} + b_1\right) * W_2\right) + b_2 \tag{3}$$

The feed-forward neural network has an input layer that takes the output of the above layer as input and two hidden layers – one with ReLU activation function and the other with the linear activation function. Predictions made using this model yield better results when compared to those of the LSTM-based models with lesser training time and complexity.

Figures 11, 12, 13, 14 and 15 basically use basic LSTM, bidirectional LSTM, attention LSTM, two-way LSTM, and feed-forward neural network for the prediction. One can see the sharp changes in the predictions in the respective figures.

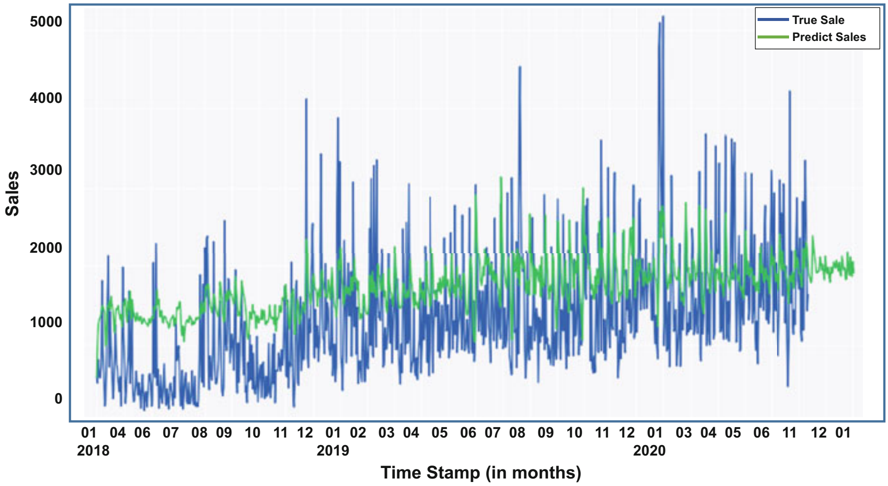



**Fig. 11** Prediction graph of basic LSTM

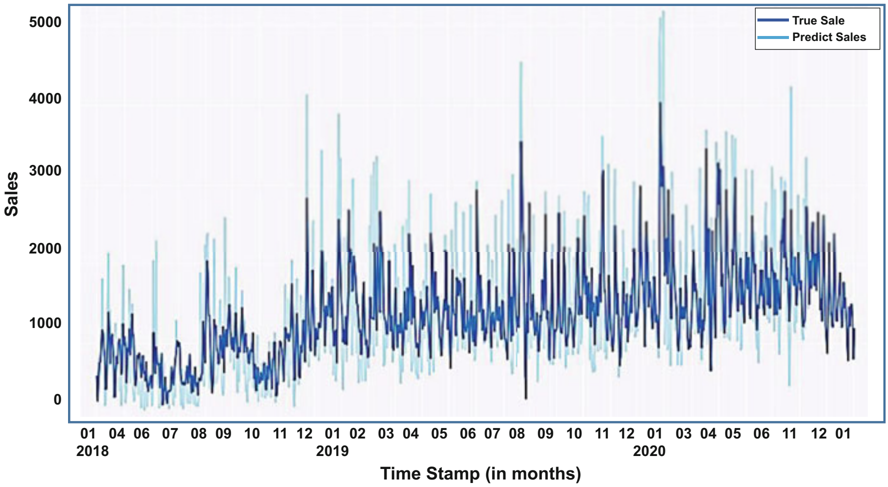



**Fig. 12** Prediction graph of bidirectional LSTM

## 7 Performance Analysis

Basic data analysis is performed based on the gathered data using Figs. 11, 12, 13, 14 and 15 for the discussed models. The data taken for prediction is that of a daily basis for predicting the sales 50 days in advance. As a next step, the anomalies in the data are detected and removed. Then, the models explained in the above sections are used for forecasting [10–12]. As far as the ARIMA model is concerned, pre-processing of data is essential to make the data stationary. This model performed

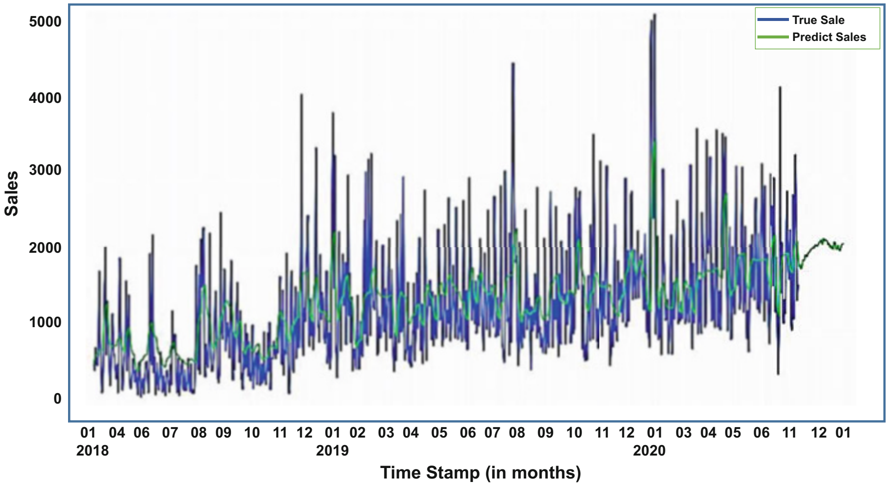



**Fig. 13** Prediction graph of attention LSTM

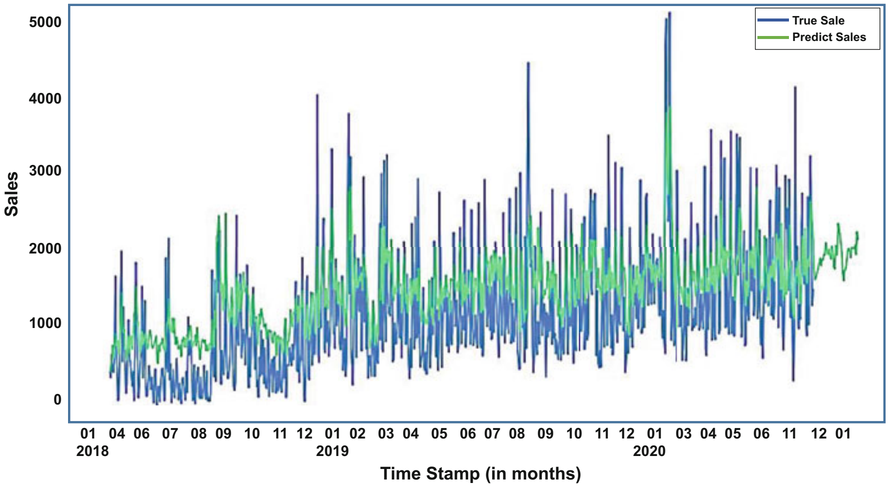



**Fig. 14** Prediction graph of two-way LSTM

better for lesser input data. As the data and randomness in the data grew, the model took longer time for forecasting, and the accuracy of the result dropped. The models are evaluated using the metrics such as root mean squared error (RMSE), MAPE, and accuracy. RMSE is the measure that gives us information about the outliers in our data.

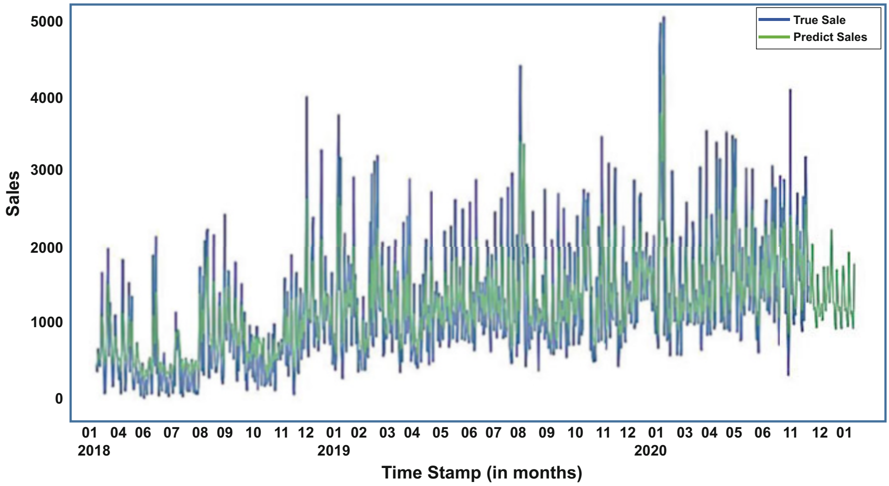



**Fig. 15** Prediction graph of attention-based feed-forward neural network

$$\text{RMSE} = \sqrt{\sum_{i=1}^{N}\left(\frac{(\text{Pred}_i - \text{Actual}_i)}{N}\right)} \tag{4}$$

MAPE is the measure that gives us information on how accurate our model is at prediction.

$$\text{MAPE} = \frac{1}{n}\left|\sum_{i=1}^{N}\frac{(\text{Pred}_i - \text{Actual}_i)}{\text{Actual}_i}\right| \tag{5}$$

RMSLE is the root mean squared logarithmic error, which is the RMSE of a log-transformed variable. This did not penalize huge differences in predicted and actual values when both predicted, and true values are huge numbers.

The data are divided into training and testing data. The above results are obtained by applying the metrics on testing data to understand and compare the prediction quality that each model would deliver (Table 2). ARIMA model has a lesser RMSLE value and the least MAPE score, and the prediction accuracy is just up to the mark. In that case, CNN has better accuracy than the basic LSTM model, but the bidirectional model shows the best accuracy among all LSTM models with the least MAPE score. Considering all the metrics, the attention-based FFNN has a better prediction with an accuracy of 76% and reasonably fewer MAPE, RMSE, and RMSLE values. The prediction of the FFNN model shows a deviation of just 5% from the actual values. The model is less complex compared to the well-performing LSTM models and consumes less time. Hence, the attention-based feed-forward neural network model is proven to be predicted with higher accuracy, lesser training time, and more accuracy compared to ARIMA and LSTM models [13]. ARIMA model also predicts with better accuracy when the prediction is for a day or two.

**Table 2** Comparison of various models based on metrics

| | RMSLE | MAPE | ACC | RMSE |
|---|---|---|---|---|
| LSTM | 0.295987 | 0.1641297 | 0.600801 | 778.513351 |
| BIDIRECTIONAL (LSTM) | 0.184005 | 0.0467932 | 0.692623 | 527.688384 |
| CNN | 0.212075 | 0.0725833 | 0.638625 | 632.191326 |
| TWO PATH (LSTM) | 0.243857 | 0.0998212 | 0.652565 | 633.024306 |
| ATTENTION (FFNN) | 0.150408 | 0.0512961 | 0.763551 | 420.290571 |
| ARIMA | 0.152942 | 0.039235 | 0.461807 | 870.527331 |

On the contrary, when the forecasting horizon increases, the model's performance goes down. One major drawback in the ARIMA model would be the pre-processing that is to be done before applying the model. LSTM gives almost promising results as an attention model, but the model complexity and training time increase as the number of layers in the model increases [14]. The proposed model seems to take comparatively less training time, but one disadvantage would be tuning the hyperparameter, which is similar to the LSTM model [15].

## 8 Conclusion

The present study considered multiple linear models and compared their performances against those of the regression models, RNN, LSTM, and others. Several data including customer details, purchase information, location, billing, discounts, and transaction information were used for prediction. Performances of various methods employed, viz., LSTM, BIDIRECTIONAL (LSTM), CNN, TWO PATH (LSTM), ATTENTION (FFNN), and ARIMA, were compared for accuracy. The FFNN model showed the best accuracy for the tested data. Comparing the estimates measured reveals that each of these methods could be capable enough to outperform each other when tested on different datasets.

Sales of a product are the major key for any business, and improving sales is the goal for every business organization by incorporating new strategies. Sales prediction is one strategy that helps them understand the market better and make better financial decisions. The data used has features affecting sales for 3 years, including the holiday details, the number of products sold daily, and the number of promotional offers on a particular day. With these features and sales value based on these features, an attention-based fusion model is presented, including an attention mechanism and a simple feed-forward neural network. The model would perform better on any sequential data, and here the performance of the model is compared with commonly used models for forecasting, which includes the ARIMA model and different variants of the LSTM model. The comparison is based on metrics such as MSE. The model is seen to train faster when compared to other models, and the training and testing errors are lesser when compared to those of other models. This model outperforms any kind of ensemble model and provides better prediction.

## References


1. Pao, J. J., & Sullivan, D. S. (2017). Time series sales forecasting. Final year project, Computer Science, Stanford University, .
2. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.
3. Pavlyshenko, B. M. (2019). Machine-learning models for sales time series forecasting. *4*, 15. https://doi.org/10.3390/data4010015
4. Khandelwal, I., Adhikari, R., & Verma, G. (2015). Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition. *Procedia Computer Science, 48*, 173–179. https://doi.org/10.1016/j.procs.2015.04.167
5. Fleurke, S. (2017). Forecasting automobile sales using an ensemble of methods. *WSEAS Transactions on Systems, 16*(37), 329–337.
6. Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLoS One, 12*(7), e0180944. https://doi.org/10.1371/journal.pone.0180944
7. Choi, J. Y., & Lee, B. (2018). Combining LSTM network ensemble via adaptive weighting for improved time series forecasting. *Mathematical Problems in Engineering, 2018*, 1. https://doi.org/10.1155/2018/2470171
8. Bianchi, F. M., Maiorino, E., Kampffmeyer, M. C., Rizzi, A., & Jenssen, R. (2017). *An overview and comparative analysis of recurrent neural networks for short term load forecasting* (Springer Briefs in Computer Science). https://doi.org/10.1007/978-3-319-70338-1
9. Wang, B., Huang, H., & Wang, X. (2012). A novel text mining approach to financial time series forecasting. *Neurocomputing, 83*, 136–145. https://doi.org/10.1016/j.neucom.2011.12.013
10. Hülsmann, M., Borscheid, D., Friedrich, C. M., & Reith, D. (2012). General sales forecast models for automobile markets and their analysis. *Transactions on Machine Learning and Data Mining, 5*(2), 65–86. https://doi.org/10.1007/978-3-642-23184-1_20
11. Adhikari, R., & Agrawal, R. K. (2014). A combination of artificial neural network and random walk models for financial time series forecasting. *Neural Computing and Applications, 24*(6), 1441–1449. https://doi.org/10.1007/s00521-013-1386-y
12. Padilla, W. R., García, J., & Molina, J. M. (2021). Improving time series forecasting using information fusion in local agricultural markets. *Neurocomputing, 452*, 355–373. https://doi.org/10.1016/j.neucom.2019.11.125
13. Tan, Z., Wang, S., Yang, Z., Chen, G., Huang, X., Sun, M., & Liu, Y. (2020). Neural machine translation: A review of methods, resources, and tools. *AI Open, 1*, 5–21. https://doi.org/10.1016/j.aiopen.2020.11.001
14. Gui, Z., Sun, Y., Yang, L., Peng, D., Li, F., Wu, H., Guo, C., Guo, W., & Gong, J. (2021). LSI-LSTM: An attention-aware LSTM for real-time driving destination prediction by considering location semantics and location importance of trajectory points. *Neurocomputing, 440*, 72–88. https://doi.org/10.1016/j.neucom.2021.01.067
15. Mou, L., Zhou, C., Zhao, P., Nakisa, B., Rastgoo, M. N., Jain, R., & Gao, W. (2021). Driver stress detection via multimodal fusion using attention-based CNN-LSTM. *Expert Systems with Applications, 173*, 114693. https://doi.org/10.1016/j.eswa.2021.114693

# Blockchain Technology: A Game Changer for Smart Healthcare Systems

**Vladimir Rocha, Arlindo Flavio da Conceição, and Dario Vieira**

## 1 Introduction

Blockchain technology, a technology for reliable recording and distributed consensus [1, 2], offers alternatives for creating interoperable, auditable, and secure systems. Blockchain technology was popularized in 2008 with the creation of the Bitcoin cryptocurrency by Satoshi Nakamoto [3], and it has been used primarily to perform financial transactions in an anonymous, auditable, reliable, and secure manner by preventing third parties (e.g., banks) from intermediating these transactions.

The concept used in the blockchain is that of *distributed ledger* and basically consists of an ordered and consistent chain of transactions distributed in several nodes of a *peer-to-peer* network. After the success of Bitcoin, other technologies were integrated into the initial version proposed by Nakamoto in order to optimize the performance of the solution [4].

This author was partially supported by SmartMed Project, Research Council of Norway, project number: 288106.

V. Rocha
Federal University of ABC (UFABC), Santo André, Brazil
e-mail: vladimir.rocha@ufabc.edu.br

A. Flavio da Conceição
Federal University of São Paulo (UNIFESP), São Paulo, Brazil
e-mail: arlindo.conceicao@unifesp.br

D. Vieira (✉)
EFREI Research Lab, Paris, France
e-mail: dario.vieira@efrei.fr



N. Gupta, S. Mishra (eds.), *Internet of Everything for Smart City and Smart Healthcare Applications*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-34601-9_3

Among these technologies are Ethereum and Hyperledger. The Ethereum [5] and Hyperledger [6] frameworks make use of smart contracts, small independent programs stored in the blockchain itself, which allow them to perform operations on the blockchain's record chain. For simplicity, one can think of smart contracts, as being similar to the existing *store procedures* in relational databases [2].

The use of smart contracts expands the power of blockchain, which in addition to storing states (e.g., balance of an account in the financial context), can now store behaviors (e.g., sending insufficient balance messages). With the use of blockchain and smart contracts, we can deal with other more general contexts, for example: verify the consistency of the unique identification of users, mediate the interoperability of data in real time, ensure the privacy of information, and make auditable all actions of access to a data.

Blockchain is estimated to have a relevant impact in the coming years [7], with applications in, for instance, input and product supply chain registration, digital banks applications [8], Internet of Things (IoT) [9], Healthcare [10, 11], self-sovereign identities [12], financial management, real estate records, asset control, certificate records (birth, marriage, death), among other categories of applications. In Healthcare, blockchain can be applied to control access and distribution of sensitive information, transparency, auditability of service delivery, and data interoperability, among other situations. Recent research, however, indicates that despite its potential, blockchain is a tool that cannot be successfully applied in all cases [13]. Thus, Healthcare professionals must learn to discern the viable use cases where the technology will really make a difference.

The goal of this book chapter is to provide healthcare professionals, students, and researchers with the necessary understanding to analyze the viability of applying blockchain technology to the healthcare field. In addition, to analyze and understand the first steps to be taken in the implementation of new projects involving blockchain. To achieve this goal, this chapter presents the following structure: Sect. 2 shows some important computational concepts for understanding blockchain technology; these concepts are prerequisites to correctly understand the potential of the technology. Next, Sect. 3 explains, step by step, how the technology works. Section 4 presents what consensus is and its relevance for distributed systems and blockchain. Section 5 compares the most widely used blockchain technologies for creating new blockchain-based applications. Section 6, the main focus of this chapter, presents the possible scenarios of blockchain applied to Healthcare. Finally, in Sect. 8, we present our final thoughts and perspectives on the future of technology.

## 2 Blockchain Properties

Blockchain technology has been widely used due to its important characteristics, such as decentralization, availability, integrity, auditability, and privacy. The main properties of this technology will be analyzed in the following:

- **Information decentralization**
  Decentralization, in the context of information, refers to the dissemination and replication of information, preventing a central or absolute entity from having power over it. In blockchain, decentralization must be observed from two sides: (i) who holds power to perform some action on information; (ii) who physically owns the information.
  In the first case, if the information is centralized in an entity (e.g., institution, organization, or person), who can perform any action on the information? Note that several institutions in our day to day work this way. For example, you own an account at a bank, but a bank manager can update your personal information anytime. In this sense, the bank has the power over information.
  In the second case, if the information is stored only on the entity's computers, this entity physically owns the information. Following the example of the bank, it stores information in its infrastructure, to which external or third parties do not have access.
  Blockchain aims that information is decentralized both from the perspective of the power to perform some action and from the perspective of physical storage. In the first case, the participants are responsible for jointly deciding what information can be modified or inserted. For this, they will need to reach an agreement (called consensus) if this information is valid. In the second case, the information is stored on the participants' computers, avoiding physical centralization.
- **Availability**
  The availability of information is closely related to physical decentralization. Once the information is stored on the participants' computers, it becomes available to be used regardless of whether one of them leaves the system. For example, if ten computers allow access to the same information, nine of them could leave the system, and the information would still be accessible on the remaining computer. On the other hand, if the information is centralized in only one computer and it fails or leaves, no one else will be able to access it.
  The concept used to provide availability is information replication. In this sense, the same information must be replicated and distributed on several computers. As mentioned earlier, blockchain replicates using the *peer-to-peer* network. In addition to replication, if new information is added, it is necessary to synchronize all replicas. For this, distributed consensus techniques are used, which will be explained in Sect. 4.
- **Privacy**
  Privacy allows all blockchain operations, called transactions, to be performed anonymously, preventing third parties from knowing exactly which real-world entity (e.g., person or institution) made the transaction. For this, encryption techniques are used, which allow an entity (in the real world) to be identified (in the blockchain) only through a number. In this sense, a third party, even having access to read the entire blockchain, will only view the operations executed identified just by numbers, without knowing who is the person or institution behind them.

An important point to mention is how to obtain this number, which corresponds to the person or institution identifier. Basically, there are two approaches. The first approach is to use a certificate authority who verifies that the person or institution, in the real world, really exists. This approach is usually used in private blockchains, detailed in Sect. 3.7, where only a few people have access to the system. The second approach is to use a certificate authority who does not verify that you really exist. This concept may seem abstract, but think that, in the virtual world, people can impersonate others or create imaginary entities. This approach is used in public blockchains, detailed also in Sect. 3.7, where anyone has access to the system.

- **Integrity**
  Integrity is an important point in blockchain. As the information can be distributed (replicated) on several computers, it is necessary to trust that this information is reliable, that is, that it has not been altered by anyone. However, how can you trust the information someone presents if you do not necessarily know the person or institution in real life? (this could happen by the anonymity mentioned before)
  To surpass this issue, blockchain uses the chain concept, which allows the linking of pieces of information. Why? Making an analogy with a real-life chain, a person can quickly identify that there was a break if a link is broken, meaning that the chain is not reliable (i.e., its integrity has been compromised).
  How to create these information links? In blockchain, a block is created with a unique identifier among all blocks that already exist. Within this block will be added operations (transactions) performed by users. The link between two blocks is created by making the last block point the identifier of the previous one, thus forming the link. Note that if someone wants to break the integrity of a chain of X-Y-Z blocks (for example, by breaking Y), he must create a new W block that points to the X block and make the Z block point to the new W block. The problem is that the creation of blocks costs time and computational power, which in practice prevents the creation of block W in an adequate time, as will be explained in Sect. 4.2.
- **Immutability**
  Blockchain immutability refers to the fact that the information (i.e., the block header information or the transactions contained in it) cannot be modified from the moment it is added to the chain.
  How is it possible to achieve immutability if the information changes? For example, a person's balance may change (due to credit and debit operations for a given amount) or a person's medical record may change over time (due to new exams), etc.
  To allow the changes, the blockchain just creates a new block, adding the changes as a new transaction. Note that with this, the chain will contain all the modifications made to the information through time, as a complete history, and not just the last state of it.
  In the case of a person's balance, consider that a person named A has just logged into the system. At that moment, a transaction representing the join is created.

Next, person B sends 10 units to A. At this point, a new transaction is created where B sends 10 units and A receives 10 units. Finally, A sends three units to B. A new transaction is created at that moment where A sends 3 units and B receives 3 units. Note that all operations are added into the blockchain, quite different from having only the last state of the final balance of A, which would be seven units.

Of course, immutability, which does not allow information to be changed or removed from the chain of blocks, is not always desired. For example, consider a blockchain-based healthcare system that adds exams to a patient's electronic records. At some point, the patient demands (supported by law) to remove all his information stored in the system. Note that immutability prevents this from happening, leading the health organization to possible legal problems if it keeps them.

- **Auditability**

  Blockchain auditability allows anyone who owns the chain to inspect and verify that all the information contained in it is valid. Auditability uses the characteristics presented above, such as privacy and integrity, among others.

  To verify that the information is valid, it is necessary to verify that both the blocks and the transactions are valid. In the first case, it is required to validate that each block points to the previous block until reaching the first block generated, following the links explained before and detailed in Sect. 3. In the second case, it is necessary to verify that all transactions of a given entity, such as a person or institution (due to anonymity, only its identifier will be shown), are consistent in the context in which the blockchain is being used.

  For example, coherence in the context of financial transactions is related to the balance owned by a person. Whoever performs the audit can verify if, at a given moment, the person has enough balance, using the immutability condition presented above, to send money.

## 3 How Does Blockchain Work?

In this section, we present the technologies that compose the blockchain and show how they interact to execute correctly. For this, we first offer an overview of how it works. Next, we explore the concept of the block, transaction, and chain and how they are entangled to form the blockchain. From there, some ideas of cryptography and *hash* will be explained, which will increase the blockchain's security and efficiency. Finally, the concept of consensus will be described, that is, how the computers that are part of the blockchain can reach an agreement on which blocks are valid and which one must be added to the chain.

## 3.1 Overview

In this section, the blockchain will be built step by step through two scenarios, the first in the context of a purchase of a good between two people and the second in the context of the health area.

In the first scenario, consider a person purchasing a particular good from another through an electronic medium, such as a website. For both to be sure that the financial transaction took place, that is, that the money was sent by one and received by the other, it will be necessary for a computer to store this transaction.

In the second scenario, consider a physician who sees a patient and records his medical consultation using an electronic medical record. So, both people (physician and patient) can view this information, and it will be necessary for a computer to store the medical record.

In Fig. 1, it is possible to conceptually visualize a block that contains a header and the transactions. Note that the concept of a transaction is very general and can be both: information from a financial transaction and information from a patient's medical record, as mentioned in the first and second scenarios, respectively. A block can contain several transactions but suppose that a block has only one transaction for simplicity.

Now, the patient has a second medical consultation with the same physician. After the service, the physician records the information in the patient's medical record, generating a new block containing a new transaction. As shown in Fig. 2, in this new block, transaction 2 will be part of the patient's record, with the change made by the new (second) medical consultation. Note that the second block points to the first, creating an interconnected blockchain.

At this point, it is necessary to ask two questions: (1) what happens if the computer that recorded the transactions no longer has access to the Internet or if its hard drive where the transactions were stored burns?; (2) what happens if someone, who does not have permission, modifies the information in the medical record?

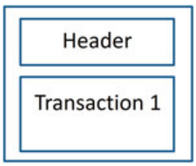



**Fig. 1** Block

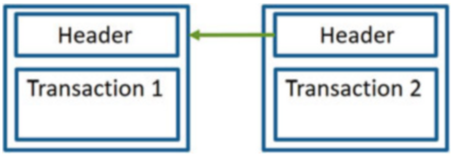



**Fig. 2** Chain of blocks

In the first question, there is a problem with the information availability. That is, transactions may be lost and cannot be recovered. In the second question, the information is available but unreliable, which means we cannot ensure its integrity.

To solve these issues, one of the alternatives would be to replicate the chain of blocks on several computers. In that way, if one computer becomes unavailable, another computer could take its place. In the case of modified information, the other replicas could check whether there was any fraud and try to reach a consensus. Both cases will be analyzed in Sect. 4.

Once a computer has the blocks, it will need to analyze whether the blocks (and the transactions contained in the blocks) are valid or not. In this sense, what information should have each block to perform the validation?

## *3.2 Block and Chain of Blocks*

As mentioned, blockchain is composed of an interconnected chain of blocks, which in turn contain one or more transactions. Now, it is necessary to understand what information composes the block.

The block is a structure composed of two modules: the header and a transaction list. The header consists of various metadata that uniquely identifies the block. The list of transactions, on the other hand, identifies the transactions contained in that block. For simplicity, in this text, we always represent the list of transactions with just one transaction but remember that lists could contain thousands of transactions.

In Table 1, it is possible to observe the main fields that determine the header of a block along with its description. Specifically, the Merkle Root, the Difficulty Target, and the Nonce fields will be explained in more detail later.

With the information shown in the aforementioned table, let us create the blocks for the second scenario (i.e., medical consultations). Consider that in the first consultation, held at 8.30, block 1 was created with transaction 1. On the second consultation, at 19:30 on that same day, a new block was created with transaction 2. In Fig. 3, we can see some header information that the system generated for both blocks.

Note that in block 2, the value of the Previous Block Hash will correspond to the value that uniquely identifies block 1 (in Sect. 3.5 will show how this value

**Table 1** Block header

| Name | Responsibility |
|---|---|
| Previous Block Hash | Point to the header of the previous block. |
| Merkle Root | Unique number determining the transactions belonging to the block. |
| Timestamp | Date and hour of the creation of the block |
| Difficulty Target | Number determining the threshold when creating a block. |
| Nonce | Number used to create a block. |

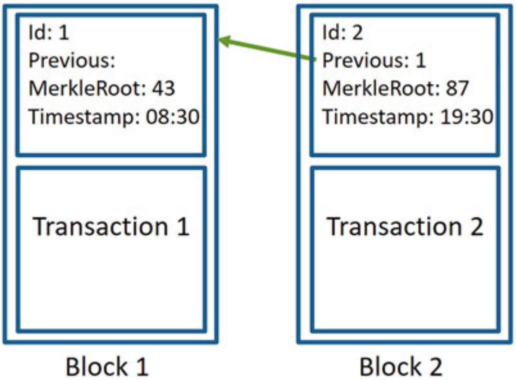



**Fig. 3** Chain of blocks with header

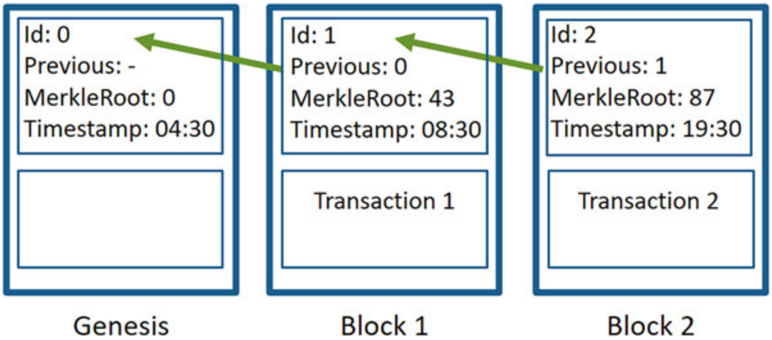



**Fig. 4** Chain of blocks with genesis

is created). Furthermore, the system will generate a value that will determine the transactions that exist in the block (i.e., Merkle Root) and the time the block was created at 19:30 (i.e., timestamp).

At this point, a question should arise. To who block 1 points in its Previous Block Hash field? Intuitively, it should point to a previous block. However, there must be some block that does not point to a previous block, thus representing the beginning of the chain. This initial block is called the genesis block.

In a blockchain system, the first block of the chain is called the genesis block, and all computers that use the blockchain must know it. In this sense, if from any block you go back to the previous one, and from there to the previous one, using the value of the Previous Block Hash field, you will finally arrive at the genesis block.

So far, the example of the second scenario can be complemented with the genesis block, as shown in Fig. 4. More technical information about the block can be found in [14].

### *3.3 Transaction and Transaction Chain*

Within each block are stored the transactions sent by the users of the system. A transaction allows showing that a given item (e.g., a monetary value, a document, an access policy, etc.) was created by a certain user and will be available to another.

Conceptually, a transaction can be seen as an event that occurred in the system. In this sense, in the first scenario, the event would be a financial transaction between two people. In the second scenario, the event would be the insertion of the patient's electronic record and the physician's permission for the patient to access and view the record.

In general, a transaction consists of a unique transaction identifier and two modules: inputs and outputs. The input represents the identifier of an entity performing the transaction, and the output represents the identifier of an entity receiving the transaction. Figure 5 shows how the transaction would be for a financial transfer of 5 units from a person with identifier ID1 to a person with identifier ID2.

However, how to validate that ID1 actually owed the transferred item (in the example above, the 5 units) before the transaction? In this sense, like blocks, transactions also need to point to a valid transaction (that is, an already existing transaction stored in some block of the blockchain), creating a chain of transactions. With that, if someone wants to check if ID1 owns the item or not at some moment, he just analyzes all the transactions linked to that identifier, using the transaction pointers for that. More technical information about transactions can be found in [15].

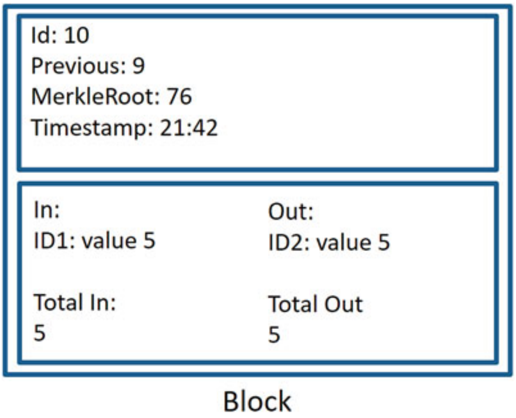



**Fig. 5** Transaction

### 3.4 Public and Private Key

We mentioned that the blockchain is composed of interconnected blocks, which in turn are composed of one or more interconnected transactions. We also mentioned that the transaction contains identifiers of system users who create and send. In this context, how can a computer validate that the identifier really belongs to the person who creates the transaction and not someone impersonating her?

To explain this validation, it will be necessary to first understand some concepts related to the area of cryptography, whose responsibility, among others, is information security. These concepts will be explained below using an example.

Consider that you have a text message that you need to send to someone using some electronic medium, such as an email or a chat application. This text will pass through several intermediary computers until it reaches its destination (this is what normally happens on the Internet). To avoid someone eavesdrops the message, you will need to encode it at the source so that intermediaries cannot tell what the message says, and decode it at the destination to retrieve the message.

In cryptography, encoding at the source is called encrypting, and decoding at the destination is called decrypting. The encoded text, which has no meaning for anyone who does not know how to decode it, is called "ciphertext." Fig. 6 shows the concepts applied to the text "Patient: John. Tuberculosis: negative."

How to encode the text? To do this, encryption uses a secret code called a key (think of it as your credit card PIN) and the use of an encryption mechanism. In the mechanism context, there are two alternatives that we can use: symmetric and asymmetric cryptography. In symmetric cryptography, the same code is used for both encoding and decoding; thus, the source and the destination will need to know the code. In asymmetric cryptography, two different codes are used, one for encoding and one for decoding. Next, we will see the second case, which is the alternative used by most blockchains.

As mentioned, asymmetric cryptography uses two keys, called public and private, which are related to each other. The most interesting thing about this approach is the property that the text encrypted by one key can only be decrypted by the other. As an example, see the mentioned steps in Fig. 7. In the figure, the text "Patient: John. Tuberculosis: negative" is encrypted with a private key (green color), turning it into

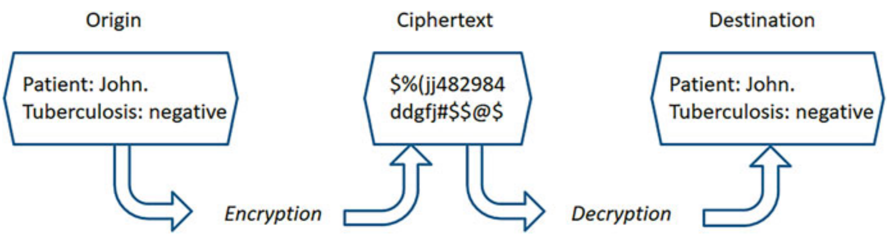



**Fig. 6** Cryptography applied to a text

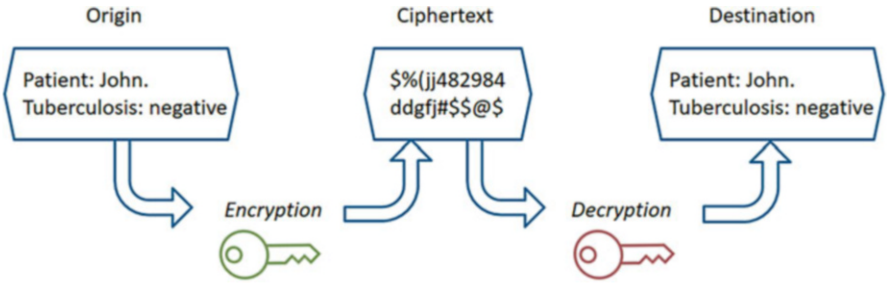



**Fig. 7** Asymmetric cryptography: flow

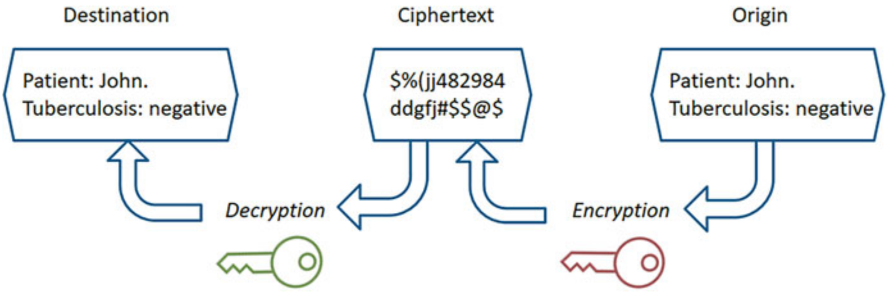



**Fig. 8** Asymmetric cryptography: inverse flow

ciphertext. Next, the ciphertext is decrypted using a public key (red key), recovering the initial text.

In Fig. 8, the reverse steps can be seen, where the text "Paciente: John. Tuberculosis: negative" is encrypted with a public key (red color), turning it into ciphertext. Next, the ciphertext is decrypted with a private key (green color), recovering the initial text.

The important point is that the private key allows for identifying uniquely the owner of the message since he is the only person who should possess the private key (in this sense, this key should not be shared with anyone). With that, the receiver will be able to identify that the message really corresponds to the owner, given that the only way to decrypt it is by using the public key related to that private key. More technical information about cryptography can be found in [16].

Finally, it is worth mentioning that generating the private key, and its corresponding public key, must be performed by a recognized (and normally trusted) entity. In real life, for example, several companies carry out this activity, such as Verisign, DigiCert, among others.

## 3.5 Hash Function

As mentioned so far, a user can create a transaction using his private key, insert it inside a block, concatenate it with a previous block, and replicate it on other computers. The other computers, in turn, will obtain these blocks and verify whether the information contained is valid or not. However, note that the more transactions there are, the more information needs to be validated. Wouldn't there be a way to validate if the information is correct just by comparing two numbers, regardless of the size of the information?

Again, cryptography comes into play, providing a mechanism for compressing information that has the following characteristics: deterministic, "avoids conflicts," and is "one-way" [17].

To understand these three characteristics, consider that we initially need to compress the text "Block 101 with 1 transaction from user John." Let us assume that an engine compressed the text to "B101-1-J." The determinism is related to the fact that the compression of the initial text will always result in the same value "B101-1-J." The "avoids conflicts" characteristic, called collision, is related to the fact that different texts should not result in the same value. Thus, for example, the text "Block 101 with 1 transaction from user Joh" (without the consonant "n" at the end) should create a completely different compressed text.

Finally, the "one-way" characteristic is related to the fact that makes it impossible to recover the original text from the compressed one. In this sense, "B101-1-J" would not fulfill this condition, as someone could understand that "B101" corresponds to block number 101, 1 corresponds to the number of transactions, and "J" corresponds to the name of some user.

The mechanism normally used by a blockchain to attest the existence of an information is called hashing, which allows compressing a text of any size into another of a fixed size. There are several algorithms that perform the mentioned compression, such as MD5, SHA1, among others. Below is an example using MD5 to understand the features mentioned above.

If the following text "Block 101 with 1 transaction from user John" (without quotes) is the input to the MD5 algorithm, regardless of how many times MD5 is applied to that input text, the result would be: 07abed8341ca66731d515fe67df4345c. In turn, for the text "Block 101 with 1 transaction from user Joh" (without quotation marks and without the consonant "n"), the code obtained will be very different: 09ff12389ffa319918eca4d0d8e2f7d8. See that, even with a small change, the value is completely different if compared with the previous one, corroborating the idea of avoiding conflicts or collisions. Finally, note that for the code obtained, there is not even a hint on how to get the text before being compressed, corroborating the one-way feature. More technical information about hash functions can be found in [16, 18].

### 3.6 Merkle Trees

A tree can be described as a data structure composed of a set of linked nodes, which start at a node (called the root) and end at one or more nodes (called the leaves). The most interesting thing about this structure is that to get from the root node to a leaf, there will only be one possible path to follow. Furthermore, each non-leaf node (called parent) may have one or more nodes (called children). For example, Fig. 9 shows a tree numbered from the root node (with the number 1), in which each node has two children (left and right), to the 4 leaf nodes (with numbers 4 to 7).

The Merkle tree [19] is a tree widely used in the blockchain to check that transactions inserted within a block are valid [3]. Before understanding how the blockchain uses the Merkle tree, let us analyze how it is built. Imagine that a block has 4 transactions, each with its unique identifier. Figure 10 has the same structure as Fig. 9, but each node of the former will store a value as explained below. As shown in Fig. 10, a *hash* value will be obtained from each transaction identifier and added to the corresponding leaf. Thus, leave 4 will have a value $v_4$ calculated with the function $H(\text{ID1})$. Leave 5 will have a value $v_5$ calculated with $H(\text{ID2})$, where $H$ is a hash function, e.g., MD5. Next, the hash value of the concatenation of a pair of transactions is created, and this value is inserted in their parent node. For example, node 2 will have a value of $v_2$ calculated from the function $H(v_4 v_5)$. The same for node 3, which will have a value of $v_3$, calculated from the function $H$ for transactions ID3 and ID4. Finally, the root node will have the value $v_1$ obtained from the hash of the concatenation of the values in nodes 2 and 3 (i.e., $H(v_2 v_3)$), thus forming the Merkle tree.

As mentioned in the previous section, we know that the hash function has certain properties that allow the value obtained to be deterministic. Note that, in the case of

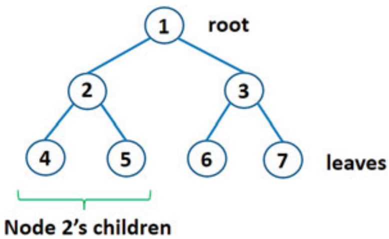



**Fig. 9** Tree

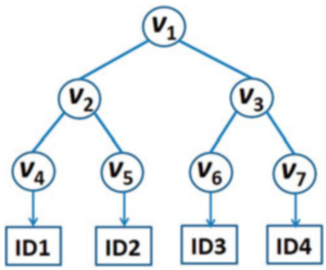



**Fig. 10** Merkle tree

the Merkle tree, we will have always the same root value if we start with the same transaction identifier values. Note that if any leaf transaction has another identifier, the root node will have also another value.

Where the blockchain uses the Merkle tree? As mentioned in Table 1, one of the header fields that must be filled when creating the block is the Merkle Root value, which determines the transactions that exist in the block. Note that the value $v_1$ of the root node in Fig. 10 represents all transactions in the block because it is a combination of them. Thus, when creating the block, the Merkle Root of the header will be filled with the existing value at the root of the Merkle tree.

## *3.7 Types of Blockchain*

According to Buterin [20], the creator of Ethereum, there are two types of blockchains: permissioned and permissionless. In a permissionless blockchain, any member can make modifications and audit the chain. In a permissioned blockchain, only authorized members can perform operations on the chain. In addition, it is common to associate the term permissionless blockchain with public blockchain and permissioned blockchain with private, federated, or consortium ones.

In a permissionless blockchain, any entity can join and leave the system at any time. Upon joining, the entity becomes a member that will be able to perform modifications and audition on the entire blockchain. As it is possible that potentially each member owns the blockchain (through replication), in this type of blockchain, there is a total decentralization of the information. Examples of permissionless blockchains are Bitcoin [3] and Ethereum [5] platforms.

In a permissioned blockchain, only certain entities will be transformed into members of the system, and only they will be allowed to perform operations on the chain. Thus, some of them will be able to read the blocks, others will be able to write, and others will be able to audit the chain. To identify and authorize the members, it will be necessary to create trusted entities responsible for managing permissions. Examples of permissioned blockchain are the Hyperledger Fabric [6] and Corda [21] platforms.

Wüst and Gervais [22] proposed a subdivision of the permissioned blockchain into public and private. The division only considers auditability, where the "public permissioned blockchain" allows any member to verify the chain's data and the "private permissioned blockchain" only allows verification for a well-defined and authorized set of members.

Blockchain applications that require non-anonymized user identification tend to be built using permissioned infrastructure. On the other hand, permissionless structures tend to offer greater anonymity. Another important issue for choosing the type of blockchain is the creation and maintenance of the infrastructure that supports the network of nodes. A private or federated blockchain is usually the responsibility of an institution responsible for keeping it operational. In turn, public blockchains concentrate less decision-making responsibility on an institution.

## 4 Consensus

As we mentioned, a chain of blocks is replicated on several computers for availability reasons (if one of the computers leaves the system, others may take its place), but see the following situation that can happen.

Consider that at the first instant, three computers A, B, and C have the same chain of blocks 0 and 1 as shown in Fig. 11.

At the second moment, as shown in Fig. 12, computer D sends a message for A, B, and C to add the valid yellow block (because it points to block 1), but only A and B receive it. Concurrently, computer E also sends the message to A, B, and C to add the valid green block, but only C receives it. A question that may arise is, why some messages are not received? On the Internet, messages could be lost, mainly due to congestion on the routers from which the message passes until reaches its destination.

At the end of the second moment, Fig. 13 shows that computers A and B have the yellow block and computer C has the green block, both pointing to block 1 (therefore, both green and yellow blocks are valid).

Consensus aims for computers to agree on a certain value [18]. In the case of blockchain, they must reach an agreement on which block (green or yellow) should be added at the end of the chain, i.e., after block 1. At the end of the agreement, all

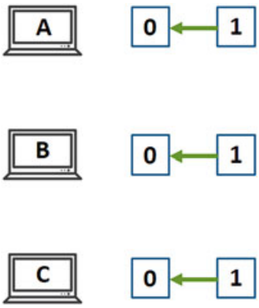



**Fig. 11** Consensus: initial moment

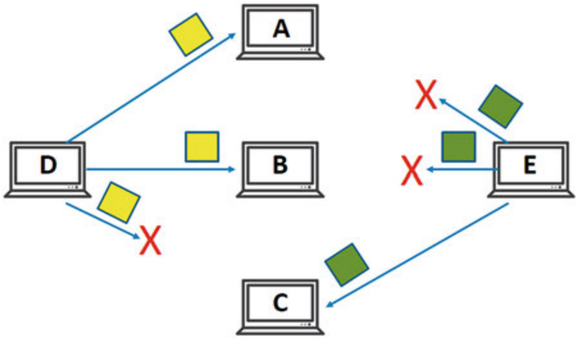



**Fig. 12** Consensus: Receiving blocks

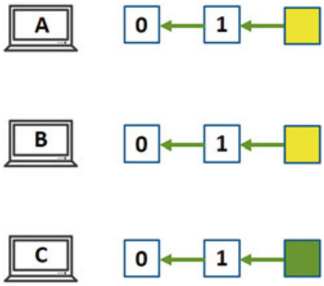



**Fig. 13** Consensus: Adding blocks

computers (in the example, A, B, and C) must have the same chain of blocks. Below we will see three strategies employed by blockchain to reach consensus.

## *4.1 Consensus Process via Byzantine Fault Tolerance (BFT)*

In BFT-like consensus, usually used in permissioned blockchains, only a few computers are responsible for carrying out the agreement process, usually a dozen or fewer. The choice of these computers can be based on different characteristics or available resources, such as their computational power, uptime, and bandwidth, among others. In order to understand the consensus process, let us assume that 7 computers were chosen.

Having the computers that will run the consensus, the first step of the process is to choose a leader among them, that is, a computer responsible for deciding which block will be inserted at the end of the chain. There are several alternatives to carry out the choice, for example: (i) the most updated computer will be elected as the leader; (ii) each computer can have a unique identifier number and the smallest/greatest of these will be chosen as the leader, etc. In our example, if the other six consensus computers see that the leader is no longer available, the computer with the second lowest identifier will be chosen as the new leader, and so on.

With the leader elected, each computer that is not part of the consensus (called client) can propose a new block to be added to the chain. For that, the client sends the proposed block to any of the seven computers responsible for the consensus, but only the leader will add it to the chain. In this sense, the consensus computer that received the block will redirect the client's request to the leader.

The leader, in turn, will receive block requests, either directly from the client or through redirection from the other consensus computers, and will order these requests. Generally, the first request that arrives is the first one to be served. After ordering the requests, the leader will add the blocks in that order and will replicate this information to the other six consensus computers.

When a consensus computer receives the order of the blocks, it will insert the blocks in its respective chain, responding to the leader that the insertion was executed. Finally, as soon as the leader gets the majority of the same responses, it will respond to the client's computer that its block was successfully added. By majority, we mean half of the consensus computers (if we do not have byzantine, or bad, behaviors, such as paxos [23]) or three quarters (if we have it).

### 4.2 Consensus Process via Proof of Work (PoW)

In PoW-like consensus, usually used in permissionless blockchains, it is possible for all computers to execute the consensus process, unlike BFT where only a few are chosen. To coordinate the potential thousands of computers, the choice of a leader is not the most adequate, considering that this leader may not be able to deal with all messages coming from all computers or even could be unreliable or untrustable, e.g., in a security sense. Thus, it will be necessary to create a process that does not rely only on one computer.

The blockchain used in Bitcoin was one of the first techniques, applied in a real system, to enable consensus on thousands of computers. The process begins with a computer (called a miner) obtaining from another computer (another miner or a full node, which also has the complete history of blocks) the chain of all blocks, with their respective transactions, existing up to that moment. To give you an idea, the complete history in 2022 has thousands of blocks, with approximately 400 gigabytes in size. In order to understand how the process works, let us consider that the chain has only 100 blocks.

When the miner (called M1) obtains the full history of blocks, it must verify that each transaction is valid (using the transaction-chain concept mentioned in Sect. 3.3) and that each block is valid (using the block-chain concept mentioned in Sect. 3.2).

After performing the validations, the M1 miner will obtain new transactions that were sent by clients (and that do not exist in any previous block), creating a new block with these transactions.

When creating a new block (in our example, block 101), it is necessary to fill in the header fields. First, the "timestamp" field corresponds to the miner's computer time, for example, 7/10/2022 1:30 PM. Next, the "Previous Block Hash" field should point to block 100, calculated through the hash of that block, using MD5, or some similar hash function. The "Difficulty Target" and "nonce" fields are numbers that the miner should use to prove to other participants that computational work was actually done to create the block.

The work is performed in the following way: "Difficulty Target" is a number calculated by the system and generated approximately every 2 weeks. This number, which allows an average of 2016 blocks to be created every 2 weeks (i.e., 1 every 10 minutes), usually starts with a number of zeros, for example, 000101827749837 (starting with 3 zeros). Next, the miner will need to find a number smaller than the "Difficulty Target," which is obtained through the hash of the block being created. However, think about the following, only with the header information (timestamp, previous block hash, and Difficulty Target), MD5 may not be able to generate a value lower than the "Difficulty Target." For example, let us assume that the MD5 of the text "02/12/2019 13:30 block100 000101827749837" gives the value 100405682589837. Note that this value is greater than 0001018277498372371. This is where the nonce field comes in. The nonce is a header attribute that can be modified so that the hash is less than the Difficulty Target. See the example in Table 2 below for different nonces, applied to the aforementioned header.

Note that the miner computer had to perform three calculations (computational work with the nonces 0, 1, and 2) to find a number less than the Difficulty Target. In real blockchain systems, such as Bitcoin, the computer performs millions or billions of calculations, hence the name Proof of Work.

After finding a suitable nonce value, the M1 miner inserts it into the header and creates the block (let us call this block yellow for illustrative purposes). After creating the yellow block, miner M1 disseminates this block to other miners, and let us call some of these, M. PoW Consensus must consider two cases: (1) miner M who received the yellow block from M1 was also trying to create a block pointing to block 100, and (2) miner M had already received a valid green block, from another miner M2, which also pointed to block 100.

For the first case, as soon as miner M receives the yellow block, he must immediately check if the received block is valid (e.g., looking if it is pointing to block 100). If it is valid, M will stop creating the current block and will add the yellow block at the end of its chain, starting the creation of a new block, and now pointing to the yellow one.

For the second case, miner M received a valid green block pointing to block 100. But how could this happen if miner M1 created the yellow block, also valid, at that instant? Note that the creation of a green block could have happened by another miner M2 (nothing prevents this situation) and disseminated before the yellow block M1 arrived at M.

Now, as Fig. 14 shows, M will have two valid blocks, one green and one yellow, both pointing to block 100. What to do? The rule for reaching consensus will be to wait for new blocks to arrive and, after a certain amount of time, choose the one

**Table 2** Example of Nonces

| Nonce | Text to be used in MD5 | Result |
|---|---|---|
| **0** | "12/02/2022 13:30 block100 000101827749837 **0**" | 100405682589837 |
| **1** | "12/02/2022 13:30 block100 000101827749837 **1**" | 320106680549244 |
| **2** | "12/02/2022 13:30 block100 000101827749837 **2**" | 000047479763563 |

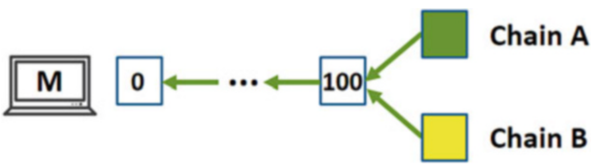



**Fig. 14** Consensus: Two chains

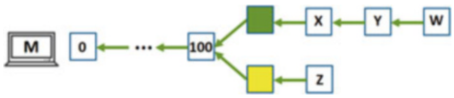



**Fig. 15** Consensus: Receiving blocks

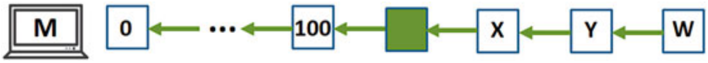



**Fig. 16** Consensus: Choosing the longest chain

with the longest chain from block 100 onward. Imagine the following case, after a certain time, M receives three new blocks that contain the green block, called chain A, and only one new block that contains the yellow block, called chain B, as shown in Fig. 15. Finally, as chain A is longer than chain B, miner M will discard chain B (which contains the yellow block and the Z block), as shown in Fig. 16. Note that the consensus will eventually occur since the longest chain rule will be followed by all miners (including M1).

## *4.3 Consensus Process via Proof of Stake (PoS)*

As in PoW, PoS-like consensus is also usually used in permissionless blockchains in which it is possible for all computers to execute the consensus process.

To coordinate the computers to reach a consensus, the process chooses randomly, at a regular interval time, a set of them, forming a committee. Also randomly, the process chooses a leader computer from the committee, which will be responsible for creating and adding a block, just as BFT-like consensus does. As soon as the leader creates the block, it sends it to all computers in the committee, which will validate the block and add it to its chains.

One question that could arise is: why is used the “stake” word? In order to be selected as a committee member, the computer must deposit an amount of money (usually in cryptocurrency value), which is called a “stake.” This money is necessary to penalize misbehavior. Indeed, if the computer had bad behavior (e.g., creating a bad block or invalidating a valid one), a percentage of the deposited money will be lost.

## *4.4 Other Consensus*

There are several consensus protocols that were developed and, some of them, deployed in blockchains beyond PoW and PoS. We can mention, for example: Delegated PoS (DPoS), Proof of Authority (PoA), Proof of Elapsed Time (PoET), Proof of Trust (PoT), and Proof of History (PoH).

For a detailed list and explanation, please see [24, 25].

**Table 3** Blockchain technologies comparison

| Name | Year | Type | Consensus | Smart Contract | Tx/s | Currency |
|---|---|---|---|---|---|---|
| Bitcoin | 2008 | Permissionless | PoW | – | 7 | BTC |
| Ethereum 1.0 | 2018 | Permissionless | PoW | Solidity | 14 | ETH |
| Ethereum 2.0 | 2022 | Permissionless | PoS | Solidity | 100,000 | ETH |
| Hyperledger Fabric 1.4 | 2017 | Permissioned | Paxos/BFT | JS, Go, Java | 3185 | – |
| Hyperledger Fabric 2.0 | 2020 | Permissioned | Paxos/BFT | JS, Go, Java | 20,000 | – |
| R3 Corda | 2016 | Permissioned | Paxos/BFT | Kotlin/Java | 1678 | XDC |
| Cardano | 2017 | Permissionless | PoS | Plutus (Haskell) | 257 | ADA |
| Diem (former Libra) | 2019 | Permissioned | BFT | Move (Rust) | 643 | DIEM |
| Stellar | 2015 | Permissioned | BFT | – | 3000 | XLM |
| Ripple | 2012 | Permissioned | BFT | Codius/Solidity | 1500 | XRP |
| Solana | 2017 | Permissionless | PoS & PoH | Rust/C/C++ | 3000 | SOL |

## 5 Technologies

Currently, there are several blockchains that can be used. Table 3 provides a brief comparison of the most used.

## 6 Blockchain Use Cases for Healthcare

In 2018, Zhang et al. [26] point out six prominent use cases for blockchain in Health: secure data sharing among patients and doctors, monitoring supply chains, aggregated observation of complex events (Big Data), patient identification, implementation of autonomous processes, and data sharing for scientific research.

With the evolution of Smart Contracts, we can use blockchain technology in more complex healthcare scenarios, including:

- Blockchain can provide a secure and tamper-proof way to store and share electronic medical records (EMRs) among healthcare providers, reducing the risk of errors and improving patient care. It can also be used to interoperate among different standards.
- Blockchain can be used to create a tamper-proof record of the supply chain of drugs, making it easier to track the movement of drugs from the manufacturer to the patient and reducing the risk of counterfeit drugs. It can also be used to emit alerts of incompatibility [27].
- Blockchain can be used to securely store and share data from clinical trials and research studies, making it easier for researchers to access the data they need and reducing the risk of errors. It can also be used to manage consent of patients about their data usage [28].

- Medical billing and claims. Blockchain can create a tamper-proof record of medical billing and claims transactions, making it easier for healthcare providers and insurance companies to process and pay claims.
- Secure communication between healthcare providers. Blockchain can create a secure and tamper-proof way for healthcare providers to communicate with each other, share patient information, and collaborate on patient care.

There are some examples of projects and initiatives that are currently using or exploring the use of blockchain technology in healthcare:

- **MediLedger**:[1] This project, which is being developed by a consortium of pharmaceutical companies, aims to create a secure and tamper-proof system for tracking the supply chain of drugs using blockchain technology.
- **Gem Health Network**:[2] This initiative, which is being developed by the blockchain company Gem, aims to create a secure and interoperable network for sharing and managing electronic medical records using blockchain technology.
- **Hashed Health**:[3] This company is using blockchain technology to create a secure and tamper-proof system for managing medical billing and claims transactions.
- **Medicalchain**:[4] Using blockchain technology to create a secure and transparent system for storing and sharing electronic medical records.
- **BurstIQ**:[5] Creating a secure and tamper-proof blockchain-based system for managing and sharing health data, with a focus on patient-centric and privacy-preserving approach.
- **SimplyVital Health** [29]: This company was using blockchain tokens to create a secure and transparent system for managing healthcare data and executing smart contracts for value-based care. The tokens seem to be discontinued.

These are just a few examples, and many more projects and initiatives are exploring blockchain technology in healthcare.

## 7 Future Work

For the future, one application is the use of blockchain technology in managing clinical trial data. By securely storing and sharing data from clinical trials on a blockchain network, researchers could more easily access the data they need to conduct their studies, speeding up the development of new treatments and therapies.

[1] https://www.mediledger.com/.
[2] https://www.gem.health/.
[3] https://hashedhealth.com/.
[4] https://medicalchain.com/en/.
[5] https://burstiq.com/.

A potential application is blockchain-based "smart contracts" in managing healthcare payments and claims. By using smart contracts, healthcare providers and insurance companies could automatically and securely process and pay claims, reducing administrative costs and improving the healthcare system's efficiency.

Another application could be in the field of precision medicine and personalized medicine, where blockchain can be used to store and share patient genetic data securely, which can be used to identify the best treatment for the patient.

Blockchain could also create an open marketplace for personal health data, where individuals could share their data with researchers and companies in return for incentives.

And lastly, blockchain technology can also help create decentralized autonomous organizations (DAOs) to manage healthcare services and data, which can improve the transparency and security of healthcare systems and reduce the cost of healthcare.

As the technology develops, innovative blockchain-based applications and products will reach the healthcare market.

## 8 Conclusions

Therefore, blockchain in healthcare systems has room to grow and can change the healthcare scenario. In general, the complexity lies in integrating systems and writing smart contracts capable of solving relevant problems in healthcare. The first step is identifying how blockchain can contribute to a use case. The second step is to define which kind of blockchain should be used.

In addition, it is worth studying the business model and the economic viability of the projects because using public blockchain may incur costs in transaction execution fees.

## References


1. Swan, M. (2015). *Blockchain: Blueprint for a new economy*. California: O'Reilly Media Inc.
2. Greve, F. et al. (2018). Blockchain e a revolução do consenso sob demanda. In *Minicursos do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (MinicursosXXSlahUndXXSBRC)* (p. 36).
3. Nakamoto, S., & Bitcoin, A. (2008). *A peer-to-peer electronic cash system* (p. 4). https://bitcoin.org/bitcoin.pdf.
4. Eyal, I., Gencer, A. E., Sirer, E. G., & Van Renesse, R. (2016). Bitcoin-NG: A scalable blockchain protocol. In *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)* (pp. 45–59).
5. Wood, G. (2014). Ethereum: A secure decentralised generalised transaction ledger. *Ethereum Project Yellow Paper, 151*(2014), 1–32.

6. Androulaki, E., et al. (2018). Hyperledger fabric: a distributed operating system for permissioned blockchains. In *Proceedings of the Thirteenth EuroSys Conference* (EuroSys '18, pp. 30:1–30:15). New York: ACM.
7. Buldas, A., et al. (2020). Blockchain technology: Intrinsic technological and socio-economic barriers. In *Future Data and Security Engineering* (pp. 3–27). Cham: Springer International Publishing.
8. Andolfatto, D., & Martin, F. M. (2022). The blockchain revolution: Decoding digital currencies. Louis Review: In *Federal Reserve Bank of St. Louis Review*
9. Huh, S., Cho, S., & Kim, S. (2017). Managing IoT devices using blockchain platform. In *Proceedings of the 2017 19th international conference on advanced communication technology (ICACT)* (pp. 464–467). New York: IEEE.
10. da Conceição, A. F., da Silva, F. S. C., Rocha, V., Locoro, A., & Barguil, J. M. M. (2018). Electronic health records using blockchain technology. In *Workshop em Blockchain: Teoria, Tecnologias e Aplicações (WBlockchain, SBRC)*, 1(1/2018).
11. Fernandes, A., Rocha, V., da Conceição, A. F., & Horita, F. (2020). Scalable architecture for sharing EHR using the Hyperledger blockchain. In: *2020 IEEE International Conference on Software Architecture Companion (ICSA-C)* (pp. 130–138).
12. Siqueira, A., da Conceição, A. F., & Rocha, V. (2021). Blockchains and self-sovereign identities applied to healthcare solutions: A systematic review. arXiv preprint arXiv:2104.12298.
13. Monteil, C. (2019). Blockchain and health. In *Digital Medicine* (pp. 41–47). Berlin: Springer.
14. Bashir, I. (2018). *Mastering Blockchain: Distributed ledger technology, decentralization, and smart contracts explained*, 2nd edn. Birmingham: Packt Publishing.
15. Antonopoulos, A. M. (2017). *Mastering Bitcoin: Programming the Open Blockchain*. California: O'Reilly Media.
16. Stinson, D. R. (2005). *Cryptography: Theory and practice*, 3rd edn. Discrete Mathematics and Its Applications. New York: Taylor & Francis.
17. Mironov, I. (2005). *Hash functions: Theory, attacks, and applications*. Technical report.
18. Tanenbaum, A. S., & van Steen, M. (2006). *Distributed Systems: Principles and Paradigms* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall Inc.
19. Merkle, R. C. (1980). Protocols for public key cryptosystems. In *1980 IEEE Symposium on Security and Privacy* (pp. 122–122).
20. Buterin, V. (2015). *On public and private blockchains—Ethereum blog* (p. 7).
21. Brown, R. G., Carlyle, J., Grigg, I., & Hearn, M. (2016). *Corda: An Introduction*. https://docs.corda.net/XXSlahUndXXstatic/corda-introductory-whitepaper.pdf.
22. Wüst, K., & Gervais, A. (2018). Do you Need a Blockchain? In *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)* (pp. 45–54).
23. Lamport, L. (2001). Paxos made simple. In *ACM SIGACT News (Distributed Computing Column) 32, 4 (Whole Number 121, December 2001)* (pp. 51–58).
24. Oyinloye, D. P., Teh, J. S., Jamil, N., & Alawida, M. (2021). Blockchain consensus: An overview of alternative protocols. *Symmetry, 13*(8), 1363.
25. Jalal, I., Shukur, Z., & Bakar, K. A. A. (2020). *A systematic literature review: A study on public blockchain consensus algorithms.*
26. Zhang, P., Schmidt, D. C., White, J., & Lenz, G. (2018). Blockchain technology use cases in healthcare. In *Advances in computers* (vol. 111, pp. 1–41). Amsterdam: Elsevier.
27. Uddin, M., Salah, K., Jayaraman, R., Pesic, S., & Ellahham, S. (2021). Blockchain for drug traceability: Architectures and open challenges. *Health Informatics Journal, 27*(2), 14604582211011228.
28. Arbabi, M. S., Lal, C., Veeraragavan, N. R., Marijan, D., Nygård, J. F., & Vitenberg, R. (2022). A survey on blockchain for healthcare: Challenges, benefits, and future directions. In *IEEE communications surveys & tutorials*.
29. Halamka, J. D., Chu, R., Hanna, K., & Kuzmeskas, K. (2018). Early experiences with blockchain and EHRs. In *Blockchain in Healthcare Today*.

# Part II
# Sustainable Approaches Towards Smart City Applications

# Securing Public Safety Mission-Critical 5G Communications of Smart Cities

**Evangelia Konstantopoulou, Nicolas Sklavos, and Ivana Ognjanovic**

## 1 5G Network Vulnerabilities

The 5G mobile network is poised to host an unprecedented number of devices, with the upcoming 5G New Radio (NR) supporting three generic services, ultra-reliable and low-latency communications (uRLLC), massive machine-type communications (mMTC), and enhanced mobile broadband (eMBB) [1], thus improving numerous areas of urban life. uRLLC provides low-latency and high-availability data services [7]. It focuses on mission-critical use cases such as intelligent transport systems and remote control. mMTC is useful in high-density connections, such as intelligent transportation, smart grid, and intelligent manufacturing. eMBB satisfies user-centric services with high bandwidth requirements for its services. Moreover, 5G will be used for human-to-human communication, as well as for human-to-machine and machine-to-machine communication.

Accordingly, in order to meet these new performance requirements, 5G will offer the advanced technologies of software-defined networking (SDN), network function virtualization (NFV), network slicing (NS), and cloud computing [2]. The SDN is a technology that aims to decouple the control plane from the data layer, thus making it more open and adaptable. NFV addresses flexibility and scalability requirements and is closely related to the cloud. Finally, NS is the implementation of network as a service (NaaS), a service-oriented concept that supports applications over self-contained network slices. While these novel IT technologies improve smart cities'

E. Konstantopoulou · N. Sklavos (✉)
SCYTALE Group, Computer Engineering and Informatics Department, University of Patras, Hellas
e-mail: e.konstanto@upnet.gr; nsklavos@upatras.gr

I. Ognjanovic
University of Donja Gorica, Donja Gorica, Podgorica, Montenegro

 
N. Gupta, S. Mishra (eds.), *Internet of Everything for Smart City and Smart Healthcare Applications*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-34601-9_4

ease of deployment, they also lead to a new set of vulnerabilities, in addition to the threats and challenges already inherited from 4G networks [3].

5G will facilitate the operation of verticals such as smart cities, healthcare IoT, smart factories, and smart grids, by connecting millions of smart computing devices. It is important to note that smart application services come with different threats for each IoT domain. Even less homogeneous devices will be configured to send data over cellular networks to cloud applications and backends. This cross-industry transformation introduces new types of attack vectors and vulnerabilities, requiring ad hoc security regulations [4]. It is important for industries to be prepared for the new privacy and security regulations that need to be implemented in cyberspace, as well as new attacks that compromise cyber-physical systems. Moreover, vertical operators with no security expertise in network security must obtain security services such as authentication, penetration testing, and intrusion detection from their service providers, following a security-as-a-service (SECaaS) approach. Critical infrastructures in this new, smart, and connected world become a very desirable target, by well-rounded individuals when it comes to technology, motivated by political and financial profits.

The need for elevated security requirements for 5G becomes evident when dealing with a constantly changing security threat landscape. A second visit to the confidentiality-integrity-availability (CIA) triad is necessary, in order to cater to the security needs of 5G key technologies (Fig. 1).

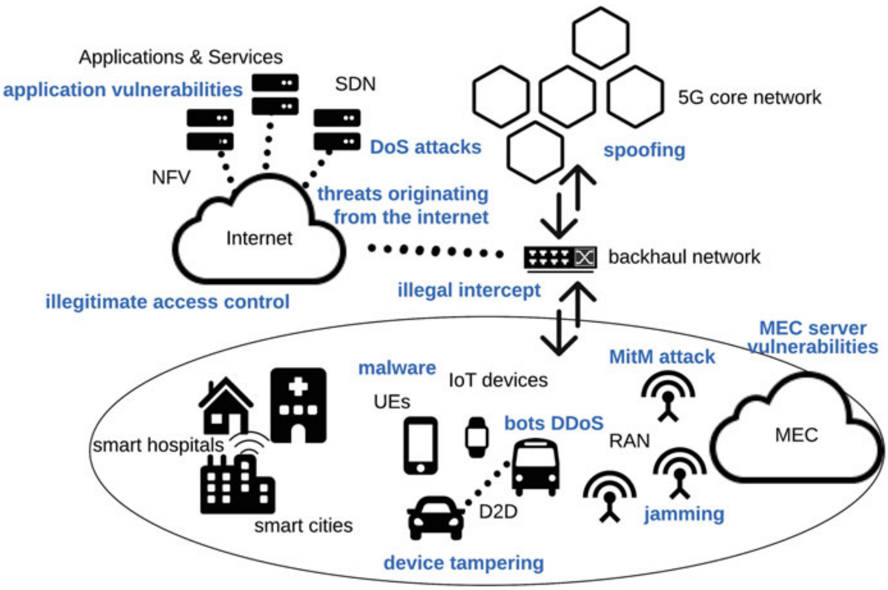



**Fig. 1** The 5G architecture threat surface

- Confidentiality: This term refers to two related ideas: firstly, data confidentiality, which protects against the disclosure of private or confidential information to unauthorized parties [9], and secondly, privacy, which guarantees that individuals are aware of the distribution, collection, and storage of their personal information. To ensure confidentiality in 5G networks, authentication, access control, and encryption are implemented.
- Integrity: This term refers to two related ideas: firstly, data integrity, which guarantees that information is only altered with, and secondly, system integrity [9]. This ensures the non-tampered execution of a system's function. In 5G New Radio, control and data channel integrity is maintained by mutual authentication, provided by protocols such as 5G-AKA [2]. Thus, unlawful tampering of the system is prevented.
- Availability: This term means guaranteeing that information is used and accessed in a timely and trustworthy manner. The eMBB can achieve at least 95% availability and reliability for all 5G applications, while mMTC can support use cases with 99.999% availability and reliability [2].
- Authenticity: This term refers to the quality of being true, verifiable, and trustworthy and having assurance in the veracity of a transmission, the integrity of a message, and the trustworthiness of a message's source [9].
- Accountability: This term refers to the facilitation of after-action recovery, legal recourse, fault isolation, non-repudiation, and intrusion detection, through the security objective that an entity's actions can be linked back to the responsible party only. To enable future forensic analysis in order to track security breaches, systems must preserve records of the system's operations.
- Centralized Policy Management: 5G is the first mobile network that uses centralized policy management, making applications and resources of 5G use cases easily accessible [2].
- Visibility: The softwarized network architecture provides high visibility across the mobile backhaul. It aids with network capacity expansion, anomaly detection, and provision of higher quality of service (QoS) and availability. These advantages should be preserved in security mechanisms since increased transparency and control of mobile networks allow the implementation of end-to-end (E2E) security methods and prevent data-driven threats [2].

## 2 Securing 5G Military Communications

5G, with its evolved radio interface, is expected to expand the technological abilities of many actors, including the military. Mission-critical communications and tactical networks, which involve a significant number of heterogeneous and resource-constrained devices, place a special emphasis on next-generation networks. Through a variety of novel radio access technologies, they deliver the required computation power with cloud computing and dependable, robust communication through new RATs. In a variety of contexts, 5G technologies may offer chances for useful military applications. What is more, a smart city domain can support decent situational awareness in disaster scenarios, by leveraging IoT capabilities. The sensor network especially can be significantly helpful for gathering intelligence. At the same time, a number of security and privacy challenges are introduced through these information sources.

## *2.1 5G and IoT Technologies for Military Applications*

Firstly, the aforementioned NFV and network slicing technologies make it possible for the military to employ a "slice" of the public 5G infrastructure as a dedicated private infrastructure or private multipurpose 5G systems, with both eMBB and uRLLC functionalities available [5]. They can enhance combat efficiency by offering distinctive services upon request for the battlefield. Military communication systems can become more capable by utilizing essential 5G technologies including multi-means collaborative optimization, low-latency and high-reliability networks, spectrum sharing, and millimeter-wave (mmWave) application [11]. The development of exclusive, low-cost military 5G systems is made possible by the use of 5G NR in shared, licensed, unlicensed, and new frequency bands [8]. Low-frequency bands can support tactical scenarios of high mobility, while mmWave frequencies can offer close proximity applications with demands of high bandwidth, as well as enable rather covert radio systems.

Moreover, 5G implementation in various sub-bands allows for military communications in reserved frequency bands. The introduction of very large multiple input multiple output (MIMO) antenna arrays allows for high-capacity tactical radio systems, through beam forming and beam steering [8]. 5G Proximity Services (ProSe) that facilitate direct connectivity in close proximity and integrated access and backhaul (IAB) that allows node communication over the air both enable the implementation of meshed networks with difficult propagation conditions and standalone 5G clusters. Lastly, 5G non-terrestrial networks (NTNs) can possibly extend military applications through airborne or satellite systems.

A smart city's IoT capabilities can also be proven useful in case a disaster hits. IoT devices, used by smart city applications, are interoperable with each other, as they are developed on open data technologies and standards [11]. This can simplify information exchange between different data sources, which in turn assists sensor positioning in smart city environments. The military can utilize data from the sensors, effectors, and actuators of the public domain for recovery operations. This includes motion and acoustic detection, unarmed air vehicles (UAVs), soldier positioning, and GPS. Private entities may also allow tactical centers to access their data as well as provide open APIs. Thus, they offer additional reliability, precision, and access to data that may still be unattainable to the military's resources.

### 2.2 Security Solutions

Future vital technologies that deal with public safety, such as emergency systems, depend heavily on data management and processing in the network. Dynamic authorization, risk analysis, and network monitoring are required for network security in untrusted 5G/6G networks. Zero trust architecture (ZTA), which provides sensitive data protection and security, is an emerging and crucial framework for information security in environments like these [6]. A ZTA provides network assurance in response to a subject's request for access to network resources under the presumption that they are unreliable even after initial authentication and authorization, due to potential internal human error or social engineering attacks.

There are a few main principles when it comes to zero trust. First off, regardless of their position in the network, all network components are regarded as being untrusted. As a result, all communications must adhere to the same security standards. For each access request, dynamic and continued trust assessment and risk analysis are performed. Furthermore, in case access is provided, it should be with the most limited rights possible. The access is only given for that particular resource and no other, and the decision to give access must be made based on a dynamic policy. All network resources and requesting subjects have their security supervised ideally constantly and in real time. In terms of their adherence to security policy requirements, an automated system evaluates the overall security of devices, users, and network assets.

When it comes to protecting smart city information for military usage, E2E encryption between the user and the sensor is necessary. When that is not possible, a secure channel must be established, where the data is collected and encrypted again. Information can also be stored securely within clouds by using encryption. Authentication, integrity, and confidentiality for secure communication between devices can be ensured by remote attestation and usage of keys or other device materials for cryptography [5]. Lastly, directional modulation (DM), a cryptographic approach for security at the PHY layer, can be used by IoT devices in order to achieve physical layer security [11].

## 3 Secure Public Safety Networks

Public safety authorities, including the police, paramedics, and fire and rescue personnel, will use the 5G network for their communications in smart cities. To cover the demands for high bandwidth and low latency, a hybrid network architecture will be deployed, consisting of both tactical network bubbles and commercial network infrastructures [5]. Commercial network infrastructure offers capacity for public safety users. Though such a network is shared with civilians, public safety users hold priority over it, operating as mobile virtual network administrators and overseeing the cloud's core network and application services. When required, or in network failure scenarios where the availability of the commercial network is compromised, tactical bubbles offer additional communication coverage. They provide 4G and 5G connectivity, containing an access network as well as core network functions. In addition, they may also strengthen security when they are safely segregated from the business infrastructure [10]. However, the implementation of a hybrid infrastructure overall makes the assurance of confidentiality, integrity, and availability of mission-critical applications and sensitive information challenging as the threat landscape widens.

## 3.1 Use Cases

Firstly, group communication within the tactical bubble is achieved with mission-critical push-to-talk [5]. Streaming is another fundamental service that is possible within the tactical bubble, such as video surveillance and audio in critical scenarios [12, 13]. A local mission site can provide UAVs' navigational data from the aerial space for their remote control within the tactical network. UAVs can offer decent connectivity for network recovery in areas under a crisis. Application servers and national databases are run remotely in the cloud in the case of mission-critical services (MCX). As remote users depend on such services, availability must be guaranteed. Special attention should be paid to sensitive information, which must remain local to not get accessed by eavesdroppers. That could include the location of users, cameras, and UAVs or operational information [5]. Mobile devices are becoming increasingly prevalent, making them a crucial element of a disaster management system (DMS). Values of a DMS are durability, trustworthiness, and energy-efficiency, and it must offer higher-quality communications [14]. In the scenario of an inaccessible commercial or dedicated network when executing a rescue operation in a smart city, public safety users can access a civilian private network. Moreover, public safety users can utilize the shared commercial network to access services in their organization's cloud. More specifically, in cases of mission-critical communications with stringent QoS constraints, a cloud-enabled small cell infrastructure has also been proposed. Its distributed orchestration architecture utilizes multi-access 5G technologies to allow first responders to share public 5G infrastructure.

## 3.2 Attack Categorization

A wide range of weaknesses, beginning with the continually changing 3rd Generation Partnership Project (3GPP) security standards' specification, might result in security breaches [8]. Verification, testing, and certification can help to reduce implementation errors—intentional ones caused by hardware backdoors as well as accidental ones [5]. Architectural failures, such as vulnerabilities resulting from interconnected systems, are caused by flaws in the planning and design of public safety communications. A process vulnerability is to blame when security is breached while a product or service is being developed. However, configuration and operational errors are the most typical kinds of root vulnerabilities. The former are brought on by typical end users, whereas the latter are administrative errors, whether deliberate or not. In situations like these, rigorous security procedures and standards are beneficial.

The attackers can also be divided into groups based on their motives and the resources at their disposal [5]. To begin with, some benign actors interfere with public safety operations without intending to do so. When it comes to hostile attackers, some are more interested in showcasing their abilities and aren't purposefully targeting public safety networks, while others intentionally compromise security from the inside. Cybercriminals that aim to thwart specific operations or do so for monetary benefit make up the fourth class. Another category consists of advanced persistent threat (APT) groups with plenty of resources, which specifically target public safety operators. Finally, there are also foreign organizations that engage in hybrid warfare and breach communication networks.

There are a few basic ways in which public safety communication networks might be attacked. Within the range of a radio access network, an adversary may launch open-air interface or local radio attacks and can be detectable while they're broadcasting [5]. It can be really difficult to protect a tactical bubble, either by a physical attacker who can access user equipment (UE) or by a remote attacker who advances with the aid of malware [6]. Insiders are hostile parties, whether intentional or not, who have gained access to the infrastructure of the operators. At the same time, non-operational attacks can occur at any time during the lifecycle of a product.

Not all attacks lead to consequences of the same severity, however. Depending on the size of the impact, each security breach can fall into a different type of risk level [5]. Small-scale incidents restrict a single individual without totally preventing access to the service. Medium-scale incidents affect a network segment or public safety service operator resources without entirely obstructing society's activities. In large-scale incidents, civilians' lives could be compromised. There is a variety of threat scenarios that stem from targeting different assets in public safety networks, including resources, services, and users.

### *3.3 Threat Scenarios*

Starting off with attacks on access networks, attackers in the form of user equipment and devices threaten through physical or radio interfaces. It's possible for eavesdroppers to learn how public safety actors' equipment operates. Adversaries may employ traffic analysis to locate UE [5]. Signals can be recognized using a variety of radio transmitters that have specific manufactural flaws. Moreover, the few public safety-specific procedures may be used by denial-of-service (DoS) attackers [13]. Prioritized malicious UEs have the potential to do more damage. The centralized database may not have information regarding the available spectrum as a result of a network attack, which would allow a commercial mobile network to continue using the frequencies needed by the tactical bubble. This might cause interferences that would make communication less efficient.

Attacks in the network domain can compromise software and data, hardware, as well as transit communications. Functions relating to user profiles and authentication data need to be highly protected, as exposure of such information might have unfavorable consequences. Threats that target availability rise in importance as the number of devices used in public safety operations increases. In that case, the scenario of exhausting authentication functions when having multiple simultaneous authentication efforts occurs [5]. In another scenario, a group can misbehave and get access to crucial assets hosted on shared hardware that belongs to many governmental or civil entities, by exploiting the virtualization layer. What is more a vulnerability of the communication infrastructure may potentially be exposed through covert attacks on the basic systems it depends upon, such as the electric supply. A threat actor may discover exploitable holes in the backhaul security, thus challenging availability in transit networks.

Group communication services, application servers in the cloud, operational data, as well as highly personal information are targeted with threats in the application domain [15]. A lot of times, 5G's key technologies are used by applications with needs of better performance and low-latency communications and operate at the edge [5]. This leads to a security vulnerability, where a trusted middle party is needed at the edge. Lastly, ineffective bandwidth sharing, denial of service or denial of information for some parties, or the leakage of classified information to a device without the proper authorization can all result from improper configuration of authorization policies.

Depending on the hardware and operating system, imposed regulations, the quality of the software, and other factors, devices have varying degrees of security. Access control, firewalls, and cryptography are some examples of security measures for devices [5, 11]. Essential services may be accessed by unauthorized parties, network services may not be available, and a device may become unusable as a result of such a takeover. It is important that less rights should be granted to unapproved devices to confine the impact of a breach. Following that is a description of threat scenarios for each use case as well as a survey of security solutions that are proposed for 5G networks and public safety communications.

### *3.4 Security Solutions*

When it comes to network security, the 3GPP-specified authentication and key agreement (AKA) protocol variations are the foundation for access management. The 3GPP-specified authentication and key agreement (AKA) protocol variations (EAP-AKA, EAP-AKA', or 5G-AKA) are the foundation for access management [2]. A USIM application, which has symmetric keys, stores user credentials. New authentication methods, such as certificate-based authentication and secondary authentication, have also been established. Following authentication, subsequent communication is secured using SNOW, AES-CTR, or ZUC cipher-based 3GPP confidentiality and integrity algorithms [5]. The 3GPP has a designated isolated operation for public safety (IOPS) mode of operation for tactical bubbles without backhaul so that, in essence, devices are supposed to store a USIM application for regular use and one for use in a closed network.

Concealing the UE identifiers that are transmitted through air is one defense against UE monitoring. 5G uses asymmetric cryptography to secure unique IDs that are utilized whenever a device enters the network [5]. Secondly, following registration, UEs are given temporary identification pseudonyms that are regularly changed. For transit communications that are exposed to remote attacks, firewalls are a necessary security measure. To ensure infrastructure security, continuous monitoring is needed to guarantee that devices are running as expected in the physical layer, the system-level layer, and the virtualization layer. Authorization based on access control policies is necessary for public safety users to be granted priority over the limited available bandwidth [6]. Statistics should be collected to monitor the security state and detect recurring attacks and abnormalities.

Identity and access control, as well as confidentiality and integrity for group communication, all contribute to end-to-end application security [11]. Additionally, certain IoT devices might have unique methods for protecting both control and data transfers, like with UAVs. Applications for mission-critical group communication make sure that only authorized users can access communication groups. By reducing the amount of data held within databases in tactical bubbles, the threat of data breaches and the leakage of authentication and authorization information can be reduced [5]. Private keys and other forms of authentication data for UEs must exist within the bubbles and be properly protected. Increased security measures inside the tactical bubble are another way to reduce the risk of a data breach.

UE security can be strengthened by discouraging the use of unreliable features and configurations in devices. Thus, security breaches and network risks can be reduced to a minimum. Secure booting is another method for reliable identification and authentication [5]. Each tier of the system verifies the integrity of the subsequent layer prior to execution to detect software tampering. The virtualized credential platform eSIM is a flexible solution to deliver credentials that grant public safety users UE access to network or application services. Physical credentials such as USBs or memory cards also allow for trustworthy authentication. Security solutions for public safety networks are summarized in Fig. 2.

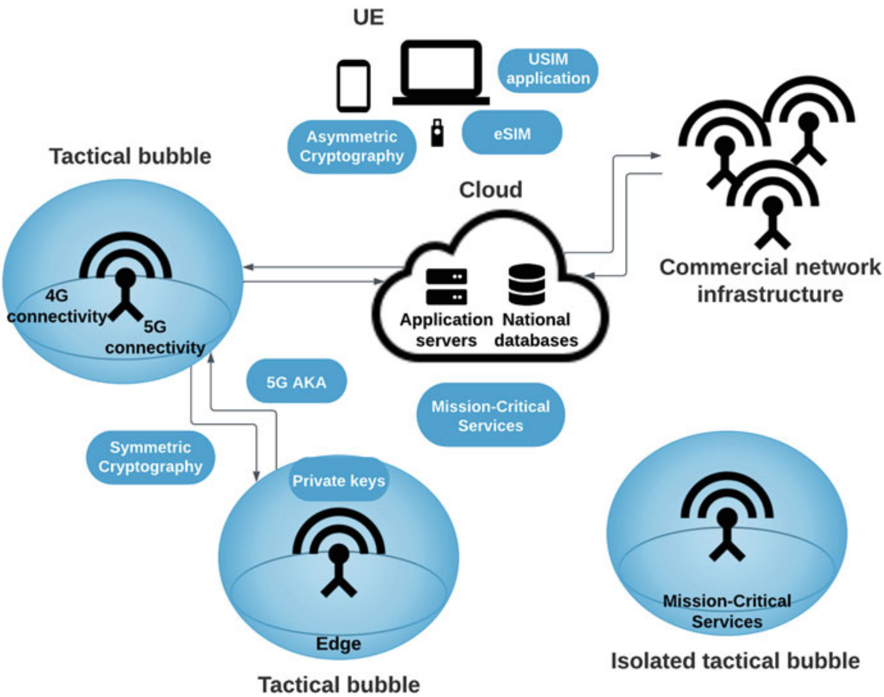



**Fig. 2** Security solutions in the hybrid architecture for public safety communications

## 4 IoT Healthcare: Privacy and Security Concerns and Solutions

State-of-the-art IoT technology can allow for healthcare to advance significantly. For example, medical devices may be relocated using passive RFID tags, thanks to radio-frequency identification (RFID) technology [16]. The real-time location system (RTLS) permits the real-time monitoring of tagged items and aids in the development of a network of interconnected sensors that continuously monitor and report any changes in their location, circumstances, and quantity. Big data technology is used to handle the huge volumes of medical data that are stored in the cloud [7]. Edge computing can facilitate rural healthcare and increase patient satisfaction by placing computational or storage resources within the RAN [16]. Moreover, edge computing devices are connected to medical databases and cloud computing servers through the Internet to store medical data, whereas fog computing provides supplementary resources and helps meet speed requirements. These technologies carry their own vulnerabilities, posing several security issues.

At the same time, IoT healthcare apps facilitate the improved collection and analysis of biometric data [16]. They have the ability to precisely track individuals and supplies, as well as also handle a variety of stakeholders, such as hospitals, and evaluate the collected sensor data. That being said, IoT-based healthcare apps must include security as a key component. Since the acquired data are sensitive and the system's availability is essential for the patient's well-being, offering privacy and guaranteeing the confidentiality, integrity, and authenticity of IoT-based systems are key in the health sector, yet a challenging task.

### *4.1 Security Requirements*

Scalability is one of the main challenges in IoT healthcare app development. As an increasing number of heterogeneous devices enter the IoT network, a security mechanism fitted for a variety of network configurations and protocols is needed. These interlinked devices are capable of exchanging data over networks such as Bluetooth, ZigBee, RFID, and 6LoWPAN. WiFi, 5G, and edge and fog computing also allow them to interact with their environment [16]. The main security-preserving concepts of confidentiality, integrity, availability, authentication, authorization, and non-repudiation are relevant in healthcare applications as well. What is more, a proposed security system for IoT healthcare must be robust and resilient to security breaches and have self-correcting capabilities when those cannot be prevented. At the same time, the IoT devices that will be used to promote healthcare have limited resources, making the design and implementation of security robust systems for the framework a challenging task [18]. Lightweight solutions are thus needed, which are area and power efficient, while offering the highest level of achievable security.

### *4.2 Security and Privacy Concerns*

The IoT network has several weaknesses that may be quickly and readily exploited by hostile attackers. Information disruptions can target a system's availability through DoS attacks, thus jeopardizing patients who rely on real-time health monitoring apps to provide them first aid in emergency situations [16]. Disruptions may also threaten the confidentiality of highly sensitive patient information or even tamper it, as well as target the authenticity of messages and confuse unsuspecting entities through spoofing. A false diagnosis might be carried through as a result of modification and fabrication of patient information, raising the risk of potentially fatal consequences for the patient. Replay attacks prevent the data from staying current, allowing for dated information to be exchanged instead. Unauthorized parties might perhaps access the transmitted data and utilize it for their own gains, violating the demanded data privacy policy. Moreover, eavesdropping and traffic-monitoring assaults are some of the most frequent attacks that could compromise IoT healthcare systems.

### *4.3 Security Solutions*

When considering security solutions for IoT health applications, we need to take into consideration the limitations that bound them. These cryptographic primitives only need minimal computations; therefore, hardware designs usually choose between either low area or high throughput. When lightweight algorithms are implemented on hardware, it is important that the security level is not decreased too much along with computational complexity. Lightweight solutions are thus needed, which are area and power efficient, while offering the highest level of achievable security [18]. Cryptographic algorithms can be chosen for each application with regard to its performance, the available resources, and the security requirements [19]. Efficient implementations of encryption/decryption algorithms result in the optimization of IoT healthcare devices' design.

Proposed security models for IoT healthcare systems focus on preventing security breaches in devices and data theft, ensuring both the confidentiality and integrity of the exchanged data from the sensors to the rest of the IoT network. To prevent unauthorized access and allow E2E security, devices and sensors will need to have limited interactions with external interfaces [16]. That can be possible with a simple operating system with defined network operations.

Blockchain can be a very useful technology when managing medical records in the network. The acquired sensor data are transmitted by wireless LAN or a comparable technology to the IoT gateway device [16]. After receiving it, the gateway device sends the data to the blockchain network. Storage of records happens within the nodes in the blockchain network, followed by a signature of a private key. After that, to maintain data integrity, modification or deletion is prohibited. Validation is executed by peers in the network. Furthermore, data privacy is enabled through access control [17]. Access to records by a node is only possible after its successful registration to the network and its authentication, while patients can use special IDs to view their medical records. Distributed healthcare applications and devices will only exchange messages across their network through defined and secured communication protocols. Figure 3 illustrates the framework of an IoT-based healthcare system and its security elements.

## 5 Summary

In conclusion, security-critical networks will play a vital role in the quality of life of users in smart cities. Firstly, the privacy and security issues that stem from 5G's novel technologies and affect critical networks must be taken into account in order to define a new set of key security objectives. Public safety and national security actors such as the military, police officers, and rescue personnel will benefit from 5G's key technologies and the IoT's vast interconnection. Mission-critical applications and sensitive data will be exposed to physical and virtual attackers as a result of

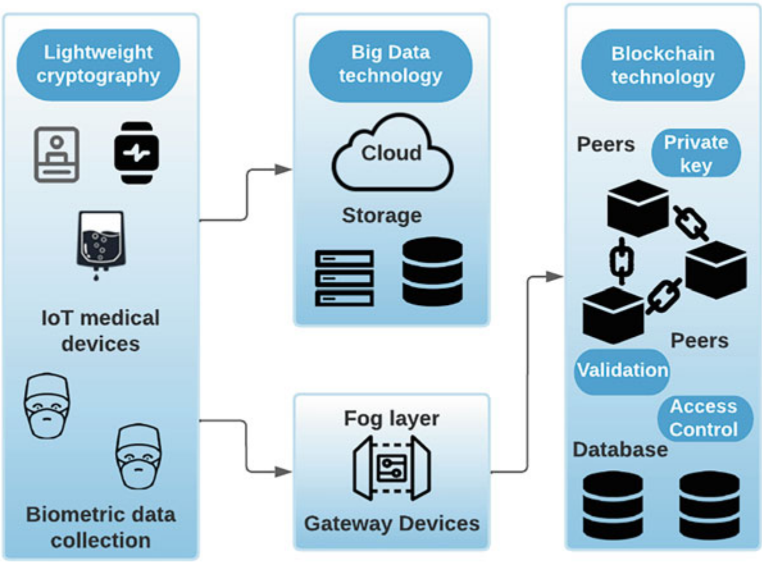



**Fig. 3** The framework of an IoT-based healthcare system

the switch from dedicated infrastructure to hybrid architecture, thus making the need to explore solutions toward safe communication imperative. Another aspect of smart cities that will directly affect the citizens' well-being is the maintenance of security in IoT healthcare applications. Key privacy properties such as anonymity, unlinkability, and undetectability must be preserved with the exchange of messages and private user data across the network.

## References


1. Khan, R., Kumar, P., Jayakody, D. N. K., & Liyanage, M. (2020). A survey on security and privacy of 5G technologies: Potential solutions, recent advancements, and future directions. *IEEE Communications Surveys & Tutorials, 22*(1), 196–248.
2. Sicari, S., Rizzardi, A., & Coen-Porisini, A. (2020). 5G In the internet of things era: An overview on security and privacy challenges. *Computer Networks, 179*, 107345.
3. Ericsson, A. B. (2021). A guide to 5G network security 2.0 (287-01 FBG 101 0955 Rev).
4. Stallings, W. (2020). *Cryptography and network security: Principles and practice* (8th ed.). Pearson Higher Education.
5. Larsen, L. M. P., Berger, M. S., & Christiansen, H. L. (2018). Fronthaul for cloud-RAN enabling network slicing in 5G mobile networks. *Wireless Communications and Mobile Computing, 2018*, 4860212.
6. Bastos, L., Capela, G., Koprulu, A., & Elzinga, G. (2021). Potential of 5G technologies for military application. In *2021 international conference on military communication and information systems (ICMCIS)* (pp. 1–8).

7. Liao, J., & Ou, X. (2020). 5G military application scenarios and private network architectures. In *2020 IEEE international conference on advances in electrical engineering and computer applications (AEECA)* (pp. 726–732). https://doi.org/10.1109/AEECA49918.2020.9213507
8. Johnsen, F. T., et al. (2018). Application of IoT in military operations in a smart city. In *2018 international conference on military communications and information systems (ICMCIS)* (pp. 1–8). https://doi.org/10.1109/ICMCIS.2018.8398690
9. Ramezanpoura, K., & Jagannatha, J. (2022). Intelligent zero trust architecture for 5G/6G networks: Principles, challenges, and the role of machine learning in the context of O-RAN. *Computer Networks, 217*, 109358. https://doi.org/10.1016/j.comnet.2022.109358
10. Suomalainen, J., Julku, J., Vehkaperä, M., & Posti, H. (2021). Securing public safety communications on commercial and tactical 5G networks: A survey and future research directions. *IEEE Open Journal of the Communications Society, 2*, 1590–1615. https://doi.org/10.1109/OJCOMS.2021.3093529
11. Ali, K., et al. (2021). Review and implementation of resilient public safety networks: 5G, IoT, and emerging technologies. *IEEE Network, 35*(2), 18–25. https://doi.org/10.1109/MNET.011.2000418
12. Chochliouros, I. P., Spiliopoulou, A. S., Lazaridis, P. I., et al. (2021). 5G for the support of public safety services. *Wireless Personal Communications, 120*, 2321–2348. https://doi.org/10.1007/s11277-021-08473-5
13. Naqvi, S. A. R., Hassan, S. A., Pervaiz, H., & Ni, Q. (2018). Drone-aided communication as a key enabler for 5G and resilient public safety networks. *IEEE Communications Magazine, 56*(1), 36–42. https://doi.org/10.1109/MCOM.2017.1700451
14. Volk, M., & Sterle, J. (2021). 5G experimentation for public safety: Technologies, facilities and use cases. *IEEE Access, 9*, 41184–41217. https://doi.org/10.1109/ACCESS.2021.3064405
15. Selvaraj, S., & Sundaravaradhan, S. (2020). Challenges and opportunities in IoT healthcare systems: a systematic review. *SN Applied Sciences, 2*(139). https://doi.org/10.1007/s42452-019-1925-y
16. Konstantopoulou, E., & Sklavos, N. (2022). Design and implementation of a lightweight cryptographic module, for wireless 5G communications and beyond. In *2022 IEEE international conference on internet of things and intelligence systems (IoTaIS)* (pp. 166–171). https://doi.org/10.1109/IoTaIS56727.2022.9975849
17. Alharam, A. K., & Elmedany, W. (2017). Complexity of cyber security architecture for IoT healthcare industry: A comparative study. In *2017 5th international conference on future internet of things and cloud workshops (FiCloudW)* (pp. 246–250). https://doi.org/10.1109/FiCloudW.2017.100
18. Bhuiyan, M. N., Rahman, M. M., Billah, M. M., & Saha, D. (2021). Internet of Things (IoT): A review of its enabling technologies in healthcare applications, standards protocols, security, and market opportunities. *IEEE Internet of Things Journal, 8*(13), 10474–10498. https://doi.org/10.1109/JIOT.2021.3062630
19. Verma, P., & Sood, S. K. (2018). Cloud-centric IoT based disease diagnosis healthcare framework. *Journal of Parallel and Distributed Computing, 116*, 27–38.

# Applications of Machine Learning and 5G New Radio Vehicle-to-Everything Communication in Smart Cities

**Raumit Raj, Amit Kumar, Abhilash Mandloi, and Raghavendra Pal**

## 1 Introduction to Machine Learning

The study of artificial intelligence, which is the art of building smart machines that can mimic human cognition and behavior, is the field that gave rise to machine learning. The 1950s and 1960s saw researchers researching the use of computers to mimic human learning and decision-making, which is when machine learning first emerged. Many of the techniques were built on the foundation of this earlier study.

Alan Turing, often regarded as the pioneer of modern computers, is one of the main figures in the development of machine learning. He created the idea of the Turing machine, a dictatorial machine that could complete any computational work when given a set of instructions. This idea aided in the creation of the modern computer and served as the basis for theoretical computer science.

Arthur Samuel, who first used the phrase "machine learning" in 1959, is another significant person in the history of the field. Samuel had created the first self-learning computer program, a checkers-playing computer program that got better over time by learning from its errors. This early program was a crucial milestone in the direction of the creation of contemporary machine learning algorithms. Machine learning continued to progress and advance in the decades that followed, with notable advancements in fields like neural networks in deep learning [1].

R. Raj
Department of Electronics Technology, Guru Nanak Dev University, Amritsar, India
e-mail: 2021010015313@gndu.ac.in

A. Kumar · A. Mandloi · R. Pal (✉)
Department of Electronics Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, India
e-mail: d21ec012@eced.svnit.ac.in; asm@eced.svnit.ac.in; raghavendrapal@eced.svnit.ac.in



N. Gupta, S. Mishra (eds.), *Internet of Everything for Smart City and Smart Healthcare Applications*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-34601-9_5

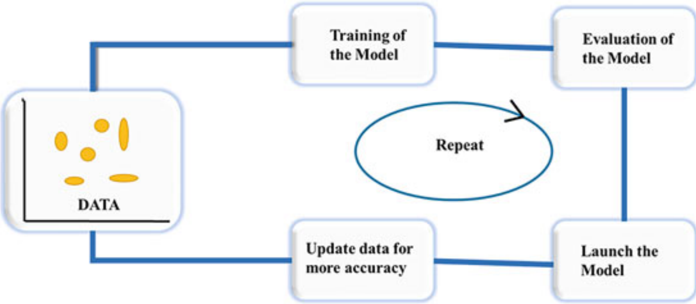



**Fig. 1** General approach for machine learning algorithm

Machine learning is a fantastic solution for issues that require a lengthy list of rules or a lot of fine-tuning. Additionally, it is a good strategy for complex problems for which a standard approach does not produce better results, for fluctuating conditions, for extremely huge volumes of data, and for many other situations. A general machine learning approach looks like as what is shown in Fig. 1.

## 2 Applications of Machine Learning

Let's learn about more machine learning applications so that we can appreciate how diverse the field is. We'll talk about a few more examples of machine learning applications in different fields [2]. Figure 2 shows the various applications of machine learning in real life.

(a) *Space Science and Technology:* Training a model to forecast the trajectory of a celestial body, such as an asteroid, based on historical data and physical models of the body's motion could be the goal of a machine learning project in the field of space science and technology.
(b) *Defense Sector:* Using historical data and knowledge of prospective dangers, a machine learning project in the defense sector may involve training a model to predict the possibility of a security threat in each location.
(c) *Agriculture:* A machine learning project in the field of agriculture would entail developing a model to forecast crop production in each location based on past data and knowledge of weather patterns, soil conditions, and other elements involved in crop cultivation.
(d) *Healthcare:* Training a model to forecast a patient's chance of contracting a specific disease using the patient's medical records and other data could be the focus of a machine learning project in the field of healthcare.
(e) *Mobile Applications:* Training a model to enhance the functionality of a personal assistant application, such as by improving its accuracy at tracking voice,

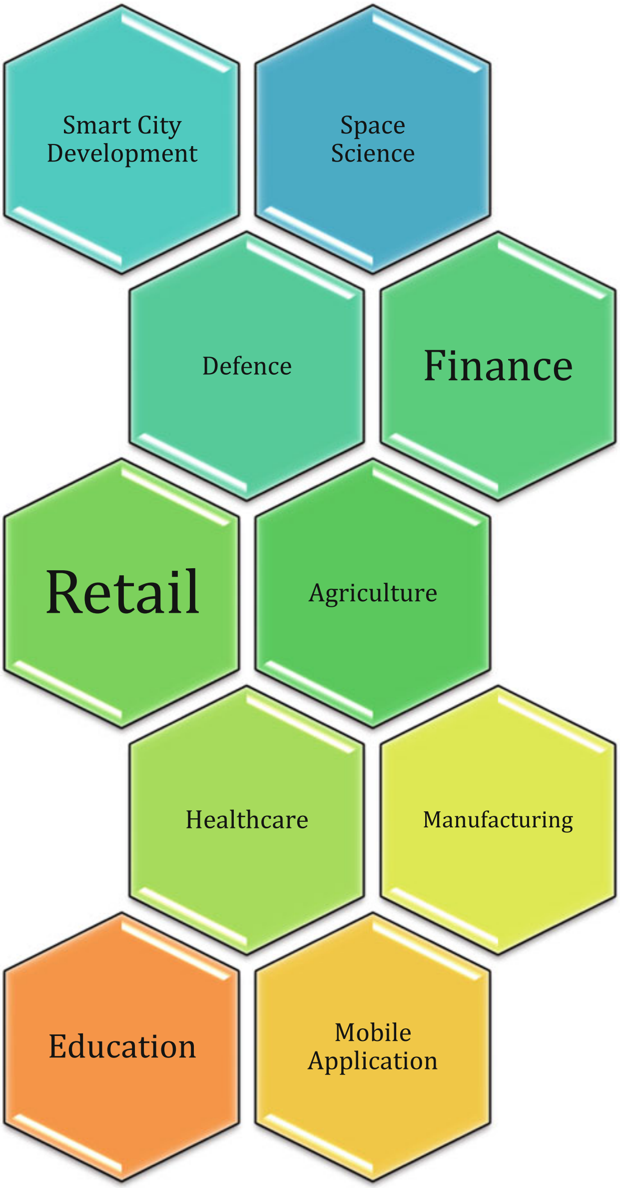



**Fig. 2** Applications of machine learning in real life

organizing speech, or comprehending natural language, could be the focus of a machine learning project in a mobile application.

(f) *Smart City Development:* In the field of smart city development, machine learning algorithms can be utilized to create intelligent transportation systems, maximize energy efficiency, and enhance public safety and security.
(g) *Finance:* The development of fraud detection systems, customer behavior analysis, and stock market trend prediction are all possible using machine learning algorithms.
(h) *Retail:* Personalized product recommendations, improved supply chain management, and better customer service can all be achieved with machine learning algorithms.
(i) *Manufacturing:* Predictive maintenance systems, production process optimization, and quality control can all be enhanced using machine learning algorithms.
(j) *Education:* Machine learning algorithms can be applied to student performance data analysis, adaptive educational technology development, and personalized learning.

## 3 Deep Learning

Deep learning is a part of machine learning, which is a branch of AI that uses algorithms to learn from data. Deep learning techniques use a hierarchy of numerous layers, each of which learns to extract and change the data from the preceding layer. This is in contrast to typical machine learning algorithms, which are designed to work on preset input and output data. Deep learning algorithms may develop a more abstract and complicated representation of the data, thanks to this hierarchical structure, which is useful for a variety of applications, incorporating speech and picture recognition, natural language processing, and perhaps autonomous vehicles [3].

Simplifying even further, deep learning takes its inspiration from the structure and operation of the brain, specifically the neural networks that comprise the brain. Large data sets are used to train artificial neural networks, enabling the network to learn and come to wise decisions on its own. Figure 3 shows the various layers in deep learning.

The term "neural network" is now introduced [4]. In a nutshell, neural networks are a collection of algorithms for pattern recognition. They take their cues from the way the human brain works, with its network of interconnected nodes that can analyze information and draw conclusions from it. Layers of connected nodes make up a neural network. Each layer processes the data before sending it to the one below. The raw data is delivered to the input layer, the first layer in the stack. The outcome is produced by the output layer, the last layer, using the intermediately processed data. There are layers that are concealed between the input and output layers; these levels process the data and transfer it to the following layer.

The network gets stronger as the number of layers and nodes rises, but it also gets harder to train and optimize. Deep learning algorithms come in a wide variety of forms, but they always have a few things in common. First, they are made to work with a lot of data that is frequently processed in parallel using specialized hardware,

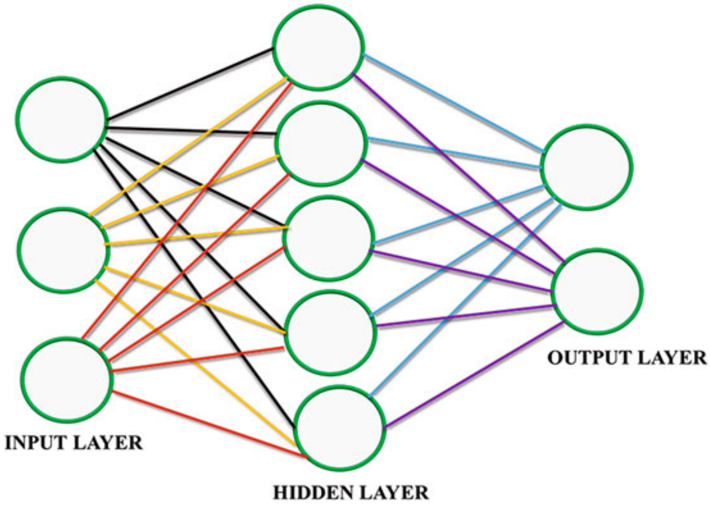



**Fig. 3** Deep learning layers

including graphics processing units (GPUs) or tensor processing units (TPUs). This enables deep learning systems to learn from data sets, especially big data sets, fast and effectively.

## 4 The Smart City Landscape

A smart city is typically a city that adopts technological advances in order to improve public services, such as transportation, energy efficiency, and waste management. By investing in such technologies, cities are able to make better use of resources, cut operating costs, create new jobs, and reduce their carbon footprint. Smart cities use digital technology to collect and analyze data, allowing for more efficient use of resources and better management of urban services like electricity, water, and transportation. Smart cities are becoming increasingly important as cities around the world face issues such as aging infrastructure, climate change, population growth, and rapid urbanization. Smart cities use the power of Internet of Things (IoT) and big data to gather and analyze data to optimize city services. It ties together different systems and data sources to provide better services and make decisions faster. Data gathered from smart systems like public transportation and water management can be used to better monitor usage levels and plan upgrades. Smart sensors also allow cities to monitor air pollution levels and take action to improve air quality.

Smart cities are essential for managing resources and assets, as they allow for the collection of data from citizens, devices, and assets. This data is then used to monitor and manage the various traffic and transportation systems, such as power plants, water supply networks, waste management, law enforcement, parking, healthcare, and smart communication [5, 6].

The creation and management of smart cities are increasingly supported by machine learning and deep learning. Due to the real-time analysis of massive volumes of data made possible by these technologies, city planners and administrators are better able to make timely choices and improve the quality of life for inhabitants.

In order to manage the transportation system or to analyze traffic data and forecast management, machine learning algorithms are widely utilized in smart cities. This enables city planners to optimize traffic flow and lower the number of vehicles on the road. This could aid in lowering emissions, enhancing air quality, and conserving energy.

Video data from security cameras can be examined using deep learning, a branch of machine learning, to look for patterns and anomalies that might point to criminal behavior or possible risks to public safety. This can aid in enhancing local security by enabling law enforcement to react to crises more rapidly.

Machine learning is also being used in smart cities to optimize how energy and other resources are used. Machine learning techniques, for instance, can be used to analyze data from sensors on water and electricity networks to find leaks and optimize resource distribution. The administration of public services is another way that machine learning is used in smart cities to minimize waste and increase utility efficiency. In order to find patterns and trends that may point to the need for additional resources or services in a given area, machine learning algorithms can be used to analyze data from a range of sources, including social media and government databases. This can assist local officials in allocating resources wisely and enhancing the quality of life.

In general, the use of machine learning and deep learning in smart cities is improving the effectiveness and efficiency of a wide range of services, from public safety and resource allocation to transportation and energy management. Cities are becoming more livable, sustainable, and receptive to the needs of their residents by utilizing the potential of these technologies [7].

## 5 The Internet of Everything

The connectivity of objects, people, processes, and data made possible by the Internet is referred to as the "Internet of Everything" (IoE). It is a notion that incorporates the expanding trend of adding Internet connectivity to a variety of gadgets and systems, from laptops and smartphones to appliances, cars, and even industrial equipment.

IoE aims to build a seamless, linked network of gadgets and systems that can communicate and cooperate to boost productivity, efficiency, and the general quality of life for people and businesses.

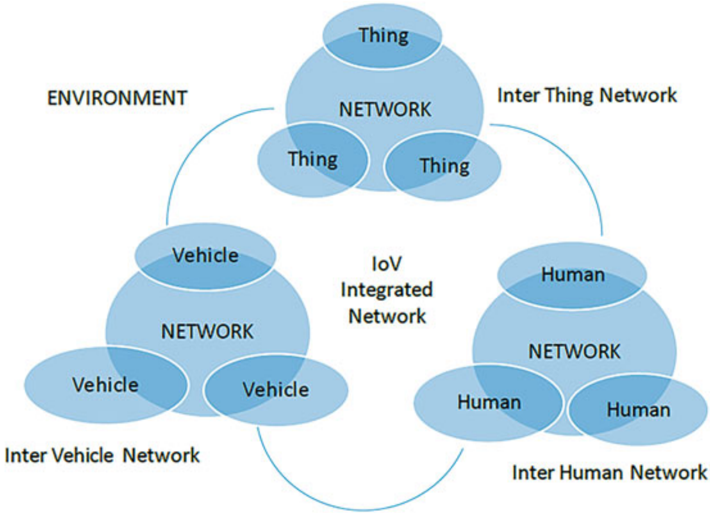



**Fig. 4** IoV network model

To do this, a range of communication technologies, including Wi-Fi, Bluetooth, and cellular networks, are used to allow devices to connect with one another and with central servers.

The ability of the IoE to collect and analyze enormous amounts of data from several sources is one of its main advantages. This information may be utilized to learn more and make better judgments, which will increase efficiency and effectiveness across a number of domains. By tracking the usage and condition of individual components, for instance, the IoE can be used to optimize the performance of industrial equipment. As a result, the equipment's lifespan may be increased, its efficiency may be increased, and downtime may be decreased.

An IoV network is made up of a combination of "human," "vehicle," and "thing" elements, which relate to different network groupings that cooperate with one another. The terms "human" and "vehicle" refer to individuals, while "thing" refers to any element that is neither "human" nor "vehicle." Both can use and/or create apps and services for the environment. Through interactions between the environment and "a vehicle," the environment and "a person," and the environment and "an object," IoV offers these services and applications over the Internet. In order to enable interaction between "vehicle" and "thing," as well as "vehicle" and "person," IoV also offers services and applications inside a "vehicle" by way of an intra-vehicle network [8]. Figure 4 depicts a model of the IoV network.

The IoE is being developed with the use of machine learning, which is a key technology. Machine learning algorithms have the capacity to analyze massive amounts of data in real time, allowing systems and devices to decide what to do with the data and how to proceed. As a result, the IoE is better able to respond to shifting circumstances and demands.

## 6 5G New Radio Vehicle-to-Everything Communication

The usage of wireless technology is now possible in a variety of multidimensional applications, thanks to recent breakthroughs in the field. Safety in cars is one of them. Intelligent transportation systems (ITS) employ the IEEE 802.11p standard for vehicular ad hoc networks, also known as dedicated short-range communications (DSRC), to improve road safety (VANET). Vehicles in the VANET periodically switch over their beacons every 100 ms in order to stay informed of everything going on around them. It is difficult for automobiles to exchange their beacons in this time-limited circumstance with a large number of passengers. As a result, a reliable medium access control (MAC) protocol is required to allow high beaconing suited for vehicles with high mobility. Some works in the VANET are given in [9–14]. Figure 5 depicts a clustering-based vehicular ad hoc network, and Fig. 6 shows the chronology of development of technologies for vehicular communication.

Vehicular communication, also known as vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication, is the use of wireless technology between vehicles and between vehicles and the infrastructure around them. This technology allows vehicles to communicate with each other and with surrounding infrastructure to better manage traffic, provide safety warnings to drivers, and enhance navigation.

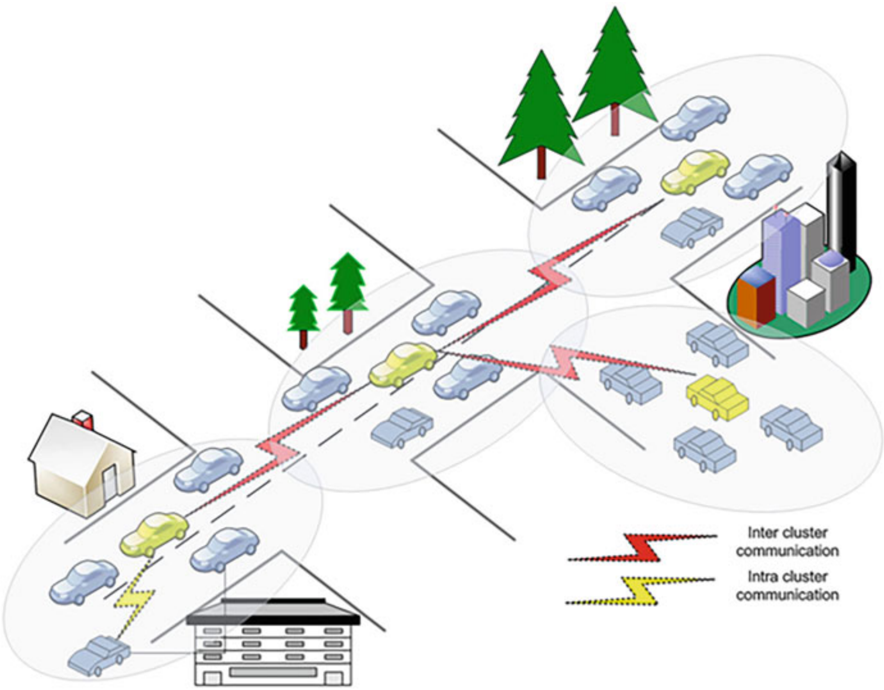



**Fig. 5** Clustering-based vehicular ad hoc network

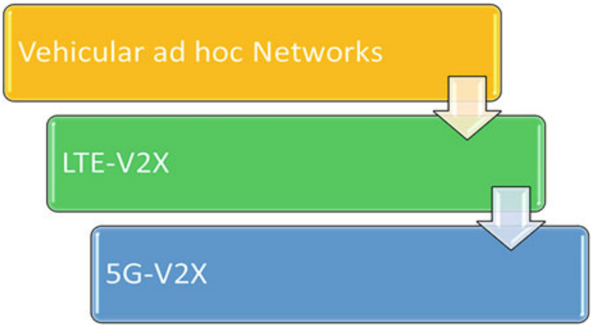



**Fig. 6** Chronology of connectivity technologies for vehicular communications

Communication is facilitated through the use of communication protocols such as dedicated short-range communications (DSRC), Wi-Fi, and 5G. V2V and V2I communication systems can help reduce the number of traffic collisions by providing drivers with real-time safety alerts, such as warnings regarding nearby vehicles braking suddenly or congested roads ahead. Additionally, this technology can help enable autonomous vehicles, as these vehicles can communicate with each other to better understand their environment and cooperate to efficiently navigate through traffic. Such communication can also enable more efficient traffic management by providing insight into the movement of cars and the associated traffic and congestion levels.

Cellular is a communication platform that utilizes C-V2X direct transmission to enhance autonomy, road safety, and Internet of Vehicles (IoV) connections with or without cellular network connections. This chipset solution is compatible with 5G and advanced driver assistance systems (ADAS) sensors and is designed to create V2I, V2V, and V2P connections [15–17]. As shown in Fig. 7, C-V2X is able to provide IoV connections. Despite these functions, user-created virtual cells are currently underutilized, and there is no efficient way to increase the services provided by 5G networks, including IoV services via Fifth-Generation New Radio Vehicle-to-Everything (5G NR V2X) communications.

It is anticipated that the 5G network will have lightning-fast speeds and high levels of reliability. If users of wireless and mobile networks use a single Internet Protocol (IP) or identity, they will be able to access the services and apps that allow interoperability. The system consists of a user terminal, many self-sufficient Radio Access Technology (RATs), and an Internet connection. The C-V2X mobile terminal, however, will need a unique radio interface for each RAT, as well as Internet connectivity and guaranteed quality of service as part of its supporting mechanisms (QoS). The control data in IP versions 4 and 6 (IPv4 and IPv6) is adequate to ensure that packets from different connections are routed correctly and following user requirements.

The most crucial point to bear in mind is that end-to-end (E2E) latency for safety applications is predicted to be under 100 milliseconds (ms) for C-V2X [18]. The fundamental time-frequency resource structures of C-V2X and LTE are the same. The frequency granularity that is the smallest is [19], and it consists of 12

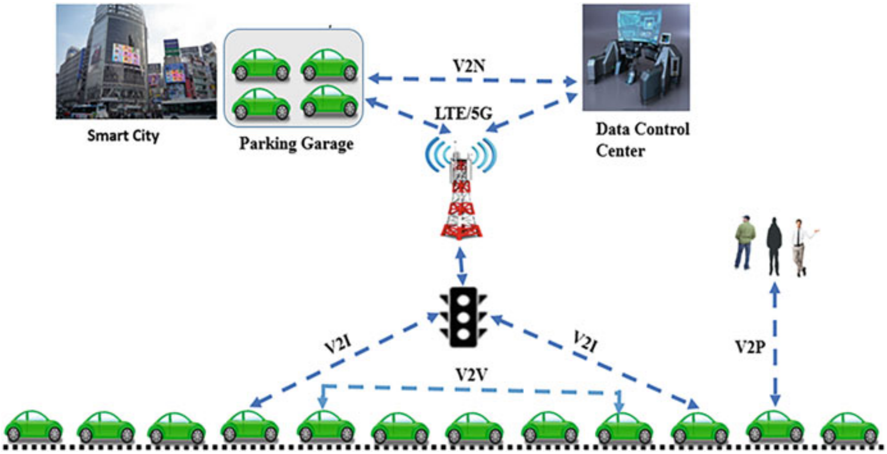



**Fig. 7** C-V2X communications

subcarriers, each of which has a frequency of 15 kilobits per second (kHz). The time allotted for C-V2X transmission is broken down into units known as subframes. Each subframe consists of 14 orthogonal frequency-division multiplexing (OFDM) subcarriers and lasts 1 millisecond. The devices employ either 16-quadrature amplitude modulation (QAM) or quadrature phase shift keying (QPSK) techniques to transmit data symbols along with reference signals and control information. In order to ensure optimum signal quality, every OFDM subcarrier is turbo coded.

## *6.1 Vehicle-to-Infrastructure (V2I)*

The roadside unit (RSU) enables vehicles incorporating V2I technology to connect to a range of infrastructure, from traffic signals to buildings and other road users such as cyclists. It can be employed as part of the transport infrastructure in the form of either an eNB for LTE or a gNB for the upcoming 5G NR and obtains data from the local vulnerable user, sensor, and application server. The RSU transmits the processed information from the application layer from the UE which supports V2I applications to connected UEs. This can be conducted through unicast, multicast, and broadcast transmission, and the RSU can select which UEs should receive the V2I application information. Additionally, the server for cellular networks can be placed flexibly [20].

## 6.2 Vehicle-to-Network (V2N)

To provide drivers with convenient in-vehicle services like traffic updates and video streaming, "vehicle-to-network" (V2N) technology connects automobiles to cellular infrastructure and the cloud. Two of the most prevalent instances of this technology are automobiles that can sync with a user's smartphone and vehicles with built-in navigation and traffic systems, such as Google Maps.

## 6.3 Vehicle-to-Pedestrian (V2P)

V2P refers to cellular-based communication between UEs that enables P—vulnerable road users like walkers, cyclists, motorcyclists, roller skaters, etc.—applications. Because of its limited direct communication range, V2P necessitates the transmission of V2P-related application information between different UEs (one for vehicles and the other for pedestrians), either directly or through infrastructure supporting V2X service, such as RSUs, application servers, etc. With UE supporting V2X service, information from a V2P application can be transmitted by a car or a pedestrian.

## 6.4 Initial Access Problem

Finding and establishing a reliable physical communication link between the BS and the UE is known as initial access (IA) [21]. This is an issue because it increases the likelihood that there will be a delay or interruption as the UE and BS look for the best beam alignment to create a direct communication link. Therefore, it's crucial to find a solution to this initial access problem that reduces latency while also enabling dependable and high-capacity communication [22, 23]. In this study, we have concentrated on solving the initial access issue. Figure 8 presents a basic scenario of beamforming in 5G-V2X communication.

## 6.5 Resource Allocation

In 5G C-V2X, single-carrier frequency-division multiple access is used. The 5.9 GHz channel width can be either 20 MHz or 10 MHz, depending on the protocol. Channels are made up of subframes in time and subchannels in frequency. Resource blocks could be used to separate the elements of a subchannel (RBs). Each subcarrier works at a frequency of 15 kilohertz, with 12 subcarriers per RB and a total frequency bandwidth of 180 kilohertz. A subchannel can have a different

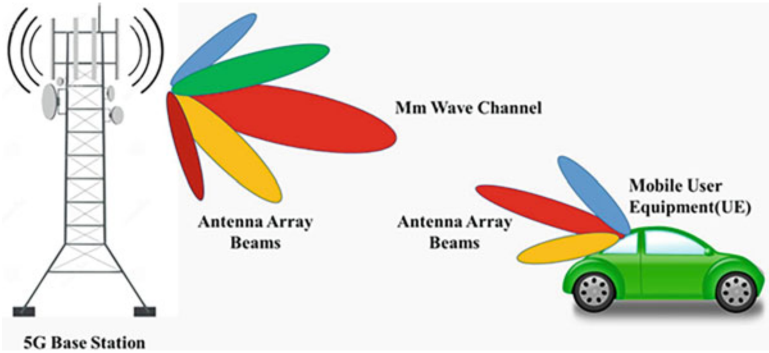



**Fig. 8** Beamforming in 5G-V2X communication

number of RBs than usual. Sidelink control information (SCI), which keeps track of the modulation and coding scheme (MCS) that is employed for decoding at the receivers, is sent over the physical sidelink control channel (PSCCH). The physical sidelink shared channel (PSSCH) is used to send data across a predetermined number of subchannels (PSCCH). Figure 9 presents the resource pool in 5G-V2X communication.

In 5G-V2X, there are two separate radio interfaces (i.e., cellular V2X and C-V2X). The PC5 interface enables sidelink communications between vehicles and user devices. Communication from a vehicle to a base station or an eNB uplink/downlink is made possible through the Uu interface (UEs). The 5G standard offers four distinct operating modes for C-V2X [24]. The only accessible modes for ultra-reliable low-latency communications are 3 and 4 (URLLC). This is because each mode may conserve resources in a unique manner. When a vehicle is beyond the range of the base station, it switches to mode 4 and uses sensing-based semi-persistent scheduling (SPS) to divide time and spectrum resources. Otherwise, vehicles function in mode 3 with two choices for resource distribution while within base station communication range.

Several studies on the distribution of physical resources focus on the scheduling and resource reservation for mode 4 transmission. However, since most vehicles are used in urban areas, typically causing a large amount of data to be shared between many UEs, mode 3's dynamic resource allocation technique is often undervalued. It has the greatest opportunity to optimize resource exploitation and QoS when used to satisfy the tough requirements of eV2X services. The decentralized structure of the SPS-based system also offers flexibility in meeting various requirements.

### 6.5.1 Resource Allocation Modes

Two new resource allocation modes are introduced in Release 14, one of which is specifically designed for vehicle-to-everything (V2X) sidelink communication.

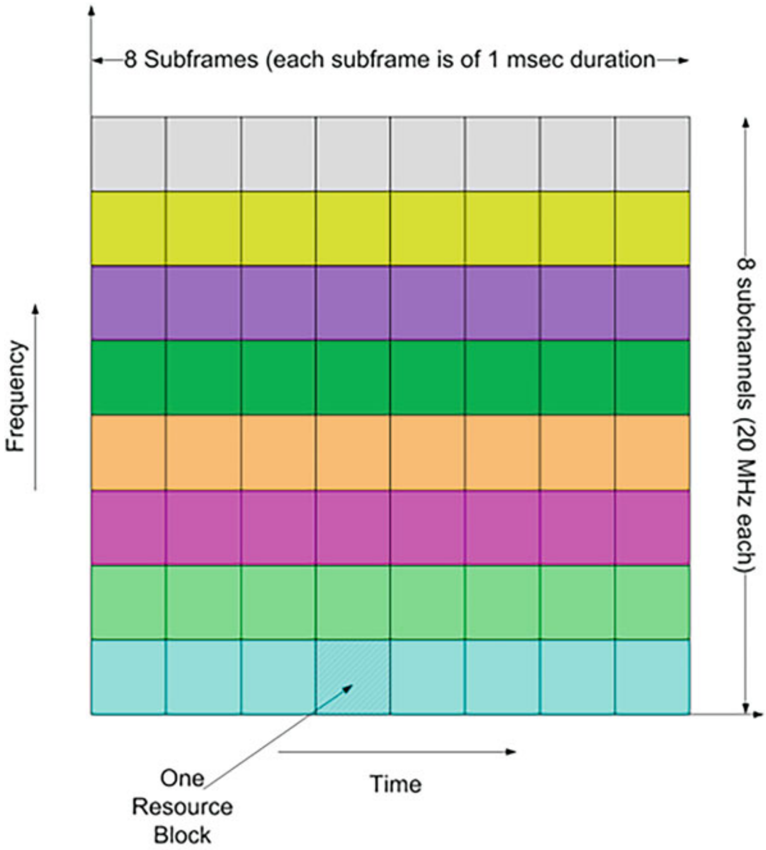



**Fig. 9** Resource pool in 5G-V2X communication

In the following flow diagram, various modes of resource allocation is described. Resource allocation is divided into two modes (mode 3 and mode 4). Both modes have different criteria for resource allocation. Dynamic scheduling and semi-persistent scheduling techniques are used in the coverage of the base station, which is operated in mode 3. In contrast, sensing-based semi-persistent scheduling is used without coverage of the base station, which is operated in mode 4. Figures 10, 11, and 12 show the various modes of resource allocation.

In mode 3, vehicles operate in the coverage of the cellular base station. The selection of subchannels in mode 3 does not provide an algorithm. In contrast, when a vehicle operates without the coverage of a base station, it uses sensing-based techniques to allocate the resources in mode 4.

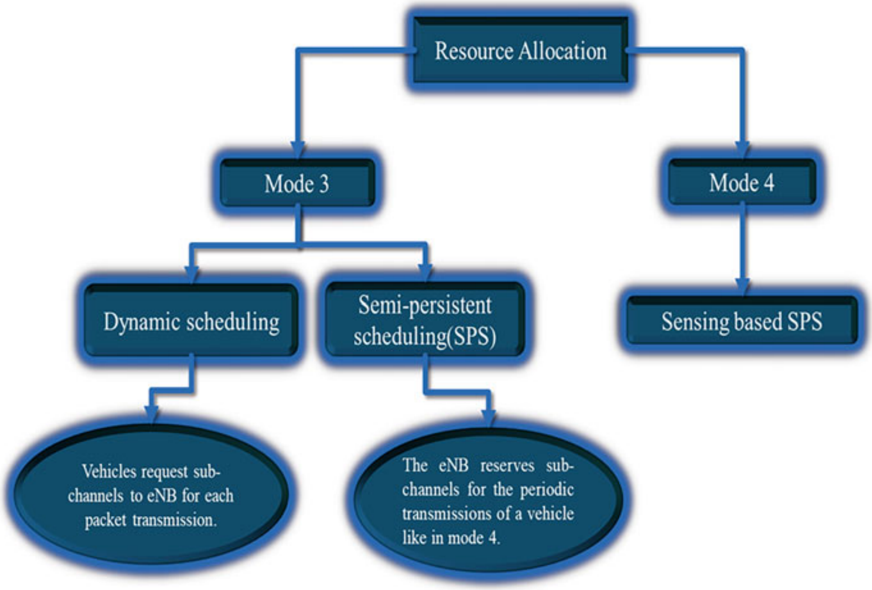
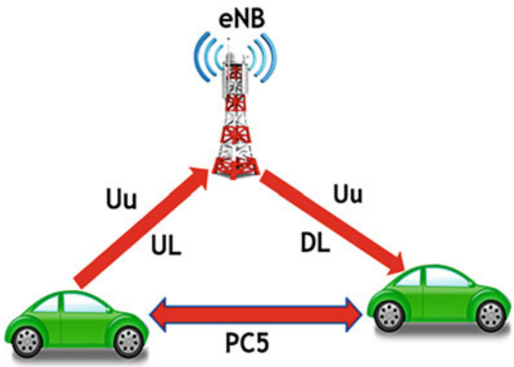



**Fig. 10** Various modes of resource allocation

**Fig. 11** Resource allocation in mode 3




## *6.6 Relay Vehicle Selection*

For a long time, the industry has explored and employed the concept of using an intermediate device or collection of devices to help in the flow of information to another device or set of devices (referred to as sources). It has been demonstrated to offer a number of benefits that can range from improved performance to increased reliability, improved security, and reduced latency. By incorporating multiple relays, the conventional three-terminal communication channel, which was traditionally comprised of just a source node, a destination node, and a single relay, has been

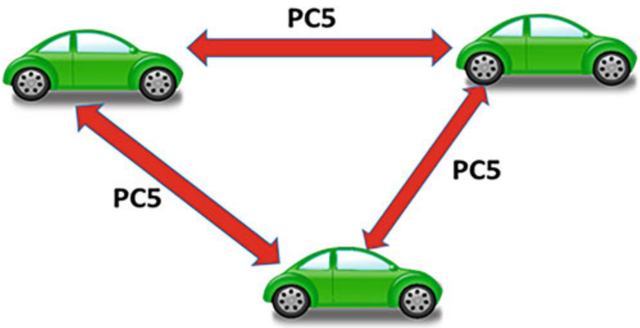



**Fig. 12** Resource allocation in mode 4

upgraded to become a multi-terminal version, with multiple relays. These intermediate devices can help in optimizing the path and information flow, thereby providing even greater efficiency and effectiveness. Additionally, with the use of multiple relays, it is possible to make the entire system more resilient and flexible, as data can be routed through multiple paths [25], multi-antenna relays, and buffered relays [26]. The diversity advantage in [27] is demonstrated through data transmission between a source and a relay. Whenever source-destination and source-relay-destination connections are present, cooperative communication is conceivable. In this case, the receiver must have a suitable combining strategy. Cooperative communication also happens when information is shared among several sources in a way that frequently makes each source act as a relay for a single source. When the reference connection is lost, relaying can still be done without the relay interacting with the sender. When a network contains a large number of relays and multi-hopping from source to destination is enabled, finding the right routing protocols becomes difficult. Relay selection techniques are used to provide road safety to vehicles under Line of Sight (LOS) or Non Line of Sight (NLOS).

## 7 Main Challenges for Machine Learning

When working with machine learning algorithms and systems, a number of difficulties can occur. The following are some of the major challenges:

(a) *Data availability and quality:* To build and test models, machine learning algorithms need a lot of high-quality data. The model's performance can be adversely affected by data that is noisy, imperfect, or biased. Furthermore, data could not be easily accessible or might be restricted for security and privacy concerns.
(b) *Algorithmic fairness and bias:* Machine learning algorithms can occasionally pick up on and encode biases that are present in their training data, which

can result in inaccurate or prejudiced diagnoses or predictions. This can be especially troublesome in fields like healthcare and criminal justice, where using biased algorithms can have disastrous results.

(c) *Explainability and interpretability:* Machine learning algorithms have the potential to generate highly accurate predictions, but they can also be complicated and challenging to comprehend. In areas like health and finance where transparency is crucial, this can make it challenging to explain the reasoning behind the forecast.

(d) *Overfitting and underfitting:* These are two frequent issues that can occur during the training of machine learning models. When a model is overfitted, the training data's noise or random fluctuations are learned by the model, which results in a poor generalization of new data. On the other side, underfitting happens when the model is too straightforward and incapable of capturing the underlying patterns in the data, resulting in subpar performance.

(e) *Efficiency and scalability:* Machine learning methods can use a lot of processing power, especially for larger and more complicated data sets. This can limit the usage of machine learning models in real-time or high-demand applications and make it difficult to train and deploy them at scale. Additionally, given that assertions must operate on low-power devices in contexts with limited resources, the computational efficiency of machine learning methods can be problematic.

## 8 Conclusion

In this chapter, first, machine learning is explained. The details of its origin to the classifications of the machine learning algorithms were discussed. Along with this, the difference between the existing problem-solving approach and the machine learning problem-solving approach was also discussed. Some applications of machine learning in real life are briefed. The application of machine learning in smart cities was elaborated. Further, the smart cities current scenario is also discussed. The connection between machine learning and smart cities is very strong, as a lot of problems related to smart cities such as monitoring, traffic management, and security will be solved efficiently using machine learning. In addition to machine learning, the vehicular communication is explained in detail. The various types of vehicular communications such as vehicular ad hoc networks, LTE-V2X, and 5G-V2X are explained. Various problems related to 5G-V2X implementation are discussed. These are majorly the initial access problem, relay vehicle selection problem, and resource allocation problem. The resource allocation problem is a big problem, and researchers are working to allocate the resources in time and frequency domains efficiently. Since the resources are complex to distribute, the use of machine learning is eminent in this area. If we try to allocate the resources using traditional methods, it may not produce good results. However, if we let the vehicles learn to allocate resources to themselves or maybe by the base station using machine learning, it will produce optimum results. The initial access issue, increased millimeter-wave attenuation, non-line-of-sight connection, etc. are some of the issues with 5G NR V2X communications. These issues are thoroughly discussed in this chapter, along with potential solutions. Additionally, Industry 4.0 breakthroughs in artificial intelligence and machine learning have made it possible to find the best solutions to the aforementioned issues. The application of machine learning techniques in the repair of issues with 5G NR V2X communications is also covered in this chapter. Each 5G NR V2X communication issue is covered in depth, along with potential fixes that may or may not involve machine learning methods.

## References


1. Ullah, Z., Al-Turjman, F., Mostarda, L., & Gagliardi, R. (2020). Applications of artificial intelligence and machine learning in smart cities. *Computer Communications, 154*, 313–323.
2. Das, K., & Behera, R. N. (2017). A survey on machine learning: concept, algorithms and applications. *International Journal of Innovative Research in Computer and Communication Engineering, 5*(2), 1301–1309.
3. Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., & Iyengar, S. S. (2018). A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR), 51*(5), 1–36.
4. Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon, 4*(11), e00938.
5. Kirimtat, A., Krejcar, O., Kertesz, A., & Tasgetiren, M. F. (2020). Future trends and current state of smart city concepts: A survey. *IEEE Access, 8*, 86448–86467.
6. Samiksha, S., Balachandran, K., & Sumitha, V. S. (2016). *A framework for smart transportation using Big Data*. Paper presented at the international conference on ICT in Business Industry & Government (ICTBIG), 2016.
7. Alatabani, L. E., Ali, E. S., Mokhtar, R. A., Khalifa, O. O., & Saeed, R. A. (2022). *Robotics architectures based machine learning and deep learning approaches*. Paper presented in 8th International Conference on Mechatronics Engineering, 9–10 August 2022.
8. Carlos, R. S., & Fatima, D. F. (2020). A survey of 5G technology evolution, standards, and infrastructure associated with vehicle-to-everything communications by internet of vehicles. *IEEE Access, 8*, 117593–117614.
9. Pal, R., Prakash, A., Tripathi, R., & Naik, K. (2019). Scheduling algorithm based on preemptive priority and hybrid data structure for cognitive radio technology with vehicular ad hoc network. *IET Communications, 13*(20), 3443–3451.
10. Verma, A., Pal, R., Prakash, A., & Tripathi, R. (2018). Information retrieval in two-tier VANET/P2P using RSU as a superpeer. *Wireless Communication Technology, 2*(1), 1–9.
11. Pal, R., Gupta, N., Prakash, A., Tripathi, R., & Rodrigues, J. J. (2020). Deep reinforcement learning based optimal channel selection for cognitive radio vehicular ad-hoc network. *IET Communications, 14*(19), 3464–3471.
12. Singh, P., Pal, R., & Gupta, N. (2016). Clustering based single-hop and multi-hop message dissemination evaluation under varying data rate in vehicular ad-hoc network. In *Advanced computing and communication technologies: Proceedings of the 9th ICACCT* (pp. 359–367).
13. Prakash, P. V., Tripathi, S., Pal, R., & Prakash, A. (2018). A slotted multichannel MAC protocol for fair resource allocation in VANET. *International Journal of Mobile Computing and Multimedia Communications (IJMCMC), 9*(3), 45–59.

14. Prakash, U., Pal, R., & Gupta, N. (2015, November 1–5). *Performance evaluation of IEEE 802.11 p by varying data rate and node density in vehicular ad hoc network*. Paper presented in 2015 IEEE Students Conference on Engineering and Systems (SCES), IEEE.
15. Kutila, M., Pyykonen, P., Huang, Q., Deng, D., Lei, W., & Pollakis, E. (2019). *C-V2X supported automated driving*. Paper presented in Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops), Shanghai, China, 1–5 May 2019.
16. Qualcomm. (2019). Cellular vehicle-to-everything. [Online]. Available: https://www.qualcomm.com/invention/5g/cellular-v2x
17. Behnad, A., & Wang, X. (2017). Virtual small cells formation in 5G networks. *IEEE Communications Letters, 21*(3), 616–619.
18. Mannoni, V., Berg, V., Sesia, S., & Perraud, E. (2019). *A comparison of the V2X communication systems: ITS-G5 and C-V2X*. Paper presented in Proc. IEEE 89th Veh. Technol. Conf. (VTC-Spring), Kuala Lumpur, Malaysia, IEEE, 1–5 April 2019.
19. Ma, H. S., Zhang, E., Li, S., Lv, Z., & Hu, J. (2016). A V2X design for 5G network based on requirements of autonomous driving, SAE Tech. Paper 2016- 01-1887.
20. Molina, M. R., & Gozalvez, J. (2017). LTE-V for sidelink 5G V2X vehicular communications: A new 5G technology for short-range Vehicle-to-Everything communications. *IEEE Vehicular Technology Magazine, 12*(4), 30–39.
21. Marco, G., Marco, M., & Michele, Z. (2016). Initial access in 5g mmwave cellular networks. *IEEE Communications Magazine, 54*(11), 40–47.
22. Cheol, J., Jeongho, P., & Hyunkyu, Y. (2015). Random access in millimeter-wave beamforming cellular networks: issues and approaches. *IEEE Communications Magazine, 53*, 180–185.
23. Antonino, O., Olga, G., Sergey, A., Osman, N. C. Y., Tuomas, T., Johan, T., & Yevgeni, K. (2018). *Improving initial access reliability of 5g mmwave cellular in massive v2x communications scenarios*. Paper presented in 2018 IEEE International Conference on Communications (ICC), IEEE, 1–7 May 2018.
24. Nestor B, Fotis F, & Paul P (2018). *Enhanced 5g v2x services using sidelink device-to-device communications*. Paper presented in 2018 17th Annual Mediterranean ad hoc networking workshop (Med-Hoc-Net), IEEE, 1–7 June 2018.
25. Gastpar, M., Kramer, G., & Gupta, P. (2002). The multiple-relay channel: Coding and antenna-clustering capacity. In *IEEE International Symposium on Information Theory* (p. 136).
26. Xia, B., Fan, Y., Thompson, J., & Poor, H. V. (2008). Buffering in a three node relay network. *IEEE Transactions on Wireless Communications, 7*(11), 4492–4496.
27. Laneman, J. N., Tse, D. N. C., & Wornell, G. W. (2004). Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Transactions on Information Theory, 50*(12), 3062–3080.

# Analysing the Challenges and Opportunities of Smart Cities

**Fezile Ozdamli and Muhammad Bello Nawaila**

## 1 Introduction

A smart city uses information and communications technology (ICT) to provide better citizen welfare and quality of government service, share information with the public and improve operational efficiency [1].

Many researchers have explained the concept of a smart city from different aspects. For example, Obringer and Nateghi [2] argued that the technology-oriented smart city definitions must reflect climate change's social and environmental effects. In contrast, the social welfare and sustainability-based smart city concepts affect climate change.

Samarakkody and Amaratunda state that six dimensions can be seen holistically within the concept of a smart city [3]. These six dimensions consist of areas where strategies and plans should be implemented to understand disaster resilience and efforts in the smart city (Fig. 1).

Researchers state that these six dimensions are areas where disaster resilience in the smart city should be understood and efforts and strategies should be implemented. Many components make a city smart. Bapat et al., on the other hand, state that the main ones are water and waste management, transportation and people's safety [4]. There are some key elements that make a city smart (Fig. 2).

Although smart homes are an element of smart cities, they do not have to be in them. Individuals can turn their homes into smart homes at their own will. However, they are associated with smart cities as they work based on smart devices and data

F. Ozdamli (✉)
Near East University, Nicosia, Cyprus
e-mail: fezile.ozdamli@neu.edu.tr

M. B. Nawaila
Aminu Saleh College of Education, Azare, Nigeria

 
N. Gupta, S. Mishra (eds.), *Internet of Everything for Smart City and Smart Healthcare Applications*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-34601-9_6

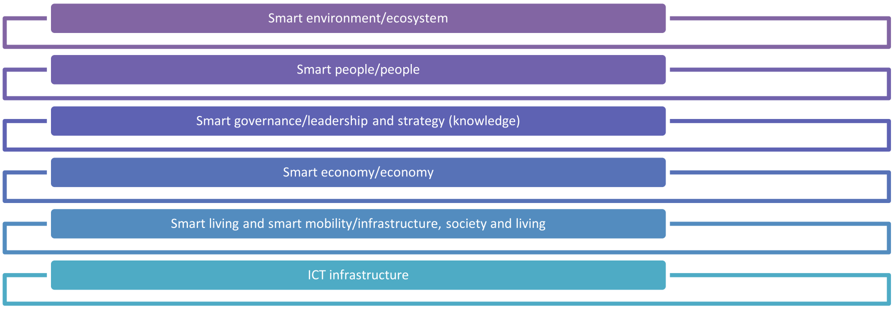



**Fig. 1** Six dimensions of smart cities. (Adapted from [3])

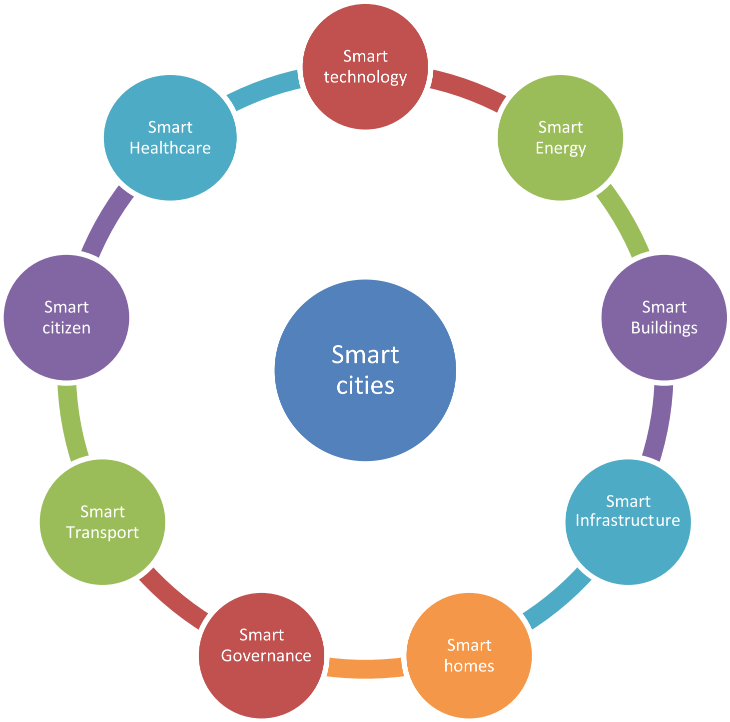



**Fig. 2** Key elements of a smart city

collection. For example, many people can manage their devices at home with their smartphones. For example, the air conditioner can be set to heat the house without going home, or the necessary settings can be made on the smartphone so that the washing machine ends at the time of returning home.

With the increase in smart cities, the number of smart buildings is also increasing. New house designers design their houses as smart buildings, and even other homes are turned into smart houses. The simplest example is that doors can be opened automatically with cameras with face recognition features. TV channels can be changed automatically according to the preferences of the people in the house. Another area where smart technologies are used is smart traffic lights, which know how long to allow vehicles and pedestrians to cross to avoid traffic jams. As life increases in cities, planning and managing energy use is important. Smart electricity meters and systems can be used to save energy. For example, by using street lamps with sensors at night, energy savings can be achieved only when vehicles and pedestrians pass by.

Today's smart cities are developing rapidly, aided by the rapid advancement in technology and tools and the huge investment and priorities placed by most countries across the globe. At the developmental stage, smart cities also consider environmental aspects such as lower power consumption, lower harmful gas emissions and producing lower carbon footprint [5].

Adhering to this new paradigm, a lot of smart cities focus more on continuous innovation and improvement in the majority of their services, with most smart cities having their pain points, such as maintenance of complex IT infrastructure [6], poor participation of citizens in city administration [7], waste management [8], energy deficiency [9], water scarcity [10], traffic congestion [11] and so on. Despite their differences in kind and scale, the aforementioned issues are present in all smart cities. Smart cities' planning, creation and administration must consider these concerns. High-end tools and technology can decrease or even solve problems for similar issues [12].

To be considered a "smart city", a place must allow residents to live and work there, utilise its resources effectively [13, 14], be a part of a technology-based infrastructure [15] and have confident and forward-thinking.

In this chapter, the authors analyse challenges faced by smart cities and various smart services and how these issues are effectively addressed by researchers with special attention to ICT and digital technologies by reviewing some selected research papers devoted to smart cities' challenges and problems and how these issues are resolved using the latest technologies and tools.

Finally, the authors recommend future works and discuss the probable solutions to undertake its full-fledged implementation while addressing the observed limitations.

As more and more new smart cities are built worldwide [11, 16, 17], smart cities are rapidly evolving. It is a well-known fact that smart cities are utilising the most recent technology and tools to address the practical issues encountered during the maintenance, operation and development of smart cities [18]. There is abundant academic literature available to address the issues and effective solutions for smart cities [1, 15, 19–23]. However, it has been noted that the majority of research papers and literature have concentrated on the individual problem statement and a specific solution based on mostly a single technology [7–10]; however, very few research works have described in detail the list of issues that smart cities face and the available technologies and tools that they employ to address those issues.

The idea of "smart cities" is quite broad, and the creation, administration and upkeep of such cities rely on a wide range of technological advancements [18, 24, 25]. Due to time and financial limitations, it is only possible to fully describe some of the tools and technology that smart cities utilise to handle their issues.

In this chapter, a holistic analysis is made of opportunities brought about by smart cities and the challenges faced while developing them and those faced by the residents. This chapter makes an effort to explore a variety of real-world issues smart cities face as well as numerous digital technology-based and ICT solutions and tools employed in smart cities to address these issues. This chapter would be a great resource for smart city policymakers and students to find pertinent information on various technologies and tools to solve smart city problems.

## 2 Opportunities

When we look at the use of smart cities in countries, a cost-sharing programme has been developed for building owners who invest in real-time energy systems with the policies designed in the USA. It is stated that with this developed programme, energy savings of 15–30% are achieved every year. In Canada, a company, one of the largest food retailers, introduced smart cart contactless shopping, skipping the payment queues of customers and allowing them to weigh products and pay, thus saving time. Barcelona, Singapore, Stockholm and Milan started smart transportation applications to make traffic flow more efficient. In a smart city project in Singapore with a just-in-time waste collection application, containers operating with solar energy, which provide eight times the capacity of traditional garbage cans and operate with sensors and compactors according to their occupancy levels, were installed. A similar study in Seoul also showed that it reduced waste collection costs by 75% [26]. In India, they focus on energy and water resources management in clean and sustainable environments in smart city projects [27]. Japan is pursuing a smart city policy that updates existing infrastructure for efficient energy use, disaster recovery and disaster prevention [28]. In addition, Japan shows great interest in smart car projects.

Also, one of the important benefits of smart cities is mobility. With the advantage of mobility, real-time job tracking and quick answers to problems can be found throughout the city via wireless communication [29].

## 3 Smart Homes

Smart homes are usually realised with IoT applications. Depending on the features of smart homes, many advantageous features are offered to their users.

Thanks to the devices in the smart home, people can drink their specially prepared hot coffee in the coffee machine when they wake up in the morning.

The refrigerator's sensors can identify the missing products and order online. In addition, it is possible to achieve lower costs with sensor lighting technologies in the homes of even those who do not switch to smart home applications. Smart lockers are also available, along with virtual assistants found in most smart homes. These virtual assistants can send alerts from water and gas sensors to homeowners and even shut off water and gas systems through disaster-preventing smart locks. Smart homeowners can control from anywhere at any time. For example, when the child of the smart homeowner returns home from school, his mother or father can open and close the house doors from the workplace. This way, the owner can be sure that the child and the house are safe. The systems that can generally be controlled from the home automation application are shown in Fig. 3.

## 4 Smart Buildings

Today, working environments are changing with the effect of the pandemic. Home office applications are increasing. In addition, the number of people who spend more time at work with the creation of flexible working environments is too high to be underestimated. Especially the developers in the technological fields are making great efforts to provide better office experiences. Thanks to smart buildings, employers reduce costs, ensure employees' comfort and ensure their facilities' confidentiality. For example, employees can easily log in and out with face recognition applications. The clean electricity produced with solar panels meets the building's electricity. Technologies, such as sensor lighting, sensor windows, etc., increase employees' comfort in business life. It also facilitates information exchange between connected devices and employees. These practices enhance the sense of community among employees.

## 5 Air Quality

One of the features of smart cities is to provide a sustainable environment. IoT sensors can capture air pollution levels. Thus, environmental authorities can plan important policies with this data. Another feature that helps improve air quality in smart cities is green public transport applications [30]. Due to the migration of citizens to cities, the necessity of improving air quality has come to the fore. Innovative technologies in smart cities are expected to reduce this problem. As they adopt clean industries and ecologically friendly architecture, smart cities typically minimise carbon emissions and pollution, thanks to clean energy sources like solar, wind and water [31].


**Fig. 3** Control features in the home automation app

## 6 Reduce Pollution

Data collected from monitoring the most populated urban areas show that pollution has an impact on not only our own lives but also the lives of future generations. Therefore, we must exercise extreme caution and work to limit the dangerous emissions we send into the environment [32]. One of the important causes of global warming is air pollution, so there is an increasing focus on solving this problem. Environmental control and pollution avoidance, two of today's most pressing issues, are increasingly under the research spotlight [25]. The Internet of Things and smart city technologies are the solutions to this issue in preventing and controlling environmental pollution. Smart cities typically offer systems for recycling garbage into ecologically beneficial products, techniques for turning waste into electricity and systems for treating rainwater, irrigation water and sewage water. Peng et al., in their study to reduce environmental pollution by analysing the data they obtained with sensors and videos to test the current environmental pollution conditions using IoT technologies, stated that the designed system successfully eliminates and monitors haze, water resources pollution and fire disasters [25]. The studies mentioned in the literature show that it is necessary to have a well-developed infrastructure to overcome crises [33]. The United Arab Emirates planned to deal with environmental problems such as climate change by using smart urban technologies in their "Masdar City" projects. The project aimed at electricity consumption by 51% and water consumption by 55% with building design, bus stops, pavements and wind towers [34].

## 7 Smart Transport

Another of the most important advantages of the smart city is that it provides efficient mobility in transportation. Traffic safety is increasing in cities using smart transport systems. With UAV applications, data on how traffic accidents happen can be collected, and measures can be taken for them. Analyses can be done with UAV vehicles [35, 36]. In cities where smart transportation is applied, IoT sensors are installed in certain parts of the city to determine the traffic situation. These sensors collect data and transfer it over the 5G network. The use of blockchain technology at this stage ensures the secure transmission of data. This data helps to find practical solutions by analysing it with data experts and traffic authorities and using artificial intelligence [26].

In addition, smart traffic systems notify drivers in real time of routes where traffic flow is fast and congested, thus saving fuel and leaving less carbon footprint. Barcelona is known as a pioneer city in smart transportation applications. There are traffic sensors, connected cars and autonomous vehicles in the city, so the big data collected can be analysed, and solutions are produced quickly.

## 8 Smart Waste Management

The massive increase in the population of smart cities will eventually come with its disadvantages, among which is the enormous increase in waste generation (waste from industries, humans and food). This huge waste collection and processing are by no means an easy job, and if the not properly managed system can be hazardous to the environment, hence the need to place more attention on waste management.

Waste management entails proper disposal of waste and controlling/monitoring, collection, processing and recycling [9].

During waste management exercises, waste prevention is always classified as the priority, such as reusing waste materials, recycling, energy generation and, finally, waste disposal [37].

Even though multiple ways currently exist for waste control, processing and recycling, with the ever-increasing waste generated, these processes are ever becoming challenging and expensive.

Currently, engineers, scientists as well as researchers are working on improving smart waste management systems using automated waste bins, automated waste collection, landfill modernisation, energy generation and environmentally friendly operation.

## 9 Smart Education

By applying smart education, Fortes et al. [37] believe that educational institutions can use the latest technology to develop education in the context of smart cities. Smart education entails the newest form of learning in this technology-driven era; this is expected to play a vital role in smart city development.

Smart education employs the Internet and other relevant smart technologies (mobile applications, social media, etc.) to aid students and stakeholders. This is used to improve educational innovation and practices. It usually generates data via the online activities of teachers, employers and students.

It also uses advanced techniques, tools and teaching practices in conjunction with the latest technologies for effective and efficient educational services. It can change pedagogy [38]. It, therefore, presents opportunities to improve teaching and to learn beyond the realms of distance education.

## 10 Smart Tourism

The quest for smart cities to address certain travel-related needs has brought about the concept of smart tourism.

Smart tourism is an avenue that engages stakeholders in the field of tourism and tourist by enabling these cities/destinations to utilise their peculiarities to maximise the tourist experience in a fun and convenient way. It employs technology, coordination and data control to create sustainable, efficient and qualitative knowledge.

Currently, various cities are tilting towards smart applications and smart tourism as an integrated system component and integrating them as part of the economic and developmental policies [39].

The emergence of smart tourism has its root in sustainability, and more cities are now prioritising smart tourism to provide the interface between the destination and visitors in tackling specific needs.

These smart tourist destinations can be distinguished by constant innovations facilitated using standard technology and interface to improve tourists' and residents' quality of life. According to [40], for a tourist destination to be smart, it must have the following:

1. Better tourist management experience.
2. High level of data use and generation.
3. Higher level of local residence engagement.
4. High level of interface between the destination and tourist.
5. Digitalisation of systems.

## 11 Smart Healthcare

Smart cities can benefit from smart healthcare in various ways, including those that affect patients, governments, the medical community and suppliers [41]. When systems and data are integrated and interoperable across various services, including basic health and public safety, housing quality, environmental health, social services, emergency services and transportation, a data-enabled, digitally connected "smart city" makes healthcare smarter [42].

Improving access to traditional healthcare and smart cities is not the only way cities can protect public health. Sustainable and environmentally friendly planned smart cities include green streets and social life centres, which contribute to forming a healthy environment. To meet the needs of the smart health ecosystem, smart health services can use the Internet of Things (IoT), sensors, big data, artificial intelligence and wearable device technologies.

## 12 Smart Governance

Rural-to-urban migration poses new challenges to governance and administration in all cities across the globe, posing more problems like economic crisis, massive stress to city infrastructure, safety and security issues, overstretching of healthcare facilities, transportation and energy and water shortages.

To tackle these, cities are digitalising; therefore, these smart cities need to be managed effectively; authorities and administration need to be properly managed.

According to the European Charter, territorial self-government and good administration are rights [43], and since power belongs to the citizens, that explains the framework for citizens' participation in governance [44].

The advancement in information and communications technology was addressed via the European digital transformation strategy, where the idea of trade, entrepreneurship and digital transformation serving the people was represented using three important bedrocks, that is, developing, deploying and adopting technology to develop a trustworthy, sustainable, democratic and open digital society.

According to Sararu [45], multiple quotas are clamouring for an accelerated pace to implementing e-governance (which is e-government facilitated by public administrators) which aims to bring forth cheaper, faster and more efficient administration in smart cities.

Despite the long-term discussion on smart city design and implementations and its relevancy, researchers in the field of smart city governance and administration, even from a legal standpoint, are limited. With issues of mixed messaging, for instance, lesser consumption can hamper the economy not tackled [46].

## 13 Challenges

In addition to the opportunities provided by smart cities, many challenges come to the fore. It is necessary to reduce energy consumption in a world with ecological difficulties, but the increasing use of digital tools can create negative environmental consequences. In addition, systems that obtain personal data also create various problems, such as protecting individual freedom.

There are challenges to overcome for all of the benefits of smart cities. These include government officials allowing widespread participation from citizens. There is also a need for the public and private sectors to align with residents so that everyone can actively aid the community. These difficulties can be seen from various perspectives, such as the design, implementations [36], policies [47], those experienced by residents or policymakers [48], systems or services.

Smart city initiatives must be open to the public and accessible through a mobile app or public data site [49]. It enables residents to interact with the data and carry out private chores like bill paying [50], locating smart transportation options [51] and determining how much energy is being used at home [52].

To avoid hacking or other unauthorised usage, this requires a reliable and secure data collecting and storage system [53]. To avoid privacy concerns, smart city data must also be anonymised.

## 14 Availability

To ensure that the data serves its purpose of monitoring the various components of the smart city infrastructure, it must be readily available in real time with dependable access. Accountability – System users must take responsibility for their behaviour and interactions with systems that handle sensitive data. Users' logs should keep track of who accesses the data to provide accountability in case of a problem. Sensitive information must be kept private and protected against unauthorised access. It could refer to the use of firewalls or data anonymity. Integrity – The information must be accurate in addition to being easily accessible. Additionally, it implies taking precautions to avoid outside tampering [54].

Many countries have already enacted laws, such as the US IoT Cybersecurity Improvement Act, to help determine and establish minimum security standards for connected devices in smart cities [55]. Cities, for instance, are at various levels of development and application of smart technology. Nevertheless, some are innovating and setting the standard for truly smart cities.

Singapore is one of the leading cities in the stages of smart city building. They carry out the monitoring of crowd density, car movements and the cleaning of public areas with IoT systems. Singapore has mechanisms in place for tracking the consumption of electricity, water and garbage. Monitoring methods and testing for driverless vehicles are also available to protect the health and wellness of senior citizens [56].

In addition to traffic monitoring systems, telemedicine, education, public services and smart touristic solutions are offered in Dubai. In Barcelona, on the other hand, besides smart systems that measure humidity, precipitation and air and noise pollution, smart parking, bus stops with Wi-Fi and USB charging facilities, bicycle sharing areas and many smart system applications with online payment facilities are used [48].

## 15 Privacy Issues

Technologies such as cameras with face recognition, sensors measuring crowd density and motion sensors to save electricity are integrated into smart cities. Smart cities make our lives easier in many areas; for example, sensors that find empty parking spaces and help drivers park their cars can be used. This application is carried out in Barcelona city. However, smart cities rely on collecting large amounts of data to work well [57].

Smart cities offer many opportunities to increase public safety, including connected surveillance systems, intelligent roads and public safety monitoring. However, smart cities need protection from data theft, hacking and cyberattacks and ensure that the reported data is accurate [58]. There is a trade-off between personal satisfaction and security breaches in any city. While everyone should be in an increasingly advantageous, peaceful and good environment, no one wants to feel like they are being watched. Cameras placed on the streets to help deter crime can help while simultaneously creating fear in residents [3].

Protecting privacy is essential because smart cities gather information about people's homes and lifestyles. Sidewalk Labs, a division of Google's parent company Alphabet, had to abandon plans to build a smart city on the Toronto waterfront after receiving harsh criticism from locals who worry about "surveillance capitalism" [26].

Implementing strategies like ID solutions [59], robust authentication management and physical data vaults [60] is necessary to manage the security of smart cities. Citizens must have confidence in the security of smart cities, which requires collaboration between network service managers, energy providers, device manufacturers, software developers, private sector businesses and the government [61, 62].

Smart cities have many advantages in reducing the crime rate, protecting the environment, providing convenience for individuals and many other factors. However, there are some risks that all individuals living in smart cities or smart homes should be aware of. Nathan states that the following adverse events may occur due to smart cities [57]:

- Facial recognition can be used to track down non-criminal civilians.
- Smart meters provide unprecedented information about household appliances used by citizens.
- Smart devices can accidentally spy on people in their own homes. For example, it was recently published that the smart robot accidentally shared images at home on the Internet.
- Self-driving cars run the risk of constantly sharing data about driver location and "transportation habits".
- Intelligent systems can make people's lives unnecessarily difficult for even minor breaches.

The potential impact of the data collected by smart city systems on citizen privacy is the possibility of this data being stolen by hackers. In addition, attackers can initiate cyberattacks using these smart technologies, and systems can crash with these attacks [63].

## 16 Infrastructure Issues

The physical and information technologies infrastructure should be well structured when the smart city project is planned. Regarding physical infrastructure, camera poles may be installed in crowded neighbourhoods. On the other hand, information technology infrastructure should be analysed and interpreted best regarding citizens' data and traffic behaviour. Information technology infrastructure needs to be expandable and flexible. To not change the infrastructure every few years, it should be able to adapt to the increasing population, new technologies and software [64]. However, the infrastructure required to get the desired efficiency requires serious financial investment. Since dozens or millions of IoT devices must connect and cooperate, connectivity is the biggest difficulty. As demand rises, this will enable the integration of services and continual improvements [65].

## 17 Policy Issues

The deployment of smart city systems could be hampered by regulations that address national benchmarks, ranging from municipal grants for businesses to legislative mandates and acts that reflect complete regulation [66].

The important challenges are deciding on applicable policies for smart cities and paving the way for research and development investments. In addition, since smart cities are based on data, deciding how to share personal data and deciding on policies to protect personal data are important challenges in smart city applications.

## 18 Social Issues

It is stated that there is a need for “smart” citizens who can actively benefit from smart city opportunities to put smart cities into practice [67]. For this reason, while designing smart city projects, it is necessary to plan user training for the benefits and use of these projects. In addition, smart city applications should be prepared to encourage social inclusion and be implemented to serve all citizens, not just technology enthusiasts [68]. For example, brochures or promotions containing explanations for using smart city applications can be made for elderly individuals. In addition to smart technologies, smart cities should be designed considering social aspects that motivate people to implement and contribute to the cultural fabric that gives them a sense of trust. It is important to convince citizens to use these applications, especially in smart cities built from scratch [15].

## 19 Network Connectivity

For the implementation and management of smart cities, sensors, accumulators and cameras that collect and send big real-time data are placed in many parts of the city. This collected data is instantly analysed for the smart city to perform its functions. For this reason, a strong and suitable Internet connection for high-speed data transfer is needed. Many towns need an Internet connection with the infrastructure to provide this data flow. For this reason, the Internet connection problem should be solved as a priority while planning. Also, a large amount of data is stored to run smart cities. This amount of data and connected smart systems can lower their defences against cyberattacks.

## 20 Financial Issues

Smart city planning and its implementation require professional information technology experts. Large financial resources are needed to plan, implement and successfully maintain smart cities. To be integrated into the smart city, the operation and maintenance of technologies need a budget. Unfortunately, the approval processes of new infrastructure projects take a long time, and these situations force projects with limited budgets. For this reason, this integration and improvement process causes temporary problems for people living in these cities.

Also, some municipal governments need to consider the maintenance costs when making smart city planning decisions and begin the project before anticipating how their demands may evolve. Once smart city projects are implemented, the system can quickly become outdated. For this reason, the infrastructure may need to be upgraded periodically. Smart cities take into account software systems that allow upgrades without the need for hardware replacement based on IoT technologies, which can reduce subsequent costs.

For smart cities to be implemented, awareness of the long-term benefits of smart cities should be created among authorities, planners, investors and other stakeholders.

## 21 Conclusion and Recommendation

With smart cities and smart home systems, the life of individuals has become easier in many ways. But these digital transformations have also brought forth several cybersecurity dangers and difficulties. To safeguard consumers, system developers must securely create these technologies. Chentouf and Bouchkaren argue that blockchain technology, a new technology based on cryptographic principles, can play an important role in ensuring the security of smart cities [68]. In many studies on data protection, it is stated that researchers are only in the early stages of suggesting the solution and that sufficient pilots still need to be conducted to prove the effectiveness of the proposed solutions [69]. Studies on the efficacy of blockchain applications regarding data security, one of the most important problems of smart cities, should be increased as soon as possible.

However, developing business models that can provide financing for public-private partnerships with limited budgets is necessary. City management policies should be reviewed, and more effective efficiency can be achieved by easily cooperating with all stakeholders. Rules created under experts' guidance that will make people feel privacy and security should be determined. Being honest with inhabitants is one of the most crucial requirements for success in smart city applications.

Governments have launched steps to address climate change as urban populations grow yearly. The issue of greenhouse gas emissions, traffic congestion and waste management is getting worse as the population grows. Because of this, local governments should start making plans as soon as possible to construct a smart city infrastructure that includes automated garbage disposal, smart parking, smart buildings and smart traffic lights. Strong collaboration between the public and private sectors is necessary for a smart city to prosper because only a portion of the effort necessary to establish and maintain a data-driven environment is under the purview of local governments. For instance, local government establishments must provide technologies like smart security cameras [35]. Professional data analysts are required for the technology and efficient analysis and assessment of the gathered big data [70]. Promoting the use of electric vehicles in metropolitan areas also helps enhance air quality. Cities may monitor and control air pollution with IoT sensors to lessen health issues from smog and other pollutants.

Urban areas can profit greatly from developing smart connected systems, which increase the quality of life and ensure sustainability and the most efficient use of resources [11].

ICT and digital technologies are used to successfully solve the many operational and development obstacles in smart cities. Cloud computing, 5G, digital twin, data analytics, blockchain, ML and AI and IoT are a few key technologies that have been crucial in tackling some of the main issues with smart cities [70].

One of the ways to reduce anxiety among citizens about smart cities is to educate them. The first step of technology integration is to identify users' perceptions of that technology and turn them into positive perceptions. For this reason, city governments should develop training programmes for these technologies and raise awareness of their long-term benefits. Companies such as IBM, Cisco, Nokia and Huawei create solutions for many challenges by partnering in smart city projects [71].

The success of smart city solutions depends on the collaboration of the private, public and citizen sectors. Additionally, for smart cities, the integration of technology systems like the Internet of Things can be realised, enhancing the quality of life for residents and boosting the rate of urbanisation [72]. Other technologies that contribute to the success of the smart city and the smart future include sensors and power generation systems.

## References


1. Gade, D. S., & Aithal, P. S. (2022). ICT and digital technology based solutions for smart city challenges and opportunities. *International Journal of Applied Engineering and Management Letters*, 1–21. https://doi.org/10.47992/ijaeml.2581.7000.0116
2. Obringer, R., & Nateghi, R. (2021). What makes a city 'smart' in the Anthropocene? A critical review of smart cities under climate change. *Sustainable Cities and Society, 75*, 103278. https://doi.org/10.1016/j.scs.2021.103278
3. Samarakkody, A., & Amaratunga, D. (2022). Characterising smartness to make smart cities resilient. *Sustainability, 14*(19), 12716. https://doi.org/10.3390/su141912716
4. Amol, B., Neeraj, K., Anjuri, A., & Sheetal, P. (2019). Internet of things and big data analysis in smart cities. *Journal of Advanced Research in Dynamical and Control Systems, 11*(8), 2922–2925.
5. O'Dwyer, E., Pan, I., Acha, S., & Shah, N. (2019). Smart energy systems for sustainable smart cities: Current developments, trends and future directions. *Applied Energy, 237*, 581–597. https://doi.org/10.1016/j.apenergy.2019.01.024
6. Djahel, S., Jabeur, N., Barrett, R., & Murphy, J. (2015). Toward V2I communication technology-based solution for reducing road traffic congestion in smart cities. In *2015 international symposium on networks, computers and communications (ISNCC)* (pp. 1–6). https://doi.org/10.1109/ISNCC.2015.7238584
7. Gade, D. S. (2021). Reinventing smart water management system through ICT and IoT driven solution for smart cities. *Technol. Manag. Google Sch. Cit. International Journal of Applied Engineering and Management. Letters A Ref. Int. J. Srinivas Univ. India. Gade, Dipak S, 5*(2), 2581–7000. https://doi.org/10.5281/zenodo.5715852
8. Humayun, M., Jhanjhi, N. Z., Alruwaili, M., Amalathas, S. S., Balasubramanian, V., & Selvaraj, B. (2020). Privacy protection and energy optimization for 5G-aided industrial internet of things. *IEEE Access, 8*, 183665–183677. https://doi.org/10.1109/ACCESS.2020.3028764
9. Gade, D. S., & Aithal, P. S. (2021). Smart city waste management through ICT and IoT driven solution. *International Journal of Applied Engineering and Management Letters, 5*(1), 51–65. https://doi.org/10.5281/zenodo.4739109
10. Razmjoo, A., Østergaard, P. A., Denaï, M., Nezhad, M. M., & Mirjalili, S. (2021). Effective policies to overcome barriers in the development of smart cities. *Energy Research and Social Science, 79*, 1–29. https://doi.org/10.1016/j.erss.2021.102175
11. Caragliu, A., del Bo, C., & Nijkamp, P. (2011). Smart cities in Europe. *Journal of Urban Technology, 18*(2), 65–82. https://doi.org/10.1080/10630732.2011.601117
12. 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. https://doi.org/10.1016/j.eng.2019.11.014
13. Kisseleff, S., Martins, W. A., Al-Hraishawi, H., Chatzinotas, S., & Ottersten, B. (2020). Reconfigurable intelligent surfaces for smart cities: Research challenges and opportunities. *IEEE Open Journal of the Communications Society, 1*, 1781–1797. https://doi.org/10.1109/OJCOMS.2020.3036839
14. Soni, N., & Banarasi, B. (2019). Smart city: A review and analysis of India. *International Journal of Computer Engineering and Applications, X*(VII), 41–46. [Online]. Available: www.ijcea.com.
15. Tiwari, A., & Jain, K. (2014). GIS steering smart future for smart Indian cities. *International Journal of Scientific and Research Publications, 4*(8), 1–5. [Online]. Available: www.ijsrp.org.
16. Yigitcanlar, T. (2015). Smart cities: An effective urban development and management model? *Australian Planner, 52*(1), 27–34.
17. Sánchez-Corcuera, R., et al. (2019). Smart cities survey: Technologies, application domains and challenges for the cities of the future. *International Journal of Distributed Sensor Networks, 15*(6). https://doi.org/10.1177/1550147719853984

18. López-Quiles, J. M., & Rodríguez Bolívar, M. P. (2018). Smart technologies for smart governments: A review of technological tools in smart cities. In *Public administration and information technology* (Vol. 24, pp. 1–18). Springer.
19. Trencher, G. (2019). Towards the smart city 2.0: Empirical evidence of using smartness as a tool for tackling social challenges. *Technological Forecasting and Social Change, 142*, 117–128. https://doi.org/10.1016/j.techfore.2018.07.033
20. Rehman Javed, A., & Bin Zikria, Y. (2021). Future smart cities: Requirements, emerging technologies, applications, challenges, and future aspects internet of multimedia things (IoMT): Opportunities, challenges and solutions view project smart cities view project. https://doi.org/10.36227/techrxiv.14722854.v1
21. Baig, Z. A., et al. (2017). Future challenges for smart cities: Cyber-security and digital forensics. *Digital Investigation, 22*. Elsevier Ltd, 3–13. https://doi.org/10.1016/j.diin.2017.06.015
22. Farahat, I. S., Tolba, A. S., Elhoseny, M., & Eladrosy, W. (2019). Data security and challenges in smart cities. In A. Hassanien, M. Elhoseny, S. Ahmed, A. Singh (Eds.) *Security in smart cities: Models, applications, and challenges*. Lecture Notes in Intelligent Transportation and Infrastructure. Springer, Cham. https://doi.org/10.1007/978-3-030-01560-2_6
23. Hamid, B., Jhanjhi, N., Humayun, M., Khan, A., & Alsayat, A. (2019). Cyber security issues and challenges for smart cities: A survey. In *MACS 2019 - 13th International conference on mathematics, actuarial science, computer science and statistics proceedings*. https://doi.org/10.1109/MACS48846.2019.9024768
24. Naranjo, P. G. V., Pooranian, Z., Shojafar, M., Conti, M., & Buyya, R. (2019). FOCAN: A fog-supported smart city network architecture for management of applications in the Internet of Everything environments. *Journal of Parallel and Distributed Computing, 132*, 274–283. https://doi.org/10.1016/j.jpdc.2018.07.003
25. Kunzmann, K. R., et al. (2019). Smart city environmental pollution prevention and control design based on internet of things. *Cities, 33*(1), 1–6. https://doi.org/10.1088/1755-1315/94/1/012174
26. Carrier, F. (2022). How are smart cities meeting the challenges of urbanization in the 21st century? *Wealth Management*. https://www.rbcwealthmanagement.com/en-asia/insights/how-are-smart-cities-meeting-the-challenges-of-urbanization-in-the-21st-century
27. Kim, J. (2022). Smart city trends: A focus on 5 countries and 15 companies. *Cities, 123*, 103551. https://doi.org/10.1016/j.cities.2021.103551
28. Lee, B. (2019). *ASEAN smart city network (ASCN): Pilot project and smart solution*. Korea Research Institute for Human Settlements (KRIHS) special report. [Online]. Available: https://library.krihs.re.kr/dl_image2/IMG/07/000000030229/SERVICE/000000030229_01.PDF.
29. Panahi Rizi, M. H., & Hosseini Seno, S. A. (2022). A systematic review of technologies and solutions to improve security and privacy protection of citizens in the smart city. *Internet of Things, 20*, 100584. https://doi.org/10.1016/j.iot.2022.100584
30. Smartcitypress. *How smart cities can improve air quality*. https://smartcity.press/smart-cities-air-quality/
31. Khalifa, E. (2019). Smart cities: Opportunities, challenges, and security threats. *Journal of Strategic Innovation and Sustainability, 14*(3). https://doi.org/10.33423/jsis.v14i3.2108
32. Toma, C., Alexandru, A., Popa, M., & Zamfiroiu, A. (2019). IoT solution for smart cities' pollution monitoring and the security challenges. *Sensors, 19*(15), 3401. https://doi.org/10.3390/s19153401
33. Alshamaila, Y., Papagiannidis, S., Alsawalqah, H., & Aljarah, I. (2023). Effective use of smart cities in crisis cases: A systematic review of the literature. *International Journal of Disaster Risk Reduction, 85*, 103521. https://doi.org/10.1016/j.ijdrr.2023.103521
34. Yigitcanlar, T., Han, H., Kamruzzaman, M., Ioppolo, G., & Sabatini-Marques, J. (2019). The making of smart cities: Are Songdo, Masdar, Amsterdam, San Francisco and Brisbane the best we could build? *Land Use Policy, 88*, 104187. https://doi.org/10.1016/j.landusepol.2019.104187

35. Garcia-Retuerta, D., Chamoso, P., Hernández, G., Guzmán, A. S. R., Yigitcanlar, T., & Corchado, J. M. (2021). An efficient management platform for developing smart cities: Solution for real-time and future crowd detection. *Electronics, 10*(7). https://doi.org/10.3390/electronics10070765
36. Evtiukov, S. A., Evtiukov, S. S., & Kurakina, E. V. (2020). Smart transport in road transport infrastructure. In *IOP conference series: Materials science and engineering* (Vol. 832, no. 1, p. 012094). https://doi.org/10.1088/1757-899X/832/1/012094
37. Fortes, S., et al. (2019). The campus as a Smart City: University of Málaga Environmental, learning, and research approaches. *Sensors (Basel), 19*.
38. Singh, H., & Miah, S. J. (2020). Smart education literature: A theoretical analysis. *Education and Information Technologies, 25*(4), 3299–3328. https://doi.org/10.1007/s10639-020-10116-4
39. Koo, C., Park, J., & Lee, J. (2017). Smart tourism: Traveler, business, and organizational perspectives. *Information Management, 54*(6), i. https://doi.org/10.1016/S0378-7206(17)30660-2
40. Khan, M. S., Woo, M., Nam, K., & Chathoth, P. K. (2017). Smart city and smart tourism: A case of Dubai. *Sustainability, 9*(12). https://doi.org/10.3390/su9122279
41. Das, S. (2021). *Smart cities need smart healthcare*. https://www.linkedin.com/pulse/smart-cities-need-healthcare-sanjay-das?trk=public_profile_article_view
42. Antunes, M., & Allen, S. (2022). Smart cities, smarter public health. *Deloitte Insights*. https://www2.deloitte.com/us/en/insights/focus/smart-city/building-a-smart-city-with-smart-digital-health.html
43. Mucha, B. (2021). Evaluation of the state of implementation of the European structural and investment funds: Case study of the Slovak Republic. *On-line Journal Modelling the New Europe, 35*, 4–24. https://doi.org/10.24193/OJMNE.2021.35.01
44. Trellová, L. (2018). *Ústavnoprávne Aspekty Územnej Samosprávy* (Vol. 232). Wolters Kluwer.
45. Sararu, C.-S. (2019). Public domain and private domain. *Administrative Law in Romania*, 84–100.
46. Srebalová, M., & Peráček, T. (2022). Effective public administration as a tool for building smart cities: The experience of the Slovak Republic. *Laws, 11*(5), 67. https://doi.org/10.3390/laws11050067
47. Badran, A. (2021). Developing smart cities: Regulatory and policy implications for the State of Qatar. *International Journal of Public Administration, 46*, 519. https://doi.org/10.1080/01900692.2021.2003811
48. Stratigea, A., Papadopoulou, C. A., & Panagiotopoulou, M. (2015). Tools and technologies for planning the development of smart cities. *Journal of Urban Technology, 22*(2), 43–62. https://doi.org/10.1080/10630732.2015.1018725
49. Abellá-García, A., Ortiz-De-Urbina-Criado, M., & De-Pablos-Heredero, C. (2015). The ecosystem of services around smart cities: An exploratory analysis. *Procedia Computer Science, 64*, 1075–1080. https://doi.org/10.1016/j.procs.2015.08.554
50. Sharida, A., Hamdan, A., & AL-Hashimi, M. (2020). *Smart cities: The next urban evolution in delivering a better quality of life* (pp. 287–298).
51. Xiong, Z., Sheng, H., Rong, W., & Cooper, D. E. (2012). Intelligent transportation systems for smart cities: A progress review. *Science China Information Sciences, 55*(12), 2908–2914. https://doi.org/10.1007/s11432-012-4725-1
52. Masera, M., Bompard, E. F., Profumo, F., & Hadjsaid, N. (2018). Smart (electricity) grids for smart cities: Assessing roles and societal impacts. In *Proceedings of the IEEE* (Vol. 106, no. 4, pp. 613–625). Institute of Electrical and Electronics Engineers Inc.. https://doi.org/10.1109/JPROC.2018.2812212
53. Wang, P., Ali, A., & Kelly, W. (2015). Data security and threat modeling for smart city infrastructure. In *2015 international conference on cyber security of smart cities, industrial control system and communications (SSIC)* (pp. 1–6). https://doi.org/10.1109/SSIC.2015.7245322

54. Karimi, Y., Haghi Kashani, M., Akbari, M., & Mahdipour, E. (2021). Leveraging big data in smart cities: A systematic review. *Concurrency and Computation: Practice and Experience, 33*(21). https://doi.org/10.1002/cpe.6379
55. New, J., Castro, D., & Beckwith, M. (2017). *How national governments can help smart cities succeed*. [Online]. Available: https://www2.datainnovation.org/2017-national-governments-smart-cities.pdf
56. Mahizhnan, A. (1999). Smart cities the Singapore case. *Cities, 16*(1), 13–18. [Online]. Available: www.moe.
57. Nathan, D. (2022). Privacy risks in smart cities – What you need to know. *VPNOverview*. https://vpnoverview.com/privacy/devices/privacy-risks-smart-cities/
58. Al-Turjman, F., Zahmatkesh, H., & Shahroze, R. (2022). An overview of security and privacy in smart cities' IoT communications. *Transactions on Emerging Telecommunications Technologies, 33*(3). https://doi.org/10.1002/ett.3677
59. Ahmadi-Assalemi, G., Al-Khateeb, H., Epiphaniou, G., & Maple, C. (2020). Cyber resilience and incident response in smart cities: A systematic literature review. *Smart Cities, 3*(3) MDPI, 894–927. https://doi.org/10.3390/smartcities3030046
60. Ouafiq, E. M., Raif, M., Chehri, A., & Saadane, R. (2022). Data architecture and big data analytics in smart cities. *Procedia Computer Science, 207*, 4123–4131. https://doi.org/10.1016/j.procs.2022.09.475
61. Mohammed, I. A. (2020). Security, privacy and risks within smart cities: Literature review and development of a smart city interaction framework. *International Journal of Creative Research Thoughts, 8*(1), 2320–2882. [Online]. Available: www.ijcrt.org
62. Habib, A., Alsmadi, D., & Prybutok, V. R. (2020). Factors that determine residents' acceptance of smart city technologies. *Behaviour & Information Technology, 39*(6), 610–623. https://doi.org/10.1080/0144929X.2019.1693629
63. Lake, J. (2019). Smart cities, cybersecurity and privacy: What are the risks? *Comparitech*. https://www.comparitech.com/blog/vpn-privacy/smart-cities-privacy-risks/
64. Valle, G. *9 Challenges every smart city faces*. https://www.builderspace.com/9-challenges-every-smart-city-faces
65. Yaqoob, I., Hashem, I. A. T., Mehmood, Y., Gani, A., Mokhtar, S., & Guizani, S. (2017). Enabling communication technologies for smart cities. *IEEE Communications Magazine, 55*(1), 112–120. https://doi.org/10.1109/MCOM.2017.1600232CM
66. Keshvardoost, S., Renukappa, S., & Suresh, S. Developments of policies related to smart cities: A critical review. In *2018 IEEE/ACM international conference on utility and cloud computing companion* (Vol. 2018, pp. 370–375).
67. UBIDOTS. *Key challenges of smart cities & how to overcome them*. https://ubidots.com/blog/the-key-challenges-for-smart-cities/
68. Nexusintegra. *How to confront smart cities challenges title*. https://nexusintegra.io/smart-city-challenges/
69. Siddiquee, S. M. S., Khan, M. M. H., Al-Ismail, F. S., Ullah, A., Alam, M. S., & Ahmed, H. (2022). Blockchain applications in smart sustainable city context—A systematic mapping study. *Energy Reports, 8*, 162–169. https://doi.org/10.1016/j.egyr.2022.10.102
70. Li, X., Liu, H., Wang, W., Zheng, Y., Lv, H., & Lv, Z. (2022). Big data analysis of the Internet of Things in the digital twins of smart city based on deep learning. *Future Generation Computer Systems, 128*, 167–177. https://doi.org/10.1016/j.future.2021.10.006
71. Akhtar, N., & Hasley, K. (2018). Smart cities face challenges and opportunities. *computerweekly*. https://www.computerweekly.com/opinion/Smart-cities-face-challenges-and-opportunities.
72. Anttiroiko, A. V., Valkama, P., & Bailey, S. J. (2014). Smart cities in the new service economy: Building platforms for smart services. *AI & Society, 29*(3), 323–334. https://doi.org/10.1007/s00146-013-0464-0

# Smart City: Transformation to a Digital City

**Pankaj P. Tasgaonkar, R. D. Garg, P. K. Garg, and Kavach Mishra**

## Abbreviations

| | |
|---|---|
| AI | Artificial intelligence |
| API | Application programming interface |
| ARM | Advanced RISC Machine |
| CCTV | Closed-circuit television |
| GPRS | General Packet Radio Service |
| GPS | Global Positioning System |
| GSM | Global System for Mobile Communications |
| HAN | Home area network |
| HDMI | High-Definition Multimedia Interface |
| IEEE | Institute of Electrical and Electronics Engineers |
| IoT | Internet of Things |
| IR | Infrared radiation |
| LCD | Liquid-crystal display |
| NAN | Neighborhood area network |
| NFC | Near-field communication |
| PIR | Passive infrared |

P. P. Tasgaonkar (✉)
Electronics and Telecommunication Engineering Department, COEP Technological University, Pune, India

Geomatics Engineering Group, Civil Engineering Department, Indian Institute of Technology Roorkee, Roorkee, India
e-mail: ppt.extc@coep.ac.in; pankajtgr@ce.iitr.ac.in

R. D. Garg · P. K. Garg · K. Mishra
Geomatics Engineering Group, Civil Engineering Department, Indian Institute of Technology Roorkee, Roorkee, India



N. Gupta, S. Mishra (eds.), *Internet of Everything for Smart City and Smart Healthcare Applications*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-34601-9_7

| | |
|---|---|
| PLC | Power-line communication |
| RFID | Radio-frequency identification |
| SD | Secure Digital |
| TDS | Total dissolved solids |
| UART | Universal asynchronous receiver-transmitter |
| UAV | Unmanned aerial vehicle |
| USB | Universal Serial Bus |
| WAN | Wide-area network |
| Wi-Fi | Wireless fidelity |

## 1 Introduction

Smart city is the city which has become a dream come true for the residents who desire an improvised lifestyle. The smart city is also called an information city which has further transformed into a digitized world. The inhabitants of the city flourish to nurture the natural as well as the created resources. The infrastructure is built in such a way that sensors collect the data periodically and process it which in turn gives results that have a vital role to play in the application.

The Internet of Things (IoT) which was just a buzzword a few years ago has become a part and parcel of life. It does not only give information but is also useful in entertainment also called infotainment. The smart city caters to maintaining a balance between demand and supply. The smart city will always have a check on sufficient power supply, transportation, water management, waste management, health monitoring, etc. Smart city converges all the technologies together to minimize the overhead of human efforts. It has sensors, actuators, communication modules, protocols, cloud computing, big data, etc. All these working together makes the application smarter [1]. Hence, a smart city has sub-parts such as smart transportation, smart energy, smart water supply, smart waste management, smart environment, smart healthcare, etc.

Data is collected, processed, stored, and analyzed with cloud computing techniques. Artificial intelligence and machine learning tools are beneficial and can predict future results based on past data. It is not a one-way system; it considers all the stakeholders of the society ranging from the local bodies like municipal corporations to the state and central governments, which allocate a budget based on the criteria that can be fulfilled by that city. The people are the consumers who benefit by getting improvised and faster end products.

Section 1 introduces the terminology of a smart city. Section 2 presents the features of the IoT for the smart city. This is followed by Sect. 3 on the objectives of the smart city. Section 4 describes smart transportation. Section 5 talks about the smart grid. Section 6 discusses smart water supply management. Section 7 elaborates the smart waste management. Section 8 describes the smart environment. The chapter concludes with a summary of the transformation to a digital city in Sect. 9.

## 2 Components of the IoT for the Smart City

The components of IoT can be classified as sensor, microcontrollers, communication module, cloud, and user interface. These components are shown in Fig. 1. All these components working collaboratively will enhance the system of the smart city. The sensors are automatized which have a faster response time to work as per the application of the smart city.

Sensors: There are various types of sensors, physical, chemical, and biological. The sensors' function is to collect data from the environment. Examples of physical sensors are accelerometers, ultrasonic sensors, RFID, GPS, etc. The chemical sensors are MQ sensors, while biological sensors are soil moisture sensors, heart rate sensors, etc. Table 1 indicates the working principle of the sensors and their application.

There are various types of sensors, but a few are listed below in Table 1.

Microcontrollers: Arduino Uno has ATmega328P microcontroller. It operates on 5 V. It has 6 analog input pins and 14 digital I/O pins. Raspberry Pi has an ARM processor. It has an HDMI socket and a USB 2.0 socket. It uses UART and SD card. It is useful in applications like home automation systems. The sensors are interfaced with the microcontroller for data processing.

Communication module: GSM/GPRS has an inbuilt SIM slot which acts as a mobile phone which can send and receive messages. The messages are the data of the sensors or any alerts that are generated. ESP8266 Wi-Fi module provides proper Internet connectivity which forwards the data to the cloud. An Ethernet shield is inserted on the Arduino board which provides a port address to the sensors. Bluetooth and Zigbee are also used for short-range communications. Zigbee is a

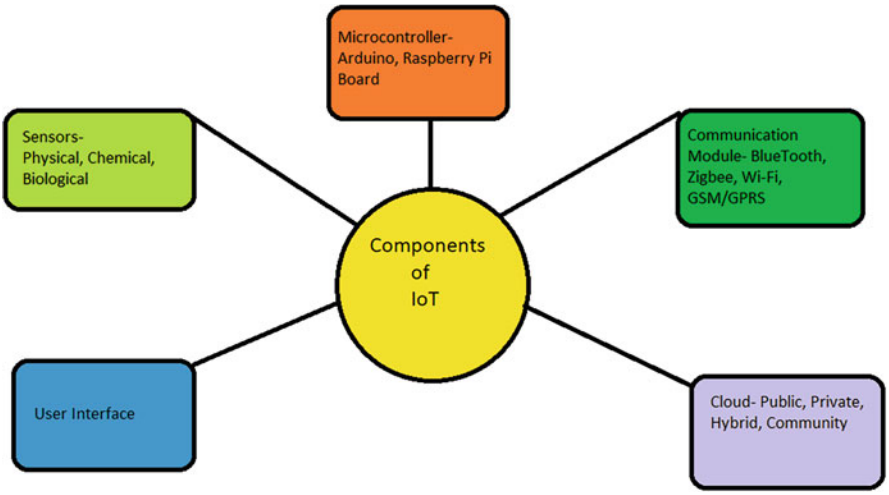



**Fig. 1** Components of IoT. (Source: Authors)

**Table 1** Working principle of sensors and their applications

| Sensor | Working principle | Applications |
|---|---|---|
| Accelerometer | A sudden force moves the ions which change the displacement | Vibration, shock analysis |
| Gyroscope | The spinning wheel rotates as per the axis of rotation | Navigation |
| PIR sensor | Change in the values of the electrodes | Motion detector |
| Temperature and humidity sensor | Negative temperature coefficient | Environment monitoring |
| Ultrasonic sensor | Reflection of the sound waves | Distance measurement |
| RFID | Tags and reader communicate through RF signal | Proximity detection |
| GPS | Triangulation of the satellites | Real-time tracking |
| MQ sensor | Change in the chemiresistors | Gas, smoke detection |
| Soil moisture sensor | Dielectric permittivity of the adjacent area | Detect soil moisture |
| Heartbeat sensors | Change in the volume of blood through any organ of the body | Calculate heart rate |
| Pulse oximeter | As per the difference in the absorption spectrum of HB | Measures oxygen level |

communication technology developed according to the IEEE 802.15.4 standard that operates on 2.4 GHz frequency. Bluetooth has a communication range of 10–100 m. The Z-Wave technology operates on a 900 MHz band and transmits data rate at 40 kbps.

Cloud computing: There are public clouds which are freely available, private clouds used by government organizations, hybrid clouds where some features of public and private clouds are merged, and community clouds where different organizations share the data on various clouds. Data is stored, processed, and analyzed on the cloud.

User interface: It is a dashboard provided to the monitoring station or end users where all the data are being interfaced between the sensors and the cloud. A proper API will make connectivity between the host and the entity. The programming is done with Visual Basic, Python, JAVA, etc.

## 3 Objectives of the Smart City

The main objective is to provide adequate resources given the increasing lifestyle of the habitants. It has aim to have ecological balance in the operation of the city. Irrespective of the environmental attributes, the smart city set a proper benchmark for human beings. A planned city will be a role model and always attract tourists. In transportation, it will offer a safe and reliable journey. The other objectives are conserving energy and utilizing efficiently. Along with that, water management will check for unnecessary wastage of water and proper utilization by the residents. Frequent checking up of the garbage bins and dumping them in a proper place is required to keep the environment clean and hygienic. It considers the citizen data which is vital in decision-making as data is being collected through a questionnaire or feedback system. It identifies the traits and upgrades of the execution plans. A transparent system is where consumers will have all the information resources and can give feedback which can improve the system. Academicians, researchers, industrialists, and government officials coming together will always build new ideas and innovations for the upliftment of the smart city.

Economy rises gradually with more job opportunities through the smart city project and cater for the resources which will be abundant for the increasing population. Accessibility with less delay is one of the prime objectives of the smart city. Sustainability offers reliable and efficient services with reduced operational costs [2].

London, San Francisco, and Barcelona are some of the top smart cities of the world which focus on smart transportation, smart energy, smart community, smart healthcare, etc. [3].

## 4 Smart Transportation

The transportation sector carries goods and services from one place to another. It considers all modes like road, rail, air, sea, etc. Sensors attached to the vehicles or placed along the path will forward the data to the monitoring station. It enhances the journey and safety of public transport.

A smart card used in multimodal transport helps in taking the exact ticket for the journey. The fare is debited from the cardholder for every travel accordingly. The card uses NFC which can be scanned by the reader. It also resolves the issues of change at every counter, and the need of waiting in a long queue is diminished. A single card will serve the purpose of train, metro, monorail, bus, auto, and cabs.

Smart parking in a commercial complex is easier with the help of RFID tags and readers. The data is stored in the database management system where it can keep a track of vehicle check-in and check-out. It will assign the fare as per the time taken for the occupancy of the parking area. IR sensors will identify whether the exact vehicle is there in the slot or the area is empty.

CCTV is installed usually at the crosswalk in the city. The surveillance through video camera can be beneficial as it checks whether the vehicles are ahead of the zebra crossing, and also it detects whether the drivers follow the traffic light or not. Video cameras installed at the toll plazas can capture footage of the vehicles passing through that point.

Smart traffic lights synchronized properly can operate as per the density of the vehicles on the road. Inductive loops measure the traffic on the road, while magnetometers classify the types of vehicles on the road.

An accident information system is provided through the data collected on the cloud. Accelerometers and gyroscopes will give the proper orientation of the vehicle. If any crash or dash happens with the vehicle, an immediate alert is given to emergency services which can save a human life. On highways, due to oil spill or any reason if an accident happens, the vehicles will get slipped, and many more accidents will happen; to avoid such circumstances, the information is being displayed on the LCD board on the highway so that drivers can change the lane at a proper interval.

Fleet management keeps a track of goods and logistics. It will check whether the products reach their destination properly on time. Tracking of the vehicles is done through GPS which continuously tracks the vehicles on the smartphone application [4]. The database keeps the information about all the products, containers, and travel time of the journey.

Pedestrian walkways, cycle routes, skywalks, and subways at the proper place are constructed with proper signalling mechanisms which help commuters to reach safely to their destination. Ring roads and multilane expressways have made things easier as despite longer distances, less time is required for travelling purposes.

RFID tags installed on the vehicles are scanned through the reader at the toll plaza for deposition of the money. The long queue of cash and pay is almost eliminated. Optical fiber cables are laid through the highway, or cellular antennas are installed so that constant Internet connectivity is maintained between the sensors and the gateway.

## 5 Smart Grid

Electricity is the backbone of all the sectors of industries. It integrates electrical, wireless technologies through IoT. It has sensors throughout the system. Electrical units are classified as generation, distribution, and transmission. The characteristics of a smart grid are consumer feedback, reliable and continuous energy to end users, it adds all the electricity generation options like hydro, thermal, gas, wind, etc.

Smart grid will be beneficial in the smart city as it will reduce carbon footprints and provide green energy in a sustainable environment. It can detect any fault immediately and take the necessary action by restoring the energy. High voltage sensor checks the exact voltage on the overhead transmission lines. Voltage sag sensors will detect the rise in the current due to the transformers, motors, etc.

Microcontrollers process the data that is collected from the sensors and process it accordingly. It processes the smart meter reading, tamper detection, and electrical parameters. It has communication through Wi-Fi, Bluetooth, Zigbee, Z-Wave, and PLC. The smart meter provides the monthly, weekly, daily, and hourly consumption of electricity at home. A power management application is installed on the mobile phone of the consumer which has a proper check of automated devices attached to it. The tariff is generated as per the power consumption. The data acquisition unit forwards the collected data from the smart meter to the management system.

An area network is formed like HAN, NAN, and WAN whenever electrical appliances are interconnected as per their geographical region. Smart grid incorporates sensor fusion, cybersecurity techniques, communication networks, distributed energy resources and distribution grid management.

## 6 Smart Water Management

This application is also one of the vital features of the smart city. It distributes, controls, and manages the water to the city in the appropriate proportion. Water quality can be measured through PH, conductivity, and total dissolved solids. Analyzer is used to monitor various parameters such as PH, conductivity, temperature, salinity, etc. The system also uses a TI controller for storing the data and then putting this data on the cloud.

The ways for water storage and management change a lot of things in water organizing. If the water is stagnant like in reservoirs, tanks, etc., then the water pressure and its level are to be taken into consideration. In the case of flowing water, as in rivers, the velocity upstream or downstream is recorded for further decisions related to its management solutions. These parameters can be easily gauged using water flow sensors and ultrasonic sensors. Ultrasonic sensors can determine the water level at what length it is being filled in a particular tank. Whenever sensor detects it relays the information to the gateway which in turn shuts down the supply to avoid overflow of the water.

A smart water management system examines the water quantity as well as the quality before reaching the end users. The quality of water plays a vital role in deciding the management technique to be employed for it. The quality chemical factors that can influence this decision include pH levels of the water that can determine its acidity or alkalinity. A balanced pH of 6–9 on a scale of 14 is usually considered to be normal in sub-tropical regions. The assimilated oxygen, nitrogen, and oxidation-reduction properties are also required to be gauged for the identification of its overall nature and response. The capability of water to eliminate the impurities by itself is termed as oxidation-reduction potential. A high value of oxidation-reduction potential signifies a decent quality of water.

The physical factors comprise the temperature, turbidity, and conduction of water. Turbidity is the opacity or cloudiness due to the infinitesimal constituents dissolved in a solution. High temperature depletes the quantity of oxygen in the water which disrupts the quality of water. The water-protecting supplements and pesticides further affect the quality of water reservoirs. A multi-level filter system will be quite efficient which will purify the water coming through the rivers and reservoirs. It removes all unwanted materials like dust, sand, extra chlorine, etc. and makes it potable for drinking purposes.

Sensors attached to the taps can detect the unwanted waste of water. In house, the tap is continuously flowing even when people are not at home, in that case, it will detect and shut the flow of water for that particular house.

Pressure sensors will maintain a proper force of water. The system is beneficial for equal distribution of the water supply.

## 7 Smart Waste Management

Smart waste management emphasizes on sensors which enable trash and recycles bins to improve the efficiency of refuse waste. Hence, the appearance of waste bins in public places improves tremendously. Smart waste management belongs to the movement of smart cities where technologies, data, and analytics are used to enhance the services and qualities of the city. Figure 2 shows the working mechanism of a smart waste management system.

Smart dust bins: If a vehicle is coming daily to collect the waste but it may be possible that bin is empty then all the efforts that have been employed are wasted which makes this method inefficient. So we need a smart system which minimizes unnecessary trips. This can be achieved by using smart-level sensors in bins or a dumpster of any size. This smart device can collect and store data depending on fill levels; by this, it is easy to predict how often bins need to be collected or emptied. Ultrasonic and IR sensors can act as a level sensor which identifies whether the bill

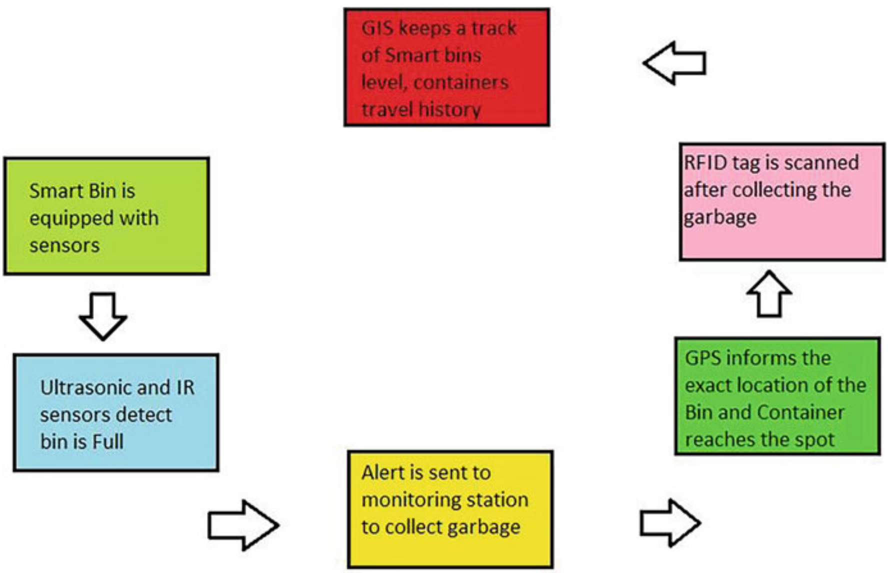



**Fig. 2** Smart waste management. (Source: Authors)

is empty or filled. With the level detector, it can inform the control station that it can collect the garbage from that place.

Smart garbage containers: Geographical information system will play an important role in all the garbage containers of the city. It will track the vehicles are in transit and whether the waste is collected and dumped can be verified. An RFID tag is attached to the smart bins which the reader scans at that location which identifies from which place the waste has been collected.

Smart weighing machines: The weighing machine will check the tons of garbage filled in the container, to avoid overweighing and unnecessary spilling out of the waste on the roads during travel.

E-waste magnetic sensors can filter out e-waste, like all electronic equipment, TV, camera, mobile phones, battery, etc. The battery should be separated since it has toxic substances like lithium which can be harmful and sometimes it may explode also. The e-wastes may be recycled and scaled up to the consumable electronics.

## 8 Smart Environment

An environment devoid of any sort of contaminant is the need of the hour to ensure a healthy and hygienic society. Only such a society forms the cornerstone of the sustainability of the human race and the growth of any nation. In this context, continuous monitoring of air and water pollutants, weather conditions, seismic events, and toxic emissions has become essential. Rapid technological developments in the fields of electronics and artificial intelligence (AI) have facilitated the production of smart environmental monitoring systems. These systems have become an indispensable part of smart city design, which is based on wireless sensor networks [5].

Modern sensors working on AI-based tracking and guidance mechanisms form the core of these networks. These along with IoT instruments are the main components of such smart environmental monitoring systems. Wireless sensor networks integrate the data acquired by various sensors and IoT instruments such as information on water and air quality, temperature, wind speed, etc.

An example of a smart environmental monitoring system can be a cloud-based water contamination monitoring and control system. It consists of smart aqua sensors deployed at various locations. They can distinguish between clean and contaminated water as they have undergone training and testing by a machine learning algorithm. The agency entrusted with the responsibility of maintaining the cleanliness of water bodies and providing safe drinking water to the citizens has access to the cloud, which contains the data collected by such sensors. Through an IoT-based analysis, it can be understood whether the water at various locations is clean or contaminated.

Another instance of a smart environmental monitoring system is a cloud-based air quality and weather monitoring system. It comprises various sensors for recording humidity, temperature, wind speed, the concentration of ultraviolet and infrared rays, and traces of pollutant gases, dust, and aerosols. These measurements are relayed to the cloud via wireless sensor networks. Then, analysts study these data, run an IoT-based analysis, and display suitable information to the public, with the possibility to update this information in real time. This information includes current humidity, wind speed, minimum and maximum temperatures, the concentration of pollutants, dust, and aerosols, and advisories such as the time duration to step out in the sun, orange or red alerts, etc.

A constellation of satellites can also be deployed to provide real-time information in the form of images for the surface of Earth or vertical profiles for meteorological variables. However, these images or profiles happen to be at a coarse spatial resolution even though the temporal resolution is very high. When coupled with a web-based interface, which can be updated in real time based on information relayed from the cloud, an efficient web-based information system can be realized. Recently, unmanned aerial vehicles or drones have become an alternative to provide high spatial resolution data for studying any Earth-based phenomenon. However, they are restricted only to smaller areas due to economic reasons. Several sensors can be mounted on a drone as per the user's requirement.

Various types of drones exist: single-rotor, multi-rotor, fixed-wing, and fixed-wing hybrid [6]. Their advantages, disadvantages, and applications are listed in Table 2.

Multi-rotor drone is the most popular out of the four and is used commonly. Depending on the number of rotors present, a multi-rotor drone can be a tricopter (three rotors), quadcopter (four rotors), hexacopter (six rotors), and octocopter (eight rotors). A quadcopter (DJI Matrice 100) available at the Geomatics Engineering Laboratory of the Indian Institute of Technology Roorkee is shown in Fig. 3. It is mounted with a multispectral camera, MicaSense RedEdge-MX, which can acquire images of the scene in five spectral bands in the visible and near-infrared regions of the electromagnetic spectrum [7].

Essential parts of a quadcopter are motor, propellers, flight controller, GPS module, electronic speed controller, power port module, sensors avoiding obstacles, three-axis gimbal, camera, battery, antennas, flight light-emitting diode, joysticks, main remote control board, and main camera board [8]. The functions of each of these parts are mentioned herein:

Motor: There are two clockwise and two anticlockwise motors, which provide stability by equalizing the turning force.

Propeller: There are propellers, one for each motor direction. It enables the drone to move up in the air by creating a pressure difference in the area above and below it through rotatory motion.

Flight controller: Inputs are fed into the flight controller from the compass, remote controller, GPS module, and sensors that avoid obstacles, which are relayed to the electronic speed controller for controlling the motors.

GPS module: This module utilizes signals from various global navigation satellite systems to determine the drone's position as well as the positional information of the scene being captured by the drone's camera.

**Table 2** Types of unmanned aerial vehicles: advantages, disadvantages, and uses

| Unmanned aerial vehicle type | Advantage | Disadvantage | Application |
|---|---|---|---|
| Single-rotor | Vertical take-off and landing<br>Hover flight<br>Possibility to mount a heavier sensor | Dangerous<br>More training is needed to fly<br>Expensive | Aerial laser scanning |
| Multi-rotor | Vertical take-off and landing<br>Hover flight<br>Easy to use and accessible<br>Efficient camera control<br>Operation in a closed area | Shorter flying time<br>Only a lightweight sensor is possible | Aerial photographs and video capture |
| Fixed-wing | Possible to cover a large area<br>Higher flight speed | Large space needed for launch and recovery<br>No vertical take-off and landing or hover flight<br>More training is needed to fly<br>Very costly | Aerial mapping |
| Fixed-wing hybrid | Vertical take-off and landing<br>Can have longer flying times | Still in the development stage<br>Does not have hovering capability | Drone delivery |

**Fig. 3** Unmanned aerial vehicle fitted with a multispectral camera and its controller in white. (Source: Authors)

Electronic speed controller: It is connected to the battery and the flight controller, receiving signals from the latter to modify the power supply to each motor.
Power port module: This module determines the power supply from the battery and its distribution to the electronic speed and flight controllers.
Sensors avoiding obstacles: These sensors provide stereo-vision capability to the drone, determining its distance from the object being imaged.
Three-axis gimbal: It enables the video acquired by the drone to remain stable and has no errors due to changes in roll, pitch, and yaw.
Camera: It is an imaging sensor, which captures the light reflected from the object and processes it into a digital image.
Battery: It is responsible for supplying power to various other parts of the drone during its flight campaign.
Antennas: They pass on information from the drone to the controller and vice versa. Found in the drone's legs, they are also accompanied by a compass which provides information on directions to the controller.
Flight light-emitting diode: These light-emitting diodes are green when the drone's back is in front of the user and are red when the drone's front faces the user.
Joysticks: They facilitate the flying movements of the drone through the main remote control board, being embedded in it.
Main remote control board: The board records information from the drone about its position and elevation and the scene being captured. It also sends flying directions to the flight controller through joysticks.
Main camera board: It converts the scene information acquired by the camera sensor into digital numbers, recording it onto the micro SD card.

Numerous cases of drones being components of smart environmental monitoring systems exist. The forest fire alert system which consists of activities like burnt area estimation due to forest fire, estimating the concentration of harmful fumes, and issuing alerts for possible future forest fire occurrences; crop insurance system which includes crop yield estimation according to acreage and dissemination of extreme weather events; and infrastructure health monitoring systems focused on rail-road infrastructure, etc.

## 9 Summary

The attributes of the smart city are based on all the stakeholders of society. The purpose will get fulfilled with proper implementation of the plans and the necessary budget allocation. The Internet of Things has changed the perspective of the smart city. It has transformed into a digitized world where all the data is available on a single platform. High-speed Internet connectivity will be quite efficient for real-time applications. The enhancement of the city will increase the economic growth of the country.

**Acknowledgments** The authors would like to thank the anonymous reviewers for their comments and suggestions, which fine-tuned the manuscript's quality. We would also like to affirm that no funding has been received for performing this research. Moreover, we declare no competing interests or personal relationship, which would impede the publication of this work.

## References

1. Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications, 6*(1), 1–15.
2. Bakıcı, T., Almirall, E., & Wareham, J. (2013). A smart city initiative: The case of Barcelona. *Journal of the Knowledge Economy, 4*, 135–148.
3. Silva, B. N., Khan, M., & Han, K. (2018). Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities. *Sustainable Cities and Society, 38*, 697–713.
4. Kim, T. H., Ramos, C., & Mohammed, S. (2017). Smart city and IoT. *Future Generation Computer Systems, 76*, 159–162.
5. Ullo, S. L., & Sinha, G. R. (2020). Advances in smart environment monitoring systems using IoT and sensors. *Sensors, 20*(11), 3113.
6. https://www.auav.com.au/articles/drone-types/. Accessed on 7 Feb 2023.
7. https://uavgarage.com/product/micasense-rededge-mx-multispectral-camera-kit/. Accessed on 7 Feb 2023.
8. https://www.dronefly.com/the-anatomy-of-a-drone/. Accessed on 7 Feb 2023.

# Bi-objective Study of Public Transport Operation in Smart Cities to Minimize On-Board Passenger Traveling Time and Stop Passenger Delay

**Yi Zhang and Anuj Abraham**

## Abbreviations

| | |
|---|---|
| EA | Evolutionary algorithm |
| GA | Genetic algorithm |
| HMCR | Harmony memory considering rate |
| HS | Harmony search |
| MINLP | Mixed-integer nonlinear programming |
| NSGA-II | Non-dominated sorting genetic algorithm II |
| NSHS | Non-dominated sorting harmony search |
| OD | Origin-destination |
| PAR | Pitch adjustment rate |

## 1 Introduction

Public transport buses are significantly affected by traffic congestion, roadwork, and accidents in the same way as other vehicles on the road network [1]. The implications for passengers at the bus stops and on-board passenger delays of buses are major and are varied from person to person. Bus passengers expect that bus service providers could do more to reduce delays by focusing on those

Y. Zhang
Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore
e-mail: zhang_yi@i2r.a-star.edu.sg

A. Abraham (✉)
Technology Innovation Institute, Abu Dhabi, United Arab Emirates
e-mail: Anuj.Abraham@tii.ae



N. Gupta, S. Mishra (eds.), *Internet of Everything for Smart City and Smart Healthcare Applications*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-34601-9_8

operational tactics anticipated to be within their control. However, the complex road environments make it harder for operators to maintain high bus reliability and thereby cause discomfort to both on-board and waiting passengers [2]. In view of this, many real-time operation methods have been studied in the past several decades to avoid bus bunching as well as minimize passenger delay resulting from fluctuated passenger demands and uncertain en route traffic [3]. The typical approaches to apply control at stops are holding strategies and stop-skipping strategies. The combined holding and stop-skipping strategy are proposed by Eberlein et al., which leads to a roughly 37% reduction compared to any of the single strategies [4]. Different with stop-skipping strategies, a holding strategy is more common due to its easy implementation, and it relatively depresses passengers less than stop-skipping, as passengers may fail to alight bus due to the bus stop-skipping [5]. Besides that, boarding limits at each stop are also considered in some literature [6–8]; however, this strategy is hard to be implemented in real practice. On the other hand, inter-stop control has also been investigated: Bus speed control is applied to minimize the headway variation in time-based model [9, 10]. Traffic signal priority [11, 12] is also implemented to provide priority to on-road buses. For the model formulated in this chapter, we combine both holding control at stops and bus speed control between stops, to form a trip-based mixed-integer nonlinear programming (MINLP) problem by incorporating bus capacity constraints, loading time constraints, volume dynamics, and bus movement dynamics. As for the objective function, many mathematical models have been developed to emphasize the stop passenger delay instead of the on-board passenger delay. Chuanjiao et al. [13] have proposed a mathematical model to minimize stop passenger delays by adjusting scheduling types. Also, a multi-bus dispatching strategy is proposed by applying holding and boarding control to minimize stop passenger delay and bus vacancy [7, 8]. For limited models considering on-board passenger delays, the weighted sum method is applied to form a single-objective function, which makes the determination of proper weights difficult [6, 14]. Also, their on-board passenger delay is defined only for the holding process without taking the traveling time into account. A bi-objective problem has been recently carried out to investigate the passenger waiting time and bus occupancy rate with the aim to satisfy requirements from both the users' and operations' perspectives [15]. However, the bi-objective studies on the conflicts between the on-board passenger delay and the stop passenger delay have seldom been discussed and shall be introduced in this chapter. In this chapter, a bus line operating system considering bus and stop volume dynamics is formulated, with the goal to reduce either the total stop passenger delay or the on-board passenger traveling time. The bus operation system incorporates both holding control and bus speed control to flexibly manage the bus movement, thereby maximizing the passengers' benefit to a large extent. Secondly, an optimization model is proposed and transformed into the mixed-integer nonlinear programming (MINLP) problem. The final results are obtained by adopting the traditional commercial solver, and numerical experiments are presented to analyze the corresponding bus movements under different objectives: minimize either the stop passenger delay or the on-board passenger delay. Furthermore, in view of the conflicts between the two types of delays, non-dominated sorting genetic algorithm II (NSGA-II) and non-dominated sorting harmony search (NSHS) algorithm, two multi-objective algorithms, are adopted to generate the Pareto graph aiming to minimize both the stop passenger delay and on-board passenger traveling time. This chapter is organized as follows. The specific description of the bus operation system is illustrated in Sect. 2. Solution algorithms for MINLP conversion, NSGA-II and NSHS, are described in Sect. 3. Finally, case studies are drawn in Sect. 4. Conclusions are summarized in Sect. 5.

## 2 Formulation of a Bus Operation System

### 2.1 Problem Statement

We define the bus line as a directed graph $G = \{S, L\}$ where $S$ is the set of bus stops and $L$ is the road segment between two connected bus stops, as shown in Fig. 1. The bus line is composed of the starting terminal, the intermediate bus stops, and the ending terminal, where the starting terminal only dispatches buses based on the specified timetable and the ending terminal can be regarded as a sink node to absorb the dispatched buses. In this chapter, we only discuss the operation side of the bus system, while the bus dispatching time from the starting terminal is assumed to be known in advance. We consider two types of the control in the system: the holding control at each bus stop and the bus traveling speed between two connected stops; also, the acceleration and deceleration of the buses are neglected for simplicity in the current formulation.

### 2.2 Model Description

#### 2.2.1 Stop Volume Dynamics

The update of the volume at bus stop $i$ with destination stop $j$ (origin-destination (OD) pair $ij$) is determined by the multiplication of the incoming passenger flow rate $f_{i,j}(k)$ and trip gap time $\mathrm{AR}_i(k+1) - \mathrm{AR}_i(k)$ and the boarding flow $B_{i,j}(k)$:

$$P_{i,j}(k+1) = P_{i,j}(k) + (\mathrm{AR}_i(k+1) - \mathrm{AR}_i(k))\, f_{i,j}(k) - B_{i,j}(k) \tag{1a}$$

$$P_{i,j}(1) = P_{i,j}(0) + (\mathrm{AR}_i(1) - t_0)\, f_{i,j}(0) \tag{1b}$$


**Fig. 1** The bus line of the proposed model

where $i$ and $j$ are the index of the bus stop, $P_{i,j}(k)$ is the number of passengers for OD pair $ij$ when the bus from trip $k$ is coming, and $\mathrm{AR}_i(k)$ is the arrival time at stop $i$ of bus dispatched at trip $k$. $f_{i,j}(k)$ represents the passenger arrival rate of trip interval $k$. $B_{i,j}(k)$ is the boarding passengers under OD pairs $ij$ of bus dispatched at trip $k$. $t_0$ is the initial time of the whole operation system.

### 2.2.2 Bus Volume Dynamics

The update of the total bus volume at stop $i$ for trip $k$ is determined by the total boarding volume $\sum_{j \in S} B_{i,j}(k)$ and the total alighting volume $A_i(k)$ at stop $i$:

$$\sum_{j \in S} V_{i+1,j}(k) = \sum_{j \in S} V_{i,j}(k) + \sum_{j \in S} B_{i,j}(k) - A_i(k) \tag{2}$$

where $V_{i,j}(k)$ is the passenger volume with destination stop $j$ on the $k$th bus when bus arrives at stop $i$ and $A_i(k)$ is the total alighting volumes at stop $i$ for $k$th bus (bus dispatched at trip $k$). For bus stops whose index is larger than stop $i + 1$, the above equation can be simplified as follows:

$$\begin{aligned} &(\forall j \in S \cap j > i + 1) \\ &V_{i+1,j}(k) = V_{i,j}(k) + B_{i,j}(k) \end{aligned} \tag{3}$$

### 2.2.3 Bus Arrival and Departure Constraints

The following equations describe the bus movement dynamics. Specifically, the sum of the departure time at the adjacent upstream stop and the traveling time between two stops determines the arrival time at the next following bus stop; also, the departure time at the bus stop is described by adding the arrival time at the current stop and the dwell time at the stop.

$$\mathrm{DP}_1(k) = \mathrm{DS}(k) + \mathrm{LT}_1(k) + \mathrm{HT}_1(k) \tag{4a}$$

$$\begin{aligned} &(\forall i \in \{S \backslash 1\}) \\ &\mathrm{AR}_{i+1}(k) = \mathrm{DP}_i(k) + \mathrm{TR}_{i,i+1}(k) \end{aligned} \tag{4b}$$

$$\mathrm{DP}_i(k) = \mathrm{AR}_i(k) + \mathrm{LT}_i(k) + \mathrm{HT}_i(k) \tag{4c}$$

$$\mathrm{AR}_i(k) \le \mathrm{AR}_i\ (k + 1) \tag{4d}$$

where $\mathrm{DP}_i(k)$ is the departure time of $k$th bus at stop $i$, $\mathrm{AR}_i(k)$ is the arrival time of $k$th bus at stop $i$, $\mathrm{LT}_i(k)$ is the loading time of $k$th bus at stop $i$, $\mathrm{HT}_i(k)$ is the holding time of $k$th bus at stop $i$, $\mathrm{TR}_{i,\,i+1}(k)$ is the traveling time between stop $i$ and stop $i+1$, and $\mathrm{DS}(k)$ is the dispatching time of $k$th bus at the starting terminal, which is known input in this chapter. Constraint (4d) is added to avoid overtaking phenomenon, as both cost (8) and constraint (1a) require a larger arrival time for trip $k+1$ compared to trip $k$.

### 2.2.4 Bus Loading Time Constraints

In this chapter, the simultaneous loading model [16] is adopted to describe the bus loading time, and the model assumes that the boarding process and the loading process happen simultaneously. Accordingly, the loading time is determined by the larger value of two loading times, either boarding time or alighting time, shown as follows:

$$\mathrm{LT}_i(k) = \max\left(\delta_1 \sum_{j \in S} B_{i,j}(k), \delta_2 A_i(k)\right) + t_{\mathrm{oc}} \tag{5}$$

where $t_{\mathrm{oc}}$ is the closing and opening time of the bus door. $\delta_1(\delta_2)$ is the unit boarding time (alighting time) for each passenger.

### 2.2.5 Bus Boarding Flow and Alighting Flow Constraints

The number of boarding passengers $B_{i,j}(k)$ depends on the demand at stop $i$, $P_{i,j}(k)$, and the remaining capacity of the targeting bus $\mathrm{Cap}_k(k) - \sum_{j \in S} V_{i,j}(k)$. Also, all passengers are allowed to alight the bus once the bus reaches its destination in this chapter, so the number of alighting passengers at stop $i$ is equal to the number of on-board passengers from all upstream bus stops with destination stop $i$.

$$B_{i,j}(k) = \min\left(P_{i,j}(k), \mathrm{Cap}_k(k) - \sum\nolimits_{j \in S} V_{i,j}(k)\right) \tag{6a}$$

$$A_i(k) = \sum_{q \in s} V_{q,i}(k) \tag{6b}$$

where $\mathrm{Cap}_k(k)$ is the capacity of the $k$th bus.

### 2.2.6 Bus Speed and Holding Time Constraints

The traveling time between two stops depends on the corresponding bus speed, and the upper bound and the lower bound are applied to the decision variables in this chapter, namely, bus speed and holding time, shown as follows:

$$\mathrm{TR}_{i,i+1}(k) = \frac{D_{i,i+1}}{v_{i,i+1}(k)} \tag{7a}$$

$$v_{\min} \le v_{i,i+1}(k) \le v_{\max} \tag{7b}$$

$$0 \le \mathrm{HT}_i(k) \le \mathrm{HT}_{\max} \tag{7c}$$

where $D_{i,\, i\,+\,1}$ is the distance between stops $i$ and $i+1$ and $v_{\min}$ and $v_{\max}$ are the minimum allowed bus driving speed and maximum bus driving speed, respectively. Similarly, the maximum allowed holding time $\mathrm{HT}_{\max}$ is also applied for the holding time at each bus stop.

## 2.3 Cost Function

### 2.3.1 Total Stop Passenger Waiting Time

To minimize the total stop passenger delay, we formulate the objective function as follows:

$$\begin{aligned} J_{\mathrm{sd}} = \min \Bigg\{ \sum_{k>1} \Bigg[ \sum_{j \in S} \frac{f_{1,j}(k-1)}{2} (\mathrm{DS}(k) - \mathrm{DS}\,(k-1))^2 \\ + \sum_{i \in S \cap i>1} \sum_{j \in S} \frac{f_{i,j}(k-1)}{2} (\mathrm{AR}(k) - \mathrm{AR}\,(k-1))^2 \Bigg] \Bigg\} + \\ \sum_{j \in S} \Bigg[ P_{1,j}(1)\,(\mathrm{DS}(1) - t_0) + \sum_{i \in S \cap i>1} P_{i,j}(1)\,(\mathrm{AR}_i(1) - t_0) \Bigg] \end{aligned} \tag{8}$$

where $\sum_{j \in S} \left[ P_{1,j}(1)\,(\mathrm{DS}(1) - t_0) \right]$ and $\sum_{j \in S} \sum_{i \in S \cap i>1} P_{i,j}(1)\,(\mathrm{AR}_i(1) - t_0)$ are the passenger delay when the first bus reaches the corresponding bus stops and $\sum_{k>1} \sum_{j \in S} \frac{f_{1,j}(k-1)}{2} (\mathrm{DS}(k) - \mathrm{DS}\,(k-1))^2$ and $\sum_{k>1} \sum_{i \in S \cap i>1} \sum_{j \in S} \frac{f_{i,j}(k-1)}{2} (\mathrm{AR}(k) - \mathrm{AR}\,(k-1))^2$ are the incremental delay between adjacent trips.

### 2.3.2 Total On-Board Passenger Traveling Time

In order to minimize the total on-board passenger delay, we formulate the objective function as follows:

$$J_{\mathrm{od}} = \min \left[ \sum_k \sum_{i \in S} \left[ \mathrm{HT}_{i-1}(k) \sum_{j \in S} V_{i,j}(k) + \mathrm{TR}_{i,i+1}(k) \sum_{j \in S} V_{i-1,j}(k) \right] \right] \tag{9}$$

where $\sum_k \sum_{i \in S} \mathrm{HT}_{i-1}(k) \sum_{j \in S} V_{i,j}(k)$ and $\sum_k \sum_{i \in S} \mathrm{TR}_{i,i+1}(k) \sum_{j \in S} V_{i-1,j}(k)$ are the total holding time and total on-road driving time, respectively.

## 3 Solution Algorithms

Although decision variables, holding times $\mathrm{HT}_i(k)$ and bus speeds $v_{i,\,i+1}(k)$, are only bounded by constraints (7b) and (7c), its value will influence the arrival time at different trips and may avoid constraint (4d) if simply adopting evolutionary algorithms (EAs) to generate population for holding times $\mathrm{HT}_i(k)$ and bus speeds $v_{i,\,i+1}(k)$. Therefore, we firstly use mathematical technique to solve the single-objective problem under either stop passenger delay or on-board passenger delay to analyze the corresponding bus movement characteristics. The proposed mixed logical model is converted into the corresponding MINLP problem, and then the commercial solver, e.g., Gurobi [17], is adopted to solve the problem to guarantee that all constraints are satisfied.

However, when tackling the multi-objective problem, the traditional mathematical programming techniques will become cumbersome, as it requires multiple runs by applying different weights on different objectives. On the other hand, EAs are naturally appropriate for solving multi-objective optimization problems due to its population-based search approach: A group of results representing a subset of Pareto set can be found in a single run. In view of this, we adopt one of the popular multi-objective evolutionary algorithms, non-dominated sorting genetic algorithm II (NSGA-II) [18], to solve our problem due to its fast ranking and sorting mechanism. Meanwhile, the corresponding individual adjustments are applied in the fitness calculation stage in order to avoid the bus overtaking phenomenon. Furthermore, harmony search (HS) algorithm is also adopted to replace the operation of crossover and mutation in genetic algorithm (GA) to form a NSHS algorithm, and the comparison between NSGA-II and NSHS is conducted in Sect. 4.

### *3.1 A MINLP Formulation for Bus Operation System*

We have formulated our bus operation system as a nonlinear cost function, (8) or (9), with associated constraints which describe the stop and bus volume dynamics (1a), (1b) and (2), bus arrival and departure constraints (4a), (4b), (4c) and (4d), bus loading time constraints (5), boarding and alighting flow constraints (6a) and (6b), and bus speed and holding time constraints (7). Among all these constraints,

constraints (5), (6a), and (7a) can be converted into the corresponding linear constraints, and details can be found as follows.

Equations (5) and (6a) can be converted into the following constraints by introducing binary variables $\tau_i(k)$ and $\rho_{ij}(k)$:

$$\tau_i(k) = 1 \iff \delta_1 \sum_{j \in S} B_{i,j}(k) \leq \delta_2 A_i(k) \tag{10a}$$

$$\rho_{ij}(k) = 1 \iff P_{i,j}(k) \leq \text{Cap}_k(k) - \sum\nolimits_{j \in S} V_{i,j}(k) \tag{10b}$$

Then the logic relationship above can be rewritten as:

$$\delta_1 \sum_{j \in S} B_{i,j}(k) - \delta_2 A_i(k) \leq M_1 \left(1 - \tau_i(k)\right) \tag{11a}$$

$$\delta_1 \sum_{j \in S} B_{i,j}(k) - \delta_2 A_i(k) \geq \epsilon + (-\epsilon - M_1)\, \tau_i(k) \tag{11b}$$

$$P_{i,j}(k) - \left(\text{Cap}_k(k) - \sum\nolimits_{j \in S} V_{i,j}(k)\right) \leq M_2 \left(1 - \rho_{ij}(k)\right) \tag{11c}$$

$$P_{i,j}(k) - \left(\text{Cap}_k(k) - \sum\nolimits_{j \in S} V_{i,j}(k)\right) \geq \epsilon + (-\epsilon - M_2)\, \rho_{ij}(k)\Big) \tag{11d}$$

where $M_1$ is a sufficiently big integer, e.g., $M_1 \geq \max_{k,\,i}(\pm(\delta_1 \sum_{j \in S} B_{i,j}(k) - \delta_2 A_i(k)))$, and, similarly, $M_2$ is a sufficiently big integer, e.g., $M_2 \geq \max_{k,\,i,\,j}$ $(\pm(P_{i,j}(k) - (\text{Cap}_k - \sum_{j \in S} V_{i,j}(k))))$, and $\epsilon$ is sufficiently small.

Based on the above relationship, we can convert our constraints as follows:

$$\text{LT}_i(k) - \delta_1 \sum_{j \in S} B_{i,j}(k) - t_{\text{oc}} \leq M_3 \left(1 - \tau_i(k)\right) \tag{12a}$$

$$-\text{LT}_i(k) + \delta_1 \sum_{j \in S} B_{i,j}(k) + t_{\text{oc}} \leq M_3 \left(1 - \tau_i(k)\right) \tag{12b}$$

$$\text{LT}_i(k) - \delta_2 A_i(k) - t_{\text{oc}} \leq M_3 \tau_i(k) \tag{12c}$$

$$-\text{LT}_i(k) + \delta_2 A_i(k) + t_{\text{oc}} \leq M_3 \tau_i(k) \tag{12d}$$

$$B_{i,j}(k) - P_{i,j}(k) \leq M_4 \left(1 - \rho_{ij}(k)\right) \tag{12e}$$

$$-B_{i,j}(k) + P_{i,j}(k) \leq M_4\left(1 - \rho_{ij}(k)\right) \tag{12f}$$

$$B_{i,j}(k) - \left(\mathrm{Cap}_k - \sum\nolimits_{j \in S} V_{i,j}(k)\right) \leq M_4 \rho_{ij}(k) \tag{12g}$$

$$-B_{i,j}(k) + \left(\mathrm{Cap}_k - \sum\nolimits_{j \in S} V_{i,j}(k)\right) \leq M_4 \rho_{ij}(k) \tag{12h}$$

where $M_3$ is a sufficiently big integer, e.g., $M_3 \geq \max_{k,\,i}(\pm(\mathrm{LT}_i(k) - \delta_1 \sum_{j\,\in\,S} B_{i,j}(k) - t_{\mathrm{oc}}), \pm(\mathrm{LT}_i(k) - \delta_2 A_i(k) - t_{\mathrm{oc}}))$, and, similarly, $M_4$ is a sufficiently big integer, e.g., $M_4 \geq \max_{k,\,i,\,j}(\pm(B_{i,j}(k) - P_{i,j}(k)), \pm(B_{i,j}(k) - (\mathrm{Cap}_k - \sum_{j\,\in\,S} V_{i,j}(k))))$.

**Proposition 1** *The same solution can be obtained after replacing Eqs. (5) and (6a) by Inequalities (11a)–(11d) and (12a)–(12h) in the model.*

For constraint (7a), we use the newly introduced variable $vs_{i,\,i+1}(k)$ to replace the speed variable $v_{i,\,i+1}(k)$, and then the corresponding constraints (7a) and (7b) are rewritten as follows:

$$\mathrm{TR}_{i,i+1}(k) = vs_{i,i+1}(k) D_{i,i+1} \tag{13a}$$

$$\frac{1}{v_{\min}} \leq vs_{i,i+1}(k) \leq \frac{1}{v_{\max}} \tag{13b}$$

**Proposition 2** *The same solution can be obtained after replacing constraints (7a) and (7b) by Inequality (13a) and (13b) in the model.*

### *3.2 Non-dominated Sorting Algorithms*

Our bi-objective optimization problem can be formulated as follows:

$$\min F(\chi) = \left(J_{\mathrm{sd}}(\mathbf{x}), J_{\mathrm{od}}(\mathbf{x})\right)^T \tag{14}$$

s.t. constraints (1a)(1b)(2)(4a)–(4d)(5)(6a)(6b)(7a)–(7c)

where $J_{\mathrm{sd}}$ and $J_{\mathrm{od}}$ are the objectives (8) and (9), respectively, and $\mathbf{x}$ denotes the vector of decision variable for our model. For a bi-objective problem, our goal is trying to find the optimal Pareto front composed of a set of solutions which cannot dominate one another, and a solution $\mathbf{x}$ is a member of the optimal Pareto front only when there are no other solutions dominating $\mathbf{x}$. Also, we say solution $\mathbf{x_1}$ dominates solution $\mathbf{x_2}$ only if the following conditions are satisfied:

$$\mathbf{x_1} \text{ dominates } \mathbf{x_2} \iff$$

$$\left[ (J_{\text{sd}}(\mathbf{x}_1) \le J_{\text{sd}}(\mathbf{x}_2)) \wedge \left( (J_{\text{od}}(\mathbf{x}_1) \le J_{\text{od}}(\mathbf{x}_2)) \right] \wedge \right.$$

$$\left[ (J_{\text{sd}}(\mathbf{x}_1) < J_{\text{sd}}(\mathbf{x}_2)) \vee \left( (J_{\text{od}}(\mathbf{x}_1) < J_{\text{od}}(\mathbf{x}_2)) \right] \right. \quad (15)$$

The whole process to implement the NSGA-II can be found in Fig. 2. Firstly, the initial populations are generated randomly, and then the fitness of each individual is calculated accordingly. After that, offspring is calculated by taking crossover and mutation operation on existing population. Next, the combined solution set is sorted to obtain their corresponding ranks and crowding distances, and those with small

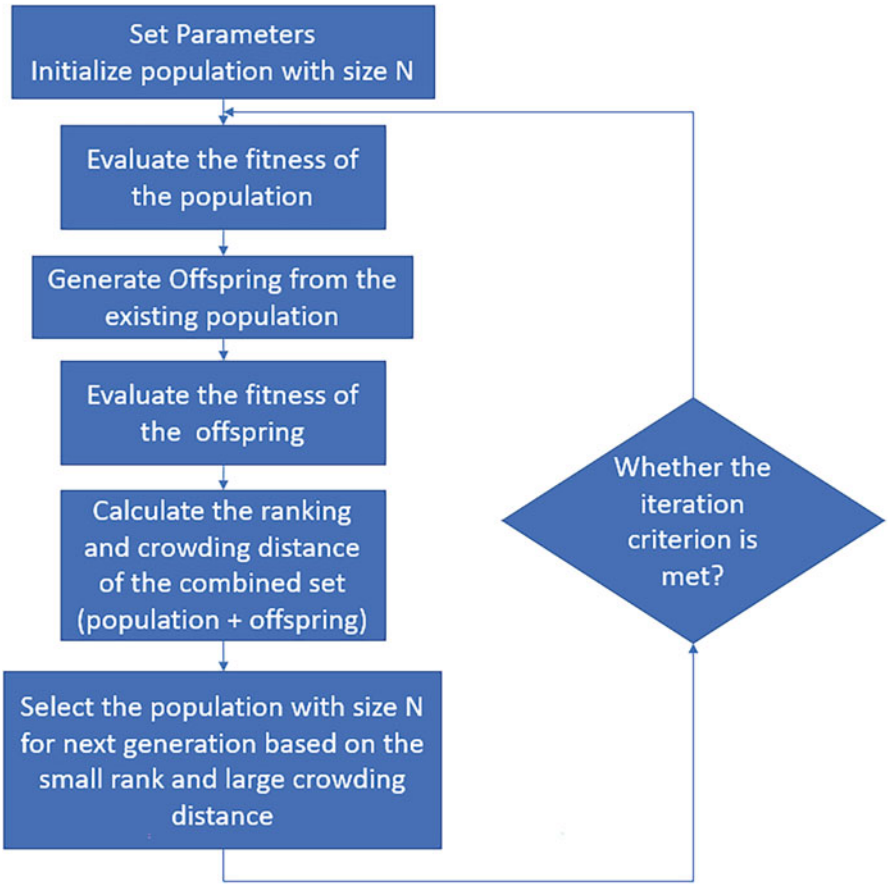



**Fig. 2** The flowchart of NSGA-II

rank are selected as individuals of the new population, if adding individuals at rank $k$ surpasses the required population size $N$ but size $N$ cannot be reached if finishing adding individuals at rank $k$-$1$, then the selection of the individuals at rank $k$ will be determined based on their crowding distance. After obtaining the new population, the above procedure repeats until the iteration time is met.

The crucial characteristic of NSGA-II is its fast sorting mechanism to determine different ranks. In its sorting stage, for each solution $x_i$, we record the number of solutions ($N_i$) that dominates $x_i$, and the solutions dominated by $x_i$, $D_i = \{x_j : x_i\}$, dominates $x_j$, $\forall\ j \neq i$. After finishing the above process for each solution in the population, the first non-dominated front layer is identified by checking whether $N_i$ is equal to 0. After the first Pareto front is selected, the domination number $N_j$ of individual $x_j$, which belongs to $D_i$, is subtracted by one if $x_i$ is selected into the first front. Then the above procedure repeats until all ranking layers are identified.

The crowding distance is developed with the aim to preserve diversity through the iterations and is considered only if we need to add solutions of a certain rank into the new population, but adding this whole rank will exceed the size of the population. Then the selection of the individuals in this rank will be based on the crowding distance. The normalized distance of solution $x_i$ which belongs to rank $k$ targeting on objective $r$, $d_i^{k,r}$, can be defined as follows:

$$d_i^{k,r} = \frac{d_{i+1}^{k,r} - d_{i-1}^{k,r}}{d_{\max}^{k,r} - d_{\min}^{k,r}} \tag{16}$$

where $d_{\max}^{k,r}\left(d_{\min}^{k,r}\right)$ denotes the maximum (minimum) value of rank $k$ in respect of the objective $r$. The summation of all normalized distances for each objective determines the crowding distance of solution $x_i$ of rank $k$ is $d_i^k = \sum_r d_i^{k,r}$. The larger the crowding distance, the higher the possibility to be selected into the new population.

Different with the traditional NSGA-II, slight adjustments are required for our problem in order to avoid bus overtaking. Once finished generating the population, decision variables $HT_i(k)$ and $v_{i,\,i+1}(k)$, in the initialization stage under constraints (7b) and (7c), the bus speeds and holding times at the corresponding trips will be modified accordingly with the update of the dynamics as trip increases in the fitness calculation stage.

It is possible that the arrival time $\text{AR}_i(k)$ is still lower than $\text{AR}_i(k-1)$ even if $\text{HT}_{\max}$ and $v_{\min}$ are applied to $\text{HT}_{i-1}(k)$ and $v_{i-1,\,i}(k)$; under such extreme case, constraints (7b) or (7c) may not be able to be satisfied in order to guarantee a positive value for $\text{AR}_i(k) - \text{AR}_i(k-1)$. This can be alleviated by adding penalty on the objective if constraints (7b) and (7c) are not satisfied and will be considered in our continued work to improve the current algorithm.

Similar to NSGA-II, NSHS only has a different update mechanism for generating new population. Instead of the crossover and mutation operation on the selected parent set, the new population in HS is generated on the basis of the harmony memory considering rate (HMCR) and pitch adjustment rate (PAR), as shown in

Algorithm 1. A new harmony vector is generated randomly according to the upper and lower bounds if a random value is larger than HMCR; otherwise, the new vector is determined by the parameter PAR: if the randomly generated value is smaller than PAR, a slight adjustment is applied on the existing vector; otherwise, a previous vector is assigned to the new vector.

## 4 Simulation Results

In this section, we firstly test the proposed model under a bus line with ten bus stops for two different objectives separately, and four different passenger arrival patterns are considered, as shown in Fig. 3.

**Algorithm 1** Generate new harmony individual

```
If (rand1() < HMCR)
x_new = x_i      (17)
    If (rand2() < PAR)
x_new = x_i ± rand () × BW     (18)
    end
else
x_new = LB + (UB − LB) × rand ()     (19)
end
```

Case 1 has the same total arrival rates $\sum_j f_{i,j}$ for each bus stop $i$, and case 2 has the largest arrival rates for the intermediate bus stops, while the largest arrival rates exist in several initial bus stops and gradually decrease as the bus stop index increases in case 3; on the contrary, the arrival rate in case 4 gradually increases, as the bus stop index increases. Next, the non-dominated Pareto fronts for stop passenger delay and on-board passenger delay are generated by adopting two proposed algorithms, NSHS and NSGA-II, under different problem scales.

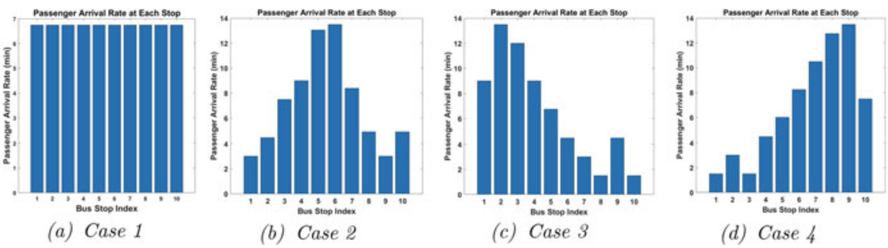



**Fig. 3** Different arrival patterns for bus stops

### 4.1 Bus Movement Analysis Under Different Objectives

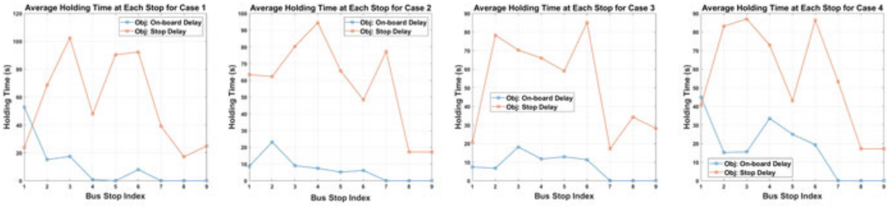
We test the proposed model on a bus line with ten bus stops, and the distance between two adjacent bus stops is randomly generated with ranges from 1.5 to 2.5 km. The bus dispatching time starts from 360 s toward 2520 s with 6 min evenly spaced. We consider 6 m/s and 14 m/s as the minimum and maximum driving speed on the road; also, each bus can accommodate 80 passengers with maximum holding time 120 s, and the details can be found in Table 1.

As the last stop is the final terminal stop without any boarding passengers, the average holding time and bus speed are only drawn for the first nine stops. Figure 4 illustrates the average holding time of seven trips at each bus stop, where four cases are drawn based on the arrival patterns shown in Fig. 3. The red line and the blue line describe the change of the holding time under the objective of the stop passenger delay and on-board passenger delay, respectively. Clearly, the holding times in all stops for all cases are certainly very low when the objective is to minimize the on-board passenger delay. The considered objective function is nonlinear, and the solver cannot guarantee global optimality. For this reason, the holding time is non-zero. Even if both objectives are nonlinear, which forms a MINLP and cannot obtain the optimal solution, the holding times under different objectives can roughly reflect the different characteristics: The red line under different arrival patterns fluctuates above the blue line. There is not an obvious distinction between the average holding times of red lines under different arrival patterns; this may result from the sub-optimal results obtained for the MINLP problem.

On the other hand, Fig. 5 captures the average bus speed of seven trips at each road segment between bus stops. Similarly, the red line and the blue line describe the change of the bus speed under the objective of the stop passenger delay and on-board passenger delay, respectively. It is obvious to see that bus speeds keep a relatively high value in all cases when the objective is to minimize the on-board passenger delay; similar to the holding time, the optimal results $v_{\max}$ for all stops are not obtained due to the nonlinearity of the objectives, which cannot guarantee global

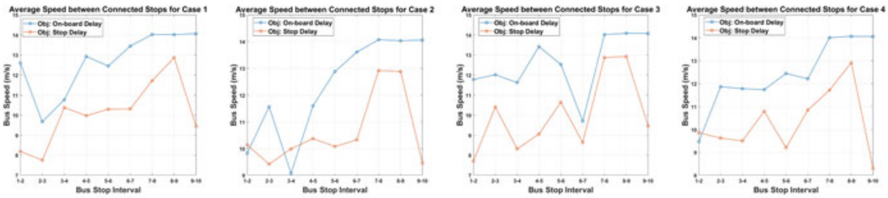
**Table 1** Parameters used in the case study

| Parameters | Descriptions | Associated values |
|---|---|---|
| $N_s$ | Bus stop number | 10 |
| $H_p$ | Planning trips | 7 |
| DS($k$) | Bus dispatching time | [360 720 1080 1440 1800 2160 2520] |
| $\text{Cap}_k$ | Bus capacity | 80 |
| $t_{\text{oc}}$ | Door open and close time | 2 s |
| $\delta_1$ | Average boarding time for each passenger | 2 s |
| $\delta_2$ | Average alighting time for each passenger | 1 s |
| $v_{\min}$ | Minimum on-road driving speed | 6 m/s |
| $v_{\max}$ | Maximum on-road driving speed | 14 m/s |
| $\text{HT}_{\max}$ | Maximum holding time at each stop | 120 s |




(a) Case 1 (b) Case 2 (c) Case 3 (d) Case 4

**Fig. 4** The average holding times under two different objectives

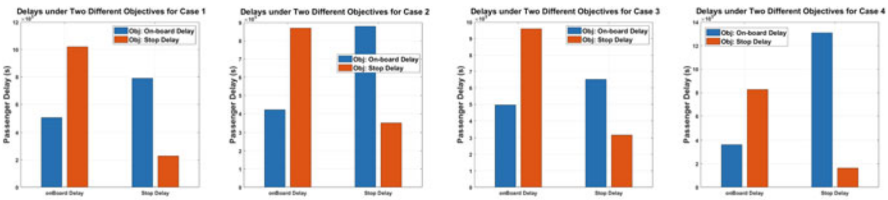



(a) Case 1 (b) Case 2 (c) Case 3 (d) Case 4

**Fig. 5** The average speeds under two different objectives

optimality. In contrast, the red line under different arrival patterns fluctuates below the blue line, which shows an opposite trend compared with the holding times in Fig. 4. This is understandable that keeping a proper headway requires the coordination of both holding time and bus speed.

Figure 6 depicts the stop passenger delay and on-board passenger delay under objectives (8) and (9), where the blue bar and the red bar denote the delays under the objective of stop delay and on-board delay, respectively, and the first group of the bars illustrates the on-board delay under two objectives, while the second group captures the stop delay under two objectives. Due to the consideration of the traveling time into the on-board passenger delay, the quantity of the on-board delay is compatible with the stop passenger delay. Also, the delay is smaller than the other case if the corresponding delay is selected as the objective to be minimized. From all four cases, we can find that the stop passenger delay conflicts with the on-board passenger delay. Once the traveling time is taken into account, the on-board passenger delay is quite significant; thereby, we cannot pay all attention on the waiting passengers at stops by completely neglecting the interests of on-board passengers. In view of this, a bi-objective analysis is conducted in the following section with the aim to find a trade-off between the stop passengers and the on-board passengers.




(a) Case 1 (b) Case 2 (c) Case 3 (d) Case 4

**Fig. 6** Stop delay and on-board delay under two different objectives

**Table 2** Parameter settings for different cases

| | Case 1 | Case 2 | Case 3 | Case 4 | Case 5 | Case 6 |
|---|---|---|---|---|---|---|
| $N_s$ | 10 | 10 | 10 | 15 | 15 | 15 |
| $H_p$ | 7 | 7 | 7 | 10 | 10 | 10 |
| PopSize | 100 | 200 | 300 | 100 | 200 | 300 |

## *4.2 Pareto Graph for Two Different Delays*

We now provide the bi-objective analysis of the bus operation system by adopting NSHS and NSGA-II under six different combinations, as illustrated in Table 2, where $N_s$ is the number of the bus stops for the entire bus line and $H_p$ is the total trips dispatched for the whole schedule. The mutation rate and the crossover rate in NSGA- II are set as 0.2 and 0.5, respectively. We set HMCR and PAR in NSHS as 0.95 and 0.5, respectively, and test both algorithms on three different population sizes, 100, 200, and 300, as shown in Table 2; also, the iteration times are all 1000 for two algorithms.

Figure 7 illustrates the Pareto front (non-dominated solutions in the first rank) of the bus operation system under two different algorithms for six different cases. The blue points are results obtained from NSGA-II, and the red points are from NSHS. The front gap between two algorithms is significantly obvious: the Pareto front with lower values is found in NSGA-II. Therefore, the result quality of NSGA-II surpasses NSHS for all cases. Also, the gap between NSGA-II and NSHS is becoming increasingly large as the problem scale increases. For the current bus model, NSGA-II is capable of finding the Pareto-optimal set; nevertheless, NSHS, on the other hand, performs well on spreading solutions, which can be reflected on all cases, and this may result from its internal mechanism to generate new vectors: the new vectors can be randomly obtained via lower and upper bounds in HS, while the new vectors can only be transformed from existing parents in GA. Also, as the population size increases, we can find that the number of results obtained from the first rank also increases accordingly, especially for NSGA-II. Furthermore, with the increase of the problem scale, the quantity of the solutions obtained from the Pareto front quickly shrinks in NSGA-II, which can be reflected by comparing case 1 and case 4, case 2 and case 5, and case 3 and case 6, respectively. The results may be influenced by the population adjustments conducted in the fitness

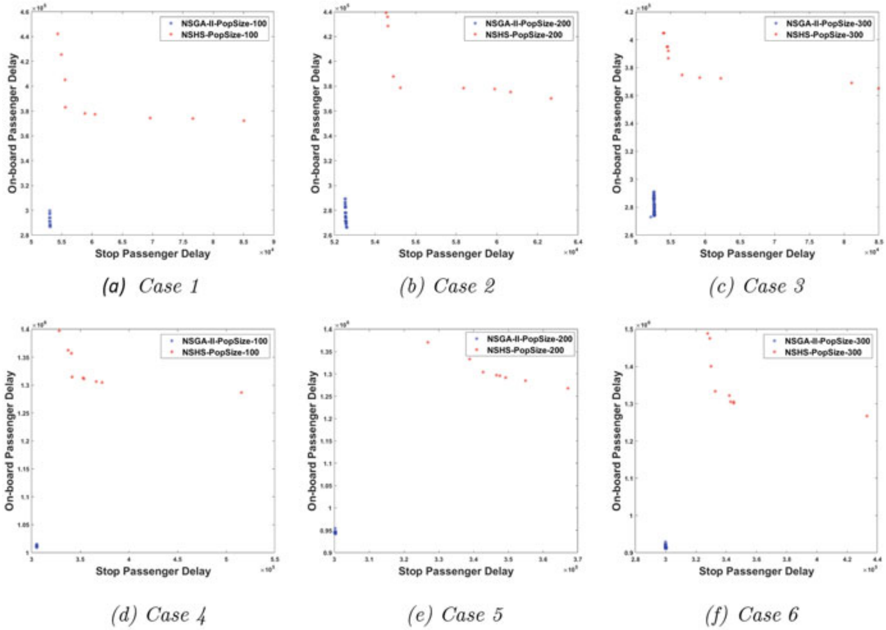



**Fig. 7** The Pareto front of the proposed model under two different algorithms

calculation stage for our problem to avoid bus overtaking; therefore, we need to consider more advanced algorithms in our future study to tackle the current issue in our bus operating optimization model with the aim to obtain sufficient, well-spread results in the first rank.

## 5 Conclusions

This manuscript proposed a dynamic bus operation system to minimize either the stop passenger delay or on-board passenger delay by adjusting the bus speed and holding time. The bus operation system is formulated as a trip-based optimization model, and the simultaneous model is adopted to describe the bus loading time. As our objective is nonlinear, one of the evolutionary algorithms, GA, is applied to solve the optimization model under two different objectives, and the corresponding bus movements are analyzed in detail to capture the impacts of the objective: the simulation outcomes indicate that the trend of the holding time is firstly falling and then rising when stop passenger delay is minimized, while the bus speed presents an opposite trend to coordinate with the holding time to maintain the headway. Furthermore, we address a bi-objective bus operation problem by minimizing both the stop passenger delay and the on-board passenger delay: The Pareto graph of two objectives based on the rank evaluation and crowding distance is obtained by adopting NSGA-II and NSHS, and the findings indicate that NSGA-II performs better than NSHS but the results from NSHS are well-spread.

## References


1. Yi, Z. (2020). *Optimization and scheduling for a large-scale urban transportation system involving human factors*. Doctoral thesis.
2. Ibarra-Rojas, O. J., Delgado, F., Giesen, R., & Munoz, J. C. (2015). Planning, operation, and control of bus transport systems: A literature review. *Transportation Research Part B: Methodological, 77*, 38–75.
3. Liping, F., Liu, Q., & Calamai, P. (2003). Real-time optimization model for dynamic scheduling of transit operations. *Transportation Research Record, 1857*(1), 48–55.
4. Eberlein, X. J., Wilson, N. H. M., & Bernstein, D. (1999). Modeling real-time control strategies in public transit operations. In *Computer-aided transit scheduling* (pp. 325–346). Springer.
5. Eberlein, X. J., Wilson, N. H. M., & Bernstein, D. (2001). The holding problem with real–time information available. *Transportation Science, 35*(1), 1–18.
6. Delgado, F., Munoz, J. C., & Giesen, R. (2012). How much can holding and/or limiting boarding improve transit performance? *Transportation Research Part B: Methodological, 46*(9), 1202–1217.
7. Zhang, Y., Rong, S., & Zhang, Y. (2020). A dynamic optimization model for bus schedule design to mitigate the passenger waiting time by dispatching the bus platoon. In *2020 American Control Conference (ACC)* (pp. 4096–4101). IEEE.
8. Zhang, Y., Rong, S., Zhang, Y., & Guruge, N. S. G. (2021). A multi-bus dispatching strategy based on boarding control. *IEEE Transactions on Intelligent Transportation Systems*.
9. Sirmatel, I. I., & Geroliminis, N. (2018). Mixed logical dynamical modeling and hybrid model predictive control of public transport operations. *Transportation Research Part B: Methodological, 114*, 325–345.
10. Abraham, A., Teja, N., Dasgupta, S., Choudhury, A., & Dauwels, J. (2021). An optimal controller synthesis for longitudinal control of platoons with communication scenarios in urban environments and highways. *SAE International Journal of Connected and Automated Vehicles, 4*(1), 81–95.
11. Liu, H., Skabardonis, A., & Zhang, W.-b. (2003). A dynamic model for adaptive bus signal priority. In *Preprint CD-ROM, 82nd Transportation Research Board Annual Meeting, Washington, DC*.
12. Zhang, Y. (2020). An adaptive pre-signal setting to provide bus priority under a coordinated traffic-responsive network. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)* (pp. 1–6). IEEE.
13. Chuanjiao, S., Wei, Z., & Yuanqing, W. (2008). Scheduling combination and headway optimization of bus rapid transit. *Journal of Transportation Systems Engineering and Information Technology, 8*(5), 61–67.
14. Koehler, L. A., Kraus, W., & Camponogara, E. (2011). Iterative quadratic optimization for the bus holding control problem. *IEEE Transactions on Intelligent Transportation Systems, 12*(4), 1568–1575.
15. Tolíc, I. H., Nyarko, E. K., & Ceder, A. A. (2020). Optimization of public transport services to minimize passengers' waiting times and maximize vehicles' occupancy ratios. *Electronics, 9*(2), 360.
16. Highway Capacity Manual. (2000). *Highway capacity manual* (Vol. 2, p. 1).
17. LLC Gurobi Optimization. (2021). Gurobi optimizer reference manual.
18. Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. A. M. T. (2002). A fast and elitist multi-objective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation, 6*(2), 182–197.

# Real-Time Traffic Accident Detection for an Intelligent Mobility in Smart Cities

**Anuj Abraham, Chetan B. Math, Shitala Prasad, and Mohit Sharma**

## Abbreviations

| | |
|---|---|
| AdaBoost | Adaptive boosting |
| ERP | Electronic road pricing |
| GTB | Gradient tree boosting |
| IoT | Internet of Things |
| RBF | Radial basis function |
| RNN | Recurrent neural network |
| SOTA | State of the art |
| SVM | Support vector machine |
| V2V | Vehicle-to-vehicle |

Anuj Abraham, Chetan B. Math, and Shitala Prasad contributed equally with all other contributors.

A. Abraham (✉) · M. Sharma
Technology Innovation Institute, Abu Dhabi, United Arab Emirates
e-mail: Anuj.Abraham@tii.ae; mohit.sharma@tii.ae

C. B. Math
Volocopter, Bruchsal, Germany
e-mail: belagal.chetan@volocopter.com

S. Prasad
Indian Institute of Technology (IIT Goa), Ponda, Goa, India
e-mail: shitala@iitgoa.ac.in

 
N. Gupta, S. Mishra (eds.), *Internet of Everything for Smart City and Smart Healthcare Applications*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-34601-9_9

## 1 Introduction

Traffic congestion occurs frequently in urban scenarios, in which accidents are one of the most common causes of death all around the world. A quick response to traffic accidents is crucial to saving a life. With connected vehicle technology, vehicle data such as acceleration, velocity, position, etc. can be periodical, i.e., these samples are collected from the vehicle periodically every second. An example of such a system is the new electronic road pricing (ERP) system execution planned for the deployment in 2022. The data collected from connected vehicles may lead to the detection of various anomalies in the traffic [1].

Traffic accident detection is complex as it can be caused by various factors such as human errors, vehicle breakdown, the bad infrastructure of road, etc. Traffic accident detection is intuitively based on the assumption that it slows down the traffic. However, due to complex traffic scenarios, various cases can have similar effects as well. For instance, traffic jams, traffic lights, lane change, and breakdown of a vehicle can also slow down the traffic [2]. The key is not just anomaly detection but also the classification of an anomaly as a traffic accident with high confidence. The specific question we address in this chapter is the following:

*Can we differentiate traffic accidents from other traffic anomalies with high confidence?*

Significant efforts have been made by researchers in the recent past to detect an accident and classification systems based on the Internet of Things (IoT). In this IoT platform, classification is based on severity level and the essential information about the accident [3]. In addition, early vehicle accident detection and notification based on smartphone technology has been seen in literature to detect and communicate crash vehicle events to emergency agencies [4]. Similarly, technological improvements in wireless communication to identify traffic accidents on roads using location and speed information based on combined vehicle-to-vehicle (V2V) communication techniques with machine learning methods have attracted interest from many researchers [5, 6].

In this regard, we compare classification modeling algorithms, such as support vector machine (SVM) and gradient boosting tree algorithms, to detect the accidents on roads. SVMs are standard supervised machine learning techniques that can be used for classification or regression. This method finds an optimal hyperplane for linearly separable patterns based on support vectors [7]. It can also be extended to nonlinear data by transforming data into new dimension space using a kernel function. Various kernels available are linear kernel, nonlinear kernel, radial basis function (RBF), sigmoid, polynomial, exponential, etc. [8]. Nonlinear classifiers yield nonlinear (curved) hyperplanes; hence, even more solutions exist. One of the main advantages of SVM is that it is effective in high-dimensional spaces even in cases where feature dimensions are greater than data samples [7, 8]. This method uses a subset of training points called support vectors, so it is also memory efficient. This method can also use common kernels or custom kernels in which different kernel functions can be specified for the decision function. If the feature dimensions are much greater than the data samples, avoiding overfitting in the choice of kernel functions and regularization term is crucial. Another limitation is that SVMs do not directly provide probability estimates. The C parameter, which is known as the regularization parameter, allows you to make “adjustments” to SVM margins on a hyperplane. The C parameter controls margin and how much to avoid misclassifying training data. Large C means a small margin hyperplane, even if it does a good job of classifying training data correctly. For a small C, it means a large margin hyperplane, even if it misclassified some training data. Misclassification of trained data leads to a chance of overfitting and underfitting if the value of C is large and small, respectively [9, 10].

GTB which is also known as a tree boosting algorithm is an ensemble tree method that can be used for feature selection [7]. This method builds trees and sum results in a sequential manner. Some of the main advantages of using GTB are that it often gives good prediction accuracy, reduces bias by minimizing the error sequentially, can be used for feature selection, and provides greater flexibility to fit model, whereas the challenges in GTB are that it is prone to overfitting due to overemphasizing on the outlier and is harder to tune [11]. In this work, we regularly inspect the evolution of training and validation errors seen in the plots. Somewhere, it is noticed that the training error is always going down, but the validation error goes up. That is the indication point to stop training to avoid overfitting [12]. Other challenges are slowing due to sequential model building and generating a prediction and harder-to-interpret results as compared to the decision tree. In the decision tree to avoid overfitting, a pruning method is used in literature [13].

In the gradient boosting tree algorithm for classification, a multinomial deviance loss function for K categories or classes is carried out. Here, initialization is just a single terminal node tree, and a commonly used loss function is deviance. Optimization of GTB is performed by using tree-specific parameters, boosting parameters (learning rate), and varying a number of estimators. Some of the tree-specific parameters include the maximum number of features, maximum tree depth, minimum samples to split, and minimum samples per leaf. In GTB, we performed tenfold cross-validation for the model testing.

In traditional machine learning, crafting features are very tedious process [14] whereas, in deep learning parameters are very important and tuning them is a crucial [10, 15]. Therefore, we need to analyze various machine learning methods based on specific tasks.

In this work, the basic objective addressed is to detect traffic accidents on roads, and information is passed to send an ambulance to save lives. In addition, other actions to clear roads for traffic flow when an accident is detected are sending tow trucks, police, etc. The chapter is outlined as follows: Sect. 2 discusses about the data collection and Sect. 3 about data aggregation, and Sect. 4 talks about several experiments performed in this chapter followed by the conclusion and scope for future work.

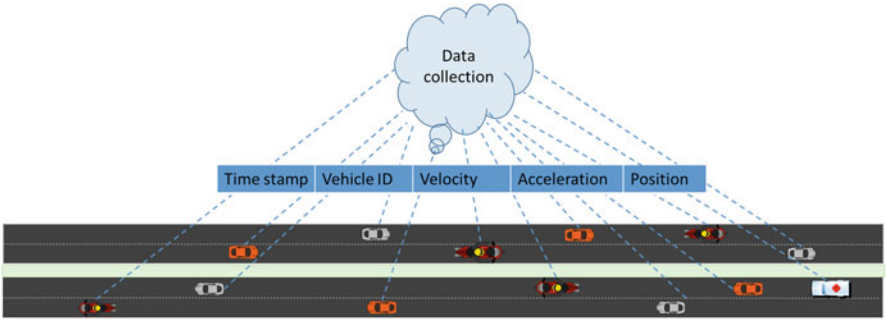



**Fig. 1** Various data features considered to detect accidents on roads

## 2 Data Collection and Processing

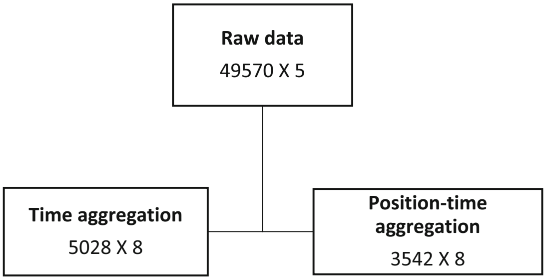
The data is generated from traffic simulation software PTV vissim. It simulates the desired connected vehicle traffic scenario. Similar to the connected vehicle system, we collect data from the vehicles periodically. Various data features considered are VEH. ID., VEHTYPE, and POS representing the vehicle ID number, vehicle type, and position of vehicles, respectively, which are shown in Fig. 1. These features are used to detect accidents on the road. These features are selected, as they are the typical data used by the connected vehicle technology standards.

The data is collected periodically using the connected vehicle technology from all the vehicles within the desired region as shown in Fig. 1. In our study, the data is collected using vissim from a highway scenario. We have simulated a 2 km four-lane highway scenario and initiated multiple accidents at different parts of the highway to collect the data. The data from the vehicle is collected every 1 second, which is typically the case for connected vehicle technology. Note that the connected vehicle technology is assumed to be reliable without any packet or data losses [16].

## 3 Data Aggregation

Traditionally, neural networks are used for time-based data. However, due to various challenges such as computational complexity, black box nature, scalability, etc. associated with neural networks, in this study, we perform aggregation of data and use SVM and gradient tree boosting approaches to detect the accidents. Aggregation of data is performed to capture the time correlation of the features. We consider time aggregation and position-time aggregation. Time aggregation is performed by aggregating vehicle’s data over a period of 10 seconds. Position-time aggregation is performed by aggregating vehicle’s data within a 50 m range over

**Fig. 2** Data aggregation to capture the time correlation of the features



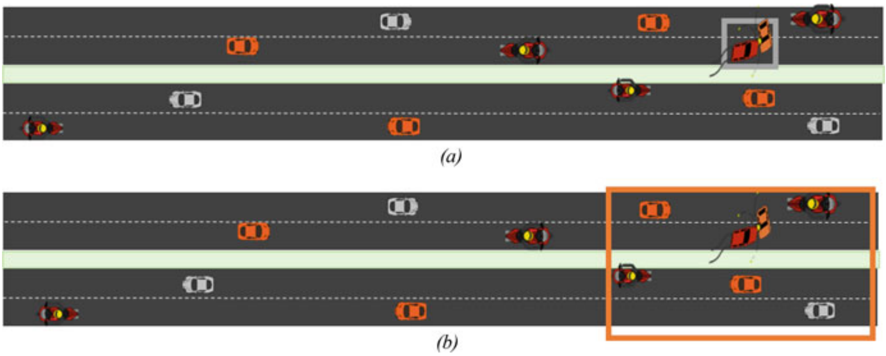




**Fig. 3** Data aggregation approaches. (**a**) Time aggregation or microscopic detection. (**b**) Position-time aggregation or macroscopic detection

a period of 10 seconds. In the aggregate data, we consider both the average and the variance of the data, i.e., average velocity, variance velocity, average position, variance position, average acceleration, and variance acceleration are considered. Aggregation of data reduces the sample size but adds more features as shown in Fig. 2.

## 3.1 *Time Aggregation*

Individual vehicle data is aggregated over a period. Hence, the accident detection using time aggregation will lead to the identification of the vehicle in the accident, i.e., microscopic detection as shown in Fig. 3a. However, accidents can be missed due to failure of connected vehicle units of the vehicle. Furthermore, as the number of vehicles increases, the time aggregation data grows, and it becomes more computationally intensive. To promote a scalable and computationally efficient solution, we proposed position-time aggregation.

### 3.2 Position-Time Aggregation

All vehicles' data within a defined range over a period is aggregated. Hence, the accident detection using position-time aggregation provides a region in which the accident has happened, i.e., macroscopic detection as shown in Fig. 3b. The number of vehicles does not affect the position-time aggregation data; however, it increases as the region under consideration grows.

Time aggregation leads to microscopic detection, i.e., individual vehicles in the accident as illustrated in Fig. 3a, and position-time aggregation leads to macroscopic detections, i.e., the region in which the accident has taken place as shown in Fig. 3b.

## 4 Simulation Results

In this section, we will discuss different road scenarios and detect road accidents using connected vehicle data under traffic constraints. We performed numerous experiments to evaluate different machine learning techniques suitable for road accident detection. For the same, we collected vehicle data from vissim simulation by creating multiple road accident scenarios, as discussed in Sect. 2 of this chapter.

This section is further sub-sectioned as SVM and gradient tree boosting which is applied on two different approaches: time aggregation and position-time aggregation.

### 4.1 Road Accident Detection Using SVM

As we discussed in Introduction, the road accidents are hard to detect in real time, and the existing intelligence systems for different scenarios are often confused. Thus, this chapter tries to summarize various state-of-the-art (SOTA) machine learning techniques in time aggregation and position-time aggregation approaches. Since SVM is a non-probabilistic, linear, binary classifier which is used for data classification by generating hyperplanes in hyperspace, we used it in road accident classification. There are different shades of SVM based on the type of kernel used such as linear and nonlinear, which is detailed further.

### 4.2 Linear SVM

In linear SVM, the data are projected to higher dimensions such that a linear hyperplane separates them out. To evaluate the performance of linear SVM for road accident detection, we set up two different train-test data distributions: 60–40%

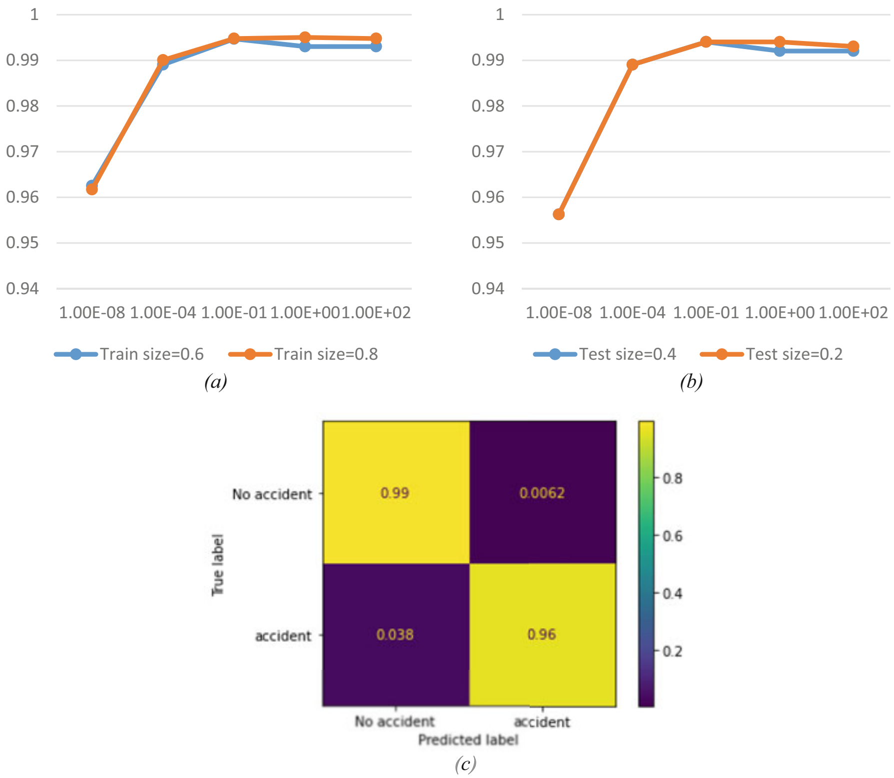



**Fig. 4** Accuracy and confusion matrix for time aggregation approach using linear SVM. (**a**) Accuracy for train data. (**b**) Accuracy for test data. (**c**) Confusion matrix

and 80–20% for time aggregation and position-time aggregation approaches. Figure 4a, b and Table 1 show the training and testing accuracies for time aggregation approach using linear SVM, respectively. As we see in Table 1, the best accuracy archived is 99.40% with regularization parameter $C = 1.00\text{E} - 01$, for both the data splits. We can see the impact of regularizations on the performance. If we reduce the C parameter, the performance is decreased drastically, while increasing the regularization parameter in linear SVM actually decreases the result slowly. We also observe from Fig. 4a, b that the 80–20% split is the best combination for road accident detection on time aggregation approach.

### 4.2.1 Time Aggregation

In Fig. 4c, we also plotted the confusion matrix for both the class for time aggregation approach using linear SVM, i.e., accident and no-accident classes. A confusion matrix is the representation of predicted label vs the true label, and it visualizes and summarizes the performance of a classification algorithm.

**Table 1** Time aggregation approach using linear SVM with regularization parameter C

| C | Accuracy Train data | | Accuracy Test data | |
|---|---|---|---|---|
| | Train size | | Test size | |
| | 0.6 | 0.8 | 0.4 | 0.2 |
| 1.00E-08 | 0.962533 | 0.961711 | 0.956262 | 0.956262 |
| 1.00E-04 | 0.989058 | 0.990055 | 0.989066 | 0.989066 |
| 1.00E-01 | 0.994695 | 0.994779 | 0.994036 | 0.994036 |
| 1.00E+00 | 0.993037 | 0.995027 | 0.992048 | 0.994036 |
| 1.00E+02 | 0.993037 | 0.994779 | 0.992048 | 0.993042 |

#### 4.2.2 Position-Time Aggregation

Next, we performed similar experiments on the position-time aggregation approach. Figure 5a, b and Table 2 show the training and testing accuracies for position-time aggregation using linear SVM, respectively. We perceive that for linear SVM on position-time aggregation, the best accuracy is 99.23% and 99.15% with regularization parameter $C = 1.00\text{E}{-}01$ for data splits 60–40% and 80–20%, respectively. In this experiment too, the best regularization performer is C = 1.00E-1. We performed the cross-test with regularization parameter and found that the results are very close, and thus, we show accuracies only for those that have variations. Also witness that the more the training data, the better the training accuracy is; see Fig. 5a, b, respectively.

In Fig. 5c, we also plotted the confusion matrix for both the class for position-time aggregation approach using linear SVM. We observed that with the proposed feature space, we achieved approximately 99.0% accuracy.

### *4.3 Nonlinear SVM*

We further extend our analysis to nonlinear SVM for both time aggregation and position-time aggregation approaches. Below are the summarizations for the same.

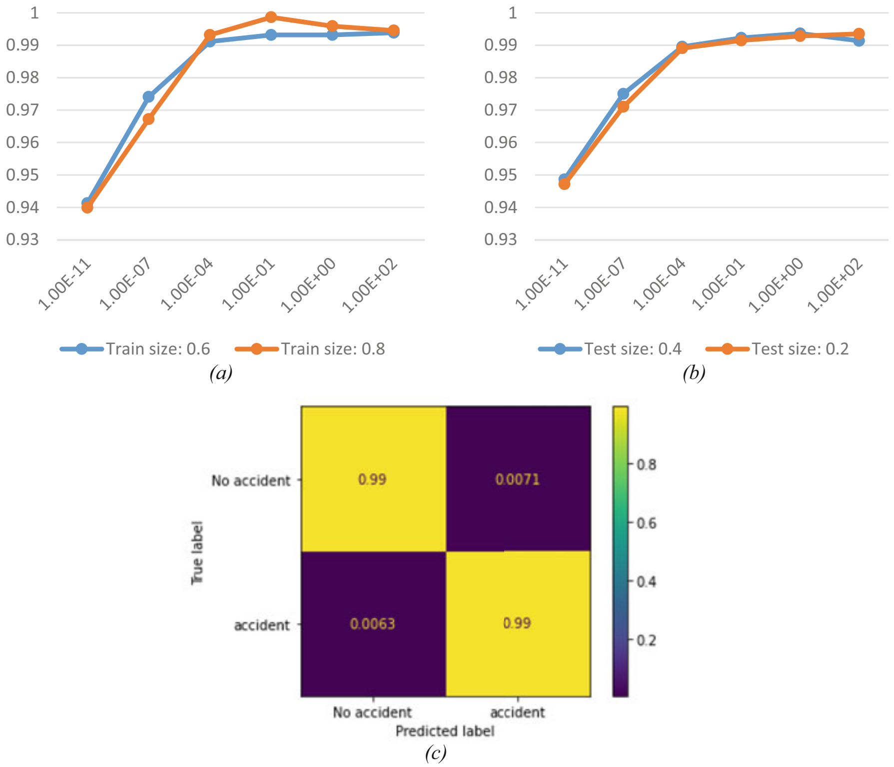



**Fig. 5** Accuracy and confusion matrix for position-time aggregation approach using linear SVM. (**a**) Accuracy for train data. (**b**) Accuracy for test data. (**c**) Confusion matrix

### 4.3.1 Time Aggregation

In nonlinear SVM, the data are projected to higher dimensions such that a nonlinear hyperplane separates them out. Figure 6a, b and Table 3 show the training and testing accuracies for position-time aggregation approach using nonlinear SVM, respectively. From Table 3, the best accuracy archived is 97.12% and 97.71% with regularization parameter $C = 1.00\text{E}{+}10$ for different data splits. Also, for nonlinear RBF kernels, the regularization parameter shows a strong impact. The difference between 60–40% and 80–20% is 0.59%. In this experimental setup, we see that the more the training data we have, the higher the test accuracy we achieve.

In Fig. 6c, we also plotted the confusion matrix for both the class for time aggregation approach using nonlinear SVM. We observed that with the proposed feature space, we achieved approximately 98.0% accuracy for no-accident classes and 94.0% accuracy for accident classes, respectively.

**Table 2** Position-time aggregation approach using linear SVM with regularization parameter C

| C | Accuracy for Train data | | Accuracy for Test data | |
|---|---|---|---|---|
| | Train size | | Test size | |
| | 0.6 | 0.8 | 0.4 | 0.2 |
| **1.00E-11** | 0.941297 | 0.939891 | 0.948613 | 0.947135 |
| **1.00E-07** | 0.974061 | 0.967213 | 0.974989 | 0.97101 |
| **1.00E-04** | 0.991126 | 0.993169 | 0.989541 | 0.989086 |
| **1.00E-01** | 0.993174 | 0.998634 | 0.992269 | 0.991473 |
| **1.00E+00** | 0.993174 | 0.995902 | 0.993633 | 0.992838 |
| **1.00E+02** | 0.993857 | 0.994536 | 0.99136 | 0.99352 |

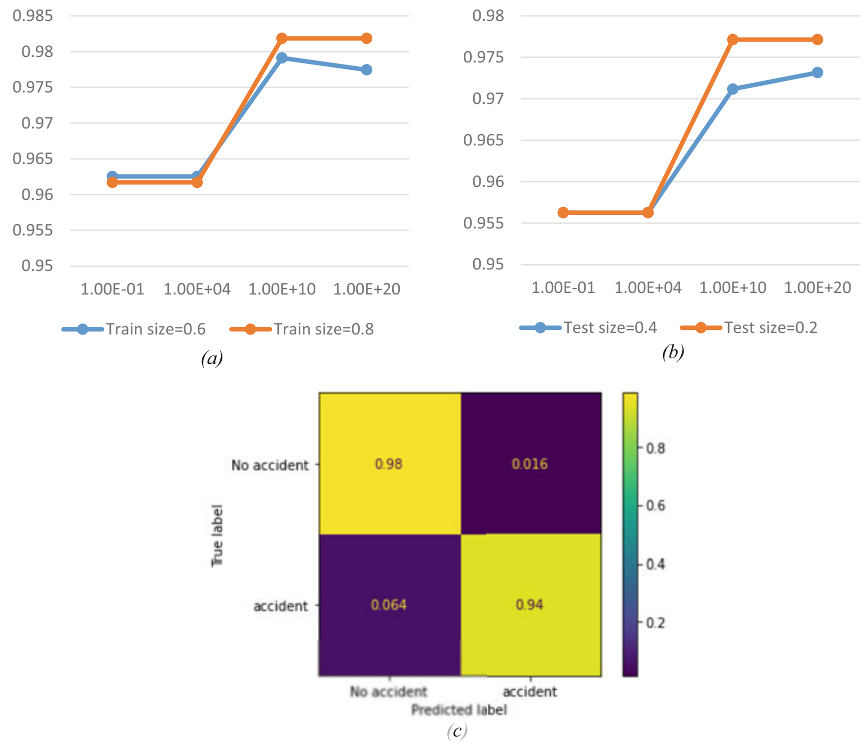



**Fig. 6** Accuracy and confusion matrix for time aggregation approach using nonlinear SVM. (**a**) Accuracy for train data. (**b**) Accuracy for test data. (**c**) Confusion matrix

**Table 3** Time aggregation approach using nonlinear SVM with regularization parameter C

| C | Accuracy Train data | | Accuracy Test data | |
|---|---|---|---|---|
| | Train size | | Test size | |
| | 0.6 | 0.8 | 0.4 | 0.2 |
| **1.00E-01** | 0.962533 | 0.961711 | 0.956262 | 0.956262 |
| **1.00E+04** | 0.962533 | 0.961711 | 0.956262 | 0.956262 |
| **1.00E+10** | 0.979111 | 0.98185 | 0.971173 | 0.977137 |
| **1.00E+20** | 0.977454 | 0.98185 | 0.973161 | 0.977137 |

### 4.3.2 Position-Time Aggregation

The performance of the test on position-time aggregation approach is shown in Fig. 7 and Table 4, respectively. From Table 4, the best accuracy is 99.66% with regularization parameter $C = 1.00\text{E}{+}04$ for 80–20% data split, which is 0.02% higher than 60–40% data split. We observe from Fig. 7 that the higher the training data we have, the higher the impact of the regularization parameter is and the higher the accuracy we achieve.

In Fig. 7c, we also plotted the confusion matrix for both the class for position-time aggregation approach using nonlinear SVM. We observed that with the proposed feature space, we achieved approximately 99.0% accuracy for no-accident classes and 96.0% accuracy for accident classes, respectively, and the result shows improvement over the linear SVM.

## *4.4 Gradient Tree Boosting (GTB): Tenfold Cross-Validation*

As we know that, the GTB is similar to adaptive boosting (AdaBoost) in which they both use an ensemble of decision *trees* to predict a target label [17]. We used gradient tree boosting to perform time aggregation- and position-time aggregation-based road accident detection. The decision tree diagram we used in this chapter is shown in Fig. 8, which is based on the feature's importance. Below are the analysis results for both approaches.

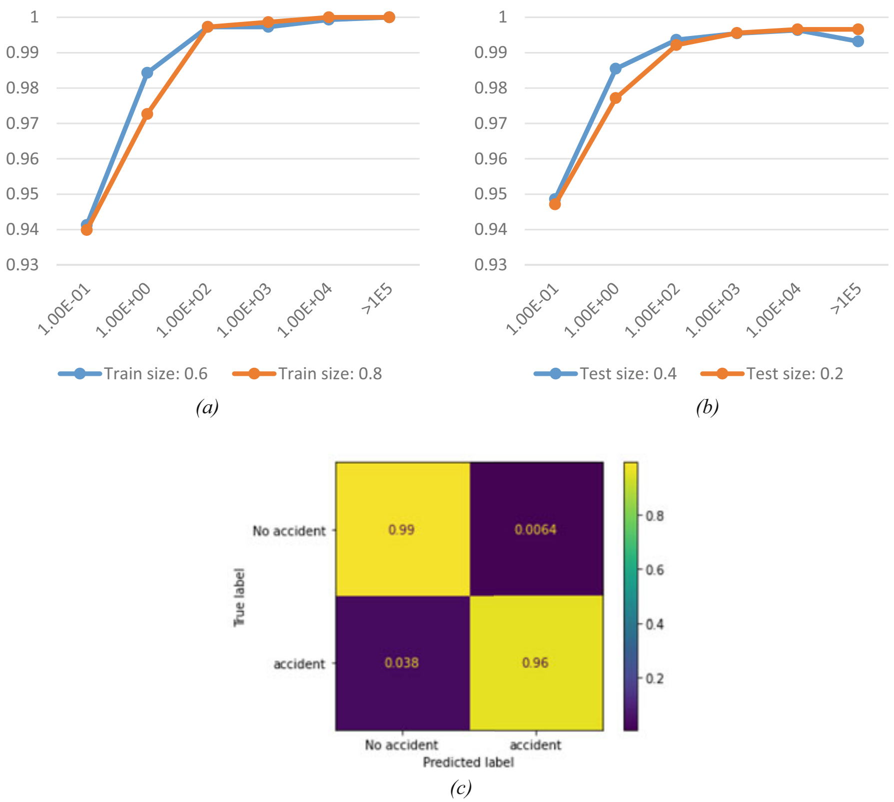



**Fig. 7** Accuracy and confusion matrix for position-time aggregation approach using nonlinear SVM. (**a**) Accuracy for train data. (**b**) Accuracy for test data. (**c**) Confusion matrix

#### 4.4.1 Time Aggregation

Using GTB, the performance of time aggregation approach is raised to 99.60% with an average F-score of 95.24%, as shown in Table 5. For this, we used 80–20% train-test setting. On the one hand, Fig. 9a shows detailed testing performance corresponding to different estimators. The best performance is achieved at an estimator value equal to five. On the other hand, Fig. 9b shows the confusion matrix for accident and no-accident classes. There is a 9% of error seen in accident detection while 0% error in no-accident. Accident detection is crucial, so the error rate is very sensitive. In Fig. 11a, we show the type of features used in this analysis of the time aggregation approach and their impact in GTB.

**Table 4** Position-time aggregation approach using nonlinear SVM with regularization parameter C

| C | Accuracy for Train data | | Accuracy for Test data | |
|---|---|---|---|---|
| | Train size | | Test size | |
| | 0.6 | 0.8 | 0.4 | 0.2 |
| 1.00E-01 | 0.941297 | 0.939891 | 0.948613 | 0.947135 |
| 1.00E+00 | 0.9843 | 0.972678 | 0.985448 | 0.977149 |
| 1.00E+02 | 0.99727 | 0.997268 | 0.993633 | 0.992156 |
| 1.00E+03 | 0.99727 | 0.998634 | 0.995452 | 0.995566 |
| 1.00E+04 | 0.999317 | 1 | 0.996362 | 0.996589 |
| >1E5 | 1 | 1 | 0.993179 | 0.996589 |

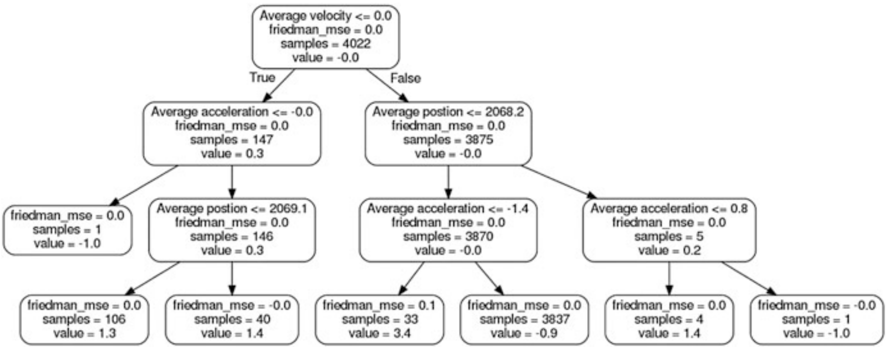



**Fig. 8** Decision tree diagram using GTB

**Table 5** Overall performance indices' comparison

| | SVM | | | | Gradient tree boosting | |
|---|---|---|---|---|---|---|
| | Time aggregation | | Position-time aggregation | | Time aggregation | Position-time aggregation |
| | Linear | Nonlinear | Linear | Nonlinear | | |
| Accuracy | **99.40** | 97.71 | 99.15 | **99.66** | 99.60 | **99.82** |
| Precision | 86.29 | 70.67 | 89.20 | 89.71 | 99.5678 | 99.54 |
| Recall | 96.18 | 93.63 | 99.37 | 99.37 | 90.91 | 99.12 |
| F-score | 90.96 | 80.55 | 94.01 | 94.29 | 95.24 | 99.34 |

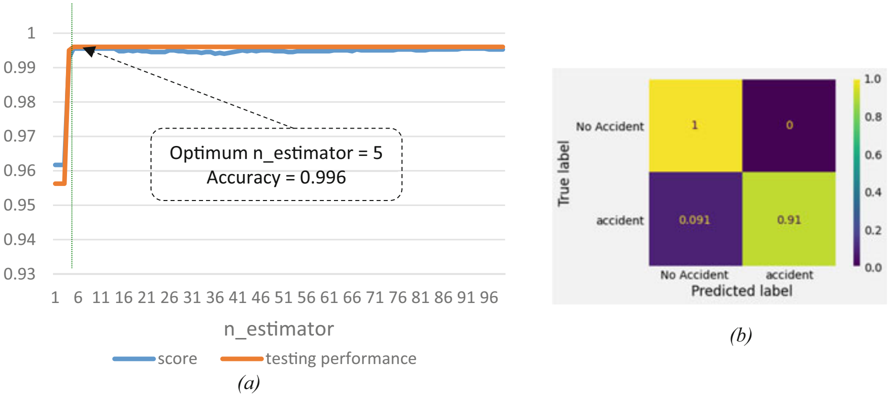



**Fig. 9** Testing performance and confusion matrix for time aggregation using GTB with tenfold cross-validation. (**a**) Optimal testing performance. (**b**) Confusion matrix

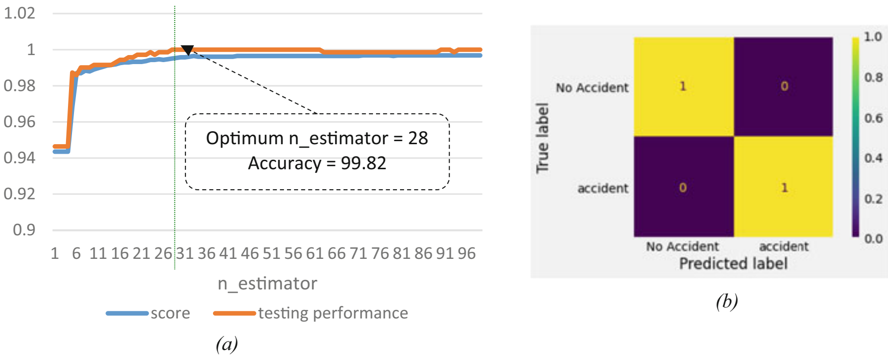



**Fig. 10** Testing performance and confusion matrix for position-time aggregation using GTB. (**a**) Optimal testing performance. (**b**) Confusion matrix

### 4.4.2 Position-Time Aggregation

In position-time aggregation, GTB performance reached 99.82% accuracy for an estimator value of 28. Figure 10a shows the testing performance accuracy graph corresponding to increasing estimators, and Fig. 10b is about the confusion matrix against accident and no-accident classes. In Fig. 11b, we show the type of features used in this analysis and their impact in GTB for time aggregation and position-time aggregation approaches. The main feature importance that has an impact on both the approaches is "average velocity," in which position-time aggregation outperforms the time aggregation by 20% as the feature value is improved from 55 to 75 that is evident in Fig. 11.

Compared to time aggregation, position-time aggregation achieves 0.22% higher accuracy, while recall rate is boosted by 8.22%, which is significantly significant;

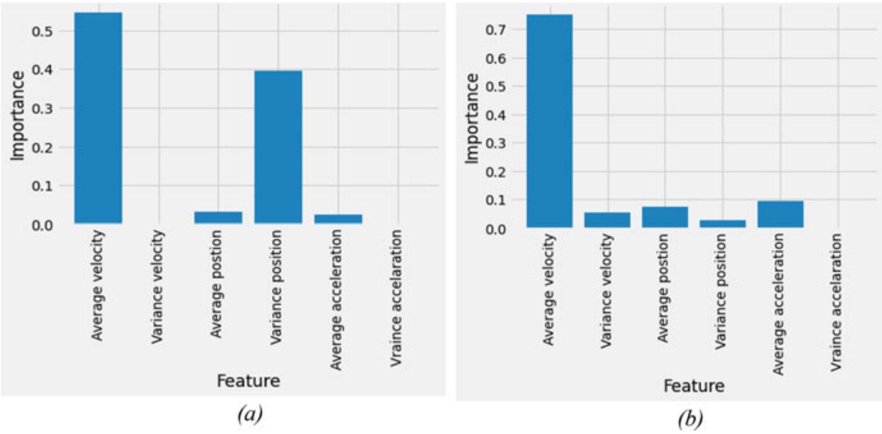



**Fig. 11** Feature importance for the data aggregation approaches using GTB. (**a**) Time aggregation. (**b**) Position-time aggregation

see Table 5. In Table 5, we also compared recall and precision rates for linear and nonlinear SVM and GTB algorithms for time aggregation and position-time aggregation approaches. Comparing time aggregation and position-time aggregation for linear SVM, we see an improvement of 2.91% and 3.19% in precision and recall rates, while for nonlinear SVM, the improvements are 19.04% and 6.01% for precision and recall rates, respectively. Compared with SVM, GTB algorithm boosts the F-score by 4.74% for the position-time aggregation approach.

As an observation from Table 5, the GTB out performances compared to SVM for both the approaches we considered in this chapter. In addition, the position-time aggregation is much better compared to time aggregation for both the machine learning algorithms.

## 5 Conclusions and Scope for Future Work

Connected vehicle technology can be used to detect accidents. In this chapter, we aggregate the data and use well-known SVM and gradient tree algorithms to detect accidents. We discuss various benefits and drawbacks of time aggregation and position-time aggregation approaches. For both the aggregations, we analyze the feature importance and notice that the aggregation affects the feature importance as well. For SVM, we use both linear and nonlinear methods and vary the test size and C parameter to find the optimal settings for both aggregations. For GTB algorithm, we vary the *n* parameter to find the optimal setting for both aggregations. We observed that the GTB method has better accuracy than SVM for position-time aggregation. Furthermore, accuracy for SVM and GTB methods was obtained as greater than 97.71% independent of the aggregation methods.

In this work, we demonstrate that SVM and GTB algorithms can provide highly accurate accident prediction with proper aggregation of data. However, further simulation studies with urban and rural areas need to be carried out in detail. We only consider SVM and GTB algorithms; further evaluation and comparison with different machine learning techniques are essential. Investigations of time correlation of the data need to be explored using recurrent neural network (RNN). We assume an ideal connected vehicle technology. Hence, connected vehicle technology challenges such as unreliable packet reception, channel congestion, etc. need to be considered. Further studies to analyze the effect of aggregation on feature importance and prediction need to be carried out. Fixed parameters such as 10 s and 50 m are considered for the aggregation. Studies to evaluate the effect of such parameters on the algorithms need to be performed.

## References

1. Nikolaev, A., Sapego, Y. S., Jakubovich, A. N., Berner, L. I., & Ivakhnenko, A. M. (2016). Simulation of automatic incidents detection algorithm on the transport network. *International Journal of Environmental & Science Education, 11*(16), 9060–9078.
2. Abraham, A., Zhang, Y., & Prasad, S. (2021). Real-time prediction of multi-class lane-changing intentions based on highway vehicle trajectories. In *IEEE International Intelligent Transportation Systems Conference (ITSC)* (pp. 1457–1462).
3. Balfaqih, M., Alharbi, S. A., Alzain, M., Alqurashi, F., & Almilad, S. (2022). An accident detection and classification system using internet of things and machine learning towards smart city. *Sustainability, 14, 210*, 1–13.
4. Aldunate, R. G., Herrera, O. A., & Cordero, J. P. (2013). Early vehicle accident detection and notification based on smartphone technology. In G. Urzaiz, S. F. Ochoa, J. Bravo, L. L. Chen, & J. Oliveira (Eds.), *Ubiquitous computing and ambient intelligence. Context-awareness and context-driven interaction* (Lecture notes in computer science) (Vol. 8276). Springer.
5. Dogru, N., & Subasi, A. (2018). Traffic accident detection using random forest classifier. In *15th Learning and Technology Conference (L&T)* (pp. 40–45).
6. Abraham, A., Teja, N., Dasgupta, S., Choudhury, A., & Dauwels, J. (2021). An optimal controller synthesis for longitudinal control of platoons with communication scenarios in urban environments and highways. *SAE International Journal of Connected and Automated Vehicles, 4*(1), 81–95.
7. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*.", Chapter 10,. Springer.
8. Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press.
9. Vapnik, V. (1995). *The nature of statistical learning theory*. Springer.
10. Abraham, A., Nagavarapu, S. C., Prasad, S., Vyas, P., & Mathew, L. K. (2022). Recent trends in autonomous vehicle validation ensuring road safety with emphasis on learning algorithms. In *17th International Conference on Control, Automation, Robotics and Vision (ICARCV), Singapore, Singapore* (pp. 397–404). https://doi.org/10.1109/ICARCV57592.2022.10004304
11. Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research, 12*, 2825–2830.
12. Natekin, A., Knoll, A., Gewaltig, M.-O., & Michel, O. (2013). Gradient boosting machines a tutorial. *Frontiers in Neurorobotics, 7*, 21.

13. Kotsiantis, S. B. (2013). Decision trees: A recent overview. *Artificial Intelligence Review, 39*(4), 261–283.
14. Chaoying, T. A. N. G., Xianghui, W. E. I., Biao, W. A. N. G., & Prasad, S. (2019). A cross-border detection algorithm for agricultural spraying UAV. *Applied Engineering in Agriculture, 35*(2), 163–174.
15. Prasad, S., Lin, D., Li, Y., Dong, S., & Min, O. Z. (2020). Rethinking of deep models parameters with respect to data distribution. In *25th IEEE International Conference on Pattern Recognition (ICPR)* (pp. 8562–8569).
16. Tangirala, N. T., Abraham, A., Choudhury, A., Vyas, P., Zhang, R., & Dauwels, J. (2018). Analysis of packet drops and channel crowding in vehicle platooning using V2X communication. In *IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India* (pp. 281–286). https://doi.org/10.1109/SSCI.2018.8628872
17. Zhang, Y., Shi, X., Zhang, S., & Abraham, A. (2022). A XGBoost-based lane change prediction on time series data using feature engineering for autopilot vehicles. In *IEEE Transactions on Intelligent Transportation Systems* (pp. 1–14).

# Part III
# Sustainable Approaches Towards Smart Healthcare Applications

# Smart E-Healthcare Business Model Using IoT

**Rachna K. Somkunwar**

## 1 Introduction

The Internet of Things, commonly known as IoT, is one of the most potent automation and control solutions available today, allowing for wireless Internet-based communication between remote devices. IoT-based healthcare systems are proven to be a good alternative since they offer the same healthcare services at a reduced cost in light of the rise in diseases and the expense of their treatment. One example of the many IoT-based healthcare gadgets being used nowadays is wearable technology and medical applications. Without being physically present, clinicians can monitor patients' conditions in real time using an IoT-based healthcare system [1].

The Internet of Things offers tremendous promise and opportunity for the health sector, particularly in the field of personal care. E-health services and products are a combination of various resources and healthcare provided by electronic and automated devices that enable the efficient and effective delivery of health information, including statistical data for healthcare professionals, governing bodies, and patients of health facilities, via the Internet and, most importantly, telecommunications as a medium. The IoT-based e-health technologies will soon change the healthcare industry [2].

The healthcare sector will benefit from new technology such as smart monitoring and health TV over the currently used conventional methods. Doctors may treat patients more accurately and carefully with the use of technology. It gathers data, which it then merges in a database that is stored on the main server with the administrative records. They function essentially as network elements that

R. K. Somkunwar (✉)
Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, India



N. Gupta, S. Mishra (eds.), *Internet of Everything for Smart City and Smart Healthcare Applications*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-34601-9_10

wirelessly connect clinics and hospitals. When configuring an infusion pump's software, the patient's weight, dosage, and category—adult or pediatric—are all taken into consideration. Infusion device software also uses a variety of libraries for patient data dependent on the size of the patient, whether they are adults or children, and compares drug dosages to the library that was picked based on the patient size. The right library must be selected by the smart infusion pump programmer during initial setup in order to guarantee that the correct dose ranges are being used. The drug library programming determines the appropriate infusion time for a medicine within the infusion rate restrictions. Another innovation is smart TV, sometimes referred to as "health TV," which provides users with IoT-based e-health solutions and acts as a platform for e-health services [3]. The availability of remote health services including at-home health coaching, health monitoring, and prescription reminders is a major emphasis of the ICT industry. Additionally, by enabling patients, such as older citizens, to control and receive medical treatment for a specific ailment from home via health TV, it promotes their independent lifestyle. The "Smart-O-Meter" is another IoT-based technological gadget that delivers a web-based patient-centered decision support system for blood pressure regulation. It employs an iterative, user-centered design approach. The viability and acceptability requirements for patient navigators and participants are met as a result [4].

## 2 System for Electronic Health Record

An electronic health record (EHR) is a computerized replacement for the conventional paper patient chart. A patient's traditional medical and treatment information is combined with additional data utilized in health and social care in an electronic health record, though. Due to the usage of different procedures, it is difficult to define electronic health record and the data content. A review found that the term "electronic health record" is applied to a variety of information systems, ranging from files compiled in specific departments to longitudinal collections of patient data. An electronic health record is utilized in all three levels of healthcare. Both secretarial workers and healthcare professionals could enter the information in EHRs. Patients themselves also record some information (validated by physicians).

Daily charting, medication administration, physical evaluation, admission nursing note, nursing care plan, referral, current complaint (such as symptoms), prior medical history, lifestyle, physical examination, diagnoses, tests, procedures, treatment, medication, discharge, history, diaries, issues, findings, and immunization are just a few of the data components that are documented in an electronic health record [5].

## 3 IoT Business Model

A well-organized and well-planned framework known as a "business model" or "canvas" is used to depict assumptions regarding the performance and accessibility of various critical resources and value chain activities. The product proposition, customer relationship management, channels necessary for the value chain, client segments to be serviced, cost and price structures, and revenue streams are also displayed along with this. In other words, it depicts how a business makes money and creates income [6]. This may be explained using the nine construction blocks listed below, which are seen in Fig. 1:

Working of business model is:

- The value proposition for a business model describes the precise services or goods you are providing to customers in a market.
- The value offer targets different client segments.
- We use the best distribution and communication methods to spread the word about our services to more people.
- A customer relationship is established and then maintained.

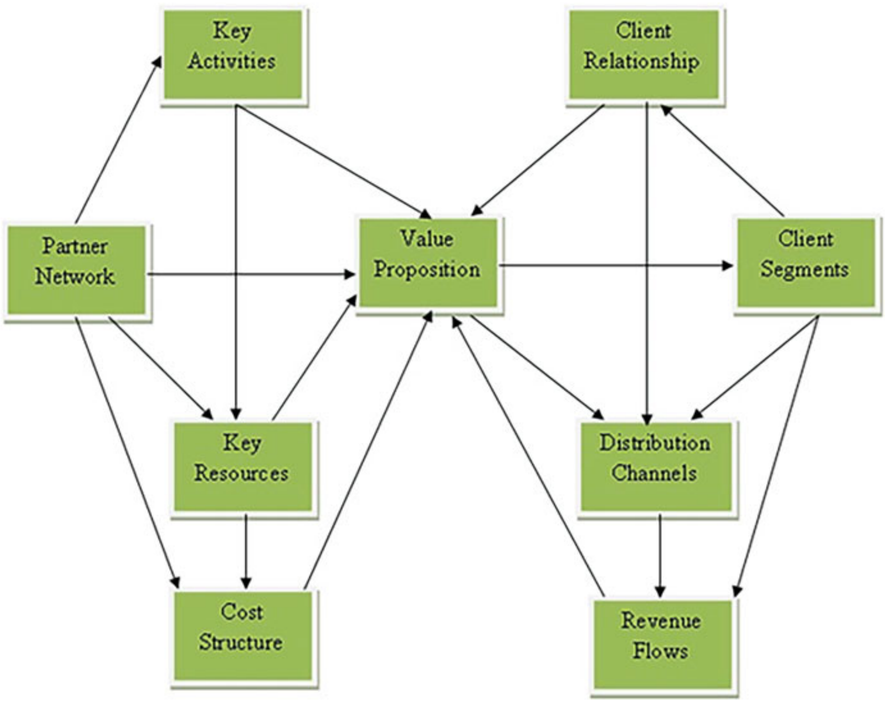



**Fig. 1** Nine-building block business model

- Key resources are needed in accordance with planning in order to increase the viability of the business model.
- The proper execution of the business model requires critical operations.
- Vendors and important partners must participate in this business model, and they must be motivated to do so.

## 4 Important Components of the Proposed IoT-Based Service Business Model

The basic objective of any corporation is to add value; thus, a thorough, carefully thought-out business model is formed [7]. It basically falls into four major categories:

1. Overall corporate identity: This comprises factors like the company's mission, vision, and reputation, all of which serve as important market differentiators for an organization.
2. Application (strategy) for the business: This classification may be used to explain how an organization's actions should translate its mission and vision. The capacity to function as a coordinator among several business units, each of which needs to play a specific and distinctive role in order to achieve planned and shared strategic goals, is a crucial aspect of this role. Planned goals, a timetable for reaching those goals, the resources needed, and unique performance indicators are further components of strategy.
3. A firm's assets for success: This will cover all internal resources that a company may use to supplement its plan. This may also include both material and immaterial assets like as human resources, organizational structures, and work cultures.
4. The setting in which the organization operates and competes: Customers, partners, potential threats from new entrants, legal information about compliance, the external availability of resources like technology resources, and all other emerging trends that might or might not have an impact on the company's position in the market are included in this list of stakeholders.

## 5 Various Business Models in the Health Sector

Financial and social factors are equally important in the healthcare sector, in addition to accessibility, cost, and quality. The six distinct business models for transforming health systems are listed below in order to fully deliver the value intended for patients and all those who are ultimately focused at the center.

### *5.1 E-Health Final Report Business Model [8]*

The components of the business model are as follows:

1. Key activities
2. Value proposition
3. Client relationships
4. Client segments
5. Key resources
6. Partner network
7. Cost structure
8. Distribution channels
9. Financial performance
10. Revenue stream source

Few key elements of the business model are:

1. Customer segment: This shows whether or not e-health services and products are beneficial to their users.
2. Value proposition: The market's offerings of e-health services.
3. Distribution channels: The routes of distribution and communication used to make these services available to customers.
4. Customer relationship: This describes the kinds of client relationships that are created and kept up in accordance with the company strategy.
5. Revenue stream: E-health services generate revenue streams.
6. Core capabilities: The fundamental functions that e-health services offer.
7. Value configuration and value offerings: The crucial procedures for developing and incorporating abilities into an effective value proposition that was expected and prepared.
8. Network of partner and vendors: The parties that will be involved in the e-health procedure.
9. Cost structure: The costs involved in offering the e-health service, both variable and fixed.

The sustainability of an e-health system is then provided based on the analysis of the mentioned parameters.

The five components that provide an overall identification and business model comparison are listed below. Additionally, it charts the effectiveness of each e-health system.

(i) Overview of the problem: Identification of operational and socioeconomic drivers is the primary emphasis of this.
(ii) Value chain: The main goal is to map out all the important partners and stakeholders.
(iii) Analysis of business model: In relation to the case study, it analyzes the development of the business model.
(iv) Analysis of impact.
(v) Identification of best practice.

## 5.2 Osterwalder and Pigneur Outline Four Elements of an Effective Company Model

(a) Customer management
(b) Product innovation
(c) Financial aspects
(d) Infrastructure management [9, 10]

In addition to this, the following elements are present:

- Technology (consists of service platforms, technical architecture, devices, and applications)
- Organization (consisting of actors, strategies and goals, roles assigned, interactions, and value chain activities)
- Service (consists of chain of values from intended value-delivered value-expected value-perceived value)
- Finance (consisting of investment sources, revenue sources, cost sources, risk sources, and pricing)
- The framework's architecture, as described below in Fig. 2, can be used to evaluate e-health projects.

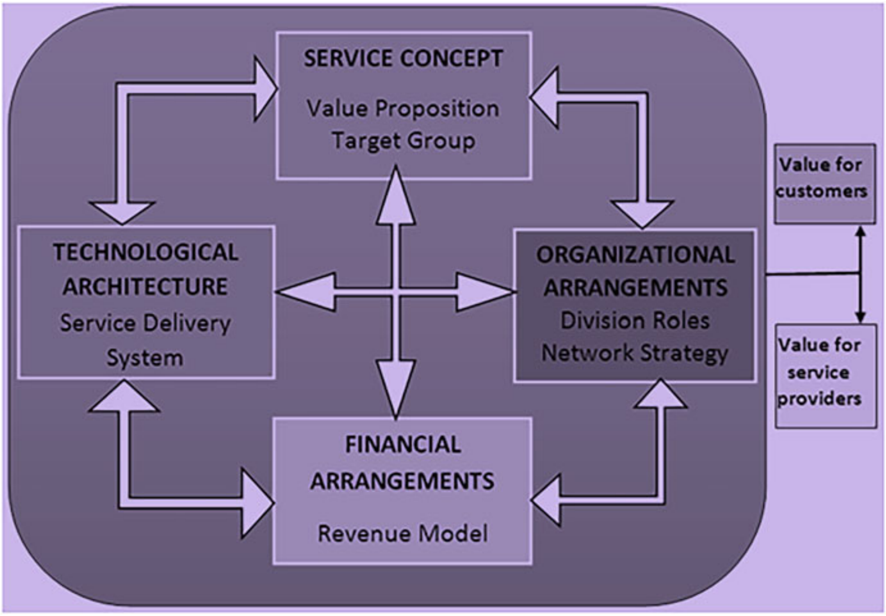



**Fig. 2** Framework of an e-health company [9]

### 5.3 There Are Six Additional Business Models for Healthcare Systems in Addition to These Two

1. *Changamka Microhealth Limited in Western Kenya—Scaling Proven Mobile Maternal Health E-Vouchers:* It is a social organization whose mission is to improve maternal healthcare services' accessibility. The Kenya Ministry of Health is connected with this organization. Mobile phones are essentially employed as a tool that is "at the center of every average person living in Kenya." The organization Samuel Agutu created intends to provide healthcare finance services to persons who are not a part of the Kenyan healthcare system. About 90% of the population is estimated to be in this. This project also distributes a variety of e-vouchers to assist low-income expecting moms in receiving care and a micro-insurance product. This enables mobile phone users to amass money to pay for healthcare facilities.
2. *Registries for ALL (Reg4ALL)—Genetic Alliance:* Genetic Alliance was founded by Ashoka Fellow Sharon Terry. One of the earliest cross-disease health registries in the world, it is controlled by the participant. It gives customers the authority or rights to choose how and with whom to share their health-related information. This approach can be used from clinical trials to obtain authorized registries and then to electronic data sharing that adheres to the stated requirements [11].
3. *Improving Access to Oral Healthcare—Sarrell Dental and Eye Centers:* Children in underserved regions will soon be able to get high-quality dental and eye care services, thanks to Sarrell Dental. Since it is a nonprofit organization, it keeps the cost to Medicaid to a minimum while still offering these services. There are 14 physical healthcare facilities. In addition to these facilities, Alabama offers rural residents a mobile dental bus service. Over 4,000,000 patients have been handled by Sarrell without a single complaint since 2004 [12].
4. *A WaterFirst Health Model—Healthpoint Services India Pvt. Ltd:* Through dual-way telemedicine services, Healthpoint Services' primary goal is to offer all essential personnel, resources, and Medicare services to those in need. These services are created in a way that makes them scalable and sustainable, allowing each household to receive preventative healthcare for a variety of issues for just Rs. 99.78 per month. With 140 operational water centers, this social group offers daily cheap healthcare services to 300 homes. By constructing 20 active water centers a month, Healthpoint Services will be able to reach more than three million Indians in the next 2 years [13].
5. *Health "Kiosks" for Kenya Slums—Access Afya:* CEO Melissa Menke and Director Duncan Goldie-Scot co-founded Access Afya, which has its headquarters in the USA. This organization's major goal is to create a high-tech network for offering paperless health centers in Kenyan slums. Registered nurses work for this company, creating and updating each patient's electronic medical record.

SMS messages are used to communicate patient care procedures, policies, and prescription directions. Additionally, health professionals sell products like water filters and clean cook stoves and coach patients on overall wellness [14].

6. *MeraDoctor—MeraDoctor:* This service focuses entirely on unrestricted medical advice from the advisory team in contact with accredited MMBS physicians. Basically, this is for low-income families all over India. This service is available by phone and is available all year long, around the clock. The cost of MeraDoctor membership is really low. The doctors are instructed to do this by asking callers about their difficulties, understanding their symptoms, and suggesting appropriate solutions. Up to this point, the general practitioners and specialists at MeraDoctor have handled about 400 illnesses over the phone. If MeraDoctor misses a call, the representative will register you and send you to the doctor as soon as feasible. There is also a REFUND service for dissatisfied consumers for improved customer experience [12].

## 6 Issues of the System

A healthcare system must guarantee continuous and secure data transfer in order to support remote patient data diagnosis and analysis. As a result, it facilitates emergency circumstances and minimizes patient and doctor round trips. Instead of PCs or other devices, smart mobile phones are the main way that Internet services are accessed in India. Also, cell phones are more efficient due to their lower power consumption and simplicity of use than computing equipment like laptops because sufficient power supplies are not always available. The system must have capabilities for Bluetooth pairing between the portable device and a mobile device. The data is also available to patients through the website. With the information, the doctor can create a prescription in the software with a special ID for the patient, which is then uploaded to the website.

The doctors, the remote center, and the web portal all have secure data communication. Smartphones are more efficient due to their lower power consumption and simplicity of use because sufficient power supplies are not always accessible for computing devices like laptops. The system needs to have Bluetooth connectivity capabilities to link the portable device with a mobile device. An important component that ensures data gathering in no-network situations, which is an issue in many rural areas of the country, is off-line data aggregation. The system must have robust, transportable interfaces.

## 7 Proposed Model

The suggested system uses a patient's health sensors to transmit data to an embedded kit. The embedded kit will use Bluetooth to transfer the health information it has gathered to the smartphone app. When there is an enough Internet connection, the mobile app can then cache the data locally and upload it to the cloud. The data is presented in the app in a readable style for humans. The doctor can obtain the patient data for analysis by logging into his account through the interface software. The data is made accessible to patients through an online site. With the information, the doctor can create a prescription in the software with a special ID for the patient, which is then uploaded to a website.

The patient can access the website to view both his or her health information and the doctor's warning. Each sensor's gathered data is unique and unstructured, making it challenging to process, store, and send. Thus, we require a sophisticated database system.

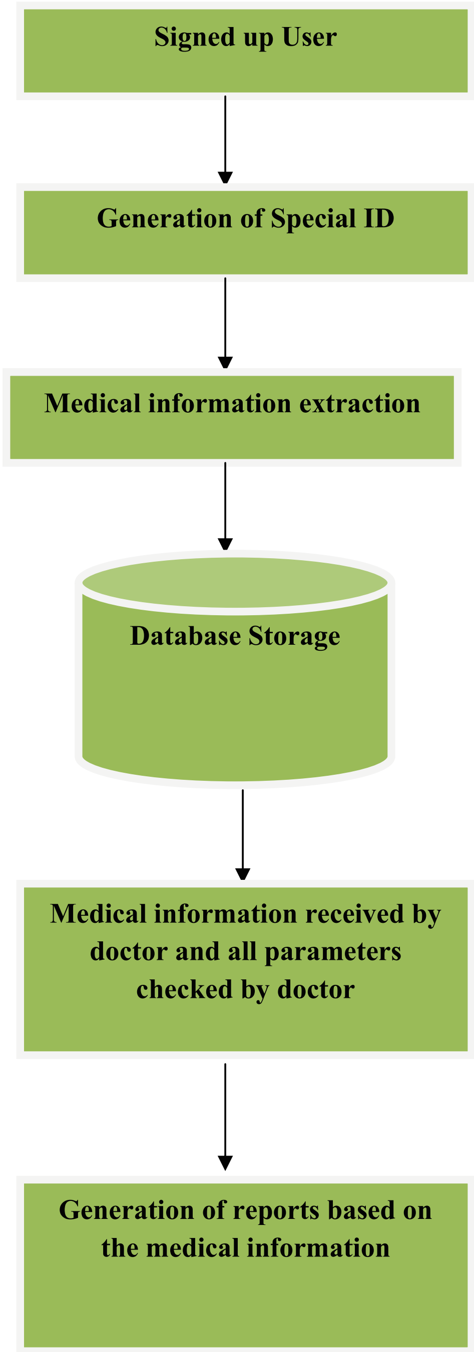
A portable embedded kit including ECG, EMG, a pulse oximeter, a blood pressure monitor, and other devices is the suggested solution. This kit collects various health data at the appropriate sample rate. It is also capable of obtaining photos of the mouth and eyes. As a result, the technology is portable and effective in locations with limited Internet and electricity.

Figure 3 depicts the proposed solution's operational principle. Bluetooth is used to link the mobile app to the hardware. When a button is pressed, the gadget receives a specific character relating to the sensor. In exchange, the gadget gathers some samples from the specific sensor and sends them to the app. When sufficient information is gathered, the connection is cut off. On the patient, one sensor is applied at a time. Following data collection, the mobile app caches the data and, depending on network availability, transfers the data to the cloud.

## 8 Summary

In this study, we are highlighting the developments of efficient and effective business models using Internet of Things to achieve success in the business. The smart systems are the driving forces nowadays to achieve excellence and success in businesses. Hence, if we could integrate the smart systems with our business functionalities and system by developing the appropriate business model, it can lead to better opportunities in the business world and also satisfying services for the people also as customer in the business models.

**Fig. 3** Proposed model



## References


1. Semwal, N., et al. (2019). An IoT based smart e-health care system. *Journal of Information and Optimization Sciences, 40*(8), 1787–1800.
2. Eysenbach, G. (2001). What is e-health? *Journal of Medical Internet Research, 3*(2), e833.
3. Sorwar, G., & Hasan, R. (2012). Smart-TV based integrated e-health monitoring system with agent technology. In *2012 26th international conference on advanced information networking and applications workshops*. IEEE.
4. Jog, Y., et al. (2015). Internet of things as a solution enabler in health sector. *International Journal of Bio-Science and Bio-Technology, 7*(2), 9–24.
5. Merilampi, S., & Sirkka, A. (2016). *Introduction to smart eHealth and eCare technologies*. CRC Press.
6. Ju, J., Kim, M.-S., & Ahn, J.-H. (2016). Prototyping business models for IoT service. *Procedia Computer Science, 91*, 882–890.
7. Balmer, J. M. T. (2001). Corporate identity, corporate branding and corporate marketing-seeing through the fog. *European Journal of Marketing, 35*, 248.
8. Gordijn, J. (2000). What's in an Electronic Business Model? Paper presented at Knowledge engineering and knowledge management: Methods, models, and tools. In *12th international conference*. Available at: http://www.cs.vu.nl/~hans/publications/EKAW2000.pdf (visited March 15, 2009); see also Gordijn, J. et al., Business modelling is not process modelling. In *Conceptual modelling for e-business and the web (ECOMO- 2000)*, Springer-Verlag, LNCS 1921, Salt Lake City, UT, (2000) October 9–12, pp. 40–51.
9. An interesting approach in this context is Buccoliero, L. (2007, May 4). A methodological and operative framework for the evaluation of an eHealth project. *International Journal of Health Planning and Management*, *23*, 3–20. (published online).
10. Kijl, B., Bouwman, H., Haaker, T., & Faber, E. (2005). Developing a dynamic business model framework for emerging mobile services. In *ITS 16th European regional conference, Porto, Portugal*.
11. Kwong, K. (2014). *Comparing genomic data sharing policies from the National Institutes of Health, Global Alliance, and Reg4All: Common ground and future directions*. Diss.
12. North, M. M., et al. (2022). Mobile applications utilization in the healthcare sector. *International Management Review, 18*(1), 88–107.
13. Torrinha, Á., et al. (2020). Application of nanostructured carbon-based electrochemical (bio) sensors for screening of emerging pharmaceutical pollutants in waters and aquatic species: A review. *Nanomaterials, 10*(7), 1268.
14. Christie, K. R. (2020). *Effects of stakeholder involvement on performance of telemedicine project in Nairobi County: A case of Access Afya*.

# Intangible Approaches to Improve Individual Health Indicators and Empower Caregivers

**Carlos R. Cunha, André Moreira, Luís Pires, and Paula Odete Fernandes**

## 1 Introduction

Populations worldwide are ageing at a faster pace than in the past, and this demographic transition will affect almost all aspects of society [1]. The number and proportion of people aged 60 years and more in the population are increasing; in 2019, the number of people aged 60 years and older was one billion; this number will increase to 1.4 billion by 2030 and 2.1 billion by 2050 [2]. Population ageing is an irreversible worldwide trend. This is the inevitable result of demographic transition, the trend towards longer lives and smaller families, which occurs even in countries with relatively young populations. In 2021, one in ten people worldwide was aged 65 and over; by 2050, this age group is forecasted to represent one in six people worldwide [3]. Globally, the number of people aged 80 years and over is increasing even faster than the number of people aged 65 years and over. By 2050, the world will have an estimated 459 million people aged 80 years and over, which is nearly triple the number in 2021, at approximately 155 million [3].

This finding was seen in all regions of the world where the number of older people increased rapidly between 1980 and 2021, a trend that is likely to continue for the next three decades [3]; even in Portugal, the proportion of older people as in other countries has been increasing steadily over the past decades. By 2050, Portugal expects an old-age dependency rate of 68.8%, which is one of the highest ratios in Europe [4]. The North interior of Portugal, namely, *Terras de Trás-os-Montes*, is following the same behaviour as what is being observed in the world and in Portugal. Since it is a region located in the northern interior of Portugal and is considered a

C. R. Cunha · A. Moreira · L. Pires · P. O. Fernandes (✉)
Applied Management Research Unit (UNIAG), Instituto Politécnico de Bragança, Campus de Santa Apolónia, Bragança, Portugal
e-mail: crc@ipb.pt; andre-moreira@ipb.pt; luica@ipb.pt; pof@ipb.pt



N. Gupta, S. Mishra (eds.), *Internet of Everything for Smart City and Smart Healthcare Applications*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-34601-9_11

region of low population density, it has been seeing an increase in the ageing index, and the ageing trends are evident. It is highlighted that in 2021, *Terras de Trás-os-Montes'* region presents an ageing index of 359.0 [5].

As fertility levels decrease, the percentage of younger people decreases, while the percentage of working-age adults and eventually older people increases. Population ageing is driven by more people living longer, healthier lives [3]. The increase in human life expectancy is a long-term trend, and survival at older ages has now become more widespread around the world. In general, an increase in the overall life expectancy reflects better health. The number of years lived in good health, or at least without severe disability, has increased in many parts of the world.

It is believed that every second person in the world holds anti-ageing attitudes, leading to poorer physical and mental health and reduced quality of life for older people, costing societies billions of dollars every year, according to the Global Report on Ageism, published by the World Health Organization (WHO) [1].

Given this concern with the significant increase in ageing, there is a need to promote and ensure people's well-being at all stages throughout the course of their lives. In this context, the role of technology has proven to be fundamental in the development of solutions for the promotion of medical health. Nevertheless, the role of technology in providing wellness care has emerged, as well as in helping caregivers, and different technologies have shown promise. The authors pointed out technologies such as the Internet of Things (IoT), extended reality (XR), virtual reality (VR), augmented reality (AR), mixed reality (MR) and machine learning (ML).

Following on from this, the purpose of this chapter is to present assignments aimed at the elderly population, prototyping technology-based solutions for creating mechanisms for measuring and promoting well-being. In addition, it presents support mechanisms for the activities of caregivers in nursing homes, as a way to empower caregivers and ensure better health and well-being.

This chapter is organized as follows: after this introductory section, a review of the relevant literature with regard to the use of emerging technologies to promote well-being and support the act of caring is presented in Sect. 2. Section 3 will present the conceptual foundations that will give rise to the prototyping approach. Next, in Sect. 4, the VR and MR prototypes, combined with ML, which were created to validate the conceptual model that was previously worked on, will be presented. Finally, some conclusions and final remarks that envision future work are presented.

## 2 Technology in the Context of Well-Being and Caregiving

The people's well-being plays a fundamental role in societies; however, few studies have been carried out on the relationship between well-being and technology [6].

In our opinion, we highlight the Internet of Things (IoT) which will support a whole re-engineering of how we can face well-being and the provision of healthcare and well-being. The use of XR has shown great potential for generating well-being –

VR has contributed to emotional and social well-being [7] and to stimulation, improvement and cognitive assessment [8], and according to [9], it can be a revolutionary tool for psychological interventions, having already been used in immersive therapies in mental disorders. MR, according to [10], is also identified as a contribution to support healthcare practices in nursing homes, where the use of MR glasses, combined with facial recognition, allows patient recognition and the automatic availability of relevant health data, to support caregivers' activities, thus ensuring caregivers' empowerment. Finally, the use of ML may allow, in a timelier manner, the detection of patterns of well-being or absence of well-being and, in this way, guarantee the increasing of proactivity in the detection of health and well-being problems, as well as providing medical or wellness care that maximizes physical and mental health.

The symbiosis between the use of technology and digital content has been transforming the health area, changing the ways of providing and obtaining healthcare, and, in this context, the use of XR, such as VR, can allow the assessment of behaviours, emotions and cognitions [7]. According to [8], using VR in experiences where comforting virtual environments are provided, which enhance calmness, seems to promote states of relaxation that, due to the little time people tend to have to relax, can be a very significant approach. In [9], a study is presented on the use of immersive and non-immersive strategies supported by VR in order to explore the emotional impacts when viewing 360° videos in age groups, highlighting the importance of positive emotions for better physical health and mental health of the elderly.

The importance of using technology is also highlighted in the caregivers' component, where several technologies have been introduced to improve the capacity to provide health and wellness services. This new reality leads to better and more personalized healthcare, but also to greater well-being for the elderly population who are often found in nursing homes. Training caregivers, formal or informal, with technology that allows them to act more proactively and efficiently should govern technology introduction strategies in the focused healthcare and well-being sector.

According to [10], MR also has great potential when we want to connect physically distributed caregivers who intend to achieve a shared goal. This is a typical scenario for situations where the caregiver needs to resort to other specialists to deal with situations that go beyond their competences – an example where MR can support collaborative work in a more immersive way.

As the area of health and well-being is information intensive, the use of artificial intelligence (AI), namely, techniques based on ML, is extremely promising due its ability to extract information and knowledge from raw data. This potential allows the detection of patterns of health and well-being translated by fluctuations and combinations of the most diverse biological and behavioural parameters. In this domain, the capacity for such detection is far superior to human capacity or to more conventional technologies based on pre-defined rules (i.e. by classical programming of computer applications).

The combination of technologies that allow creating immersive environments, rich in easily manipulated information and manipulated in a non-intrusive way to

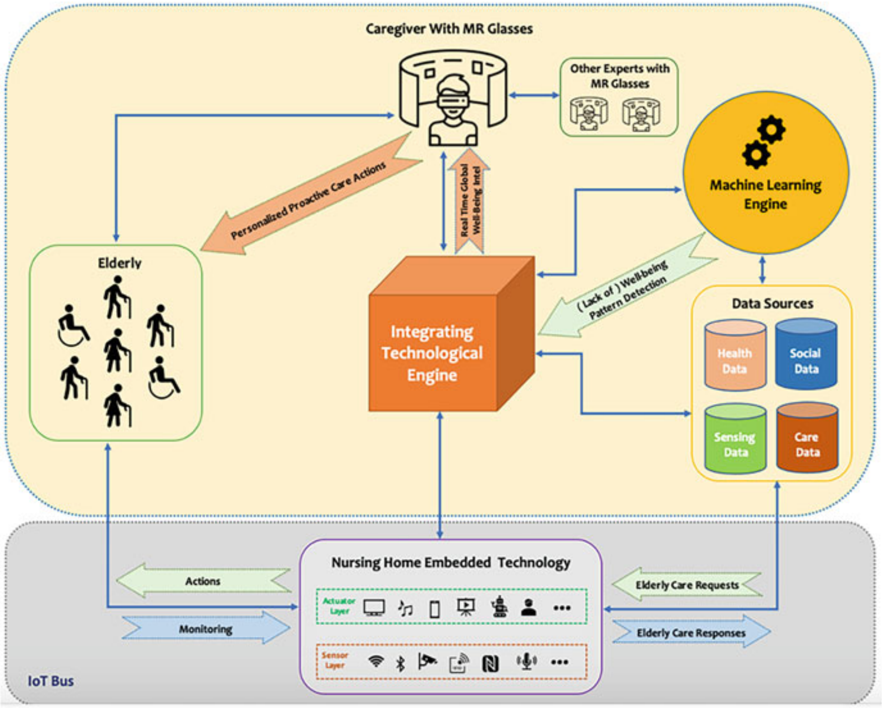



**Fig. 1** Conceptual model. (Source: Authors' own elaboration)

the normal performance of caregivers – VR, AR and MR – combined with the use of ML, will dictate, in our opinion, the way technology will be symbiotically allied with the health sector and the provision of healthcare and well-being. In this context, it is essential to think and design new information systems that take advantage of new technologies, creating innovative experiences that support new health and well-being practices that translate into increased well-being and empowerment of caregivers.

Next, the foundations that led to the development of a prototype are presented. These fundamentals present a conceptual model already presented in Fig. 1, which is summarized here in order to frame the developed prototype.

## 3 Conceptual Foundations

The conceptual foundations of this work – the conceptual model that have guided the development of a prototype – have already been published and are referred to as a way of better understanding the global vision that have supported the development of the prototype presented in the following chapter.

The conceptual model shown in Fig. 1 translates an integrated and cooperative vision of technology and support for cooperation between the different actors in the field of health and well-being. It is focused on the elderly and their caregivers, so it uses the concept of a nursing home as a test bed for its concepts.

The conceptual model starts from the credible principle that technology populates and will increasingly populate health and wellness spaces (e.g. nursing homes), making them increasingly smart spaces, where ubiquity and technological systems embedded in physical spaces will be a reality. In this context, it focuses its vision on the cooperative use of immersive technologies (e.g. MR glasses, such as the HoloLens® 2) capable of helping caregivers to access information and knowledge that allows them to act proactively. In addition, it proposes the use of immersive technologies (e.g. VR, AR) capable of creating immersive experiences of well-being in the elderly, contributing to the generation of emotional well-being.

This vision will tend to contribute to a better index of physical and emotional well-being that, along with all readings of biological parameters (medical acts), will feed full databases of information and knowledge about the elderly, as well as precious information for caregivers.

With the accumulation of data collected from the most varied sources, ideal scenarios will be created for the use of ML, which will enhance the detection of patterns of well-being or malaise. This detection will make it possible to personalize healthcare and anticipate scenarios of discomfort in a timely manner. In this way, we will be more efficient in providing healthcare and well-being.

## 4 Developed Prototype

This section presents the prototype that was developed to test the fundamentals of the conceptual model proposed in the previous chapter. Thus, Fig. 2 shows its general architecture. This prototype brings together a set of experiences using VR, AR and MR technologies combined with the use of ML. The architecture and implementation of various components of the prototype are explained in fine detail.

### *4.1 Database Layer*

Although it is possible to have our own servers hosting the databases, thus giving total control over them, cloud services are used for this prototype. Among the databases used, the main one is a graph database system.

**Neo4j (Graph Database)** Is a NoSQL database with features common to this type of database such as flexibility, easily scalable and absence of pre-defined rigid structures as in relational databases, an aspect however that differentiates them from other NoSQL databases is the fact that graph databases are strongly relational, with

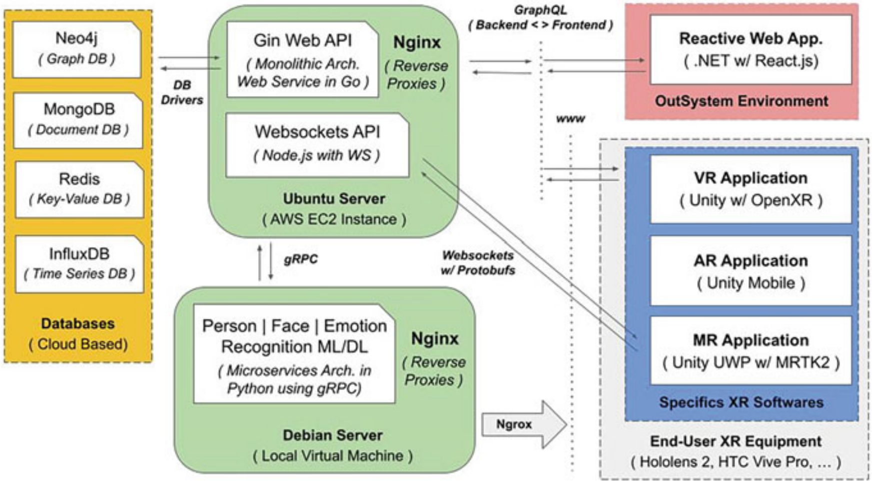



**Fig. 2** Prototype overall architecture

an additional value being given to relations not seen in relational databases, that is, since they are based on the structure of graph data, they can also store “data”, which would be the equivalent of weight in graphs [11].

Among the providers of graph databases, the more popular and first on the market is Neo4j®, created in 2007, together with the concept of graph database, by Emil Eifrem, which, in addition to services such as a database, also aims to provide data science services, that is, taking advantage of the acquired data and its structuring to convert it into knowledge and predictions; some examples are from the use of spanning tree and the use of algorithms such as Leiden for communities’ detection.

One of the disadvantages of using graph databases in favour of relational databases is that they are limited to storing large volumes of text or numbers, as well as storing non-relational data.

**MongoDB (Document Database)** Is a NoSQL database, being one of the most known and adopted in its category, more specifically, MongoDB® is a document database, treating data as JSON documents and grouping them into collections, unlike of relational databases, this type of database does not have a direct way of implementing relationships between data, although it is possible with the cost of storage, but with a gain in the speed of reading data. This type of database, although often used as a direct substitute for relational databases, is however best used when storing non-relational data; in this prototype, MongoDB® is used to store data relating to 360° hotspots used in the application of VR, as well as large volumes of data that cannot be stored in the graph database.

**Redis (Key-Value Database)** Is a key-value database, also considered NoSQL, it is mainly used to store small volumes of data and assigned it to a key, its structure can be compared to the data structure known as a dictionary, with great speed to access data and perform CRUD operations. Its use focuses on the implementation of VR session control through a Web application.

**InfluxDB (Time Series Database)** Is a time series database whose objective is to store temporal data for future analysis, being widely used for IoT environments, by storing data captured from various systems. The use of InfluxDB is foreseen to store data related to VR sessions whose objective is to capture temporal data, including sessions using eye tracking.

## *4.2 Server Layer*

For the development of this prototype, two servers were prepared, the local one using Debian 11 operating system, recognized for its stability, while the one presented on the cloud, using the Amazon Web Services (AWS), runs the Ubuntu Server 22.04 LTS, while planned to use both servers with Debian 11 OS, there were problems to establish a remote development environment with AWS running an instance with the Debian 11 OS. The two servers have different focuses, and the main services configured on each of them are presented below.

**AWS EC2 Instance** Is an instance on Amazon Web Services, using Amazon's cloud services, it is running two Web APIs on different ports, both built on a monolithic architecture, with request control being carried using the reverse-proxy tool, Nginx®.

*Gin Web API* Is a Web API developed on top of Go/Golang using the HTTP Web framework Gin. This Web API is used by all applications developed for data transition between these and the databases, using a GraphQL architecture, which is considered a direct and simplified alternative to the REST architecture; among the differences between the two, some more highlighted reside on the fact that GraphQL only uses POST requests and only one route [12], it is known as a query language, and it is also based on graphs, thus creating a synergy with graph databases and solving some of the main problems attributed to GraphQL, in particular the N+1 Problem [13].

*WebSockets API* Is an API built on Node.JS® and using the WS library, although it is possible to use WebSockets in Go, at the time this component was developed, the library for using WebSockets was archived and without maintenance, thus choosing an option that had support as well as good performance [14]. Currently, WebSockets are used to communicate with the application in MR, so that it can send the captured frames to be analysed by ML and then the inference results are sent. This communication is done consistently through the use of Protocol Buffers, also known as Protobufs, which result in smaller data volumes and faster data transition compared to other commonly used serialization methods like JSON [15].

**Local Virtual Machine** Is a virtual machine running on a computer. This second server is needed due to the limitations of the free version of AWS as well as to create an environment closer to what real use will be. This virtual machine is exposed to the World Wide Web using Ngrok which exposes SSH access, for remote development, as well as a TCP port. In this server was used a microservice architecture, in contrast to the monolithic architecture present in the AWS EC2 instance, this approach was adopted since this server consists mainly in the use of ML models, each one with is focus and dependencies, requiring then scalability, easy maintenance and independence between them [16], being then developed a set of microservices in Python focused on the use of ML, as well as a "gateway". These microservices communicate with each other using gRPC. This means of communication between services is also used for communication between the microservice described as "gateway" and the previously mentioned WebSockets API by using the exposed TCP port, which, after receiving the frame from the application in MR, will make a proto request to the "gateway", to this will therefore be analysed by the ML models.

Although not used in this prototype, it is possible to use Nginx® for load balancing, in order to distribute requests across different gRPC servers, for example, when using clusters, in order to make better use of available resources. This server contains:

*"Gateway"/Image Inference* Is a Python microservice that receives the request from the WebSockets API and is responsible for the flow of the frame analysis, that is, this microservice will communicate with the others to make the inference on the frame and it will format the final response to be sent to the WebSockets API and therefore to the MR application. The defined flow is purposeful and planned in order to fulfil each ML requirement as well as to avoid unnecessary inferences.

*Person Detection* Microservice makes use of the YOLOv7 [17] model, from the YOLO family (You Only Live Once), state-of-art in real-time object detection and the most recent at the time of the development of the microservice; this will detect people, mobile phones and televisions, then filtering detections of people inside these electronic devices by comparing the bounding box, this filtering will avoid possible repeated detections of the same person as well as minimize people to be analysed by ML. Since during the development process of this prototype YOLOv8 was released (which demonstrates better results than its predecessor), an update is planned soon.

*Face Recognition* Is the second ML in the flow; this will receive the frame as well as the result of the inference made by the YOLOv7 model. For the development of this microservice, the DeepFace library was used, which provides various analysis methods around the Face, from identifying the person through comparison with a "database" of images [18] to predicting age or gender, using state-of-the-art models [19].

For this prototype, in order to receive only the necessary data, as well as optimize the code in terms of performance, and to take advantage of gRPC and asynchronous programming, parts of this library were reprogrammed, and, in the end, the ArcFace

[20] model was used to identify people in the frame within the bounding boxes defined by the person detection microservice. This detection will result in the acquisition of UUIDs corresponding to existing patients in the graph database. This same library is used for the detection of facial expressions, limited to anger, fear, neutral, sadness, disgust, happiness and surprise. After face detection and UUID acquisition, the person detection results, that is, the bounding boxes, are filtered so that only those that have people in the database continue to the next phase.

*Emotion Recognition* Is the last microservice in the flow using ML, this makes use of a PyTorch implementation of a CNN (convolutional neural network) by Abhishek Tandon based on the methodology proposed in the same article where the EMOTIC dataset is presented, this article proposes the recognition of emotions through the analysis of the person's body as well as the environment that surrounds him, giving importance to the context, thus resulting in two types of emotional detections, continuous and categorical, the latter making use of a list of 26 emotions while continuous refers to three emotional dimensions, valence, arousal and dominance, each evaluated from 0 to 10 [21]. This model makes use of the YOLOv3 model as well as CNN developed on the Places 365 dataset [22] with a focus on the analysis of the environment.

To make use of this model, it is necessary to have the person's bounding box predefined; thus, using the bounding boxes resulting from person detection and filtered by face recognition, after inferring the frame about these same bounding boxes, the emotions detected are then returned to the "gateway".

This model as well as the EMOTIC dataset cannot be used for commercial purposes; however, its use for non-commercial research is permitted.

In the final phase of the flow, the "gateway" will format the various data acquired so that they can be serialized to Protocol Buffers and sent to the WebSockets API; the returned data boils down to the UUID of the detected person; the list of emotions, continuous and categorical; and the bounding boxes for the person's face and body. Although the model used for the detection of emotions used is more specific, the diversity of possible emotions makes it more vulnerable to false detections, so it is planned to combine it with the result of emotional detection from face recognition in order to increase the possibility of accurate detections.

### *4.3 End-User Applications*

At the end-user level, four applications are being developed, three of which fall under the category of XR applications with different users in focus, while the fourth is a Web application created for general use. The main equipment used to create immersive experiences are presented in Fig. 3.

**Web Application** Is being made using the low-code OutSystems platform; this allows the rapid development of the most common functionalities without losing the visual level; for this prototype, the most recent model is used, reactive Web

**Fig. 3** HTC® VIVE and HoloLens® 2 used equipment

application, which makes use of ReactJS with .Net, implementing functionalities that allow the execution of CRUD operations on the project's databases. Other more advanced features to be developed are:

- CRUD operations on the Faces database in the face recognition microservice
- Control over VR sessions through WebSockets
- Creation and manipulation of 360° hotspots as well as points of interest to be used in VR sessions using cloud storage and MongoDB®
- Visualization of data coming from data science results on Neo4j®

This application is focused on being used by different types of users, from caregivers in homes and doctors in hospitals to family members of patients.

**VR Application** Is a VR application being developed using OpenXR™ in Unity®, OpenXR™ being an API developed by Khronos® and supported by some of the main companies in the XR market, such as Meta®, HTC®, Valve™ and Microsoft®, with the objective simplify the development of cross-platform XR applications, that is, regardless of the type of headset used by the user. This application focuses on patients with the aim of conducting sessions to analyse their mental and health status.

Although in the current development phase the sessions are focused on the use of 360° hotspots with eye tracking, the objective is to implement the application in a way that can support other types of sessions such as using 3D scenarios; for this, a dynamic process is being planned to download all of the assets necessary for carrying out the session from a third server or cloud storage; these assets will vary from the 360° hotspots created in the Web application to 3D scenarios created in another context.

**AR Application** Refers to an application in AR; the purpose of this application is to be more accessible and for general use, since it can be used on most of the modern mobile phones.

Although the intention is to make use of AR, some of the foreseen functions are the use of Spectator View, that is, so that other people can see the same as caregivers without the need to use MR glasses as well as visualization of data from a closer to a mobile application.

**MR Application** Is a MR application developed in Unity® using MRTK2 (Mixed Reality Toolkit) developed by Microsoft®; although MRTK2 incorporates OpenXR™, the application is being developed with the HoloLens® 2 glasses as a target.

The application aims to assist medical caregivers, through the visualization of data corresponding to patients, among other functionalities that facilitate their work, as well as make use of the potential of technology. This application is structured in order to use mainly static classes and through the use of a managers/handlers architecture with a strong use of callbacks and actions.

As this is an application under active development and more advanced, some of the features, functional and non-functional, already implemented include:

*Embedded Database* refers to an object-oriented database called Realm developed by the MongoDB® team. The use of this database allows storing basic data in the application in order to be accessible even in situations where it is not possible to have an Internet connection; this type of concept is comparable to the use of SQLite in Android applications; however, SQLite is a relational database. Its use is being considered in the VR application, given the use of MongoDB® to store data related to VR sessions and the Realm SDK feature of synchronizing app data with MongoDB® Atlas.

*Pooling* is a term used in game development; it resumes with the reuse of GameObjects, such as 3D objects, that is, to minimize the use of the instantiate (GameObjects creation) and destroy (GameObjects destruction) methods that are found among the heaviest methods in terms of resource usage. Although it is an application and not a game, MR application makes use of this concept to reuse UI elements.

*UI Stackers* is the implementation of a logic equivalent to the operation of the UI in mobile applications, where the interfaces work as layers in stackers.

When developing games in Unity® and other game engines, UIs tend to be predefined and created from scratch with a static perspective, with only its content being changed, i.e. the flow of navigation UI works as if by a script, the same applies to mobile applications, however XR applications allow having N number of windows open, all of which can make use of the same interface, as well as with Web pages, in this way the existing interfaces in the MR application, as well as its components (buttons, input blocks, etc.) are pre-defined in a scriptable object that works as a dictionary of interfaces, from which the data referring to the interface needed to be instantiated are retrieved, among the stored data are references to the components, which are then updated the text and assigned Actions to the buttons, when this interface is no longer needed, more specifically the GameObject containing the interface. Will be put in "recycling" mode, that is, instead of being destroyed or given as non-existent, it will be placed in the Pool, and reused the next time that interface is needed instead of creating a new one. This method allows to have a dynamic UI and minimize the consumption of resources, however this method alone does not create the mentioned logic by itself, to achieve this objective, whenever an instance is created this is also stored in a Stack data structure, when navigating via the interface, if a new window is opened, it will create a new Stack or replace the previous window, saving a reference to it in a variable in order to be able to create the interaction or navigation flow.

A more specific aspect at the level of development of interfaces for MR with MRTK2 that differs from other scenarios resides in the fact that the equipment used, in this case the HoloLens® 2 glasses, may be unable to provide a good user experience if sprites are used, a method common when assembling interfaces in Unity®, due to the difficulty of rendering them; in this way, the interface must be assembled using GameObjects with materials assigned, and these materials will then have the sprite as texture.

*QR Code Reading allows* the reading of QR codes through the use of the Microsoft.MixedReality.QR plugin present in the NuGet package manager; this functionality is currently used for user authentication; however, its potential is unlimited, and it can be used for situations from the simplest ones, such as viewing the list of patients present in a hospital room in a new window by reading the QR code placed at the door, to controlling IoT systems.

*Web API Client* is the implementation that allows the Unity® application to communicate with the databases, for this implementation was used the asset Best HTTP/2, developed by Tivadar György Nagy and present in the Unity Asset Store, the communication involves since POST requests in GraphQL till WebSocket connections.

*Notifications/Alert System* is a notification system that makes requests to the Web API, currently referring to medications to be given to patients under the caregiver's or institution's responsibility, the response to the request will be primarily stored in the embedded database and therefore created coroutines that run in the background as timers, once the time has elapsed, an interface will be instantiated as a pop-up alerting the timer, the possibility of implementing a geolocation system that allows the caregiver to find the patient referring to the notification is being analysed, although this moment system is used to alert about medication to be used, it is possible to integrate with IoT systems or create a means of communication between the patient and the caregiver, in scenarios, for example, where the patient needs help and alerts the caregivers through the use an application or an IoT system. When data is saved in the embedded database, the caregiver will not be limited to the Internet connection.

*Patients and Emotion Recognition* is the main and most advanced functionality implemented in the MR application, and consists of capturing frames in byte array format, that is, the scenario in front of the user using the MR headset, which is serialized to Protocol Buffers and sent to the WebSockets API where it will go through the analysis process in ML, a process exposed in previous sections, after the inference result is returned, also in Protocol Buffers, the message is deserialized, and through the data referring to the bounding box's an interface is instantiated at the level of the patient's face demonstrating a material referring to the main emotion and a button through which a window referring to the patient's profile can be opened through the UUID.

Frame capture is performed through the MediaFrameReader class provided by the Windows Runtime API; this API is accessible given the fact that applications and games to be developed for HoloLens® 2 must be compiled with Universal Windows Platform (UWP) build settings; although there are other methods such as the use of the PhotoCapture class in Unity®, this method is the least expensive in resource consumption, with little effect on the application's performance; on the other hand, it delays development because it is necessary to compile the application in order to be tested, that is, it is not possible to test in Unity® itself.

After having the inference result, three methods are used to find the 3D location of the bounding box, the first method makes use of RayCast, however, although this method returns the correct direction, the distance is defined through the collision between the RayCast and the Spatial Mesh, called the Hit Point, however, while it identifies flat objects such as tables, chairs and computers, it does not recognize the shape of people, thus returning a false distance. The direction of the RayCast corresponds to the centre of the bounding box, which is converted to a 3D position; this method is based on the method used in the project made as proof of concept for the thesis "Mixed Reality Task Assistance for Pharmaceutical Laboratories using Markerless Object Tracking and Cloud Services" by Severin Pereto, named Lab Assist Vision, having been updated due to the use of deprecated methods and attributes.

The second method is used to find the distance, this method is based on a sample project provided by Microsoft®, Holographic Face Tracking, however, this project makes use of FaceAnalysis API from Azure, as well as it is limited to one Face, this method complements the previous one given the fact that while the identified distance is close to the real one, the detected direction does not correspond to reality, this may, however, be due to the way the inference results are returned, not having been tested.

The third and final method boils down to merging the results of the two previous methods in order to receive a 3D position closer to reality; this method returns the position on the axis between the user and the hit point, with the same distance to the user as the distance between this and the position resulting from the second method; and this method can be compared with the method Lerp provided by Unity®.

Currently, this functionality is working; however, there are problems related to an offset on the axis Y depending on the distance between the user and the patient.

In the period of time between requests to the WebSockets API, computer vision is used to create a tracking system, currently using CSRT trackers provided by the OpenCV for Unity asset, developed by Enox Software, present in the Unity Asset Store, the tracking however is unstable in longer periods of time.

While in the first stages of the development, the WebSockets API was made in Python using the FastAPI framework and serializing the data in JSON, to improve the performance it was planned to make use of gRPC for direct communication between the MR application and the "gateway" microservice, and thus minimize the

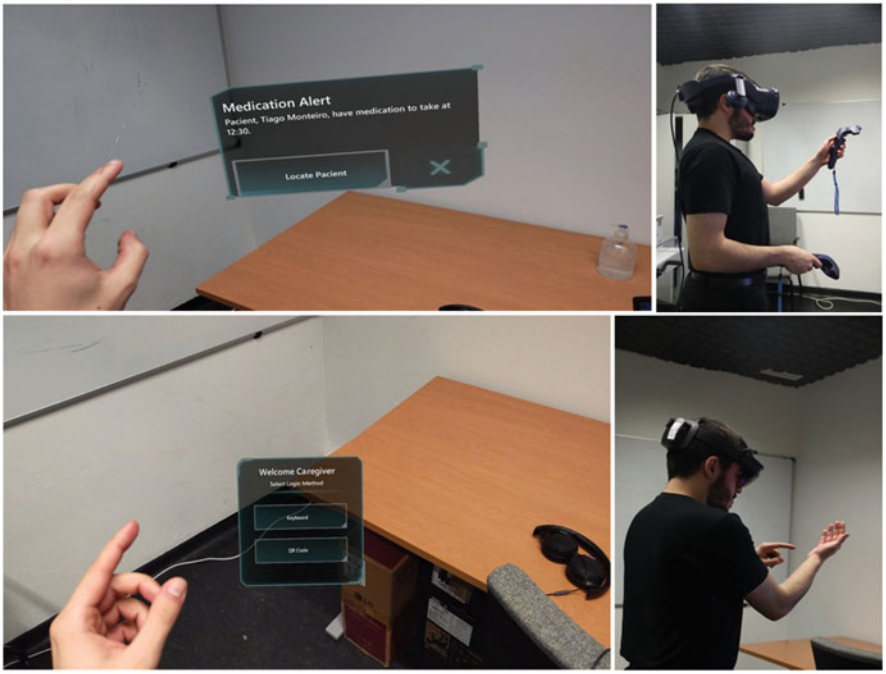



**Fig. 4** Some pictures of the prototype used in a laboratory context

time between requests, however given deprecations and incompatibilities due to the. NET version used by Unity®, it was used Node.js® with Protobufs, at the moment its being solved problems related to the deserialization of the binary message in Unity®, when forwarded from the WebSockets API to the MR application.

In addition to the already developed and under development features mentioned, some plans include refactoring blocks of code for parallel computing using Unity's job system package in order to take advantage of the HoloLens® 2 hardware, as well as the implementation of a method that allows the caregiver using the MR glasses to visualize the events in the patient's VR session.

Some illustrative images of the prototype in operation, in a laboratory context, can be seen in Fig. 4.

## 5 Conclusion and Final Remarks

The rapidly growing number of people reaching older ages highlights the relevance of promoting health and disease prevention and the treatment of diseases over the life cycle, as conditions experienced earlier in life can have a substantial impact on a person's life health and well-being in older ages. The concept of "healthy ageing" outlines the importance of maintaining functional capacity as people age to enable their ongoing engagement in society.

The development of solutions capable of generating immersive environments has great potential to be used by the elderly to generate experiences that enhance emotional well-being. In addition, they allow the assessment of emotional and mental states. In a complementary perspective, the creation of applications based on XR to help caregivers is an approach to empowering these same caregivers.

In a world where the use of embedded technology in physical spaces grows and where technology begins to be an extension of our body, it is important to re-engineer thinking in the creation of new and innovative solutions that help the management and self-management of emotional states and mental and physical health, as well as help caregivers.

This work presents a prototype that has its conceptual foundations in a previously designed model, which takes advantage of alternative realities combined with ML, to demonstrate what can be done for an innovative vision in actors linked to health and well-being.

The next step will be testing in a real environment (i.e. in a nursing home), in order to adjust and validate the developed prototype, to measure its impact on improving the physical and emotional well-being of the elderly as well as to understand the level of help it represents for caregivers.

**Acknowledgement** This work is funded by the European Regional Development Fund through the Regional Operational Program North 2020, within the scope of Project GreenHealth-Digital strategies in biological assets to improve well-being and promote green health, Norte-01-0145-FEDER-000042. The authors are grateful to the FCT Portugal for financial support by national funds FCT/MCTES to UNIAG, under project no. UIDB/04752/2020.

## References

1. W. H. Organization. (2020). *United Nation's decade of healthy ageing (2021–2030)*. World Health Organization. *[Google Scholar]*,.
2. W. H. Organization. (2022). *WHO guideline on control and elimination of human schistosomiasis*. World Health Organization.
3. United Nations. (2023). *World social report 2023: Leaving no one behind in an ageing world*, New York.
4. Eurostat, E. (2019). *Ageing Europe—Looking at the lives of older people in the EU*. EU, Stat. book.
5. PORDATA. (2022). *Ageing indicators*. Available online: https://www.pordata.pt/en/municipalities/ageing+indicators++according+to+census-1055
6. Kafaee, M., Ansarian, Z., Taqavi, M., & Heidari, S. (2021). Design for well-being: The fourth generation of technology development. *Technology in Society, 67*, 101775. https://doi.org/10.1016/j.techsoc.2021.101775
7. Valmaggia, L. R., Latif, L., Kempton, M. J., & Rus-Calafell, M. (2016). Virtual reality in the psychological treatment for mental health problems: An systematic review of recent evidence. *Psychiatry Research, 236*, 189–195. https://doi.org/10.1016/j.psychres.2016.01.015

8. Riches, S., Azevedo, L., Bird, L., Pisani, S., & Valmaggia, L. (2021). Virtual reality relaxation for the general population: A systematic review. *Social Psychiatry and Psychiatric Epidemiology*. https://doi.org/10.1007/s00127-021-02110-z
9. Liu, Q., Wang, Y., Yao, M. Z., Tang, Q., & Yang, Y. (2019). The effects of viewing an uplifting 360-degree video on emotional well-being among elderly adults and college students under immersive virtual reality and smartphone conditions. *Cyberpsychology, Behavior and Social Networking, 23*(3), 157–164. https://doi.org/10.1089/cyber.2019.0273
10. Marques, B., Teixeira, A., Silva, S., Alves, J., Dias, P., & Santos, B. S. (2022). A critical analysis on remote collaboration mediated by augmented reality: Making a case for improved characterization and evaluation of the collaborative process. *Computers and Graphics, 102*, 619–633. https://doi.org/10.1016/j.cag.2021.08.006
11. Kundu, G., Mukherjee, N., & Mondal, S. (2021). Building a graph database for storing heterogeneous healthcare data. In S. Tomonobu, P. N. Mahalle, T. Perumal, & A. Joshi (Eds.), *Information and communication technology for intelligent systems* (pp. 193–201). Springer Singapore. https://doi.org/10.1007/978-981-15-7062-9_19
12. Brito, G., & Valente, M. T. (2020, March). REST vs GraphQL: A controlled experiment. In *2020 IEEE international conference on software architecture (ICSA)* (pp. 81–91). https://doi.org/10.1109/ICSA47634.2020.00016
13. Hartig, O., & Pérez, J. (2018). Semantics and Complexity of GraphQL. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18* (pp. 1155–1164). https://doi.org/10.1145/3178876.3186014
14. Tomasetti, M. (2021). An analysis of the performance of Websockets in various programming languages and libraries. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3778525
15. Popic, S., Pezer, D., Mrazovac, B., & Teslic, N. (2016, October). Performance evaluation of using Protocol Buffers in the Internet of Things communication. In *2016 international conference on smart systems and technologies (SST)* (pp. 261–265). https://doi.org/10.1109/SST.2016.7765670
16. Alshuqayran, N., Ali, N., & Evans, R. (2016, November). A systematic mapping study in microservice architecture. In *2016 IEEE 9th international conference on service-oriented computing and applications (SOCA)* (pp. 44–51). https://doi.org/10.1109/SOCA.2016.15
17. Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2022, July). *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*. Accessed: 17 Feb 2023. [Online]. Available: http://arxiv.org/abs/2207.02696
18. Serengil, S. I., & Ozpinar, A. (2020, October). LightFace: A hybrid deep face recognition framework. In *2020 innovations in intelligent systems and applications conference (ASYU)* (pp. 1–5). https://doi.org/10.1109/ASYU50717.2020.9259802
19. Serengil, S. I., & Ozpinar, A. (2021, October). HyperExtended LightFace: A facial attribute analysis framework. In *2021 international conference on engineering and emerging technologies (ICEET)* (pp. 1–4). https://doi.org/10.1109/ICEET53442.2021.9659697
20. Deng, J., Guo, J., Yang, J., Xue, N., Kotsia, I., & Zafeiriou, S. (2018, January). ArcFace: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*. https://doi.org/10.1109/TPAMI.2021.3087709
21. Kosti, R., Alvarez, J., Recasens, A., & Lapedriza, A. (2019). Context based emotion recognition using EMOTIC dataset. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42*(11), 1–1. https://doi.org/10.1109/TPAMI.2019.2916866
22. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2018). Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(6), 1452–1464. https://doi.org/10.1109/TPAMI.2017.2723009

# Edge Computing and Network Softwarization for the Internet of Healthcare Things

**Christiano A. P. Rodrigues, Victória Tomé Oliveira, Dario Vieira, Marciel Barros Pereira, and Miguel Franklin de Castro**

## 1 Introduction

The population of the globe is ageing. We can attest to this by analyzing global data on increased life expectancy and reduced birth rate. Along with this, new socioeconomic challenges emerge from the desire for a healthy and well-being long life. Naturally, this desire has driven the development of researches focused on improving the elderly Quality of Life (QoL) [1]. These researches become even more relevant during health crises like the recent pandemic.

However, the term QoL, although long-discussed, is still very broad and can be observed from many perspectives. Elderly QoL may be related to the absence of chronic diseases, perception of loneliness, physical well-being, understanding of the aging/death process, among others. In this work package, we take as reference the definition of QoL proposed by the World Health Organization (WHO) described

C. A. P. Rodrigues (✉) · V. T. Oliveira · M. B. Pereira
Efrei Research Lab, Paris, France

Federal University of Ceará (UFC), Fortaleza, Brazil

Group of Computer Networks, Software Engineering and Systems (GREaT), Fortaleza, Brazil
e-mail: christiano.rodrigues@efrei.fr; victoriat.oliveira@alu.ufc.br; marcielbp@ufc.br

D. Vieira
Efrei Research Lab, Paris, France
e-mail: dario.vieira@efrei.fr

M. F. de Castro
Federal University of Ceará (UFC), Fortaleza, Brazil

Group of Computer Networks, Software Engineering and Systems (GREaT), Fortaleza, Brazil
e-mail: miguel@ufc.br

 
N. Gupta, S. Mishra (eds.), *Internet of Everything for Smart City and Smart Healthcare Applications*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-34601-9_12

as the individual perception of your life in a sociocultural context and concerning goals, expectations, and personal standards.

Along with the increase in the number of older adults, the number of devices with computing power, communication, and sensory capabilities surrounding us also grows, particularly those provided with wireless communication technologies. These have significantly multiplied, which has made it easier to get more data on many scenarios. Although the current network was not thought to enable this situation, the Internet of Things (IoT) paradigm is familiar to academia and industry.

The Internet of Things' significance in daily life is growing due to its wide use in various applications, such as supplying patients with emergency services, healthcare monitoring systems, and E-health applications for a variety of purposes, including the early identification of medical problems, emergency alerting, and computer-assisted rehabilitation [2].

Technologies for continuous monitoring and assistance allow systems and applications to provide data on the behavior of patients outside a hospital environment and within their daily lives. Through the collection and analysis of these data, issues related to aging and measures of quality of life can be used to monitor, guide, and provide the necessary knowledge for decision-making.

This chapter is organized as follows: the Sect. 2 will present the IoT, Cloud Computing (CC), and Fifth-Generation Mobile Networks (5G) concepts we consider noteworthy for the evolution of health systems. After presenting these notions and how they relate to the healthcare scenario, we present in Sects. 3, 4, and 5 some of the most relevant technologies for 5G, edge computing, intelligent edge, and network slicing, respectively. Then, we present in Sect. 6 the healthcare scenario and how all the technologies mentioned earlier relate to it. Finally, we conclude this chapter in Sect. 7.

## 2 Improving Healthcare Systems

In this section, we will discuss three main technologies, the IoT, Cloud Computing, and the 5G. These three, in our opinion, are the primary building blocks for modern healthcare system improvement.

### *2.1 The Internet of Things*

Kevin Ashton initially coined the IoT to describe a system where the physical world is connected to the Internet via ubiquitous sensors [3]. IoT has been receiving broad definitions since the end of the 1990s. It now encompasses a wide range of sectors, including household, business, healthcare, and utilities, to mention a few.

If hastily reduced to its acronym, IoT would conceal its numerous facets. That is the reason why there are still many different definitions for the Internet of Things, such as:

- According to [4], there are three paradigms that IoT can be acknowledged: Internet-oriented (middleware), things-oriented (sensors), and semantic-oriented (knowledge). A knowledge base is necessary to interact and carry out these three visions. The main focus for creating and integrating the objects into the network will be the intersection of these visions. No matter the network, things will actively participate in business, information, and social processes.
- IoT was also described by Singh et al. [5] as the outcome of three visions: Internet vision, things vision, and semantically oriented vision. The fact that we can monitor everything using sensors and widespread technology such as Radio-Frequency Identification (RFID), according to the authors, represents the first vision. The necessity to link smart objects and ensure that they adhere to TCP/IP architecture has been pushed by the Internet-oriented vision. Semantic-oriented vision is fueled by the fact that there will be an enormous number of sensors accessible, and an enormous amount of data will be collected. As a result, a significant volume of information has to be processed.
- Finally, according to [6], IoT is a global infrastructure for the information society that enables improved services by connecting (physical and virtual) objects using current and developing interoperable information and communication technologies. The Internet of Things is a paradigm that connects physical and virtual entities that can be recognized and included in communication networks.

IoT is a cutting-edge technology with broad applications in smart cities, industrial control, retail, waste management, emergency services, security, and logistics. Most significantly, IoT is an attractive paradigm that has the potential to completely transform the way healthcare is provided today through a range of innovative and highly personalized solutions. Additionally, IoT-based solutions have the potential to lower costs through enhanced user experience, shorten the time needed for remote health service, and improve care quality.

These solutions include remote health monitoring, remote diagnosis, chronic disease management, elderly independent care, monitoring of patient adherence to treatment, medicine authentication, drug supply monitoring, and scheduling to ensure their best use in large scales [7].

Over the past several years, our environment has become increasingly digital, with biomedical and wearable sensors, cell phones, and electronic devices all producing vast amounts of health data. Not only health-related, but IoT has the potential to connect almost everything on our planet to everything else. However, IoT systems are intricately designed and have a certain amount of storage and retrieval space. Numerous IoT applications and services might benefit from the Cloud Computing model [8] that we will discuss next.

## 2.2 Cloud Computing

Cloud computing (CC) is a service delivery model that has been successfully used in recent years. According to [9], a vast set of virtualized resources (such as hardware, development platforms, or services) are available and useable through cloud computing. These resources can be dynamically reconfigured to adjust to a variable load, allowing optimized use of resources. Cloud providers typically use a pay-per-use model, and the provider must offer infrastructure guarantees through customized service-level agreements.

Because of its benefits, cloud computing has been used by businesses in many different endeavors, like healthcare. One advantage of using cloud resources is that it enables businesses to manage their information and communication technology resources with greater agility, increasing their competitiveness as a result. In addition to this benefit, cloud computing enables small and new businesses to lower the entrance barrier to the market, allowing operating costs to increase proportionately with revenue growth.

Cloud has the potential to become the foundation of IoT healthcare systems [10]. Additionally, cloud computing makes it possible for health professionals, carers, and patients to share information in a more structured and organized manner, reducing the loss of medical records [11]. Consequently, healthcare systems have much to gain from IoT and CC.

Applications built using cloud concepts benefit from resource-efficient multi-tenancy techniques, elastic growth, and ease of updating and maintenance. Building cloud-centric applications still present several difficulties. These challenges involve data privacy, placing applications in the cloud, and migrating them geographically to guarantee the quality of service.

## 2.3 5G Network

The IoT is familiar to the academy and the industry, although the current network was not thought to support this scenario. The first solution to several needs that will enable IoT to become a reality is the 5G wireless network. In addition, a new network is necessary due to the growing applications' diversity and needs. Considering these obstacles, the fifth generation is anticipated to have much higher speeds, support a massive number of devices, and lower energy consumption and latency.

5G will not only help IoT but has already created various opportunities for applications and services. 5G is considered a network that will support a diverse set of performance and service requirements, which will aid in the development of a wide range of vertical markets.

Future 5G will combine evolutionary- and service-oriented viewpoints. While the second view aims to provide a variety of different services with different requirements, which will include both human-type communications and machine-type communications, the first view concentrates on scaling up and improving the efficiency of mobile networks by increasing traffic volume, the number of devices supported, throughput, etc. [12].

These two viewpoints form the basis of the 5G desired capabilities, which are difficult to achieve because the current network is based on a one-size-fits-all concept. In contrast, 5G should be adaptable, programmable, and use open interfaces. As a result, a multi-tenant ecosystem will be created, promoting the development of new business markets [13].

As was previously stated, the fifth generation of mobile communications, or 5G, is more than merely an incremental improvement over earlier generations. With previously unheard-of data rates, latency, connectivity, dependability, and energy efficiency, the fifth generation is intended to be a cutting-edge network [14]. For 5G, there are eight major prerequisites [15, 16]:

- 10 Gbps—It is an increase of almost ten times compared to the traditional Long Term Evolution (LTE) network's data rate.
- Round-trip time (RTT) of 1 millisecond—A ten-fold reduction when compared to the LTE network.
- High bandwidth within the same zone—Allows several connected devices with greater bandwidths to stay connected for a longer time in a specific zone.
- Billions of linked devices—5G networks must be able to connect a large number of devices.
- 99.999% availability—The network will essentially always be available with 5G.
- Nearly full coverage—IoT aspires to provide access to anything, anytime, everywhere. Therefore, 5G wireless networks must ensure complete coverage wherever users may be.
- Approximately, 90% less energy will be used—With 5G wireless's high data speeds and widespread connection, energy utilization will become even more crucial. It is also necessary to take green technology development into account.
- Long-lasting battery—Devices must consume less power in 5G networks.

The LTE network was based on a one-size-fits-all idea, convenient for everyone or every scenario. Nevertheless, 5G must be versatile and adaptive to offer the best support for the many use cases and situations. The specifications for various scenarios have been categorized into three primary groups by the International Telecommunication Union (ITU) and Fifth Generation Public Private Partnership (5G-PPP) [17] and are presented next. Performance standards for these categories are shown in Fig. 1 and Table 1:

- **Enhanced mobile broadband (eMBB)**—It incorporates what we already have while taking into account improved performance and user experience. This group of examples includes access points and wide-area coverage, among others. Continuous coverage, medium to high mobility, increased data rate, and medium to broad-range coverage are desired. Given that mobility is restricted, the access points must support locations with high user densities and a significant volume

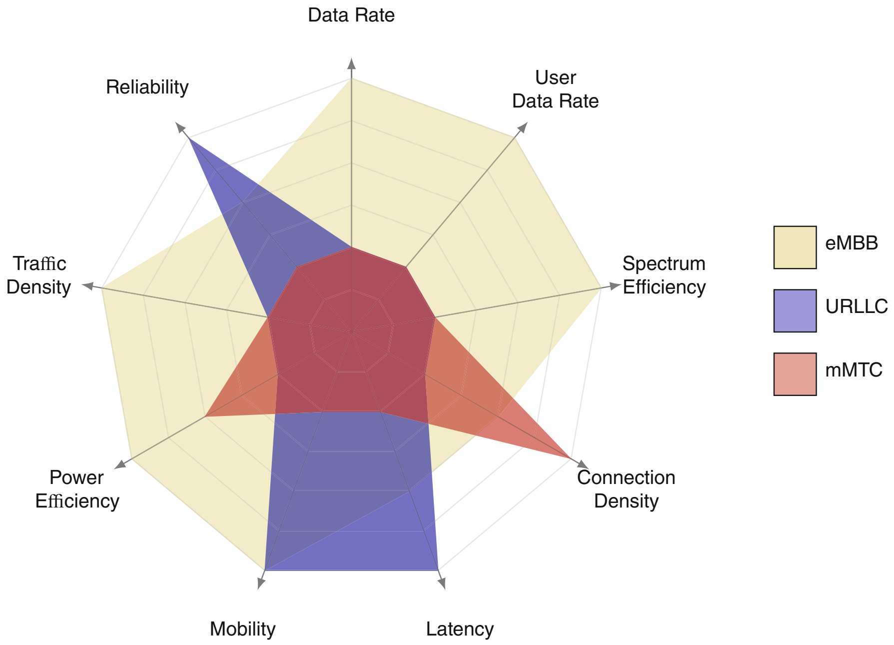



**Fig. 1** Different requirements in different use cases

of data traffic. Additionally, the user data rate is higher than the wide-ranging coverage.

- **Ultra-Reliable and Low Latency Communications (URLLC)**—Flow, latency, and availability requirements for this use case are rigorous. This use case includes, among other things, automated distribution in a smart grid, wireless control in industrial manufacturing, intelligent transportation systems, remote medical surgery, public safety, and emergency support.
- **Massive Machine Type Communications (mMTC)**—This final category includes scenarios where there is a massive number of connected devices, often sending a small amount of non-delay sensitive data. The devices in this scenario ought to be reasonably priced and have long-lasting batteries.

5G is ideally suited to offer the connectivity required among the many healthcare system stakeholders. Ubiquitous and pervasive wireless technology in healthcare systems has dramatically increased thanks to wearable computing, bio-engineering, and mobile device developments. The URLLC scenario introduced by 5G is a crucial one that can enable a wide range of applications that depend on mission-critical networks, such as remote diagnosis and surgery, with haptic feedback [19]. Another important scenario for healthcare is the mMTC since it can facilitate all the Internet of Healthcare Things (IoHT) devices to exchange data, voice, text, images, and other health-rated files.

**Table 1** Key performance indicators (KPIs) for the 5G use cases [18]

| KPI | Use cases | Values |
|---|---|---|
| Maximum data rate | eMBB | DLL: 20 Gbps, Uplink (UL): 10 Gbps |
| Maximum Spectral Efficiency | eMBB | DLL: 30 bps/Hz, UL: 15 bps/Hz |
| Latency of user plane | eMBB, URLLC | 4 ms for eMBB and 1 ms for URLLC |
| Control Plane Latency | eMBB, URLLC | 20 ms for eMBB and URLLC |
| Density of connection | mMTC | 1,000,000 devices /km$^2$ |
| Energy Efficiency | mMTC | Able to withstand long sleep periods (sleep mode) to allow low power consumption when there is no data |
| Reliability | URLLC | Succession of 1–$10^{-5}$ to transmit a 32-byte data unit in a Layer 2 protocol in 1 ms with the quality of the channel at the edge of the cover |
| Mobility | eMBB | Up to 500 km/h |
| Mobility interruption time | eMBB, URLLC | 0 ms |

The advancement of fifth-generation wireless technology, which offers significantly higher throughput, low latency, and personalized services, is projected to continue helping the development of healthcare systems and applications. However, it is essential to emphasize that the requirements for these different use cases shown in Fig. 1 and Table 1 are remarkably challenging and diverse. That is why Edge Computing and Network Slicing have recently gained so much attention. Both technologies will be discussed next.

## 3 Edge Computing

The continued growth of the IoT brings more connected devices to collect and consume data on the network. Due to the resource constraints of these devices, a common approach is to use the seemingly abundant resources that exist in Cloud Computing (CC), to run, analyze, and process the data remotely. However, some IoT application areas such as factory automation, smart grids, and transportation systems have critical end-to-end low-latency requirements as they require real-time responses, which are hardly provided by resources offered by CC. Edge Computing emerges as a viable solution to overcome the existing limitations of CC, supporting performance-critical Internet of Things applications.

Several similar concepts have emerged in academia and industry to overcome some limitations of just using cloud computing. Among these concepts are Cloudlet [20], Fog Computing [21], and Mobile Edge Computing/Multi-access Edge Computing [22]. The common denominator of these concepts is the extension of cloud computing, bringing cloud services and resources closer to end users on geo-distributed nodes at the edge of the network [23]. We use the term edge computing

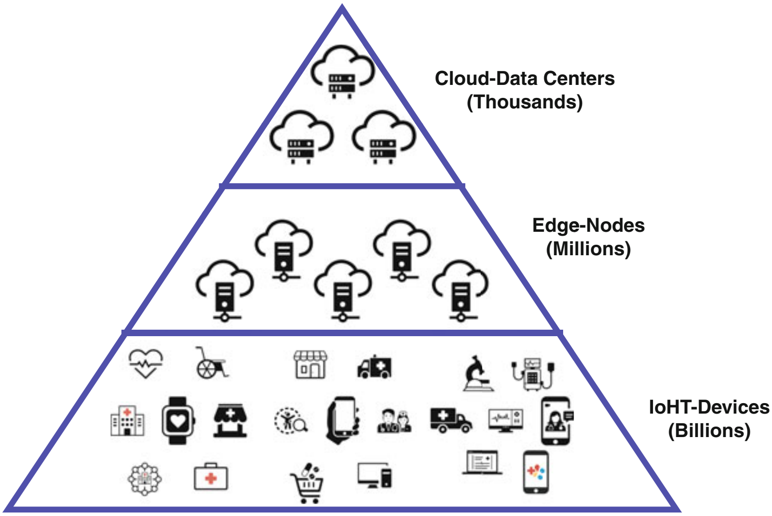



**Fig. 2** Simple three-tier architecture for IoT-based healthcare systems

to encompass these partially overlapping and complementary concepts. In general, edge computing adds a new layer of general-purpose distributed computing nodes between end-user devices and cloud data centers, as shown in Fig. 2.

In recent years, Edge Computing (EC) has gained considerable popularity in academic and industry circles. It serves as a key enabler for many future technologies, such as 5G, Internet of Things, augmented reality, and vehicle-to-vehicle communications, connecting Cloud Computing facilities and services to [24] end users.

Edge Computing is a new concept in the computing landscape. It brings the services and utilities of cloud computing closer to the end user and is characterized by fast processing and application response times. Internet-enabled applications developed today, such as surveillance, virtual reality, and real-time traffic monitoring, require fast processing and fast response time [24, 25]. End users typically run these applications on their mobile devices with limited resources, while the core service and processing run on *cloud* servers. Leveraging *cloud* services over mobile devices results in high-latency and mobility-related problems [26, 27]. EC addresses the application requirements mentioned above by bringing processing to the edge of the network.

Despite the similarities, EC has distinct characteristics from CC, as shown in Table 2. The main difference between EC and CC is in the location of their nodes [28]. Edge nodes are widely distributed geographically, while CC has nodes in some centralized locations. On the other hand, CC uses large centralized data volumes with high resource capacity as nodes, while the edge nodes are more heterogeneous and limited in terms of resource capacity. Due to this limited capacity, edge nodes take up less space than cloud nodes and can be located one or a few hops away

**Table 2** Comparison between Cloud computing and Edge computing [Author]

| Feature | Cloud computing | Edge computing |
|---|---|---|
| Infrastructure | Centralized | Distributed |
| Geographic distribution | Locally Clustered | Distributed |
| Resource capacity | High | Low |
| Heterogeneity | Low | High |
| Latency | High | Low |
| Distance from End Users | Far | Near |
| Mobility | Limited | Supported |

from the end user's device network [29]. This proximity to users allows *Edge Computing* to offer significantly lower latencies when compared to CC. In addition, edge nodes can provide real-time, local contextual information to applications (e.g., user mobility and network status) that is not available or limited in the CC [30].

Despite presenting promising ideas to enable the implementation of IoT networks, EC also faces many challenges, especially those related to resource management in a vast, distributed, dynamic and heterogeneous scenario. One relevant challenge for resource management is service allocation, which can be defined as the decision-making process that will determine where to allocate services in the EC infrastructure, based on a set of constraints, requirements, and performance goals. In this way, this decision-making process can be extended to include and solve other network problems, such as load distribution and service migration. Service allocation and migration and load distribution are crucial challenges that must be addressed to achieve efficient resource management in EC [31, 32].

## 4 Intelligent Edge

Artificial intelligence is revolutionizing almost every branch of science and technology. Given the ubiquity of smart mobile devices and IoT devices, for the future, most smart applications are expected to be deployed at the edge of wireless networks. From this, strong interest has been generated in developing an "Intelligent Edge" to support applications developed using AI on various edge devices. Thus, a new area of research has emerged, called the Intelligent Edge, which crosses and revolutionizes two disciplines: wireless communication and artificial intelligence [33].

Intelligent edge proposes to incorporate artificial intelligence into edge computing for dynamic and adaptive management and maintenance of applications at the edge. With the development of communication technology, network access methods are becoming more diverse. EC infrastructure acts as an intermediate medium, making the connection between smart end devices and the cloud [34].

Thus, smart end devices, edge, and the cloud tend to merge into a community of shared resources.

However, managing and maintaining such a large and complex global architecture involving wireless communication, networking, computing, storage, etc., is a major challenge [35]. Typical and more conventional network optimization methodologies are based on static mathematical models; however, it is difficult to accurately model rapidly changing edge commuting network environments and systems. Artificial Intelligence (AI) is expected to deal with this problem: when faced with complex and cumbersome network information, AI can respond with its powerful learning and dynamic reasoning capabilities to extract valuable information from data and make adaptive decisions, achieving intelligent maintenance and management accordingly [36].

With the advancement of new technologies, patients today demand a personalized, intelligent, sophisticated, and advanced healthcare framework that is tailored to meet their needs. In conjunction with 5G and intelligent IoT sensors, edge computing provides real-time, intelligent healthcare solutions. Recently, many researchers have proposed methods based on hierarchical computing to leverage machine learning and deep learning techniques for the distribution and allocation of inference-based tasks among edge nodes, which could tremendously improve the computing power and computational capabilities of edge devices [37–39].

EC can be used with numerous edge devices and local servers in order to do collaborative and efficient processing of healthcare sensor data. Intelligent edge architectures can be fully or partially trained at the edge level, while additional processing can be distributed between edge and fog nodes, or cloud processing can be done for computationally intensive applications. From the rise of smart sensors and IoT devices, the Internet of Everything (IoE) [40] has become viable. Edge intelligence can also be applied in platforms for Industry 4.0 and Healthcare 4.0, making IoT architectures smarter and more resilient [41].

## 5 Network Slicing

As stated before, 5G is treated as a network that will support various sets of service requirements, which will aid in developing a wide range of vertical markets, including healthcare. However, it is crucial to emphasize that the requirements shown in Fig. 1 and Table 1 are challenging and varied. Due to this, a one-size-fits-all network will not be able to meet all of these needs.

Using the network as a service model, or Network as a Service (NaaS), is one technique to overcome the difficulty of supporting various use cases. Put another way, logical (virtual) networks are built on the same physical infrastructure and tailored to function for a particular use case. For instance, eMBB applications require a large bandwidth to offer services with high data rates, such as streaming high-resolution videos. Applications URLLC (like autonomous driving) require reliable, secure communication and low latency. For each of these categories, a virtual network may be built tailored to applications' needs.

There is a broad consensus that natively flexible and programmable networks that allow network slicing and softwarization are essential for implementing 5G's various services. The whole network architecture may be virtualized using Software Defined Networking (SND) and network Function Virtualization (NFV) technologies to create end-to-end virtual networks [12].

Various architectures were proposed and studied in the literature [12, 42, 43], and even though there is not a standard one, these solutions have similarities that can be mapped, at least at a high level, in three main layers and an entity as described below [12]:

- Infrastructure/resource layer. This layer provides a pool of computational, storage, network, and radio resources that can be virtualized and allocated to slices.
- Network function layer has a set of virtual and real network functions that run on the resources provided by the infrastructure layer.
- Service/application/business layer. High-level descriptions of end-to-end network services are provided by this layer, along with interfaces for adding new services.
- Management and orchestration entity. This entity is in charge of putting slices into action utilizing the infrastructure resources and network functions in a chain. Additionally, it oversees these slices' full life cycle.

The decoupling of network infrastructure and network services is made possible by network slicing, which can increase resource efficiency. This implies that businesses, such as Mobile Network Operators (MNO), are not required to manage and maintain infrastructure and services independently. It is feasible to lease network slices to multiple tenants, for example, Service Providers (SP) may construct services and provide them to end users without owning their network equipment.

Figure 3 shows four networks. At the bottom is a physical network composed of several devices and links between them. The top three networks are virtual and fulfill different requirements. Each virtual network uses resources from a pool of available resources to compose a slice that will meet the needs of a specific 5G use case.

A network slice is an end-to-end logical network composed of different resources from a physical infrastructure that can belong to different companies, and its resources can be shared among different tenants. Each network slice is an instance of an end-to-end logical network that is independent and isolated from each other. A slice can be thought of as a set of subsets, which means we have allocate enough resources on all parts of the network to meet the desired requirements for that specific slice [19].

Network slicing is promising to handle and offer reliability for healthcare applications [44]. Network slicing enables the virtualization of the physical network into discrete virtual subnetworks, providing flexibility, quick adaption, and affordable operation. In the next section, we will discuss the healthcare industry in more detail.

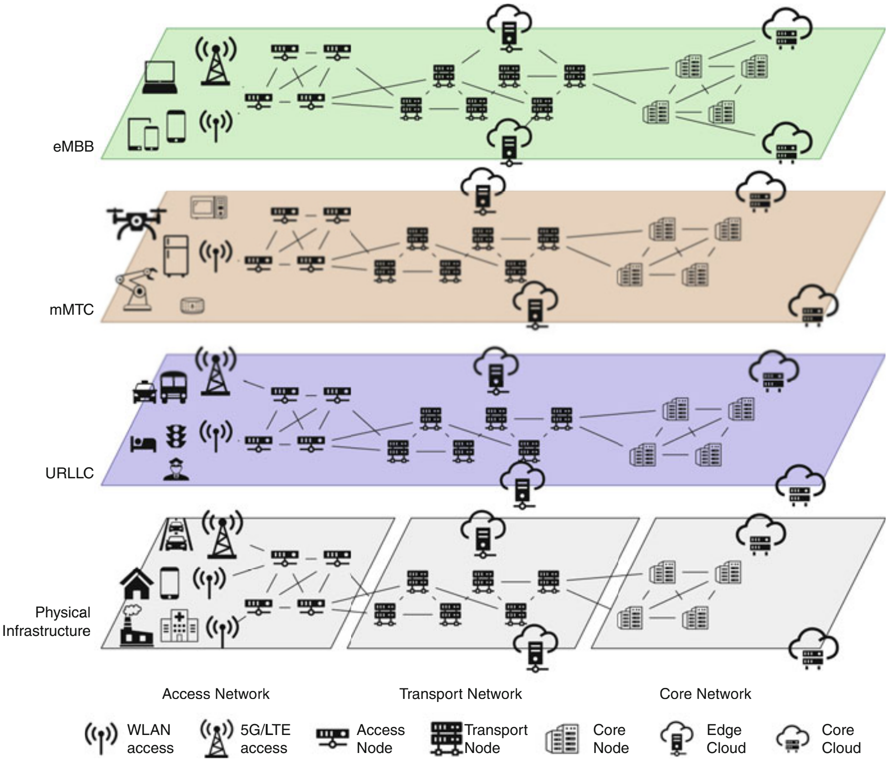



**Fig. 3** Multi-tenancy sliced network

# 6 Healthcare

At the beginning of the twentieth century, multiple technologies, such as the radio and telephone, improved and made medical information easier toward the development of a remote health-related service with the help of telecommunication and electronic information technologies, known as telemedicine [45].

Although many professionals adopted telecommunications media to exchange valuable medical information, the presence of a real-time paradigm could have been higher. However, by the 1960s, some critical experimental real-time applications in healthcare emerged: the National Aeronautics and Space Administration (NASA)'s space-flight program adopted telemetry to monitor human behaviors in space, and the Nebraska Psychiatric Institute in Omaha and the Norfolk State established the first interactive video links in clinical settings [46].

These two examples introduced important features of real-time data acquisition and processing, even by physicians. In the following years, the emergence of the Internet improved telemedicine regarding the easiness of creating links between medical staff with each other and patients. Also, the development of new body sensors and their improvement allowed the inclusion of this data in medical records, which eventually became digital records. As a communication system, telemedicine proposes a new model for producing, transmitting, and controlling medical data and services [47].

In 2004, the World Summit on the Information Society stated the *promotion of collaborative efforts of governments, planners, health professionals, and other agencies, along with the participation of international organizations in creating reliable, timely, high-quality, and affordable healthcare and health information systems*. In following years, the advances in many technologies have revolutionized the medical practice to its current state, bringing an immersive experience for those healthcare providers around the world and simultaneously eliminating the perception of distance, marking an optimal transformation from an in-person clinical visit to a synchronized/unsynchronized virtual reality.

In recent years, there has been a crescent employ of permanent medical records and data links to provide people with a better quality of life by allowing the development of new products to aid medical diagnosis and communication between health agents.

Many subgroups might benefit from the quality of life improvements considering the adoption of aid technologies. For example, the WHO estimates that more than one billion of the world's population live with some form of disability, of whom almost 200 million experience considerable difficulty in their daily lives, and it is to be expected that, in the future, disabilities will rise because of the increasing population of older persons and the risk that disability is more significant among older persons [48]. Also, the world population is aging due to the global increase in life expectancy and falling fertility rates. According to the [49], more than 16% of people in the world will be over 65 by 2050, while in 2019, this ratio was less than 10%. Within this context, it is crucial to consider the recommendations that universal design principles offer to build accessible and age-friendly environments [49].

### *6.1 Health 4.0*

The contemporary adoption of devices that aid medical attention has evolved telemedicine to the IoHT in the context of Health 4.0 Systems. While the first Health 1.0 refers to the primary patient-clinician encounter, the following phases for the Health industry present an evolution of generated data, treatment strategies, clinicians interactions, and the spread of study subjects involved with the development of health solutions.

While Industry phases developed over a broad period of more than one century, the Health industry phases presented a more significant gap between the beginning of the first phase—Health 1.0—and the following ones: indeed, the Health Industry tends to be more conservative. Therefore, Health 4.0 follows the principles of Industry 4.0 for the digitization of laboratories and to implement automation in

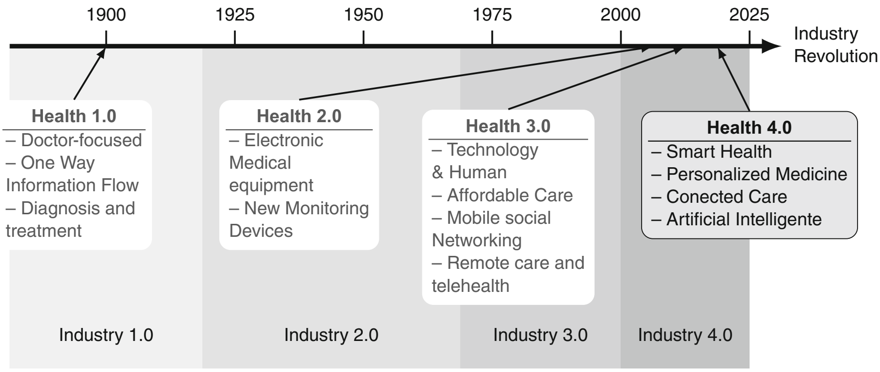



**Fig. 4** Historical timeline of Health 1.0 to Health 4.0

numerous processes used in the general health sector and hospitals [50]. Figure 4 presents a comparison of timeline evolution of the Health industry, with the main paradigms that have emerged in each phase, compared to its counterpart in industry phases.

The advancements in computer science, the electronic industry, and telecommunications have improved network throughput and computing processing technologies. As a result, the Health industry could exponentially adopt diagnostic tools in the healthcare sectors toward ubiquitous monitoring of the patients, which aids in the early detection of disorders and the implementation of a proactive treatment plan [51].

From Health 1.0 to Health 2.0, along with significant health, life science, and biotechnology development, numerous new medical equipment and devices have been invented, developed, and tested to be used in healthcare [52]. However, the Health 1.0 knowledge needed to be more elementary regarding treatment methods, small data generation, and poor contribution between clinicians and a single discipline strand. The central alignment of Health 2.0 was the development of technologies mirroring the Web 2.0 paradigm in the mid-2000s, focused on interoperability and participatory user-generated content. Therefore, Health 2.0 aligned with more team collaboration, adopting more complex treatments, multiple subjects, and a more significant amount of generated data.

Health 3.0 is a counterpart of Web 3.0, which consists of users accessing personalized content, followed by further development and implementation of information systems, e-health, and medical records since the 2010s. The main character in Health 3.0 is the crescent generation of Electronic Health Records (EHRs) by different electronic medical devices and available to health agents, physicians, or researchers. In addition, using available computer networks, remote care, and telehealth have become possible, and electronic visits replaced some face-to-face encounters, especially after the COVID-19 outbreak.

The health industry faced an increasing amount of generated data by medical devices, which promoted the adoption of technologies such as the Machine Learning

| | Health 1.0 | Health 2.0 | Health 3.0 | **Health 4.0** |
|---|---|---|---|---|
| **Data** | Small | Big | Huge | |
| **Subjects** | Single | Multiple | Community | |
| **Treatment** | Simple | Complex | Intelligent | |
| **Professionals** | Solo | Team | Network | |

**Fig. 5** Comparison of characteristics from Health 1.0 to 4.0

(ML), edge and cloud computing, especially after 2015. Finally, this new phase—the Health 4.0—created a revolution in healthcare. Such a value-based system enables Healthcare Industry to improve in providing the Quality of Service (QoS) with well-informed decisions [53]. Furthermore, the Health 4.0 phase expects an evolution to more intelligent treatment aided by the massive amount of generated information with the collaboration of multidisciplinary professionals. Figure 5 summarizes the significant paradigms in each phase for the health industry previously described considering the four main paths for healthcare enhancement: the treatment technologies, the clinicians' involvement, the multidisciplinary skills, and the amount of data.

## *6.2 5G and IoHT Applications in Health 4.0*

In the medical care sector, IoT performs a vital role in several Health 4.0-related applications. The adoption of automated medical data collection prevents human errors during the data retrieving stage, which improves the quality of diagnosis and eliminates possible risks due to human errors that risk the patient's health due to false diagnosis. The dense IoT environments can be categorized into different subgroups regarding a specific application. The devices employed in the medical domain are known as the IoHT devices, including sensors, communication technology, and mobile computing, all coordinating to provide healthcare services.

The 5G and IoHT are responsible for producing many applications in the Health 4.0 indoor or outdoor, in medical and non-medical environments. Some important application scenarios for IoHT in Health 4.0 are the *Remote Health Monitoring*, *Home care*, *Haptic Communication*, *Virtual Online Consultation*, *Ambulance Services*, and the *Telesurgery*, whose overview is illustrated in Fig. 6, are discussed as follows:

1. **Remote Health Monitoring**: This application consists of the evolution of the telemedicine approach by adopting technologies such as smartphones and several sensors to monitor patients' vital information in medical environments. By employing IoHT devices in the medical environment, it is possible to establish frameworks to collect a set of bio-signals such as the Eletrocardiogram (ECG) signal, blood pressure, body temperature, glucose level, and among others, through the 5G-powered wearables or other remote monitoring and rehabilitative

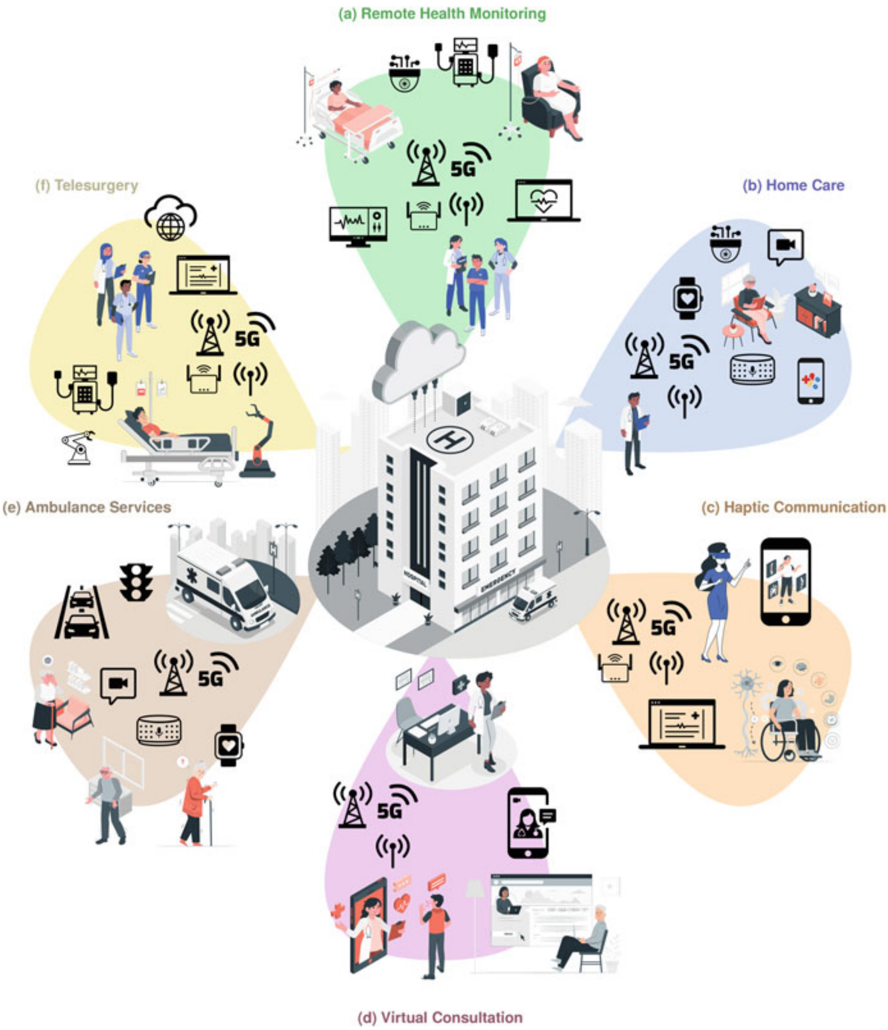



**Fig. 6** Summary of IoHT scenarios with 5G use cases deployment

devices. Remote health monitoring considers, in general, less invasive sensors and even the remote analysis of bio-samples, such as the blood sample analysis. A deployment example is presented in Fig. 6a.

2. **Home care:** is an enhancement of remote health monitoring in non-medical environments. An example is the Ambient Assisted Living (AAL) for home care, which combines IoHT and fitness devices, which address the health of the elderly, aged, or incapacitated individuals [54]. The AAL environment, as shown in Fig. 6b, commonly combines the employ of AI with many data-driven devices, such as video cameras, wearable, and other human–computer interface

devices, that demand a high network broadband, low latency, high reliability, and availability, especially for vision-based applications. Moreover, these scenarios belong to the eMBB and URLLC 5G use cases.

3. **Haptic Communication**: This use case presented in Fig. 6c involves adopting devices to perform non-verbal communication, primarily tactile actions, to improve QoL of people with disabilities. Therefore, the IoHT devices allow people to interact with clinicians, caregivers, or others using active or passive wearable or other devices implanted in the body. Furthermore, the Haptic ways of communication will enable them to express the information through touch or imagination. Thus, the Haptic devices or sensors convert the human input into the tactile input using tactile coding [55]. Moreover, haptic devices can also represent invasive sensors during a surgery-based treatment. Therefore, haptic communications belong to the use case URLLC, considering it is necessary to achieve very low latency and high reliability and availability.
4. **Virtual Online Consultation**: This scenario has received particular attention after the COVID-19 outbreak because it provides virtual follow-up visits between the patient and the doctor or the medical staff, especially for low-gravity health cases. These services sometimes demand privacy and high network requirements for proper functionality, such as the requirement for 4K/8K video streaming with low latency and low jitter [56]. The high-quality video streaming allows doctors to consult many patients without needing a physical meeting, which also avoids the spread of unwanted diseases by the patient's attendance at the hospital. Although video streaming is the leading technology for this scenario, collecting vital or other health data from patients using devices described in the remote health scenario may also be necessary. Moreover, there is a demand for very high data rate and channel capacity, high reliability, and security under the eMBB use case. Hence, the interaction between patient and medical staff can be done over mobile devices, as illustrated in Fig. 6d.
5. **Ambulance Services**: The paradigm of the Ambulance use case, presented in Fig. 6e, consists in integrating a set of mobile health monitoring systems with 5G-powered IoHT sensors fitted to the ambulance. The convergence of all these technologies will enable the continuous real-time streaming of patients' vital signs from the moment the emergency ambulance approaches the incident scene until the arrival at the hospital [57]. Additionally, it is possible to combine home care features for better emergency treatment and avoid delays due to unforeseen mechanical failure by real-time tracking and monitoring vehicle parts [58]. Although the ambulance services scenario belongs to the 5G URLLC, mMTC, and eMBB category requirements, this use case must satisfy high reliability and availability while present very low latency and an average data rate and capacity.
6. **Telesurgery**: This scenario, also called Robotic for Surgery, combines haptic devices with a set of demanding technologies to provide an environment to perform remote surgeries with the sense of need for ultra-low latency and precision and reliability. The combination of Augmented Reality (AR), Virtual Reality (VR), and AI provides vital information about the disease condition to the valuable information for the surgery as well as establishing an environment

for training with surgery simulation. The real-time high-quality video feedback introduces the data rate demand for a 5G scenario. Moreover, hospitals can have several assistive-connected robots to aid patients in critical care. With this regard, Stable 5G networks have made robotic surgeries a reality in some countries with a demonstration in many countries that automated procedures using 5G can safely and effectively be conducted [59]. This scenario facilitates the collaboration in real time between clinicians and surgeons around the world. This scenario requires very low latency and high data rate, availability, and reliability provided by URLLC and eMBB use cases, as shown in Fig. 6f.

As described in this section, each of the Health 4.0 applications presents specific requirements regarding the 5G use cases considering conditions such as latency, data rate, and reliability. Moreover, Table 3 presents a summary of 5G key capabilities for the 5G and IoHT healthcare applications.

## *6.3 Open Challenges for IoHT in 5G&B*

The major challenges for the Health Industry in the context of Fifth-Generation Mobile Networks and Beyond (5G&B) need to address two areas: the core technologies in 5G communication, which consist of the infrastructure, wireless channel models, security, latency, and reliability, and the healthcare-specific technologies for medical needs [61].

The hospital deployment scenario consists of an indoor environment that might present different path loss and delay spread than a typical indoor environment [62], thus requiring another channel modeling to identify. The channel models impact directly on the eMBB, URLLC, and mMTC use cases.

Regarding the security challenges, 5G expects the adoption of ML techniques to identify vulnerabilities and security flaws in such complex network architecture. The prevention of well-known network attacks, such as the Denial of Service (DoS), is essential to maintain the 5G availability and reliability in health scenarios. It is particularly challenging under the mMTC use case due to the massive number of connected devices. Moreover, there is a concern about the user's privacy and data leakage.

Also, reliability, low latency, and bandwidth-increasing challenges are present in streaming high-definition video for online consultation and remote surgery applications. The latency reduction and network reliability are crucial for the proper execution of a remote medical procedure. Doubtless, providing a highly reliable and shallow latency channel for these applications still needs to be solved in future networks. Also, the fulfillment of these indicators might provide an improvement in the patient monitoring remote care systems.

Regarding the medical Research & Development (R&D) opportunities, there is a chance to introduce new vision-based ML techniques for upcoming high-quality video streaming to aid in achieving better diagnosis.

**Table 3** Key healthcare components and the corresponding 5G dimensions [58, 60]

| Healthcare Component | 5G use cases | Key capabilities | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Data rate | Reliability | Traffic density | Power efficiency | Mobility | Latency | Conn. density | Spectrum efficiency | User data data rate |
| Remote health monitoring | eMBB, URLLC | | X | | | X | X | | | |
| Home care | eMBB, URLLC, and mMTC | | X | | | X | X | | | |
| Haptic communication | eMBB, URLLC | | | | | X | X | | | X |
| Virtual consultation | eMBB, mMTC | | X | | | | X | | | X |
| Ambulance services | URLLC, mMTC | X | | | | X | X | | | |
| Telesurgery | URLLC | X | X | | | | X | | | |

## 7 Conclusion

During this chapter, we presented the IoT, CC, 5G, and how these concepts relate to the healthcare scenario. We also present the junction of edge computing and artificial intelligence, forming the intelligent edge. The intelligent edge provides real-time intelligent health solutions with dynamic and adaptive management. Another relevant topic for 5G is the flexibilization and softwarization of the network, which allows providers to sell virtual networks that offer flexibility, adaptation, and affordable operation cost for different use cases, such as healthcare. Finally, we discussed the healthcare scenario and how all the technologies mentioned in this chapter relate to and can improve it.

The Health industry evolution in recent years has contributed to the consolidation of Health 4.0, which incorporates many technologies paradigms to improve healthcare in many scenarios. Corroborated by the previously presented open challenges, the evolution of network capabilities in 5G architectures contributes to the growing adoption and development of various new solutions for Health 4.0 and its challenges within the 5G scenario.

## References

1. Anthony Berauk, V. L., Murugiah, M. K., Soh, Y. C., Chuan Sheng, Y., Wong, T. W., & Ming, L. C. (2018). Mobile health applications for caring of older people: review and comparison. *Therapeutic Innovation & Regulatory Science, 52*(3), 374–382.
2. Selvaraj, S., & Sundaravaradhan, S. (2020). Challenges and opportunities in IoT healthcare systems: A systematic review. *SN Applied Sciences, 2*(1), 1–8.
3. Sheng, Z., Yang, S., Yu, Y., Vasilakos, A. V., McCann, J. A., & Leung, K. K. (2013). A survey on the IETF protocol suite for the Internet of Things: Standards, challenges, and opportunities. *IEEE Wireless Communications, 20*(6), 91–98. ISSN 15361284. https://doi.org/10.1109/MWC.2013.6704479.
4. Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks, 54*(15), 2787–2805. ISSN 13891286. https://doi.org/10.1016/j.comnet.2010.05.010.
5. Singh, D., Tripathi, G., & Jara, A. J. (2014). A survey of Internet-of-Things: Future vision, architecture, challenges and services. In *2014 IEEE World Forum on Internet of Things (WF-IoT)* (pp. 287–292). New York: IEEE. ISBN 978-1-4799-3459-1. https://doi.org/10.1109/WF-IoT.2014.6803174. https://ieeexplore.ieee.org/document/6803174/.
6. International Telecommunication Union (2012). Overview of the Internet of things. In *Series Y: Global information infrastructure, Internet protocol aspects and next-generation networks—Frameworks and functional architecture models* (p. 22). https://www.itu.int/rec/T-REC-Y.2060-201206-I.
7. Latif, S., Qadir, J., Farooq, S., & Imran, M. A. (2017). How 5g wireless (and concomitant technologies) will revolutionize healthcare? *Future Internet, 9*(4), 93.
8. Sadeeq, M. M., Abdulkareem, N. M., Zeebaree, S. R., Ahmed, D. M., Sami, A. S., & Zebari, R. R. (2021). IoT and cloud computing issues, challenges and opportunities: A review. *Qubahan Academic Journal, 1*(2), 1–7.
9. Vaquero, L. M., Rodero-Merino, L., Caceres, J., & Lindner, M. (2009). A break in the clouds. *ACM SIGCOMM Computer Communication Review, 39*(1), 50. ISSN 01464833. https://doi.org/10.1145/1496091.1496100.

10. Sultan, N. (2014). Making use of cloud computing for healthcare provision: Opportunities and challenges. *International Journal of Information Management, 34*(2), 177–184.
11. Islam, S. R., Kwak, D., Kabir, M. H., Hossain, M., & Kwak, K. S. (2015). The Internet of Things for health care: a comprehensive survey. *IEEE Access, 3*, 678–708.
12. Foukas, X., Patounas, G., Elmokashfi, A., & Marina, M. K. (2017). Network slicing in 5G: Survey and challenges. *IEEE Communications Magazine, 55*(5), 94–100. ISSN 01636804. https://doi.org/10.1109/MCOM.2017.1600951.
13. Afolabi, I., Taleb, T., Samdanis, K., Ksentini, A., & Flinck, H. (2018). Network slicing and softwarization: A survey on principles, enabling technologies, and solutions. *IEEE Communications Surveys & Tutorials, 20*(3), 2429–2453.
14. Shafi, M., Molisch, A. F., Smith, P. J., Haustein, T., Zhu, P., De Silva, P., Tufvesson, F., Benjebbour, A., & Wunder, G. (2017). 5G: A tutorial overview of standards, trials, challenges, deployment, and practice. *IEEE Journal on Selected Areas in Communications, 35*(6), 1201–1221. ISSN 0733-8716. https://doi.org/10.1109/JSAC.2017.2692307.
15. GSMA Intelligence. (2014). Understanding 5G: Perspectives on future technological advancements in mobile. In *White paper* (pp. 1–26).
16. Agiwal, M., Roy, A., & Saxena, N. (2016). Next generation 5G wireless networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials, 18*(3), 1617–1655.
17. International Telecommunications Union. (2016). Report on Standards Gap Analysis. https://goo.gl/qXlsm9.
18. ITU-R. (2017). Minimum requirements related to technical performance for IMT-2020 radio interface(s). Technical report, ITU-R. https://www.itu.int/md/R15-SG05-C-0040/en.
19. Rodrigues, C. A., Vieira, D., & de Castro, M. F. (2021). 5G network slice—a URLLC resource allocation perspective. In *2021 11th IFIP International Conference on New Technologies, Mobility and Security (NTMS)* (pp. 1–5). https://doi.org/10.1109/NTMS49979.2021.9432660.
20. Satyanarayanan, M., Bahl, P., Caceres, R., & Davies, N. (2009). The case for VM-based cloudlets in mobile computing. *IEEE Pervasive Computing, 8*(4), 14–23.
21. Bonomi, F., Milito, R., Zhu, J., & Addepalli, S. Fog computing and its role in the Internet of Things. In *Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing (MCC '12)* (pp. 13–16). New York: ACM. ISBN 978-1-4503-1519-7. https://doi.org/10.1145/2342509.2342513.
22. MECISG ETSI. (2016). Mobile edge computing (MEC); framework and reference architecture. In *ETSI, DGS MEC* (p. 3).
23. Roman, R., Lopez, J., & Mambo, M. (2018). Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges. *Future Generation Computer Systems, 78*, 680–698. ISSN 0167-739X. https://doi.org/10.1016/j.future.2016.11.009.
24. Hassan, N., Gillani, S., Ahmed, E., Yaqoob, I., & Imran, M. (2018). The role of edge computing in Internet of Things. *IEEE Communications Magazine, 56*(11), 110–115.
25. Liu, M., Yu, F. R., Teng, Y., Leung, V. C., & Song, M. (2018). Distributed resource allocation in blockchain-based video streaming systems with mobile edge computing. *IEEE Transactions on Wireless Communications, 18*(1), 695–708.
26. Ahmed, E., Akhunzada, A., Whaiduzzaman, M., Gani, A., Ab Hamid, S. H., & Buyya, R. (2015). Network-centric performance analysis of runtime application migration in mobile cloud computing. *Simulation Modelling Practice and Theory, 50*, 42–56.
27. Pace, P., Aloi, G., Gravina, R., Caliciuri, G., Fortino, G., & Liotta, A. (2018). An edge-based architecture to support efficient applications for healthcare Industry 4.0. *IEEE Transactions on Industrial Informatics, 15*(1), 481–489.
28. Khan, W. Z., Ahmed, E., Hakak, S., Yaqoob, I., & Ahmed, A. Edge computing: A survey. *Future Generation Computer Systems, 97*, 219–235 (2019)
29. Yousefpour, A., Fung, C., Nguyen, T., Kadiyala, K., Jalali, F., Niakanlahiji, A., Kong, J., & Jue, J. P. (2019). All one needs to know about fog computing and related edge computing paradigms: A complete survey. *Journal of Systems Architecture, 98*, 289–330.

30. Gedeon, J., Brandherm, F., Egert, R., Grube, T., & Mühlhäuser, M. (2019). What the fog? edge computing revisited: Promises, applications and future challenges. *IEEE Access, 7*, 152847–152878.
31. Maia, A. M., Ghamri-Doudane, Y., Vieira, D., & de Castro, M. F. (2021). An improved multi-objective genetic algorithm with heuristic initialization for service placement and load distribution in edge computing. *Computer Networks, 194*, 108146.
32. Maia, A. M., Ghamri-Doudane, Y., Vieira, D., & de Castro, M. F. (2020). Dynamic service placement and load distribution in edge computing. In *2020 16th International Conference on Network and Service Management (CNSM)* (pp. 1–9). New York: IEEE.
33. Zhu, G., Liu, D., Du, Y., You, C., Zhang, J., & Huang, K. (2020). Toward an intelligent edge: Wireless communication meets machine learning. *IEEE Communications Magazine, 58*(1), 19–25.
34. Yang, Y. (2019). Multi-tier computing networks for intelligent IoT. *Nature Electronics, 2*(1), 4–5.
35. Li, C., Xue, Y., Wang, J., Zhang, W., & Li, T. (2018). Edge-oriented computing paradigms: A survey on architecture design and system management. *ACM Computing Surveys (CSUR), 51*(2), 1–34.
36. Wang, X., Han, Y., Leung, V. C. M., Niyato, D., Yan, X., & Chen, X. (2020). Convergence of edge computing and deep learning: A comprehensive survey. *IEEE Communications Surveys & Tutorials, 22*(2), 869–904.
37. Abdellatif, A. A., Mohamed, A., Chiasserini, C. F., Tlili, M., & Erbad, A. (2019). Edge computing for smart health: Context-aware approaches, opportunities, and challenges. *IEEE Network, 33*(3), 196–203.
38. Uddin, M. Z. (2019). A wearable sensor-based activity prediction system to facilitate edge computing in smart healthcare system. *Journal of Parallel and Distributed Computing, 123*, 46–53.
39. Zhang, Y., Ma, X., Zhang, J., Hossain, M. S., Muhammad, G., & Amin, S. U. (2019). Edge intelligence in the cognitive Internet of Things: Improving sensitivity and interactivity. *IEEE Network, 33*(3), 58–64.
40. Miraz, M. H., Ali, M., Excell, P. S., & Picking, R. (2015). A review on Internet of Things (IoT), Internet of Everything (IoE) and Internet of Nano Things (IoNT). In *2015 Internet Technologies and Applications (ITA)*(pp. 219–224)
41. Pazienza, A., Polimeno, G., Vitulano, F., & Maruccia, Y. (2019). Towards a digital future: an innovative semantic IoT integrated platform for Industry 4.0, healthcare, and territorial control. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)* (pp. 587–592). New York: IEEE.
42. Su, R., Zhang, D., Venkatesan, R., Gong, Z., Li, C., Ding, F., Fan, J., & Zhu, Z. (2019). Resource allocation for network slicing in 5G telecommunication networks: A survey of principles and models. *IEEE Network, 33*(6), 172–179.
43. Chahbar, M., Diaz, G., Dandoush, A., Cérin, C., & Ghoumid, K. (2020). A comprehensive survey on the E2E 5G network slicing model. *IEEE Transactions on Network and Service Management, 18*(1), 49–62.
44. Vergutz, A., Noubir, G., & Nogueira, M. (2020). Reliability for smart healthcare: A network slicing perspective. *IEEE Network, 34*(4), 91–97.
45. Haleem, A., Javaid, M., Singh, R. P., & Suman, R. (2021). Telemedicine for healthcare: Capabilities, features, barriers, and applications. *Sensors International, 2*, 100117. ISSN 2666-3511. https://doi.org/10.1016/j.sintl.2021.100117. https://www.sciencedirect.com/science/article/pii/S2666351121000383.
46. Zundel, K. M. (1996). Telemedicine: History, applications, and impact on librarianship. *Bulletin of the Medical Library Association, 84*(1), 71–79.
47. Shen, Y. T., Chen, L., Yue, W. W., & Xu, H. X. (2021). Digital technology-based telemedicine for the covid-19 pandemic. *Frontiers in Medicine, 8*, 646506. ISSN 2296-858X. https://doi.org/10.3389/fmed.2021.646506. https://www.frontiersin.org/articles/10.3389/fmed.2021.646506.

48. World Heath Statistics (2022). Monitoring health for the SDGs, sustainable development goals, 2022. License: CC BY-NC-SA 3.0 IGO.
49. Department of Economic and Social Affairs of United Nations. (2019). *World population prospects 2019: Highlights*.
50. Elhoseny, M., Abdelaziz, A., Salama, A. S., Riad, A. M., Muhammad, K., & Sangaiah, A. K. (2018). A hybrid model of Internet of Things and cloud computing to manage big data in health services applications. *Future Generation Computer Systems, 86*(C), 1383–1394. ISSN 0167-739X. https://doi.org/10.1016/j.future.2018.03.005.
51. Qadri, Y. A., Nauman, A., Zikria, Y. B., Vasilakos, A. V., & Kim, S. W. (2020). The future of healthcare Internet of Things: A survey of emerging technologies. *IEEE Communications Surveys & Tutorials, 22*(2), 1121–1167. https://doi.org/10.1109/COMST.2020.2973314.
52. Li, J., & Carayon, P. (2021). Healthcare 4.0: A vision for smart and connected health care. *IISE Transactions on Healthcare Systems Engineering, 11*(3), 171–180. https://doi.org/10.1080/24725579.2021.1884627.
53. Kumari, A., Tanwar, S., Tyagi, S., & Kumar, N.. Fog computing for Healthcare 4.0 environment: Opportunities and challenges. *Computers & Electrical Engineering, 72*, 1–13 (2018). ISSN 0045-7906. https://doi.org/10.1016/j.compeleceng.2018.08.015. https://www.sciencedirect.com/science/article/pii/S0045790618303860.
54. Arora, D., Gupta, S., & Anpalagan, A. (2022). Evolution and adoption of next generation IoT-driven Healthcare 4.0 systems. *Wireless Personal Communications, 127*(4), 3533–3613. ISSN 1572-834X. https://doi.org/10.1007/s11277-022-09932-3.
55. Gupta, R., Tanwar, S., Tyagi, S., & Kumar, N. (2019). Tactile-Internet-based telesurgery system for Healthcare 4.0: An architecture, research challenges, and future directions. *IEEE Network, 33*(6), 22–29. https://doi.org/10.1109/MNET.001.1900063.
56. Siriwardhana, Y., Gür, G., Ylianttila, M., & Liyanage, M. (2021). The role of 5G for digital healthcare against covid-19 pandemic: Opportunities and challenges. *ICT Express, 7*(2), 244–252. ISSN 2405-9595. https://doi.org/10.1016/j.icte.2020.10.002. https://www.sciencedirect.com/science/article/pii/S2405959520304744.
57. Doukoglou, T., Gezerlis, V., Trichias, K., Kostopoulos, N., Vrakas, N., Bougioukos, M., & Legouable, R. (2019). Vertical industries requirements analysis & targeted KPIs for advanced 5G trials. In *2019 European Conference on Networks and Communications (EuCNC)* (pp. 95–100). https://doi.org/10.1109/EuCNC.2019.8801959.
58. Dananjayan, S., & Raj, G. M. (2021). 5G in healthcare: How fast will be the transformation? *Irish Journal of Medical Science (1971), 190*(2), 497–501. ISSN 1863–4362. https://doi.org/10.1007/s11845-020-02329-w.
59. Pandav, K., Te, A. G., Tomer, N., Nair, S. S., & Tewari, A. K. (2022). Leveraging 5G technology for robotic surgery and cancer care. *Cancer Rep (Hoboken), 5*(8), e1595.
60. Ahad, A., Tahir, M., Aman Sheikh, M., Ahmed, K. I., Mughees, A., & Numani, A. (2020). Technologies trend towards 5G network for smart health-care using IoT: A review. *Sensors, 20*(14). ISSN 1424-8220. https://doi.org/10.3390/s20144047. https://www.mdpi.com/1424-8220/20/14/4047.
61. Georgiou, K. E., Georgiou, E., & Satava, R. M. (2021). 5G use in healthcare: The future is present. *JSLS: Journal of the Society of Laparoscopic & Robotic Surgeons, 25*(4), e2021.00064. https://doi.org/10.4293/jsls.2021.00064.
62. Kyro, M., Haneda, K., Simola, J., Nakai, K., Takizawa, K. I., Hagiwara, H., & Vainikainen, P. (2011). Measurement based path loss and delay spread modeling in hospital environments at 60 Ghz. *IEEE Transactions on Wireless Communications, 10*(8), 2423–2427. https://doi.org/10.1109/TWC.2011.062211.101601

# Health Care 4.0: Challenges for the Elderly with IoT

**Henrique Gil and Maria Raquel Patrício**

## 1 Digital Literacy: Reflections About the Concept and Implications for the Elderly

When trying to define the concept of digital literacy, we are often faced with different opinions and dimensions, given the polysemy that is associated with this expression. To try to make this question simpler, we decided to divide the expression into "digital" and "literacy." Regarding "digital," we can easily associate it with something that is related to the computer, information and communications technologies, and the media and, as Goodfellow [1] says, which is supported by a web-based environment. On the other hand, when in this context it intends to define "literacy," we are led to consider what is proposed by Lankshear and Knobel [2] when they state that it is more appropriate to refer to "literacies" in the plural form to express an expanded concept of literacy that emphasizes the diversity of social and cultural practices that are covered by the term.

Despite all the differences related to the definition of the concept, we must mention Gilster ([3], p. 1) because this author was the first to advance this definition many years ago: "the ability to understand and use information in multiple formats from a wide range of sources when it is presented via computers." This means

H. Gil (✉)
Age.Comm – Polytechnic University of Castelo Branco, Castelo Branco, Portugal
e-mail: hteixeiragil@ipcb.pt

M. R. Patrício
Age.Comm – Polytechnic University of Castelo Branco, Castelo Branco, Portugal

Centro de Investigação em Educação Básica (CIEB), Instituto Politécnico de Bragança, Bragança, Portugal
e-mail: raquel@ipb.pt



N. Gupta, S. Mishra (eds.), *Internet of Everything for Smart City and Smart Healthcare Applications*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-34601-9_13

that Gilster [3] seems to emphasize the differences between digital information media and other more conventional media (e.g., newspapers, radio, TV). So that we can have a more current view, we decided to focus on the definition presented by the European Commission within the framework of key competences for all the European citizens, in particular ([4], p. 6): "Digital competence involves the confident and critical use of Information Society Technology (IST) for work, leisure, and communication. It is underpinned by basic skills in ICT: the use of computers to retrieve, access, store, produce, present and exchange information, and to communicate and participate in collaborative networks via the Internet." From this point of view, digital literacy is in line with the opinions of Lanksheare and Knobel [5] who mention that in this perspective there is interaction with information and with its evaluation and validation, with its credibility and reliability.

In recent years, because of the increasing digitization of society, different definitions have emerged. As stated by Buckingham [6], this greater difficulty in defining the concept is a consequence of the increasing and gradual domestication of the Internet and the constant emergence of new technologies, the so-called emerging technologies that create new practices and new uses different from the first ones that were carried out when the technologies were appearing. For this purpose and with the aim of providing a broader view of different conceptions, only a few examples of these variants will be presented. Given this diversity of interpretations, we can say that we may be dealing with the concept of "multiliteracy" or "literacies." For example, Gourlay et al. [7] and Joosten et al. [8] have an opinion that can be more related to a functional perspective of technology use by highlighting the associated concepts of computer literacy or media literacy. From a different point of view that could be called "complementary," as Stordy [9] proposes, it will be the combination of technological skills with skills that promote more cognitive dimensions that lead to a set of practices of a more social nature in the use resources and digital platforms. In this regard we agree with Martin [10] who combines and advocates the need for citizens to have access and to be able to manage, evaluate, analyze, and integrate the already mentioned resources and digital platforms so that they can build new knowledge and share it with the others to enable constructive social action. In this sense, Lankshear and Knobel [5] state that digital literacy involves "mastering ideas," not "key-strokes." In this line of reasoning, this perspective is reinforced by Lankshear and Snyder [11] by the fact that users of digital technologies have practices that have a specific cultural and critical mark that can be considered as "ways of doing things" instead of being merely questions "operational" or "technical." In this sense, we agree with Pangrazio and Sefton-Green [12] that we must have a clear notion about the scope and depth in which the concept of digital literacy is referenced: "meaning digital citizenship conveys a different set of practices from simply casting a vote at a polling booth or participating in civil society to participating in online discussion."

What is more pertinent and important is that we must have a clear perception of digital literacy about the implications and direct consequences for the quality of life of citizens. The digital medium or resource or even the dimension to which it reports and on which the citizen is supported will not be decisive or fundamental. What matters to reflect and decide is to know and be able to choose this or that resource or digital medium so that, in a conscious and safe way, it can bring about improvements in the daily routines and in the quality of life and health of citizens. To this end, we share the opinion of Lanksheare and Knobel [5] when placing the emphasis on this concept of digital literacy by including “only” two fundamental dimensions: creation and communication. However, as mentioned by Gilster [3], the use of digital technologies brings and redefines practices associated with writing and with reading in these new digital resources. At the level of creation, citizens are supposed to be able to access digital resources to select, criticize, reflect, and be able to decide which information can be used to be transformed and adapted to generate and create information that must be original and that is relevant and most appropriate. And at the same time, as Casey and Bruce [13] state, have the ability to communicate meaning.

On the other hand, in terms of communication, the citizen is supposed to have the knowledge and digital skills to be able to use digital resources to be able to share and disseminate correctly and appropriately using digital social networks or other resources (e.g., digital platforms, web pages). In this way, the citizen becomes active, moving from a passive consumer to an active producer increasing their levels of self-confidence and social belonging, which is reflected in greater levels of digital and social inclusion that will allow for a better exercise of citizenship. In this regard, it is important to refer and reflect on the social networks commonly referred to as social media. In this context, Tan [14] reflects in a way that proposes to promote an extension of the concept of digital literacy to what he designates as a “multimodal outlook” or even as a “new textual landscape” to which citizens have increasingly joined. Also in this regard, the European Commission [15], within the scope of the “Digital Agenda for Europe,” reinforces the premise that there can only be an adequate and complete political and social participation if citizens are info-included because this new and present reality requires digital skills.

In a perspective of critical synthesis, it is important to remember the warning and need mentioned by the OECD [16] almost two decades ago that, to have sustainable development and good and adequate social cohesion, it is necessary for citizens to have all the skills that they need and allow them to be socially included and, for this purpose, digital skills are essential. This dimension is fundamental, and digital literacy must be seen as having causal efficacy to generate outcomes in society and in the world. This is where the importance of, as stated by Lanksheare and Knobel ([5], p. 12): “When one «has» digital literacy good things can happen; when one lacks digital literacy, one is vulnerable and undesirable things can happen.” In this context, it is important to reinforce the fact that these consequences are for each citizen and, therefore, for society in general. We all lose out if the info-exclusion rate continues to be high among the elderly. Well, the elderly represents an estate and a rich history of life experiences and learning that we all want to be shared to achieve a better world. For these reasons, we support the opinion of Tamborg et al. [17], who feel that it is essential to reflect carefully on what is necessary for safe use in a technology-rich society in a wider context. Throughout this chapter, we have been referring, albeit in a more superficial way, to the info-exclusion issues of the elderly. We cannot forget and ignore that the elderly has a long history of personal and professional life full of experiences and experiences that took place, mostly, in analogical contexts. And, fortunately, most were successful because of their more formal academic training and also of their lifelong training that included formal and non-formal contexts. The question that we now present for reflection has a direct connection with formal academic contexts: school. At school, the elderly was taught and trained for linear and pre-defined tasks and activities, with an emphasis on individual work and where memorization was the priority so that later they could reproduce this acquired knowledge, and the support was solely based on paper (books, documents, worksheets). Another characteristic had to do with the fact that this knowledge lasted for years, given that flexibility and updating were something marginal. On the contrary, as mentioned by Castilla et al. [18], the use of the Internet becomes very complex for the elderly as it relies on a new technology: hypertext. In this new type of digital document, users must be able, in addition to reading it, to select this information, which parts of the text to select, so that the "new reading" makes sense, given that the reading is no longer linear. At the same time, the elderly must deal with the fact that these documents have video, sound, and image links, which correspond to a new context for the elderly: multimedia. These new contexts make the reader active and participatory, instead of their initial training where everything was foreseen and where participation was passive. This new dimension also brings another type of metaphors and another type of terminology without a direct connection with the analogical dimension that the elderly dominates.

The European Union (EU) is fostering digital inclusion to ensure that everybody can contribute to and benefit from the digital world, such as Digital Competence Framework for Citizens (DigComp), the EU-wide framework for developing and measuring digital competence. However, EU has recognized multiple barriers to digital inclusion: "For some people, the digital world is not yet fully accessible. For others, it is not affordable, and others were not taught the skills to participate fully" [19]. In 2021, 25% of the EU-27 population aged 65–74 years had at least basic overall digital skills, with wide variation among countries, e.g., 57% in Iceland and 56% in Switzerland. In contrast, only 4% of people aged 65 to 74 in Romania had digital skills, followed by those in Bulgaria (6%) and Poland (10%). Portugal is in the seventh position of the European countries with the lowest index of digital skills of people between 65 and 74 years old, with only 17% [20]. In this context, this issue can become even more sensitive, since, as stated by Almeida et al. [21], "the progressive aging of the older population itself, as the relative important of the very old is growing at a faster pace than any other age segment."

Recently, on 22 March 2022, the Joint Research Centre of the European Commission published the latest update of the Digital Competence Framework for Citizens, DigComp 2.2, with updated examples of knowledge, skills, and attitudes. The update takes account of emerging technologies such as artificial intelligence, the Internet of Things, datafication of Internet services and apps, virtual and augmented reality, robotization, or misinformation and disinformation. DigComp 2.2 will contribute to achieving the EU objectives set out by the Skills Agenda, the Digital Education Action Plan, the Digital Decade, and the Pillar of Social Rights and its action plan, for instance, achieving a minimum of 80% of the population with basic digital skills [22].

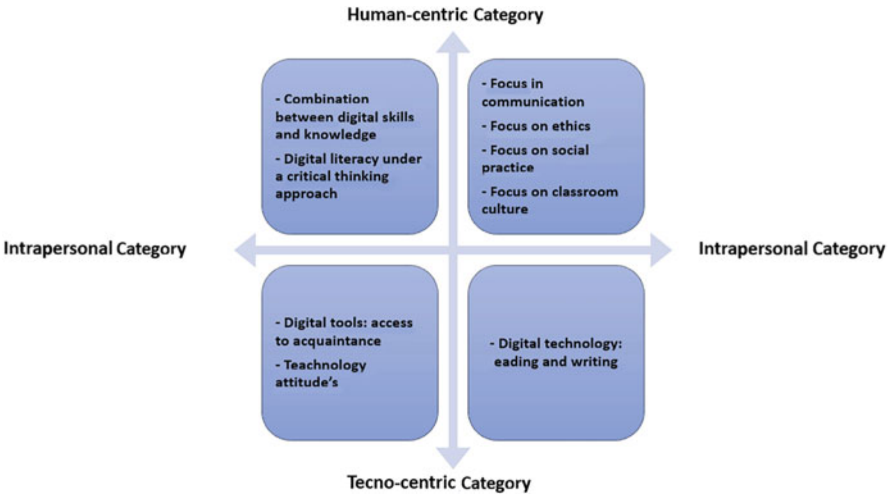



**Fig. 1** Adapted/reformulated model proposed by Tamborg et al. [17] related to the different dimensions of the concept of digital literacy

Against this background, it is urgent to empower elderly people with digital literacy. Before doing so, however, it is important to continue presenting different conceptions of digital literacy to better put into perspective its implications for the elderly.

The research carried out by Tamborg et al. [17] presents a simple and objective way of giving a clear view of the different conceptions of digital literacy. To this end, Tamborg et al. [17] propose a model that includes eight categories distributed in four quadrants according to two axes: intrapersonal/interpersonal and human-centric/techno-centric (see Fig. 1).

At this stage, we will present a critical reflection based on the research of Tamborg et al. [17] with the aim of emphasizing the potential implications for the elderly. Starting this critical-reflexive analysis with the "human-centric and techno-centric" axis, it seems to emphasize the fact that digital technologies are related as means or resources for the production and consumption of information. However, as stated by Andrejevic [23], there is a tendency to assume that going digital only implies the automatic use of "mechanized" techniques, which can lead to an "automated" and "impersonal" dimension that should be caution and avoid. This assumption may also include concerns associated with the appropriate use of technologies that imply appropriate ethical behavior. In this context, we agree with the point of view of Shin and Seger ([24], p. 22) cited by Tamborg et al. [17] when interpreting digital literacy as "discursive practices that are shaped by one's social, cultural, and political access." Now making a critical-reflexive analysis in relation to the "intrapersonal and interpersonal" axis, we are led to conclude that the implementation of digital literacy will be determined by each citizen, individually. In this, when we think about elderly citizens, we feel a greater concern because it

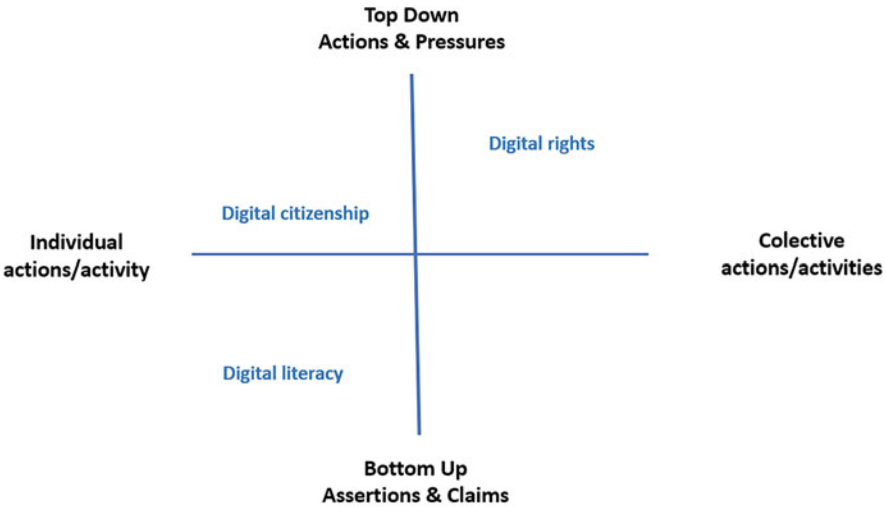



**Fig. 2** Adapted/reformulated model proposed by Pangrazio and Sefton-Green [12] that includes the different dimensions of rights, citizenship, and digital literacy

will depend on their level or degree of info-inclusion. This conception is reinforced by the opinions of Thrane et al. [25] because at present the approaches are focusing on the issues of the digital divide and the strategies that must be implemented that will have to go beyond simple access, so that true inclusion becomes a priority and a reality. For this reason, as reported by international statistical data, the age group of the elderly is the one that is more info-excluded. In consequence, the levels of concern are higher because, in extreme situations, they can jeopardize their social inclusion. In their investigations and reflections, Pangrazio and Sefton Green [12], are of the opinion that there is a very close relationship between digital rights and digital citizenship and the concept of digital literacy to which they associate the concept of "agency" when citing Richardson [26]. However, it should be noted that "agency" has some philosophical complexity, which as mentioned by Pangrazio and Sefton-Green ([12], p. 23) is "the capacity of an individual to act freely in the world, the concept is entwined with assumptions about free will, the structural constraints which limit individuals' actions and the relationship between an individual and their society." In this conception and assumption, that there must be a given "agency" leads to the overarching "power" being needed to create value in an increasingly complex digital world. Figure 2 presents the model proposed by Pangrazio and Sefton-Green [12] which includes the different dimensions of rights, citizenship, and digital literacy.

In order to reach satisfactory levels of a good and adequate use of current digital resources, whose clues have been evidenced (Figs. 1 and 2), Lee et al. [27] identified the following factors that can be associated with the elderly: (1) intrapersonal factors, which include motivation and self-efficacy; (2) functional limitations, related to

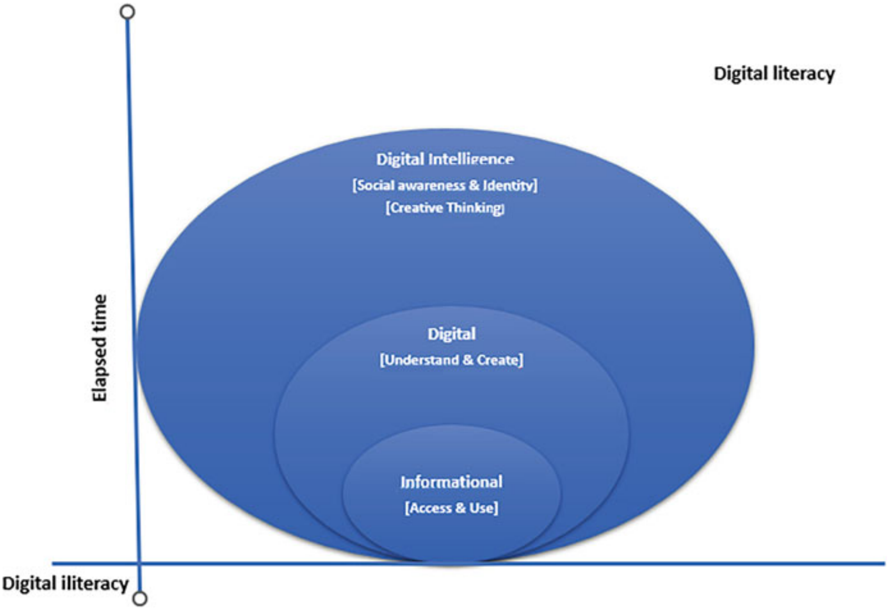



**Fig. 3** Digital literacy process of the elderly. (Adapted/reformulated from [29])

cognitive decline or spatial orientation; (3) structural limitations that prevent normal and easy access to digital resources; and (4) interpersonal limitations, which may be closely related to lack of social support.

Continuing this critical reflection on digital literacy, we will rely on another proposal by Rahman et al. [28], cited by Martinez-Alcalá et al. [29], that refers to a new dimension: digital intelligence. In this context, digital intelligence corresponds to a set of several skills that include the social, emotional, and cognitive dimensions that will allow the elderly to be able to respond to the current challenges of society and better adapt to their needs as citizens. Figure 3 presents the model proposed by Martinez-Alcalá et al. [29]:

In Martinez-Alcalá et al.'s proposal [29], there are three levels that increase in quality and depth from the most basic level (informational) related to use and access, moving on to an intermediate level (digital) where you will already have there must be skills related to understanding and creation and, finally, the final level (digital intelligence) where the elderly will already have skills that will allow them to be creative, social awareness and a digital identity. It will be an ongoing process that must be supported by specialists and that must be updated given the constant evolution of technologies, including the Internet of Things (IoT) and emerging technologies (e.g., AR, VR, AI, etc.). To further explore this model, we rely on the opinions of Martinez-Alcalá et al. [29] regarding the four skill levels that are presented in Fig. 3: (1) instrumental, related to a practical use associated with hardware usage skills and software; (2) socio-affective, skills that allow creating and establishing social bonds with other individuals through communication for socialization purposes; (3) axiological, understanding and putting into practice ethical and responsible attitudes in the use of the Internet; and (4) cognitive-behavioral, which represents a set of more reflexive and more critical skills in the sense of being able to recreate new information through a previous process of research, selection, analysis, and interpretation.

Digital skills training to improve digital literacy among the elderly is increasingly becoming a key priority in the digital society. This can empower the elderly to effectively, safely, and securely use and benefit from the opportunities offered by digital technologies and the Internet, thus achieving better health and quality of life [30].

The United Nations Economic Commission for Europe (UNECE) recently highlighted policy priorities for the digital inclusion and empowerment of the elderly in the digital era: ensure equal access to goods and services involving digital technology; enhance digital literacy to reduce the digital skills gaps; leverage the potential of digital technologies for active and healthy aging, well-being, and empowerment of the elderly; and ensure the protection of human rights of the elderly in the digital era [31]. Another area of EU policy is active and assisted living for fostering digital inclusion and aging well, through meaningful projects that promote innovative products, services, and systems based on information and communications technologies to help older people age well at home, in the community, and at work.

## 2 Ambient Assisted Living [AAL] and Internet of Things [IoT]: Implications and Consequences on Daily Routines

It is no longer a novelty that the world is gradually aging and there are already areas and/or countries where the aging rate is already very worrying. In addition to the great advantage of being able to prolong longevity, we have, on the other hand, a less pleasant panorama because the elderly is more prone to situations associated with the loss of motor, cognitive, and emotional faculties, which are caused by higher rates of loneliness and of isolation. As stated by Thomas [32], more than half of the elderly over 75 years old live alone, and the impact of this loneliness has emotional and psychological consequences that substantially worsen health and well-being. Consequently, these elderly people will have to be supported by their family or caregivers. As mentioned by the United Nations report in 2015, 7 caregivers are already allocated for each elderly person, on a global scale, but in the year 2030, the value of only 4.9 caregivers for each elderly person is projected [33]. In Europe, where the rate of aging is higher worldwide, Abdi et al. [34] report that in the year 2030, there will be only about 2.2 caregivers for each elderly person. As is easy to understand, the situation tends to become dramatic, and it is urgent to find solutions that reverse this trend.

Another very important aspect mentioned by Salovaara et al. [35] has to do with the fact that the elderly, in terms of comparison with other age groups, are those who mostly use the Internet to carry out research related to health. This data, in itself, implies that efforts are made so that more elderly people become info-included so that they can keep themselves better informed and, as a consequence, increase their quality-of-life levels. This reflection is corroborated by Heo et al. [36] when reviewing research results that show that greater use of the Internet by the elderly has provided better social support, reduced loneliness, better levels of satisfaction, and also a better psychological state.

Considering the reflections presented, it is necessary to find clues and resources that allow us to help the elderly to be able to live in their homes, as defended by the perspective of "aging in place," independently and with the highest possible indices of quality encouraging the elderly to remain recognized as vital to their community [37]. For, according to what is stated by Syed et al. [38], a survey carried out in the United States, and which can be extrapolated to other countries, about 90% of the elderly reported wanting to remain in their homes. For all the reasons, we agree with Mainetti et al. [39] in stating that treating and caring for the elderly is a great challenge, but it also represents a question that we must answer. However, we cannot forget that it is necessary for caregivers and medical experts to know and be able to motivate and accompany the elderly in their homes through digital resources [37].

To this end, the concept of ambient assisted living (AAL) is based on the objective of creating conditions for elderly people with limitations (motor, cognitive, or others) to be supported. This support aims to increase their levels of autonomy and security so that they can independently carry out their daily routines [40]. As reported by Dohr et al. [41], the applications associated with AAL are used for the elderly to have and/or maintain their levels of quality of life, well-being, and health, in a safe environment. As reinforced by Dohr et al. [41], the benefits of AAL are reflected in three different levels: (1) at the individual level (well-being and safety), (2) at the economic level (more effective and efficient management of resources that are already very limited), and (3) at the social level (improvement and increase in standards of living). In this regard, we have the same opinion as Mainetti et al. [39] in assuming that AAL can promote and create conditions of greater safety for the elderly with the detection of falls and immediate response mechanisms and through video surveillance systems.

In global terms, we can say in a way that seems to be consensual that the main needs of the elderly that must be supported in an AAL context will be the following: health (perhaps the most important and most pertinent), safety, relaxed environments (the elderly person is not subject to high levels of stress, so peace of mind will be a goal), autonomy, independence, and interpersonal relationships (privileging social contacts and relationships). To address these different levels, the AAL contexts, in the opinion of Kunze et al. [42], must be structured in three levels: (1) hardware (wireless network; sensing), (2) middleware (IT integration; data capture; data safety), and (3) services (application-oriented processes; biosignal processing; community services). In other words, they will have to be the conditions for the creation of an "ambient intelligence," as stated by Dohr et al. [41]. Therefore, an ambient where technologies and digital resources coexist with citizens and with their context (house, equipment, and spaces where activities are carried out) where the inclusion of the Internet of Things (IoT) seems to be relevant [43]. Over the last few years, the European Commission has been paying attention to this issue and has recently made efforts to speed up the integration of IoT applications and resources to respond assertively not only to the elderly but to future generations [44]. As mentioned by Mainetti et al. [39], AAL can make use of different resources that make it possible to create safer environments and that allow the elderly to remain in their homes with greater indices of autonomy, where we can highlight GPS, radio-frequency identification, and Bluetooth. With these digital resources, the aim is for the elderly to remain active if possible, maintaining their social relationships and independence. From the point of view of Almeida et al. [21], AAL has been evolving toward a more global concept called the city-wide approach. This new approach intends to see the elderly not as people with disabilities or special needs, but as individuals who are part of a given community, with their networks of contacts and who are able to continue to be active either at the level of their homes or at the level of your city/community.

IoT is now recognized, and a large majority of the population knows what it means. Other expressions such as big data, artificial intelligence (AI), and 5G technology are also starting to form part of the conversations between info-included citizens. In recent years, there has been an evolution from the so-called ICT to emerging technologies because of this new digital context that we are experiencing. There seems to be, as proposed by Gigli and Koo [45], a tendency for digital technologies to become congruent and to attempt a certain unification of the different digital resources and technologies around the IoT. In other words, the conditions are created for the establishment and growth of IoT ecosystems that will create the basis for a new and upcoming "industrial revolution" with a strong focus on healthcare and with an innovative character.

Dohr et al. [41], state that IoT will generate new contexts, more innovative contexts according to three dimensions: (1) ubiquitous communication/connectivity in an "anywhere, anytime" perspective; (2) pervasive computing: giving "power" to things/objects, a way to make the environment like and each thing/object become a computer; and (3) ambient intelligence: each thing/object is considered to be "smart objects" because they have the ability to interact and to be able to promote changes in the surrounding physical environment. In short, with the IoT, the "smart objects" can communicate with each other and this time to be able to create networks of "things/objects" through their connection to the Internet. In this new paradigm, the concept of "anytime, anywhere" gains a new dimension evolving into "anyone" or, rather, evolving into "anything" [46]. In this context, the European Commission [44] states that the IoT, by its nature, creates new networks that are independent and that operate on their own infrastructures that provide new services and new forms of communication between citizens and things and between themselves. Accordingly, we can refer to the concept of "smart cities" because many elderly people live in urban areas and it is necessary to find proposals that narrow and

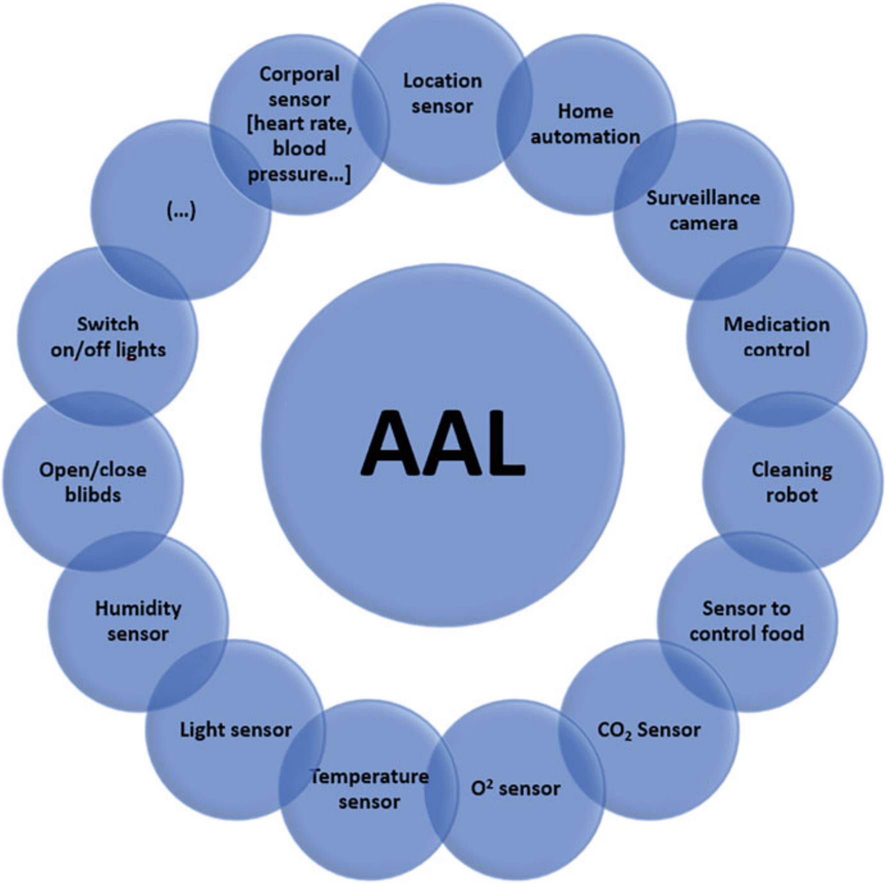



**Fig. 4** An example of AAL elderly people's residence or house full of sensors

facilitate communication between citizens and the city, considering the real needs of the elderly [21].

Associating IoT with AAL is important and pertinent for a context that has better levels of quality for the elderly has to do with the fact that IoT establishes a bridge between the analog/physical world of "things" and technologies contextualized with the social environment. Therefore, as reported by Wu et al. [47], conditions are created for a human-computer interaction that integrates what these authors call smart allocation of resource and intelligent provision of several services. As also mentioned by Dohr et al. [41], what is intended is that IoT can provide relevant information in an active and passive way, actively through local decision-making and passively through data collected by sensors. Figure 4 presents an example of intelligent communication architecture for AAL that uses artificial intelligence for processing the information collected from several types of sensors, which could be installed at different locations in the environment, depending on the parameters to be measured and the events to be monitored [48].

More recently, technologies have been using the cloud to make procedures more ubiquitous and in real time [49]. In this new technological and digital dimension, Syed et al. [38] introduce the concept of IoMT (Internet of Medical Things) which should be considered as something more focused on infrastructures that include software, medical devices, healthcare systems, and computing systems, capable of acting in daily routines and in risk and medical emergency situations. In summary, IoMT will interconnect wearable sensors, patients, healthcare providers, and other caregivers using different communication software.

In this new context, AI requires an active attitude where decision-making becomes fundamental and where IoT promotes a passage or a bridge between a physical world related to “things,” the digital world, and the social component so that a more intimate, more fluid, and more natural relationship can be established: human-computer interaction. If these dimensions are articulated, it is possible to talk about ambient assisted living (AAL) where the conditions of well-being and better quality of life and health for the elderly will be increased, based on the more general concept of smart home [50]. As reported by Suciu et al. [51] with ambient assisted living cloud applications, monitoring is carried out inside and outside the homes of the elderly with a multiplicity of sensors whose main objective is to improve the quality of life and health through the information that is reported to family members, caregivers, and health staff. For instance, some of the sensors that can be installed inside the house associate the elderly person’s gender, their location, relative to the environment, humidity, temperature, and lighting can be monitored, the steps, the use of the telephone, the incoming and outgoing messages, social media, noise level and sleep monitoring. More examples can still be presented, as proposed by Maskeliunas et al. [52] in the sense of having a positive influence to overcome or improve different dimensions that lead to higher levels of autonomy and independence: limitations in daily activities, risk of falling, chronic diseases, dementias, depression, social isolation, and medication management. Because some of this monitoring can become invasive, despite having good intentions and noble principles, it is essential to ensure that the elderly can gain more autonomy and independence, but with dignity, respect, and privacy. However, we agree with Choi et al. [30] when they state that the daily accumulation of data by digital devices and platforms will increase diagnostic accuracy, produce innumerable quality of data, and save time. In this sense, conditions are promoted to provide advantages to medical facilities and convenience in an individual and personalized way for each elderly person.

Although there is a very positive general feeling for these IoT and AAL proposals, as with everything, there are always limitations or obstacles that can make the objectives not be fully achieved. In this, security is one of the main concerns because it is very difficult to have digital platforms that can resist malicious attacks by hackers. For this reason, the elderly must be informed about this possibility, and they will have to decide whether to adopt the technologies. Another difficulty that is reported by Villari et al. [53] has to do with the interoperability between different platforms and sensors and data collection systems. This lack of integration can generate conflicts in the collection of data that can lead to less adequate decision-making. To these limitations must also be added the fact that there may be some lack of precision and accuracy in the sensors, which will lead to an incorrect assessment of reality [54]. Regarding limitations, Syed et al. [38] also mention that there are aspects that digital technologies still cannot resolve: reliance on others, inadequacy of trained caregivers, healthcare costs, and increase in diseases with no cure (e.g., Alzheimer and Parkinson diseases).

Balancing the advantages and limitations, we agree with Ajami et al. [55] and with Rubi and Gondim [56] in arguing that this new solution proves to be very relevant because it incorporates AI methods with the possibility of becoming proactive through the entire process that includes observation, learning, adaptation, prediction, and decision-making within an intelligent environment. In this way, the IoT combined with AI creates new perspectives and new types of intelligent pervasive systems and platforms because this new type of intelligence is dynamic and interactive, making it more capable of adapting to the changes that take place. However, Syed et al. [38] are of the opinion that success depends on the degree of acceptance of the elderly. In this sense, it is important that the elderly know which spaces have sensors because, for them, aspects related to their privacy and confidentiality must be guaranteed. From a more extreme perspective, a “smart home” can lead to processes that increase the loneliness of the elderly because this new digital environment can meet practically all their needs. And, for this reason, the elderly will not need to collect these aids in a space outside their homes.

## 3 Reflections and Proposals for the Healthcare with Emergent Technologies

Not being, from the beginning, a novelty or a surprise, there is a general feeling that it is urgent and pertinent to create all the conditions for an increase in the levels of info-inclusion of the elderly. Various investigations have sought to find clues that may provide viable proposals to fulfill this objective. Perhaps the simplest, as suggested by Castilla et al. [18], is that the elderly feel the need, on their own, to access and use digital resources. However, as Castilla et al. [18] also mention, the elderly must feel self-confident, but previous training and the existence of a support team will be essential, not only for more technological purposes but fundamentally for emotional reasons. Therefore, Wang and Chen [57] are of the opinion that all variables should be taken care of so that these first experiences are positive and pleasant and that they also have a practical sense in the routines of the elderly to increase their levels of well-being and health.

In this sense, as stated by Zaid et al. [58], it is important to increase the opportunities and levels of exposure of the elderly to technologies because in the recent past, research has shown positive effects in reducing loneliness and more depressive states. And, therefore, the elderly increases their levels of health and well-being. At the same time, the increase in their digital skills has also been promoting the participation of the elderly in lifelong learning programs within the scope of the so-called active aging [59]. For this reality to be even more expressive, Zaid et al. [58] are of the opinion that to increase the rates of adoption of technologies by the elderly, it is also very important and may even be very decisive for the elderly to feel the immediate benefits of these technologies for their daily routines. However, we must pay attention to a detail that can make all the difference: digital technologies are designed by young people and, for that reason, are designed and created to meet their needs, which, in general, are different from the needs of the elderly [60]. For this order of reasons, Martinez-Alcalá et al. [29] reinforce the need for these technologies to be more user-centered by involving the elderly in their design and development: "require useful, functional, user-friendly and meaningful technology." To this end, it is important to adopt a "user-centered approach" view, which we believe is a way to put into practice and which is supported by Merkel and Kucharski ([61], p. 18) "involving older people in the process of developing new gerontechnologies leads to a better acceptance and/or use of the innovative products." In this way, the elderly did not feel that they were being treated as guinea pigs, and they could feel uncomfortable because they felt that their home had become a laboratory and not the place where they "only" live with the safeguard of their dignity and privacy. For these goals to be achieved, it is necessary for the elderly to be aware that the expansion of health services with the integration of IoT/IoMT will allow health professionals to collect data that are treated and analyzed, while the elderly remain in the comfort of their own home [30]. And, given the large volume of data, it is possible to make predictions and diagnoses that are more correct and more adjusted to the needs of each elderly person.

As Maskeliunas et al. [52] mention and with which we agree, the motivation of the elderly is fundamental for carrying out activities if they require the least possible effort and the elderly feel that these activities have meaning and are of practical use for their daily routines. This motivation is based on a social perspective that includes their family members and their grandchildren, in what is usually called intergenerational relationships. In this regard, Sun et al. [62] report that family support and encouragement can significantly promote the use of the Internet by the elderly. Still from the perspective of Maskeliunas et al. [52], it is increasingly possible to customize digital resources, which makes the fear of using them for those who are info-excluded diminish. In the same vein, the latest digital capabilities tend to be more intuitive with low technical demands. In this way, it is also easier to adapt these digital resources for seniors who have some obvious disabilities.

According to recent studies and investigations, there is an increasingly evident trend toward the primordial role of digital technologies/platforms within the context of AAL, which Choi et al. ([30], p. 198) present as:

- IoT + architecture = smart home
- IoT + healthcare = daily healthcare devices
- IoT + healthcare + architecture = ALL healthcare platform

Also, as stated by Choi et al. [30], it is not enough to think and focus only on the digital-technological dimensions. On the contrary, what is fundamental and absolutely a priority is that AAL platforms will be based on the behaviors and desires of the elderly in their homes.

In general terms, it can be said that there is a low adjustment of the elderly toward digital technologies, which is often related to lack of incentives, economic problems, and digital skills associated with the lack of a policy that promotes the training of the elderly. From the point of view of Roupa et al. [63], training programs for the elderly should be designed according to their needs and priorities, and, for this purpose, we can rely on their families and grandchildren in an intergenerational perspective. Being aware that the elderly group is the most heterogeneous social group, especially because of their different life experiences (e.g., academic, professional, economic, social, religion, etc.), it becomes more difficult to find a model of training that can be understood as the most suitable for this elderly population. However, there are data from different investigations that allow giving clues to which elderly people will potentially be info-excluded. In this context, as stated by Sun et al. [62], attention should focus on the following variables: socioeconomic and demographic factors that include age, sex education, income status, health literacy, and urban and rural conditions. Consequently, several investigations have shown that older adults who are male, are younger, live in urban areas, and are of higher-income status and education use the Internet much more compared to the rest of the elderly [64].

According to Sun et al. [62], when the quality of life increased by a grade, Internet use increased 2.241 times, and, in this context, one cannot forget that the quality of life of a citizen is multidimensional, which includes not only physical health but also psychological health and social health. In this perspective, we also agree with Sun et al. [62] when stating that the greater social interaction among the elderly should also be encouraged to enable them to reap the benefits of the digital age.

Digital identity is another dimension that must be acquired during the construction of the digital identity because it is not enough just to know how to adhere and use digital resources, that is, knowing how and for what effect they use digital resources. As Blažic and Blažic [60] mention, this digital identity will allow the elderly to be more easily included and recognized in the digital context. In this way, conditions will be met for a fairer and more inclusive society.

The future of healthcare is exceptionally encouraging, seeing the rapid development in technology, IoT, AI, machine learning, and AAL systems. The elderly have new opportunities as well as challenges and significant changes that must be mastered. Digital literacy is fundamental for the enhancement of the overall user experience of Health Care 4.0. Even though many seniors are digitally vulnerable because they are info-excluded, we believe that with assertive measures of lifelong learning in knowledge-oriented digital skills and using new technologies, they will impact how seniors will deal with smart healthcare in AAL. Increasing the use of digital health among seniors requires understanding these technological barriers as well as the unique health needs of seniors to help them find the right ambient assisted living technologies and use them confidently and safely.

In this regard, helping seniors adopt technology involves more than just providing connected devices. It requires providing educational programs to increase digital health literacy, as well as a personalized approach to communication preferences with health professionals, and thus guiding individuals throughout their personal care experiences to healthier outcomes. Besides, when older people engage with technology, there are also potential benefits. The process of learning new skills can on the one hand contribute to increased well-being and independence; on the other hand, technology can help older adults overcome many issues that impact their physical and mental health, prevent some symptoms of health conditions, independent living support, social participation, and well-being.

To conclude, taking the above into consideration, we present some proposals in the field of training digital competence, based on the literature but mainly on our own experience in digital inclusion activities for the elderly: the elderly should be included from the earliest stages in a process of co-design not only of training program development but also of applications designed for them; identify each person's interests with regard to digital technology to respond to these interests; content, methods, and strategies should be varied, combining class time with self-study and blended peer learning with self-study, individual support, and also external visits; and training programs should be practical, flexible, and inclusive, with ongoing sharing of best practices, resources, and experiences among participants for the successful implementation of the training. Family support and motivation is moreover important.

The constant and rapid technological and scientific developments in healthcare, whether in the digitalization of health services or in healthcare assistive environments, require a continuous approach to the development of digital and health literacy competencies in a lifelong learning perspective. Therefore, empowering the elderly with literacy skills could enable them to benefit from the use of various digital health sources to improve their health quality and enable them to live independently for longer.

## References

1. Goodfellow, R. (2011). Literacy, literacies and the digital in higher education. *Teacher in Higher Education, 16*(1), 131–144.
2. Lankshear, C., & Knobel, M. (2003). *New literacies: Changing knowledge in the classroom*. Open University Press.
3. Gilster, P. (1997). *Digital literacy*. John Wiley.
4. European Commission. (2006). *Recommendation on key competences for lifelong learning. Council of 18 December 2006 on key competences for lifelong learning, 2006/962/EC, L. 394/15*. Retrieved September 10, 2022, from: http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:32006H0962&qid=1496720114366
5. Lankshear, C., & Knobel, M. (2015). Digital literacy and digital literacies: Policy, pedagogy and research considerations for education. *Nordic Journal of Digital Literacy, 2006-2016*, 8–20.
6. Buckingham, D. (2015). Defining digital literacy: What do young people need to know about digital media? *Nordic Journal of Digital Literacy, 4*, 21–34.

7. Gourlay, L., Hamilton, M., & Lea, M. (2013). Textual practices in the new digital literacies. *Research in Learning Technology, 21*, 21438.
8. Joonesten, T., Pasquini, L., & Harness, L. (2012). Guiding social media at our institutions. *Planning for Higher Education, 41*(1), 125–135.
9. Stordy, P. (2015). Taxonomy of literacies. *Journal of Documentation, 71*(3), 456–476.
10. Martin, A. (2006). A European framework for digital literacy. *Nordic Journal of Digital Literacy, 1*(2), 151–161.
11. Lankshear, C., & Snyder, I. (2001). *Teachers and technoliteracies*. Allen & Unwin.
12. Pangrazio, L., & Sefton-Green, J. (2021). Digital rights, digital citizenship and digital literacy: What's the difference? *Journal of New approaches in Educational Research, 10*(1), 15–27. Retrieved September 11, 2022, from: Digital Rights, Digital Citizenship and Digital Literacy: What's the Difference? | Pangrazio | Journal of New Approaches in Educational Research (ua.es)
13. Casey, L., & Bruce, B. (2011). The practice profile of inquiry: Connecting digital literacy and pedagogy. *E-Learning and Digital Media, 8*(1), 76–85.
14. Tan, E. (2013). Informal learning on Youtube: Exploring digital literacy in independent online learning. *Learning Media and Technology, 38*(4), 463–477.
15. European Commission. (2010). *A Digital Agenda for Europe.* Brussels. Retrieved September 10, 2022, from: http://ec.europa.eu/information_society/digital-agenda/index_en.html
16. OECD. (2005). *Definition and Selection Key Competencies – Executive Summary*. Retrieved September 10, 2022, from: https://www.oecd.org/pisa/35070367.pdf
17. Tamborg, A., Dreyøe, J., & Fougt, S. (2018). Digital literacy – A qualitative systematic review. *Lǽring & Medier, 19*, 1–29.
18. Castilla, D., Botella, C., Miralles, I., Bretón-López, J., Dragomir-Davis, A., Zaragoza, I., & Garcia-Palacios, A. (2018). Teaching digital literacy skills to the elderly using a social network with linear navigation: A case study in a rural area. *International Journal of Human-Computer Studies, 118*, 24–37.
19. European Commission. (2022). *Digital inclusion*. Retrieved October 3, 2022, from: https://digital-strategy.ec.europa.eu/policies/digital-inclusion
20. Eurostat. (2022). *Individuals' level of digital skills*. Retrieved October 3, 2022, from: https://ec.europa.eu/eurostat/databrowser/view/ISOC_SK_DSKL_I21__custom_3494070/default/table?lang=en
21. Almeida, A., Mulero, R., Rametta, P., Urosevic, V., Andric, M., & Patrono, L. (2019). A critical analysis of an IoT-aware AAL system for elderly monitoring. *Future Generation Computer Systems, 97*, 598–619.
22. Vuorikari, R., Kluzer, S., & Punie, Y. (2022). DigComp 2.2: The digital competence framework for citizens - with new examples of knowledge, skills and attitudes, EUR 31006 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-48882-8, https://doi.org/10.2760/115376, JRC128415.
23. Andrejevic, M. (2020). *Automated Media*. Routledge. Retrieved September 11, 2022, from: https://doi.org/10.4324/9780429242595
24. Shin, D., & Seger, W. (2016). Web 2.0 technologies and parent involvement of ELL students: An ecological perspective. *The Urban Review: Issues and Ideas in Public Education, 48*(2), 311–332.
25. Thrane, L., Shelley, M., Shulman, S., Beisser, S., & Larson, T. (2004). E-political empowerment: Age effects or attitudinal barriers. *Journal of E-Government, 1*(4), 21–37. Retrieved September 12, 2022, from: https://doi.org/10.1300/J399v01n04_03
26. Richardson, J. (2015). *Law and the Philosophy of Privacy*. Routledge. https://doi.org/10.4324/9780203516133
27. Lee, B., Chen, Y., & Hewitt, L. (2011). Age differences in constraints encountered by seniors in their use of computers and the internet. *Computers in Human Behavior, 27*, 1231–1237.
28. Rahman, T., Amalia, A., and Aziz, Z. (2021, January). From digital literacy to digital intelligence. In *4th international conference on Sustainable Innovation 2020–Social, Humanity, and Education (ICoSIHESS 2020)* (pp. 154–159). Atlantis Press. 13th-14th October 2020. Makarova.

29. Martinez-Alcalá, C., Rosales-Lagarde, A., Pérez-Pérez, Y., Lopez-Noguerola, J., Bautista-Diaz, M., & Agis-Juarez, R. (2021). The effects of Covid-19 on the digital literacy of the elderly: Norms for digital inclusion. *Frontiers in Education*. https://doi.org/10.3389/feduc.2021.716025
30. Choi, D., Choi, H., & Shon, D. (2019). Future changes to smart house based on AAL healthcare service. *Journal of Asian Architecture and building Engineering, 18*(3), 190–199.
31. UNECE. (2021). *Ageing in the digital era*. Policy Brief on Ageing No. 26. Retrieved October 3, 2022, from: https://unece.org/sites/default/files/2021-07/PB26-ECE-WG.1-38.pdf
32. Thomas, J. (2015). *Insights into loneliness, older people and well-being*. Retrieved September 30, 2022, from: https://backup.ons.gov.uk/wp-content/uploads/sites/3/2015/10/Insights-into-Loneliness-Older-People-and-Well-being-2015.pdf
33. UN. (2015). *World Population Prospects*. Retrieved September 30, 2022, from: https://population.un.org/wpp/Publications/Files/WPP2015_DataBooklet.pdf
34. Abdi, J., Al-Hindawi, A., Ng, T., & Vizcaychipi, M. P. (2018). Scoping review on the use of socially assistive robot technology in elderly care. *BMJ Open, 8*(2), 1–20. https://doi.org/10.1136/bmjopen-2017-018815
35. Salovaara, A., Lehmuskallo, A., & Hedman, L. (2010). Information technologies and transition in the lives 55-65 years old: The case colliding life interests. *International Journal of Human Compared Studies, 68*(1), 803–821.
36. Heo, J., Chun, S., & Lee, S. (2015). Internet use and well-being in older adults. *Education Psychology Cyberpsychology Behavior and Social Network, 18*(5), 268–272.
37. Rashidi, P., & Mihailidis, A. (2013). A survey on ambient assisted living tools for older adults. *Journal of Biomedical Health Informatics, 17*(3), 579–590.
38. Syed, L., Jabeen, S., Manimala, S., & Alsaeedi, A. (2019). Smart healthcare framework for ambient assisted living using IoMT and big data analytics techniques. *Future Generation Computer Systems, 101*, 136–151.
39. Mainetti, L., Patrono, L., Secco, A., & Sergi, I. (2016). *An IoT-aware AAL system for elderly people*. Retrieved October 4, 2022, from: https://ieeexplore.ieee.org/abstract/document/7555929
40. Sankar, S., Srinivasan, P., & Saravanahumar, R. (2018). Internet of things based ambient assisted living for elderly people health monitoring. *Research Journal Pharmacy and Technology, 11*(9). https://doi.org/10.5958/0974-360X.2018.00715.1
41. Dohr, A., Modre-Osprian, R., Drobics, M., Hayn, D., & Schreider, G. (2010). *The internet of things for ambient assisted living*. Retrieved October 1, 2022, from: https://www.academia.edu/6478405/The_Internet_of_Things_for_Ambient_Assisted_Living
42. Kunze, C., Holtman, C., Schmidt, A., & Stork, W. (2007). *Kontextsensitive Technologien und Intelligente Sensorik für Ambient-Assisted-Living-Anwendungen*. Retrieved October 1, 2022, from: https://www.researchgate.net/profile/Andreas-Schmidt 11/publication/229422957_Kontextsensitive_Technologien_und_intelligente_Sensorik_fur_Ambient-Assisted-Living Anwendungen/links/0912f50c7aa1d1130a000000/Kontextsensitive-Technologien-und-intelligente-Sensorik-fuer-Ambient-Assisted-Living-Anwendungen.pdf
43. Harum, N., Abidin, Z., Shah, W., & Hassan, A. (2018). Implementation of smart monitoring system with fall detector for elderly using IoT technology. *International Journal of Computing, 17*(4), 243–249.
44. European Commission. (2021). *Next-generation internet of things and edge computing*. Retrieved October 1, 2022, from: file:///C:/Users/Downloads/ng_iot_and_edge_computing_executive_report
45. Gigli, M., & Koo, S. (2011). Internet of things: Services and applications categorization. *Journal of Advanced Internet Things, 1*, 27–31.
46. International Telecommunication Union [ITU]. (2005). *ITU internet reports 2005: The internet of things - executive summary*. Retrieved October 28, 2022, from: https://www.itu.int/osg/spu/publications/internetofthings/InternetofThings_summary.pdf
47. Wu, Q., Ding, G., Xu, Y., Feng, J., & Long, K. (2014). *Cognitive Internet of Things: A new paradigm beyond connection*. Retrieved October 6, 2022, from: https://ieeexplore.ieee.org/document/6766209

48. Lloret, J., Canovas, A., Sendra, S., & Parra, L. (2015). A smart communication architecture for ambient assisted living. *IEEE Communications Magazine, 53*(1), 26–33. https://doi.org/10.1109/MCOM.2015.7010512
49. Suryadevara, N., & Mukhopadhyay, S. (2014). *Determining wellness through an ambient assisted living environment*. Retrieved October 7, 2022, from: https://ieeexplore.ieee.org/abstract/document/6813391
50. Maskeliunas, R., Damaševicius, R., & Segal, S. (2021). A review of internet of things technologies for ambient assisted living environments. *Future Internet, 11*, 259. https://doi.org/10.3390/fi11120259
51. Suciu, G., Vulpe, A., Craciunescu, R., Butca, C., & Suciu, V. (2015). *Big data fusion for eHealth and ambient assisted living cloud applications*. Retrieved October 7, 2022, from: https://ieeexplore.ieee.org/abstract/document/7185095
52. Maskeliunas, R., Damaševicius, R., & Segal, S. (2019). A review of internet of things technologies for ambient assisted living environments. *Future Internet, 11*, 259. https://doi.org/10.3390/fi11120259
53. Villari, M., Al-Anbuky, A., Celesti, A., & Moessner, K. (2016). Leveraging the internet of things: Integration of sensors and cloud computing systems. *International Journal of Distributed Sensors Network, 12*, 7. Retrieved October 7, 2022, from: https://journals.sagepub.com/doi/epub/10.1177/155014779764287
54. Duijnhoven, J., Aarts, M., Kort, H., & Rosemann, A. (2021). External validations of a non-obtrusive practical method to measure personal lighting conditions in offices. *Lighting Research & Technology, 53*(4), 285–310.
55. Ajami, H., Mcheick, H., & Mustapha, K. (2019). *A pervasive healthcare system for COPD patients*. Retrieved October 9, 2022, from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6963281/
56. Rubi, J., & Gondim, P. (2019). *IoMT platform for pervasive healthcare data aggregation, processing, and sharing based on OneM2M and OpenEHR*. Retrieved October 9, 2022, from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6806104/pdf/sensors-19-04283.pdf
57. Wang, C., & Chen, J. (2015). Overcoming technophobia in priory-educated elderly – The HELPS – Seniors service-learning program. *International Journal of Automation Smart Technology, 5*(3), 173–182.
58. Zaid, N., Rauf, M., Ahmad, N., Zainal, A., Hanis, F., & Shahdan, T. (2022). Learning elements for digital literacy among elderly: A scoping review. *Journal of Algebraic Statistics, 13*(3), 2850–2859.
59. Wang, C., & Wu, C. (2021). Bridging the digital divide: The smart TV as a platform for digital literacy among the elderly. *Behavior & Information Technology*, 1–14. https://doi.org/10.1080/0144929x.2021.1934732
60. Blažic, B., & Blažic, A. (2020). Overcome the digital divide with a modern approach to learning digital skills for the elderly adults. *Education Information Technology, 25*(1), 259–279.
61. Merkel, S., & Kucharski, A. (2019). Participatory design in gerontechnology: A systematic literature review. *Gerontologist, 59*(1), 16–25.
62. Sun, X., Yan, W., Zhou, H. *et al.* (2020). Internet use and need for digital health technology among the elderly: A cross-sectional survey in China. *BMC Public Health* **20**, 1386. https://doi.org/10.1186/s12889-020-09448-0
63. Roupa, Z., Nikas, M., Gerasimou, E., Zaferi, V., Giasyrani, L., Kazitori, E., & Sotiropoulou, P. (2010). The use of technology by the elderly. *Health Science Journal, 4*(2), 118–126.
64. Levy, H., Janke, A., & Langa, K. (2015). *Health literacy and the digital divide among older Americans*. Retrieved October 14, 2022, from: https://link.springer.com/content/pdf/10.1007/s11606-014-3069-5.pdf

# Segmentation of Lung Lesions Caused by COVID-19 in Computed Tomography Images Using Deep Learning

**Saul Barraza-Aguirre, Jose Diaz-Roman, Carlos Ochoa-Zezzatti, Boris Mederos-Madrazo, Juan Cota-Ruiz, and Francisco Enriquez-Aguilera**

## Abbreviations

| | |
|---|---|
| AI | Artificial intelligence |
| CNN | Convolutional neural networks |
| COVID-19 | Coronavirus disease |
| CT | Computed tomography |
| DICE | Sorensen-Dice similarity coefficient |
| DL | Deep learning |
| GGO | Ground-glass opacity |
| ICT | Information and communications technologies |
| IoU | Intersection over union |
| ML | Machine learning |
| ReLU | Rectified linear unit |
| ResNet | Residual network |
| SegNet | Segmentation network |

## 1 Introduction

Smart cities are transforming the way people live, work, and socialize. Smart cities offer a wide range of opportunities to improve the quality of life for citizens, and smart healthcare is one of the most promising areas. In this regard, smart

S. Barraza-Aguirre · J. Diaz-Roman (✉) · C. Ochoa-Zezzatti · B. Mederos-Madrazo · J. Cota-Ruiz · F. Enriquez-Aguilera
Autonomous University of Ciudad Juarez, Chihuahua, Mexico
e-mail: david.roman@uacj.mx



N. Gupta, S. Mishra (eds.), *Internet of Everything for Smart City and Smart Healthcare Applications*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-34601-9_14

cities can play a key role in providing a robust infrastructure and data platform for healthcare innovation. For example, the implementation of sensor technologies and data analytics in smart cities can enable healthcare professionals to gather valuable information about population health, identify disease patterns and trends, and act preventively to address public health challenges. In addition, smart cities can integrate telemedicine technologies to provide remote, real-time medical care, which can be especially useful in remote or hard-to-reach areas. Furthermore, the internet of things and big data technologies can be merged with smart health to address some of the issues with increasing the accessibility, affordability, and availability of healthcare. The connection between smart cities and smart healthcare has the potential to significantly improve healthcare and the well-being of the population [1, 2].

Smart cities require the implementation of systems capable of monitoring, processing, and communicating information from different areas, such as environmental, security, mobility, and healthcare, among others. Such systems must allow an adequate analysis of the information to support decision-making in a more accurate way and even take action automatically to control the variables that are being monitored. In healthcare, intelligent systems offer the physician a wide variety of applications to support patient diagnosis, monitor the evolution of the disease or recovery during therapy, and even choose the best treatment option [3].

In the last decade, most intelligent healthcare systems have employed some form of artificial intelligence, either machine learning or deep learning, in processing tasks that enable signal and medical image enhancement, pattern recognition, outlier identification, classification, and segmentation, among others. These tasks have been used in different applications such as cardiac arrhythmia detection, epileptic event identification, cancer diagnosis, cellular classification, and tumor region segmentation, to name a few [4].

The coronavirus disease 19 (COVID-19) pandemic has prompted the development and implementation of advanced healthcare technologies, including artificial intelligence (AI), to diagnose the virus and monitor its evolution. AI-based automatic diagnostic systems represent an almost mandatory tool for smart hospitals, as they allow timely and effective decision-making. These systems use machine learning techniques to analyze large amounts of data and recognize specific patterns associated with COVID-19. By integrating these systems into smart hospitals, doctors and healthcare staff can make a faster and more accurate diagnosis, helping to reduce the spread of the virus and improve treatment outcomes for affected patients. In addition, AI can also help predict the evolution of the virus and improve the effectiveness of treatments. In summary, the use of artificial intelligence techniques for the diagnosis and automatic monitoring of COVID-19 has become a fundamental tool in the fight against this disease and its sequelae [4, 5].

### *1.1 Coronavirus Disease 19*

COVID-19 primarily attacks the respiratory tract and commonly causes lesions that affect the lungs. Imaging studies of patients with COVID-19 have shown abnormalities related to virus-induced organizing pneumonia. This condition represents damage to lung tissue that is characterized by the presence of fibroblasts that fill the alveolar sacs and ducts, giving rise to the formation of granular tissue [6]. Similarly, in severe cases of COVID-19, there is a high risk of developing pulmonary fibrosis [7, 8]. For this reason, different techniques, algorithms, and research have been developed to optimize the diagnosis, treatment, and monitoring of this disease.

In order to understand the lung condition caused by this disease, a series of investigations were developed based on the findings observed in computed tomography (CT) images, which can be related to the use of methods based on deep learning (DL) techniques, producing a great contribution as an alternative for diagnosis. In studies like the one by Uysal et al. [9], evidence of lung lesions was discovered using CT in asymptomatic patients with a positive diagnosis of COVID-19, similar to those discovered in patients with illness symptoms. Vascular enlargement, ground-glass opacity (GGO), air bronchogram, pulmonary consolidation, pleural effusion, and mixed GGO with crazy-paving pattern or consolidation are the most common abnormal findings on a chest CT of a patient with COVID-19 [9, 10]. The fast scan time and high resolution of the acquired image, which are helpful in identifying and classifying lung abnormalities, are benefits of CT diagnosis.

On the other hand, the segmentation of medical images is a technique used mainly for the diagnosis and monitoring of diseases. In general terms, it consists of the partitioning and labeling of regions that have common characteristics within an image, such as texture, color, or intensity; this procedure can be manual, semi-automatic, or automatic. This technique is relevant in the medical field since it facilitates visualization and quantification within a medical image, allows anatomical regions to be classified to detect anomalies such as deformities and injuries, or assesses the therapeutic response of a treatment to a specific disease [11]. It is important to segment the lesions caused by COVID-19 to have a more precise criterion of the state of health of a patient; this would help to make better decisions for the treatment and monitoring of the disease. The objective of this work is the development of a deep learning-based system for the segmentation of lung lesions caused by COVID-19 in chest CT images.

## 2 Background

There are a variety of approaches that use machine learning (ML) methods, particularly deep learning (using convolutional neural networks), to automatically diagnose COVID-19 and detect areas with COVID-19 lesions in CT scans. One of them is the work carried out by N. Dey et al. in 2020, where they designed a structure based on machine learning with two stages: the first consisted of image segmentation to detect COVID-19 infection, and the second served to classify the state of the lung, between healthy and affected by COVID-19. First, a thresholding of three levels of social group optimization Kapur's entropy was used; later, the K-means clustering technique allowed them to separate the original CT image into three different regions, between the fundus, the lung section, and the artifact. Finally, the affected area was extracted by means of morphological filters. Once the image characteristics were obtained, a machine learning scheme was used to classify the patient's health status, between positive for COVID-19 and negative; an overall accuracy of 91% was obtained [12].

In another investigation, S. Nurmaini et al. proposed a system in 2020 to semi-automatically identify lesions caused by COVID-19 on CT images. Initially, the infected areas were delineated manually with the help of a professional radiologist; then, using convolutional neural networks (CNN), they trained the system and detected infection in each of the CT scan slices. This process was iterated with the help of radiologists to validate the affected areas; this is to obtain better performance, with results of up to 99% accuracy and 100% specificity. The authors indicated that the use of CNN is capable of optimizing the results for the detection of lung lesions caused by COVID-19, with greater sensitivity than reverse transcription polymerase chain reaction (RT-PCR) tests that have shown a sensitivity range between 60% and 97%, which could generate false-negative results, converting the diagnosis by means of CT in an alternative when there is clinical evidence of an infection caused by COVID-19 [13].

On the other hand, X. Xu et al. in 2020 suggested a strategy to distinguish between COVID-19-positive cases and healthy ones using deep learning in CT images. The images used in this work were preprocessed to obtain the area of interest to be analyzed later, then the affected areas were segmented using a 3D CNN model, and with this segmentation, the images were classified into three categories: COVID-19, influenza A viral pneumonia (IAVP), and irrelevant to infection. The infections caused by COVID-19 have unique characteristics that allow them to be differentiated from other diseases such as IAVP or from totally healthy cases. Such characteristics are based on the intensity of the opacity present in the CT images, where the opacities caused by COVID-19 are faint in contrast to the opacities caused by IAVP which are more intense opacities. In the cases of healthy patients, no type of opacity is present in the entire lung tissue, this being a critical characteristic for image classification. The classification task was performed using the ResNet-18 structure, which is a pre-trained CNN for image classification. The results showed an overall accuracy of 86.7%, making it a promising algorithm for the detection of COVID-19 cases [14].

In another work, carried out by J. Suri et al. in 2021, the performance of 3 different artificial intelligence models was evaluated and compared on 5000 CT slices. The images were binarized delimiting and segmenting the area of the lungs; this processing was performed with the ImgTracer™ software under the supervision of two specialized radiologists, thus generating two masks for each image. The three deep learning models used were Pyramid Scene Parsing Network (PSPNet), Visual Geometry Group Segmentation Network (VGG-SegNet), and Residual Network (ResNet-SegNet). The model with the highest performance proved to be ResNet-SegNet, having a percentage difference of 0.4% compared to the segmentation performed by the radiologists [15].

In a research carried out by A. Saood and I. Hatem in 2021, a comparison was made between two methods of lung segmentation in CT images based on deep learning (U-Net and SegNet), with the aim of detecting and labeling lung tissues infected by COVID-19. Although both techniques are built based on deep learning, their structure is different; on the one hand, SegNet is a deep neural network initially designed to segment scenes; on the other hand, U-Net from its origin was created with a medical application approach. For training, binarized images were used to discriminate infected from healthy tissue. Each network was trained with 72 images and tested with 18 images, this being a limiting aspect in the training and validation of the algorithm since only 90 images were used in total. The ability of both structures to distinguish between infected and non-infected images was high; in both cases, they exceeded a 95% precision rate regarding the segmentation of infected areas. U-Net obtained a sensitivity of 96%, while SegNet was 94%, both showing promising results and concluding that the use of deep learning in the detection of COVID-19 is highly viable [16].

Another approach to lesion detection in medical imaging is the use of radiomics. This consists of the acquisition of data on the characteristics of medical images with the aim of extracting patterns, structures, or distributions that are not easy to identify with the naked eye, helping the timely diagnosis of diseases. Z. Chen et al., in 2021, developed an alternative to segment and analyze textural features in lung parenchymal regions on CT images to help clinicians in the diagnosis of COVID-19. In the research, they used a three-dimensional V-Net through a spatial transformation network to perform the segmentation; first, they performed a binary thresholding to subsequently apply the network named SP-V-Net and segment the infected regions; with the segmented image, radiomics features are extracted, and COVID-19 is detected. The results were evaluated by sensitivity and specificity, reaching values of 96.7% and 92.7%, respectively. The clinical focus of this research brings a great contribution to the area of radiology since the interpretation of medical images is a daily activity in this area and forms part of patient-directed treatment decisions. The SP-V-Net network performed encouragingly well, and the authors mention the ability to transfer the application to other segmentation areas [17].

Although there have been many contributions to the diagnosis of COVID-19 and the automatic segmentation of lung lesions caused by this disease, there is still room for improvement in these areas. One area where this is possible is by exploring alternative training techniques and loss functions for deep learning models in order to achieve better performance on the segmentation task.

## 3 Healthcare in Smart Cities

Nowadays, the progressive adoption of information and communications technologies (ICT) in the healthcare sector has led to the practically generalized implementation of electronic medical records – something that makes it possible to improve health prevention – and, more recently, to the appearance of new medical services that can be accessed through mobile devices, making up what is known as e-health or eHealth. Currently, the concept of e-health is evolving in parallel to that of smart cities, which also aims to introduce the widespread use of ICT in cities, in order to improve the quality of life of its inhabitants, taking into account various indicators such as sustainability, energy consumption, mobility, health, and governance, among others. Smart cities make extensive use of sensors of all kinds, which provide information such as temperature, humidity, pollution, allergen concentration, and traffic conditions, among a plethora of others; this information constitutes what we can call the "context" of the city that can help us understand the environment in which a citizen develops their daily activities and how the environment affects each citizen. If we use this information appropriately, health applications can be developed that adapt their behavior to this context, giving rise to the concept of smart health as a result of the synergy between existing e-health applications and the information that a smart city can provide. A simple example could be one in which a patient receives information on his mobile device about pollen concentrations and other elements of the city to which he is allergic, recommending a route that avoids the most conflictive points, as well as nearby pharmacies where to buy the respective medicines. In addition, the application could continuously monitor their vital signs and generate alarms for immediate medical attention if necessary. The development and generalization of the smart health concept would be beneficial to society as a whole. Considering the aging population and the high costs of the healthcare system, both public and private, e-health applications would improve the quality of life and make the patient and the citizen in general less dependent on face-to-face medical services, reducing hospital visits and making many medical services cheaper.

To achieve this, it is clear that multidisciplinary research is needed, which is why large technology companies and research centers around the world are already doing so, in fields such as the integration of sensory information, data security, massive information storage and analysis, psychology, citizen adoption of new technologies, and critical situation management, as can be seen in the following example, shown in Fig. 1.

On the other hand, artificial intelligence is becoming an increasingly important technology in the healthcare field, and its integration into smart cities can have a significant impact on smart healthcare. The use of advanced technologies in healthcare can help improve access to healthcare services, reduce costs, and improve the quality of care for patients. AI-based healthcare applications can analyze large amounts of data and help healthcare professionals detect and prevent diseases more efficiently and accurately. For example, AI can be used to analyze the movement patterns of people in a city and predict the spread of contagious diseases, such as COVID-19. It can also be used to analyze medical images, identify patterns, and predict disease diagnoses. In addition, AI can help improve the efficiency of healthcare systems by automating tasks such as appointment scheduling, medical record management, and medication administration. In short, the integration of AI in smart cities can have a significant impact on smart healthcare by improving disease

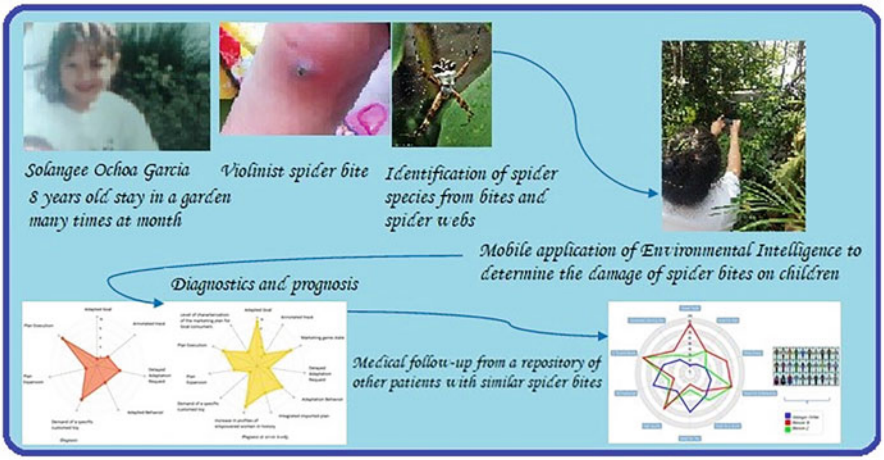



**Fig. 1** Example of a smart care-type intelligent system

detection, prevention, and treatment, which can improve people's quality of life and reduce healthcare costs [4].

# 4 Theory

This section includes an overview of the fundamental ideas of computed tomography and also reviews various topics in artificial intelligence, with special attention to deep learning architectures for image segmentation.

## *4.1 Computed Tomography*

Computed tomography is a diagnostic medical study that consists of the reconstruction of cross-sectional images of the human body through the emission of X-rays. The main elements that make up a tomography equipment are [18, 19]:

- X-ray source: It is responsible for emitting energy into the human body, which affects the human body in the form of a thin beam as can be seen in Fig. 2.
- Detectors: They are sensitive and responsible for capturing X-rays; they absorb the remaining energy after passing through the patient and measure it to transform it into electrical energy to be processed and generate the image reconstruction.

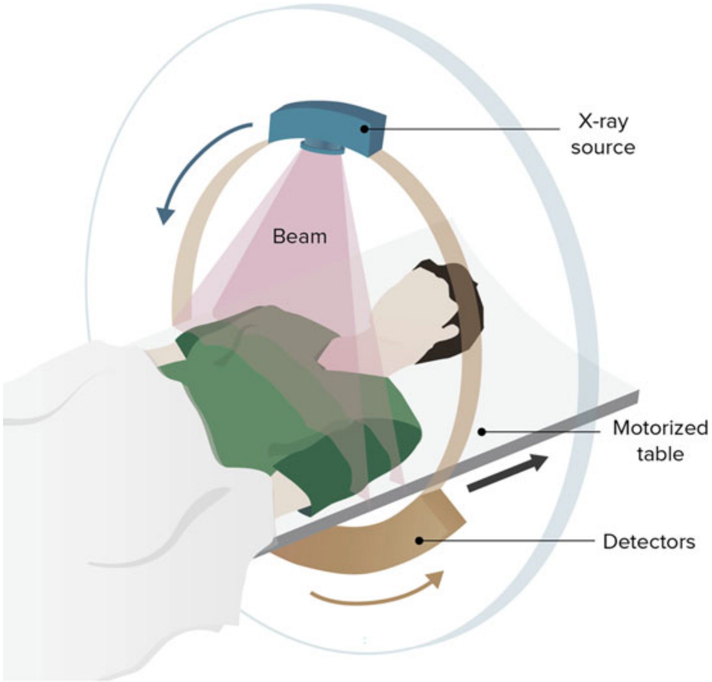



**Fig. 2** Main parts of a tomography equipment [18]

- Table: In addition to keeping the patient in a stable position, it also fulfills the function of moving the patient to cover more sections of the human body if necessary for the study.

The image is reconstructed using attenuation coefficients; these numbers are calculated with the energy perceived by the detectors and are calculated in relation to water in Hounsfield units (HU) of the corresponding material or tissue. Equation 1 used to determine it in Hounsfield units is expressed as follows:

$$HU_{mat} = \frac{\mu_{mat} - \mu_{water}}{\mu_{mat}} \bullet 1000 \quad (1)$$

where $\mu$ represents the linear attenuation coefficient; this corresponds to the ability of a material or tissue to prevent the passage of photons ($\mu_{water}$ is taken as the zero-reference point on the scale of Hounsfield units). The different tissues of the human body have different $\mu$ because each of them interacts to a greater or lesser extent with X-rays. There are reference charts, like the one shown in Fig. 3, where denser tissues like bone have a higher value on this scale and appear whiter on a CT image; less dense tissues, such as the lungs, tend to be darker on CT images [20].

The interaction between the lungs and X-rays that exists in a CT scan under normal conditions turns out to be low because a large part of the lungs is made up of

| Material | HU |
|---|---|
| Air | -1000 |
| Lung | -830 to -200 |
| Fat | -30 to -250 |
| Water | 0 |
| Heart | 10 yo 60 |
| Brain | 20 to 40 |
| Blood | 20 to 80 |
| Liver | 20 to 80 |
| Muscle | 35 to 50 |
| Spleen | 40 to 60 |
| Bone | 150 to 500 |
| Bone (dense) | 350 to 1000 |
| Metal | > 2000 |

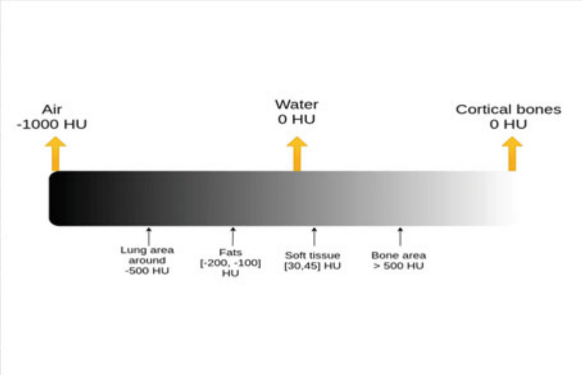



**Fig. 3** Hounsfield units of some tissues and materials of the human body

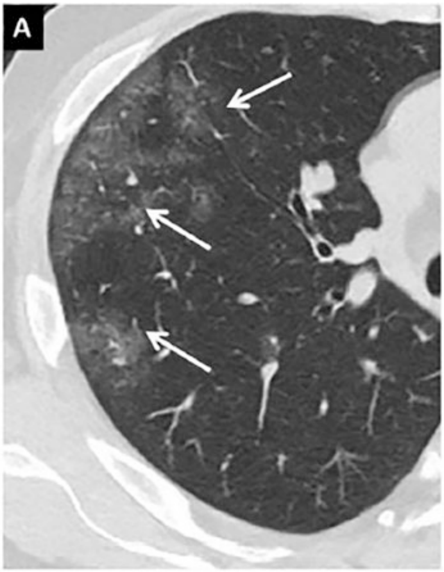



**Fig. 4** Presence of ground-glass opacities (indicated by arrows) in a patient with COVID-19 on a CT image

air spaces; however, when it is affected by COVID-19, it has been discovered that the most common pattern in confirmed cases is the appearance of ground-glass opacities due to lung involvement. These changes in the lung mean that the interaction of the X-rays that hit the lungs is not the same as in a healthy lung, resulting in ground-glass opacities on CT images, which are described as clear attenuations in internal regions of the lung as can be seen in Fig. 4, caused by partial filling of the air space, inflammation, and fibrosis of the lung tissue [21, 22].

## 4.2 Artificial Intelligence

Artificial intelligence (AI) is a term used to refer to the use of computation, technology, and algorithms to simulate intelligent behavior, generally comparable to that of humans. There are several applications of AI in our daily lives, such as assisted driving, facial recognition, voice recognition and it is even used in the commercial field with behavioral and purchase predictions. In the medical field, it is used to analyze complex data to be used in diagnoses, treatments and predictions in clinical cases. In medicine, it could be divided into two very general areas:

- Virtual: These are systems that require the use of electronic and computerized systems to assist in decision-making.
- Physical: These are intelligent devices such as surgical assistance robots or intelligent prostheses.

There are several applications in the area of healthcare such as health monitoring, medical data management, disease diagnosis, surgical treatment, drug development, etc. The use of these technologies offers some advantages such as improving accuracy, efficiency, and productivity, decreasing manual workload, saving money, and optimizing diagnostics by implementing AI [23].

## 4.3 Deep Learning

Deep learning is a subset of machine learning, which itself is a field of artificial intelligence. In regular programming, you generally have the inputs and set up the rules and logic necessary to arrive at a desired result; however, in machine learning, the inputs and the desired result are available, but the algorithm capable of reaching that result is not established; it is determined by the algorithm itself. One of the most widely used models in machine learning is artificial neural networks, which in general terms resemble the structure of the human neural system. Their basic structure can be seen in Fig. 5. These are capable of learning patterns based on observations of a dataset. The main elements of an artificial neural network are:

- Input layer: It is the connection that exists with the external data and information that the system will use to learn; this layer is connected to the hidden layer.
- Hidden layer: This layer is in charge of performing the calculations based on the information received from the input layer; there can be an indefinite number of hidden layers.
- Output layer: These are the conclusions drawn by the system from the calculations of the hidden layers; the number of outputs varies according to the application.

Each of the layers is made up of nodes which are called neurons, which are interconnected between layers, and each of them has a weight that transforms the

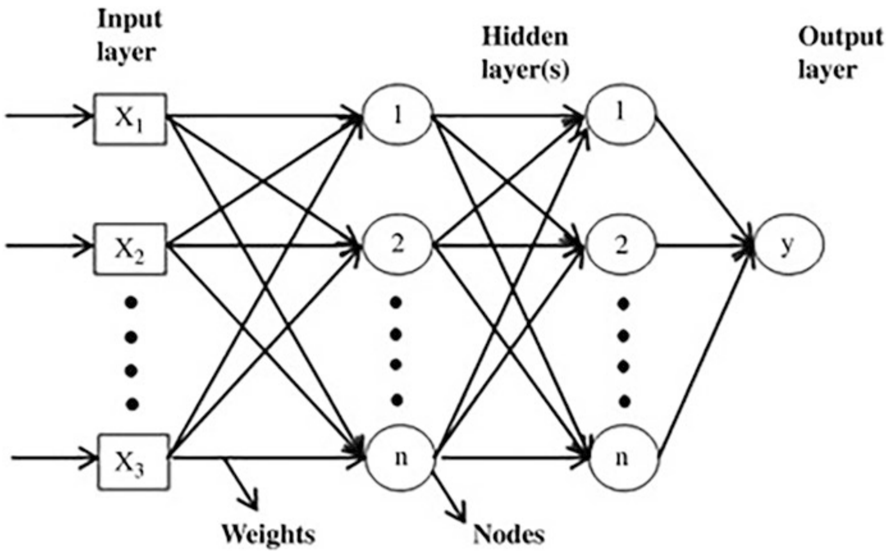



**Fig. 5** General structure of an artificial neural network

number value received from the previous layer and continues to the next layer. Initially, these values or weights will not obtain the best results; however, machine learning allows these neurons to adjust their weights according to the variation of the output until they reach or approach the desired result.

## *4.4 Convolutional Neural Networks*

Convolutional neural networks are based on the mathematical operation between matrices called convolution. This type of architecture has an application focused on detecting patterns in a large set of images and making patterns in a large set of images and making predictions in new images. The convolution operation requires a kernel, which is a matrix of smaller size than the image with specific values depending on the image with specific values according to the application you are looking for.

An important feature of convolutional neural networks is that convolution is performed sequentially and features are extracted at each convolution layer. In general terms, it can be said that the architecture of a convolutional neural network consists of five stages [24]:

1. Input: In this layer, the values of each pixel of the original image are received.
2. Convolution stage: The image is analyzed, and features are extracted by means of convolutions, and a set of features is obtained from each image.
3. Subsampling layer: This step is performed to gradually reduce the size of the data matrix while maintaining the most essential features for classification.
4. Activation function layer: It represents a limiting or threshold function, which modifies the result value or imposes a limit that must be exceeded in order to proceed to another neuron. There are many activation functions, and one of the most common is the rectified linear unit (ReLU).
5. Fully connected or connection layers: This section is similar to conventional neural networks, where all neurons are interconnected, through calculations the input image is classified.

## *4.5 CNN Architectures for Image Segmentation*

Some of the most commonly used architectures in image segmentation are [25]:

- U-Net: A convolutional neural network focused on medical image segmentation; its architecture has a U-shape as shown in Fig. 6. In general terms, its architecture is composed of two parts: The first part is the shrinkage, which consists of the convolution with $3\times3$ filters followed by a ReLU activation function and a $2\times2$ max pooling-type subsampling layer. The second part is the expansion, where the segmentation is built by means of convolutions and increases in image size [26].

- DenseNet: It is a convolutional neural network that connects each layer with all the others as shown in Fig. 7. These are faster to train and work better if there are short connections. Each layer takes the features generated by the previous layers and takes into account the current ones for the following layers [27].

- V-Net: It is an architecture similar to U-Net; it also consists of a compression path where the image resolution is reduced and an expansion path where the image is returned to the original dimensions. The difference with respect to the original U-Net architecture lies in the fact that the operations are applied on 3D image data instead of 2D; therefore, this model represents a framework for 3D medical image segmentation. The diagram of this architecture can be seen at [25].

The use of deep learning has proven to have numerous applications in image recognition, thanks to its structure that allows machine learning, extracting the main features in the training phase to subsequently identify them through neural networks in new images. A specific application is the segmentation of diseases through medical images, being these characteristics of a CT study; therefore, the aim is to develop an automatic segmentation of CT images to avoid manual analysis and speed up this procedure, allowing to improve the time in decision-making in the diagnosis, treatment, and follow-up of a patient with COVID-19.

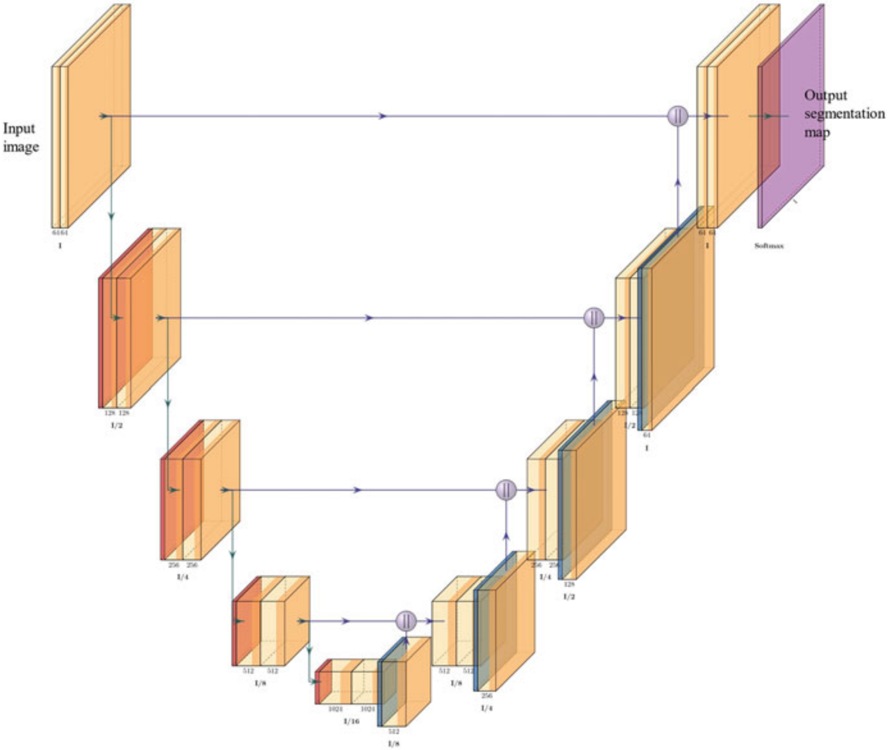



**Fig. 6** Architecture of the U-Net

# 5 Methodology

This section describes the experimental scheme for the development of the segmentation model of lung lesions caused by COVID-19, the model training strategy, and the evaluation of the model performance. It begins with a description of the database used.

## *5.1 Dataset Description*

The database was obtained from the "COVID-19 CT Lung and Infection Segmentation Dataset" repository [28]. The database is divided into four folders and separated by cases. The first folder contains the original images of the CT slices; the second, the infection masks; the third, the infection masks of the lungs as a whole; and the fourth, the lung masks. In this work, the folders with the original images of the CT

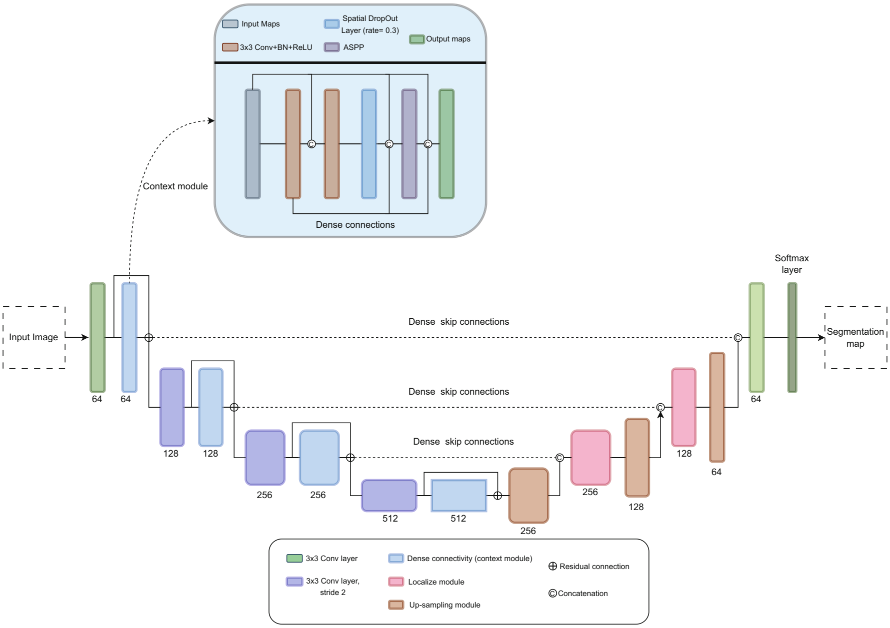



**Fig. 7** Example of DenseNet architecture for image segmentation

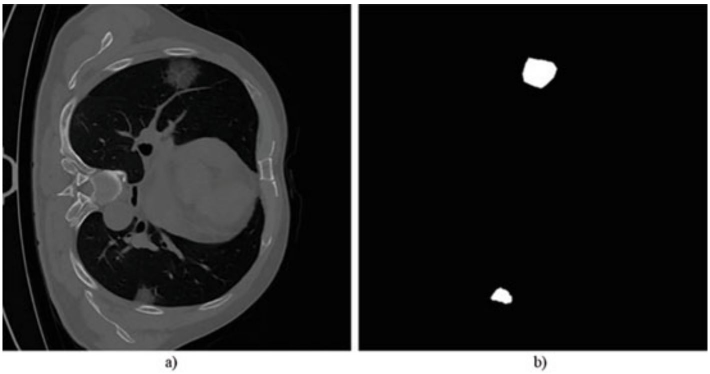



**Fig. 8** Example of database images. A CT slice is presented in (**a**), and its corresponding segmentation mask of the lung lesion is shown in (**b**)

slices and the infection masks made by the experts were used. An example of these images can be seen in Fig. 8.

The total number of CT images in the dataset is 2581, of which 1351 images showed infection and the rest (1230) did not show signs of lung injury. The size of the images is 512 × 512 pixels, and they were stored in NifTi format. The size of the computed tomography images and their segmentation masks were modified; in both cases, it was reduced by half, remaining at a size of 256 × 256 pixels. The resize function of the Cv2 library was used for this purpose.

The CT images with the presence of infection and their infection masks were divided into two sets, one for training and one for testing. Eighty percent of the images (1080 images) were used as training set, and the remaining 20% (271 images) were reserved for testing the model; this was done randomly.

## *5.2 Design of the Neural Network Architecture*

The U-Net convolutional neural network was selected because it has demonstrated good performance in segmentation tasks, more specifically in the segmentation of medical images [29]. The architecture of the U-Net neural network is divided into two phases, encoding and decoding. Figure 5 was used as a guide to establish the operations that the network will carry out and includes nine processing blocks:

- Block 1: In this block, the input image is received to be processed by 2 convolution stages with a kernel of size 3×3 and ReLU activation function, and 64 filters are used.
- Block 2: A max pooling layer is applied to the output of block 1 with a kernel of size 2×2, and then 2 convolutions are performed with a kernel of size 3×3 and ReLU activation function, and 128 filters are used.
- Block 3: A max pooling layer is applied to the output of block 2 with a kernel of size 2×2, and then 2 convolutions are performed with a kernel of size 3×3 and ReLU activation function, and 256 filters are used.
- Block 4: A max pooling layer is applied to the output of block 3 with a kernel of size 2×2, and then 2 convolutions are performed with a kernel of size 3×3 and ReLU activation function, and 512 filters are used.
- Block 5: A max pooling layer is applied to the output of block 4 with a kernel of size 2×2, and then 2 convolutions are performed with a kernel of size 3×3 and ReLU activation function, and 1024 filters are used. Subsequently, a transposed convolution is performed with a kernel of size 2×2.
- Block 6: The output of block 4 is concatenated with that of block 5, and 2 convolutions are performed with a kernel of size 3×3 and ReLU activation function, and 512 filters are used. A transposed convolution with a kernel of size 2×2 is applied to its output.
- Block 7: The output of block 3 is concatenated with that of block 6, and 2 convolutions are performed with a kernel of size 3×3 and ReLU activation function, and 256 filters are used. A transposed convolution with a kernel of size 2×2 is applied to its output.
- Block 8: The output of block 2 is concatenated with that of block 7, and 2 convolutions are performed with a kernel of size 3×3 and ReLU activation function, and 128 filters are used. A transposed convolution with a kernel of size 2×2 is applied to its output.

- Block 9: The output of block 1 is concatenated with that of block 8, and 2 convolutions are performed with a kernel of size 3×3 and ReLU activation function, and 64 filters are used. The output of this block delivers the segmentation of the image in two channels, the first one indicating the probability of each pixel to belong to the infection group and the other channel indicating the probability of each pixel to belong to the non-infection group.

### 5.3 *Strategies for Training the Model*

Experiments were carried out where four different training strategies of the segmentation model were tested in order to find the one with the best performance. All models were trained with the same images (training set), 200 epochs (*eps*), a 10% validation set (contained in the training set), a batch size of 16, and the Adam optimizer with a learning rate fixed to $10^{-3}$. However, some settings were modified seeking to optimize the model and compare it with its previous version.

- Model 1: The categorical cross-entropy loss function was used to adjust the network weights, and the accuracy metric was monitored during training. The model evaluated with the test data was the one obtained after training for the 200 epochs. For training this model, we use a callback for the automatic saving of the model when the value obtained by the loss function improves (a lower value of the loss function means that the model has improved). This callback is executed at the end of each epoch, where the value obtained from the loss function (using the validation data) is compared with the value obtained from the last saved model; if the current value is lower, it means that the model has improved and the new model is saved. The model evaluated with the test data is the last one saved and not necessarily the one obtained at the end of the 200 epochs.
- Model 2: A callback similar to the previous one was used to obtain this model, except that the *IoU* metric was compared with the validation data to save the model. If this metric increases, it means that the segmentation achieved by the model is better than the model saved at some previous epoch.
- Model 3: A custom loss function was implemented consisting of the average of the categorical cross-entropy loss calculation and the Sorensen-Dice coefficient (as a loss function). The callback implemented saved the model by monitoring the improvement of this custom loss function.
- Model 4: An adjustment of the learning rate of the optimizer was programmed. It was adjusted so that the learning rate (Lr) decreased exponentially every 20 epochs, according to Eq. 2. Initial Lr was set to $10^{-3}$. The custom loss function used in model 3 and the callback programmed in model 2 were used to train this model.

$$\mathrm{Lr} = \mathrm{InitialLr} \bullet e^{-\left\lfloor \left( {}^{\mathrm{epoch}}/_{20} \right) \right\rfloor} \tag{2}$$

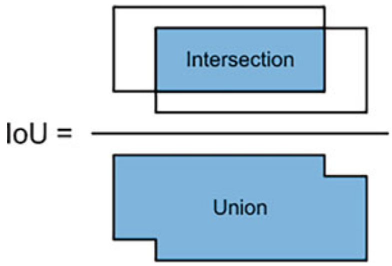



**Fig. 9** Representation of the Jaccard index calculation

### *5.4 Validation of the Model*

Model validation was performed by calculating two evaluation metrics: the Jaccard or intersection over union (*IoU*) index and the Sorensen-Dice similarity coefficient (DICE).

The Jaccard index (*IoU*) is used to compare the similarity between two volumes; it is widely used to evaluate programs focused on vision tasks, especially on segmentation tasks. This index is obtained with the area of intersection or overlap that exists between the proposed segmentation performed by a program and the expected one, taking into account the region of union between both segmentations, as shown in Fig. 9.

The Jaccard index can be calculated by means of Eq. 3.

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP}}, \tag{3}$$

where TP represents true positives and FN and FP represent false negatives and false positives, respectively.

The Sorensen-Dice similarity coefficient is used to quantify the similarity between two regions; it is also widely used to evaluate image segmentation programs. It represents the relationship between the true region (A) and the computationally segmented region (B), as shown in Fig. 10.

## 6 Results

In this section, the results obtained from the evaluation of the segmentation models using the test set images are discussed. Table 1 shows the average values of the metrics used to evaluate the performance of the five models implemented in the development of this project.

The learning curve with respect to the customized loss function obtained in model 4 is shown in Fig. 11, and the learning curve with respect to the *IoU* metric is shown in Fig. 12.

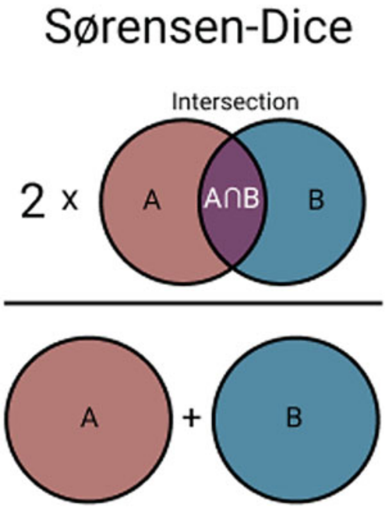



**Fig. 10** Representation of the Sorensen-Dice coefficient calculation

**Table 1** Average values of the DICE and Jaccard index of the evaluated models

| | IoU | DICE |
|---|---|---|
| Model 1 | 0.7646 | 0.8491 |
| Model 2 | 0.7721 | 0.8706 |
| Model 3 | 0.7787 | 0.8817 |
| Model 4 | **0.7899** | **0.8929** |

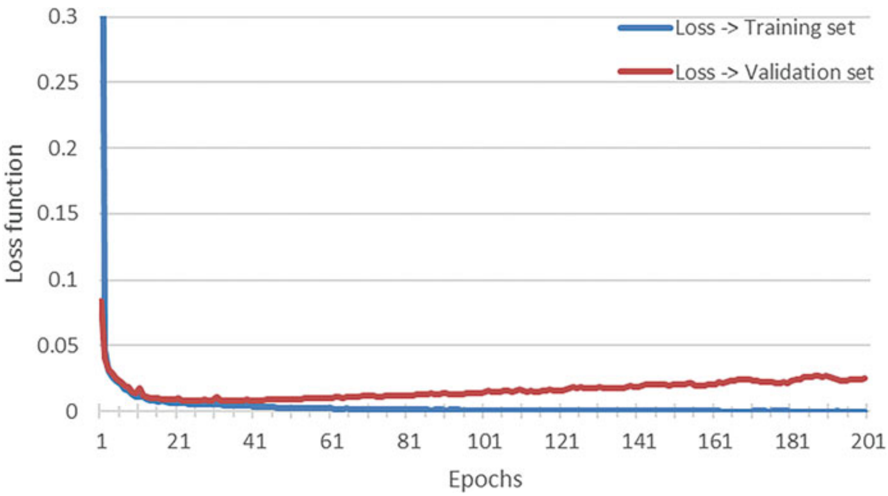



**Fig. 11** Learning curve of model 4 with respect to the custom loss function

As can be seen in the learning curves, taking into account the data from the training set, the model reaches its maximum performance (IoU = 1) at epoch 160. However, with the data from the validation set (10% of the training data), the model

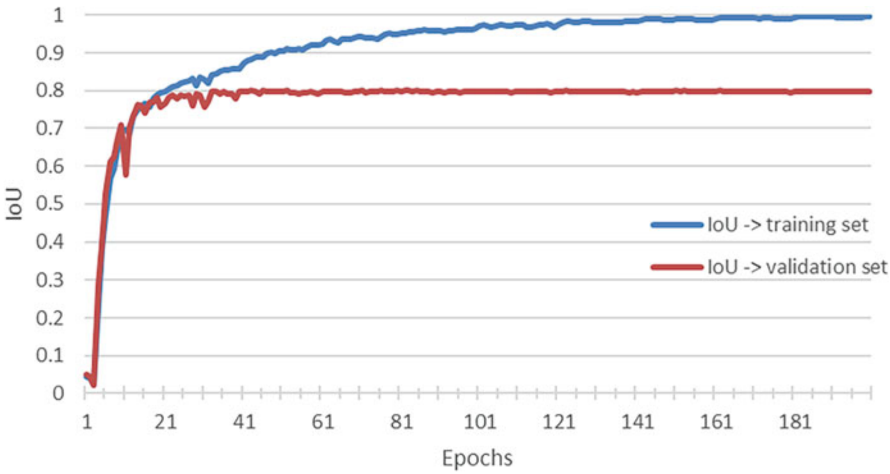



**Fig. 12** Learning curve of model 4 with respect to the *IoU* metric

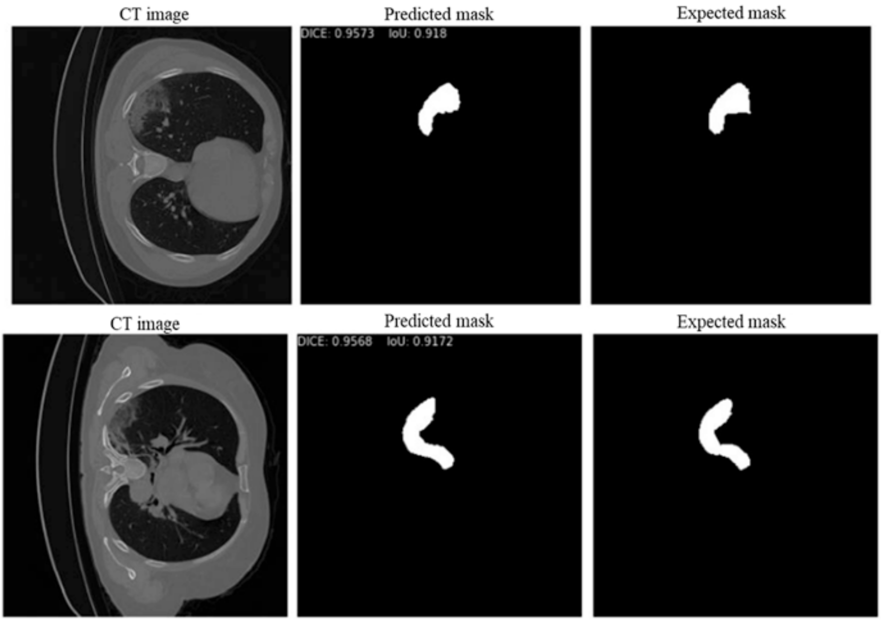



**Fig. 13** Examples with high values in Jaccard and DICE indices between the mask made by model 4 and the mask made by the experts

stops improving from epoch 41, when the model is saved and later evaluated with the test data.

Figure 13 shows two examples with high performance in the segmentation performed by model 4.

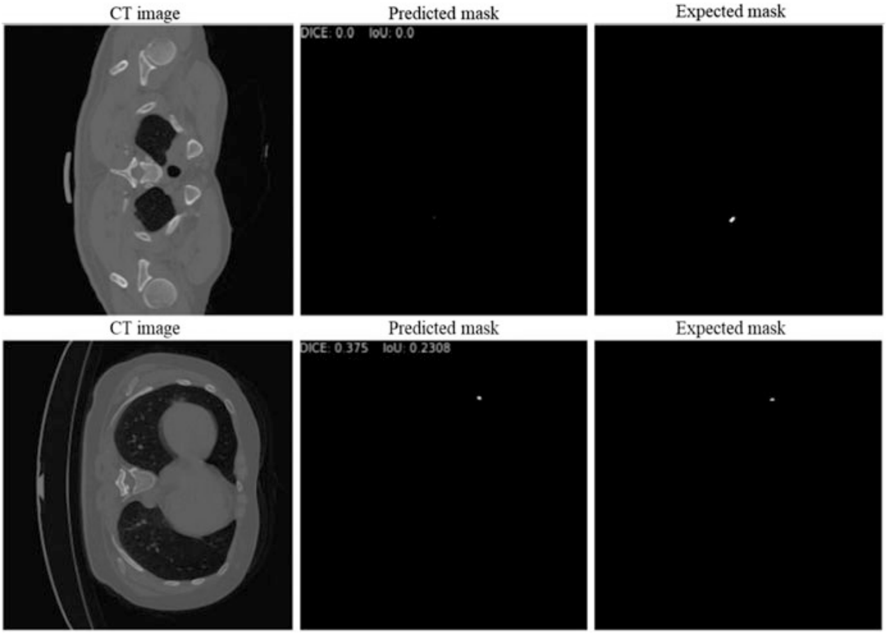



**Fig. 14** Two examples with low values of the Jaccard and DICE indices between the mask made by model 4 and the one made by the experts

Figure 14 shows two examples with a low value of the Jaccard and DICE indices. Both examples coincide in that the lesion is very small compared to the lesions present in the other CT slices, so the model has difficulties in identifying small lesions within a CT image.

Figure 15 shows two examples where the following are distinguished by colors: true positives, false positives, and false negatives. The predominant color in Fig. 15 is light blue, which means that the model built satisfactorily identifies lesions where there are lung lesions. However, there are false negatives (yellow color) where the patient has a lesion, but the model predicts that there is no lesion, as well as a few pixels marked as false positives (pink color) which means that the patient does not have a lesion, but the model predicts that there is a lesion.

In order to compare the outcomes of the model suggested in this work with those of other works that have used the same database, a search was made for works that dealt with the issue addressed in this project. This comparison can be seen in Table 2.

As can be seen, our suggested model obtains better results than those found in previous studies. Although other models have used a similar architecture to the one suggested in this study, it is crucial to note that the training method based on a customized loss function, composed of the categorical cross-entropy and the DICE index, produces a model considerably superior to the other approaches.

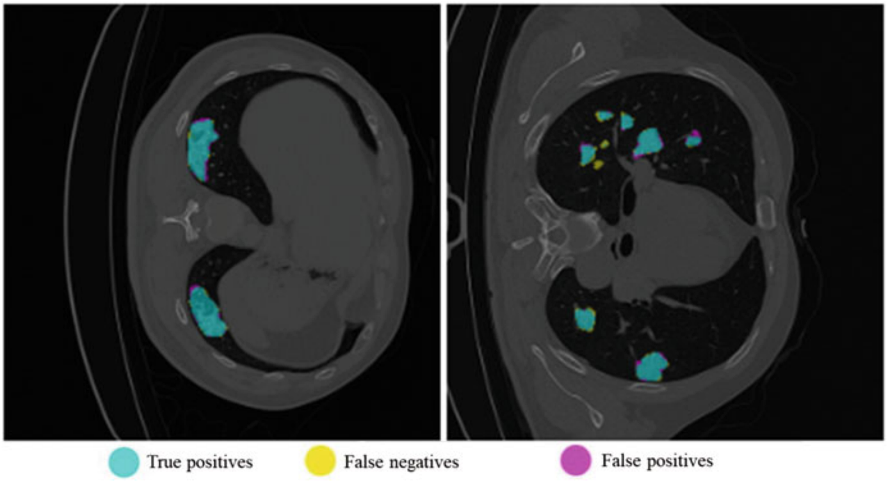



**Fig. 15** Two visual examples of the prediction proposed by the constructed model (model 4), compared to the expected prediction

**Table 2** Comparison with other related models using the same database

| Authors | Network | Metric (mean) | |
|---|---|---|---|
| | | DICE | Jaccard |
| Saood and Hatem [16] | U-Net | 0.733 | – |
| | SegNet | 0.749 | – |
| Khalifa et al. [30] | Encoder-decoder | – | 0.7807 |
| Yin et al. [31] | U-Net | 0.8544 | 0.7508 |
| | DeepLab v3+ | 0.8571 | 0.7546 |
| | U-Net++ | 0.8602 | 0.7592 |
| | DS-U-Net | 0.8696 | 0.7702 |
| Our model | U-Net (custom loss function) | **0.8929** | **0.7899** |

The development of a smart healthcare system that connects directly with a tomography equipment for the acquisition of patient tomographic studies and incorporates automatic image segmentation using the developed model is the next step in this research. The segmented image can then be displayed in a user-friendly manner for the radiologist, allowing for more accurate diagnoses. Additionally, it is proposed that this system be integrated into a smart city framework that includes the patient's CT image history. This integration can be used to assess the progression of the patient's disease and allow remote consultation with the treating physician. With the use of this smart healthcare technology, patients can receive more efficient and effective medical care while also allowing for the data to be analyzed and used to improve the overall health of the population within the smart city.

## 7 Conclusions

An automatic lung lesion segmentation system would represent a valuable tool in healthcare that can be implemented in smart hospitals as decision support systems that would be used to assist and improve the diagnosis of patients with COVID-19 and thus select the best treatment scheme for it. In the present work, a deep learning model based on U-Net architecture was built for the segmentation of lung lesions due to COVID-19 in computed tomography images. Different convolutional neural network learning methods were evaluated with the aim of obtaining remarkable performance. The training technique enabled the development of a model that performed on average at 0.892 and 0.789 for the assessment metrics DICE score and Jaccard index, respectively. This kind of model can be used to support the diagnosis and tracking of the COVID-19 evolution in the health services of a smart city. It is a tool that can provide an effective result with reliable performance, which can help the work of the physician in the analysis of computed tomography studies of patients with this disease.

As future work, it is proposed to evaluate other loss functions that allow the identification of small areas of interest, since in the present work, most of the false negatives occurred in images with small lesions. It is also proposed to test architectures other than U-Net.

## References


1. Zeadally, S., Siddiqui, F., Baig, Z., & Ibrahim, A. (2020). Smart healthcare: Challenges and potential solutions using internet of things (IoT) and big data analytics. *PSU Research Review, 4*(2), 149–168.
2. Tian, S., Yang, W., Le Grange, J. M., Wang, P., Huang, W., & Ye, Z. (2019). Smart healthcare: Making medical care more intelligent. *Global Health Journal, 3*(3), 62–65.
3. Pramanik, M. I., Lau, R. Y. K., Demirkan, H., & Azad, M. A. K. (2017). Smart health: Big data enabled health paradigm within smart cities. *Expert Systems with Applications, 87*, 370–383.
4. Nasr, M., Islam, M. M., Shehata, S., Karray, F., & Quintana, Y. (2021). Smart healthcare in the age of AI: Recent advances, challenges, and future prospects. *IEEE Access, 9*, 145248–145270.
5. Swayamsiddha, S., Prashant, K., Shaw, D., & Mohanty, C. (2021). The prospective of Artificial Intelligence in COVID-19 Pandemic. *Health Technology (Berl), 11*(6), 1311–1320.
6. Chérrez-Ojeda, I., Gochicoa-Rangel, L., Salles-Rojas, A., & Mautong, H. (2021, February). Seguimiento de los pacientes después de neumonía por COVID-19. Secuelas pulmonares. *Revista Alergia México, 67*(4 SE-Artículos de revisión).
7. George, P. M., Wells, A. U., & Jenkins, R. G. (2020). Pulmonary fibrosis and COVID-19: the potential role for antifibrotic therapy. *The Lancet Respiratory Medicine, 8*(8), 807–815.
8. Spagnolo, P., et al. (2020). Pulmonary fibrosis secondary to COVID-19: a call to arms? *The Lancet Respiratory Medicine, 8*(8), 750–752.
9. Uysal, E., et al. (2021). Chest CT findings in RT-PCR positive asymptomatic COVID-19 patients. *Clinical Imaging, 77*, 37–42.
10. Abou Ghayda, R., et al. (2021). Chest CT abnormalities in COVID-19: a systematic review. *International Journal of Medical Sciences, 18*(15), 3395.

11. Pham, D. L., Xu, C., & Prince, J. L. (2000). Current Methods in Medical Image Segmentation. *Annual Review of Biomedical Engineering, 2*(1), 315–337.
12. Dey, N., Rajinikanth, V., Fong, S. J., Kaiser, M. S., & Mahmud, M. (2020). Social group optimization–assisted Kapur's entropy and morphological segmentation for automated detection of COVID-19 infection from computed tomography images. *Cognitive Computation, 12*(5), 1011–1023.
13. Nurmaini, S., et al. (2020). Automated detection of COVID-19 infected lesion on computed tomography images using Faster-RCNNs. *Engineering Letters, 28*(4), 1295–1301.
14. Xu, X., et al. (Oct. 2020). A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering (Beijing, China), 6*(10), 1122–1129.
15. Suri, J. S., et al. (2021). Inter-variability study of COVLIAS 1.0: Hybrid deep learning models for COVID-19 lung segmentation in computed tomography. *Diagnostics, 11*(11), 2025.
16. Saood, A., & Hatem, I. (2021). COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet. *BMC Medical Imaging, 21*(1), 1–10.
17. Zhao, C., et al. (2021). Lung segmentation and automatic detection of COVID-19 using radiomic features from chest CT images. *Pattern Recognition, 119*, 108071.
18. C. N. de Excelencia Tecnológica en Salud. (2004). *Guía Tecnológica No. 6: Tomografía Computarizada*. Secretaría de Salud, Mexico, pp. 1–29.
19. Bourne, R. (2010). *Fundamentals of digital imaging in medicine*. Springer Science & Business Media.
20. Seeram, E. (2015). *Computed tomography-e-book: Physical principles, clinical applications, and quality control*. Elsevier Health Sciences.
21. Mehrabi, S., et al. (2020). Pitfalls of computed tomography in the coronavirus 2019 (COVID-19) era: A new perspective on ground-glass opacities. *Cureus, 12*(5).
22. Álvarez, J., et al. (2020). Hallazgos de imagen en Covid-19. Complicaciones y enfermedades simuladoras. *Revista chilena de radiologica, 26*(4), 145–162.
23. Malik, P., Pathania, M., & Rathaur, V. K. (2019). Overview of artificial intelligence in medicine. *Journal of Family Medicine and Primary Care, 8*(7), 2328.
24. O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks, *arXiv Prepr. arXiv1511.08458*.
25. Sengupta, D. (2019). *Deep learning architectures for automated image segmentation*. University of California.
26. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234–241).
27. Qamar, S., Ahmad, P., & Shen, L. (2011). Dense encoder-decoder–based architecture for skin lesion segmentation. *Cognitive Computation, 13*(2), 583–594.
28. Yang, X., He, X., Zhao, J., Zhang, Y., Zhang, S., & Xie, P. (2020). *COVID-CT-Dataset: A CT scan dataset about COVID-19*, pp. 1–11.
29. Yin, X.-X., Sun, L., Fu, Y., Lu, R., & Zhang, Y. (2022). U-Net-Based medical image segmentation. *Journal of Healthcare Engineering, 2022*, 2022.
30. Khalifa, N. E., Manogaran, G., Taha, M. H. N., & Loey, M. (2022). A deep learning semantic segmentation architecture for COVID-19 lesions discovery in limited chest CT datasets. *Expert Systems, 39*(6), e12742.
31. Yin, S., Deng, H., Xu, Z., Zhu, Q., & Cheng, J. (2022). SD-UNet: A novel segmentation framework for CT images of lung infections. *Electronics, 11*(1), 130.

# Index





N. Gupta, S. Mishra (eds.), *Internet of Everything for Smart City and Smart Healthcare Applications*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-34601-9