IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 June 2023, accepted 27 July 2023, date of publication 4 August 2023, date of current version 30 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3302262

TOPICAL REVIEW

Comprehensive Review on Development of Smart Cities Using Industry 4.0 Technologies

MARIEH TALEBKHAH[1], (Member, IEEE), ADUWATI SALI[1], (Senior Member, IEEE), MEISAM GORDAN[2], SHAIFUL JAHARI HASHIM[3], (Member, IEEE), AND FAKHRUL ZAMAN ROKHANI[3], (Member, IEEE)

[1]Wireless and Photonics Networks Research Centre of Excellence (WiPNET), Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Seri Kembangan, Selangor 43400, Malaysia
[2]School of Civil Engineering, University College Dublin (UCD), Belfield, D04 V1W8 Ireland
[3]Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia (UPM), Seri Kembangan, Selangor 43400, Malaysia

Corresponding authors: Marieh Talebkhah (marieh.t2018@gmail.com) and Aduwati Sali (aduwati@upm.edu.my)

This work was supported in part by the Projects EMOSEN-Energy Efficient MIMO-Based Wireless Transmission for SWIPT-Enabled Network (UPM/800-3/3/1/GPB/2019/9671600, Vote No: 9671600); in part by the Geran Putra Berimpak BIDANET: Parametric Big Data Analytics Over Wireless Networks under Grant UPM.RMC.800-3/3/1/GPB/2021/9696300 (Vote No.: 9696300); and in part by IGNITE-Interference Modeling for 5G and FSS Coexistence at mmWave With Climate Change Considerations in the Tropical Region under Grant FRGS/1/2021/TK0/UPM/01/1, Grant 03-01-21-2375FR, and Grant 5540509.

ABSTRACT Smart Cities (SCs) have recently opened new lifestyles as they introduce effective approaches for improving urban management. These cities integrate Industry 4.0 technologies and new organizations into a social and technical system. A Smart City (SC) finally aims to integrate urban business, transportation, water, energy, and other subsystems by analyzing the data collected from sensors and Information and Communication Technology (ICT). In this regard, the current article proposes a hybrid structure based on Cloud Computing (CC), cloud processing, and internet technology to develop and manage SCs and urban planning. Moreover, the development and analysis of sensors to construct an SC, as well as supporting technologies, are addressed. Therefore, the present research is focused on defining and reviewing large data and objects of the Internet and CC to identify the current challenges and limitations. This research, thus, introduced a novel approach in the context of the Internet of objects, whose data are gathered from different geographic locations through devices and sensors to be used in a new system. CC services in a city can analyze and make smart decisions to optimally manage SCs and improve the welfare of citizens. This study consequently sets the path for further investigation into the problems and difficulties associated with Big Data (BD) applications in SCs.

INDEX TERMS Smart Cities (SCs), Big Data (BD), Internet of Things (IoT), cloud computing (CC), architecture and implementation of Smart Cities, Industry 4.0.

NOMENCLATURE

5G	Fifth-Generation Wireless Communication.
5V's	Velocity, Volume, Variety, Value, and Veracity.
ACID	Atomicity, Consistency, Isolation, and Durability.
ACLs	Access Control Lists.
AI	Artificial Intelligence.
AMQP	Advanced Message Queuing Protocol.
API	Application Programming Interface.
AR	Augmented Reality.
AWS	Amazon Web Services.
BD	Big Data.
BDA	Big Data Analytics.
BDM	Big Data Management.
BIM	Building Information Modelling.
CC	Cloud Computing.
CDR	Call Detail Records.

The associate editor coordinating the review of this manuscript and approving it for publication was Agostino Forestiero.

CoAP	Constrained Application Protocol.
CPSS	Cyber-Physical-Social System.
CPU	Central Processing Unit.
CSP	Cloud Service Provider.
CSV	Comma Separated Values.
DBMS	Database Management Systems.
DM	Data Mining.
DL	Deep Learning.
EC	Edge Computing.
EC2	Elastic Compute Cloud.
FC	Fog Computing.
FTP	File Transfer Protocol.
ETL	Extract, Transform, and Load.
GHG	Greenhouse Gas.
GPS	Global Positioning System.
HDFS	Hadoop Distributed File System.
HTTP	Hyper Text Transfer Protocol.
IaaS	Infrastructure-as-a-Service.
ICT	Information and Communication Technology.
IoT	Internet of Things.
IR1.0	First Industrial Revolution.
IR2.0	Second Industrial Revolution.
IR3.0	Third Industrial Revolution.
IR4.0	Fourth Industrial Revolution.
IT	Information Technologies.
LD	Linked Data.
LDA	Linear Discriminant Analysis.
LoRaWAN	Long-Range Wide-Area Network.
LPWAN	Low-Power Wide-Area Network.
M2M	Machine-to-Machine.
ML	Machine Learning.
MQTT	Message Queue Telemetry Transport.
MR	Mixed Reality.
NFC	Near-Field Communication.
NLP	Natural Language Processing.
NoSQL	Not only Structured Query Languages.
OSI	Open Systems Interconnection.
OWL	Web Ontology Language.
PaaS	Platform-as-a-Service.
PANs	Personal Area Networks.
PCA	Principal Component Analysis.
Pcap	Packet Capture.
PLC	Programmable Logic Controller.
PLM	Product – Lifecycle – Management.
PoI	Points of Interest.
QoL	Quality of Life.
QoS	Quality of Services.
REST	Representational State Transfer.
RDBMS	Relational Databases Systems.
RDD	Resilient Distributed Data.
RDF	Resource Description Framework.
RFID	Radio Frequency Identification.
SaaS	Software as-a-Service.
SC	Smart City.
SCDAD	Smart City Data Analytics Dashboard.
SCDA	Smart City Data Analytics.
SCs	Smart Cities.
SMN	Social Media Network.
SQL	Structured Query Language.
UAVs	Unmanned Aerial Vehicles.
VA	Visual Analytics.
VPL	Visual Programming Language.
VR	Virtual Reality.
WEEE	Waste Electrical and Electronic Equipment.
WLAN	Wireless Local Area Networks.
WS	WebSocket.
WSN	Wireless Sensors Networks.
XML	Extensible Markup Language.

I. INTRODUCTION

The rapid transition from conventional desktop computing to sophisticated computing [1] and the significant enhancement in connected devices and sensors have further facilitated living in an SC. Several smart environments have been recently developed, among which, smart homes [2], [3], grids [4], transportation [5], healthcare [6], and cities [2], [7] can be mentioned.

BD applications [8] are implemented by governments [9] to enable SCs to achieve the essential sustainability and standards. SCs exploit diverse technologies to enhance health [10], transportation [11], energy [12], education [13], and water services. These goals involve cost and resource consumption reduction along with more effective and active engagement of the citizens. Big Data Analytics (BDA) has exhibited the promising potential to enhance SC services. Such platforms can be provided by relying on CC to support BD management and applications in SCs. Fig. 1 illustrates how CC supports BD collection, storage, and analysis across cloud nodes and facilities [8].

Recently emerged ICT paradigms including data-intensive computing (Big Data), Open Data [14], [15], Large-scale Distributed Systems [16], IoT [17], Physical-Cyber-Social Computing [5], [18], Service-Oriented and CC [19], are essential elements of the SCs. These paradigms work in concert to enable the creation of applications and information systems for SCs using a variety of architectures (centralized, decentralized, and their hybrids) and infrastructures (such as middleware and IoT platforms) [8], [9], [10], [11], [12].

Innovative services and applications are crucial to the realization of the SC concept. Applications in this discipline range from those pertaining to city operations (such as real-time traffic control) to entertainment and tourism (i.e., augmented reality). These applications should be able to fulfill a variety of city criteria, such sustainable development, gradual change, collaboration among a variety of city stakeholders, and privacy considerations when dealing with citizen data. An SC offers a platform to transfer a huge deal of data among the involved actors through complicated supply chains [20]. The leverage of these data flows facilitates the development of SC applications considering economic, environmental, energy, water, waste, public

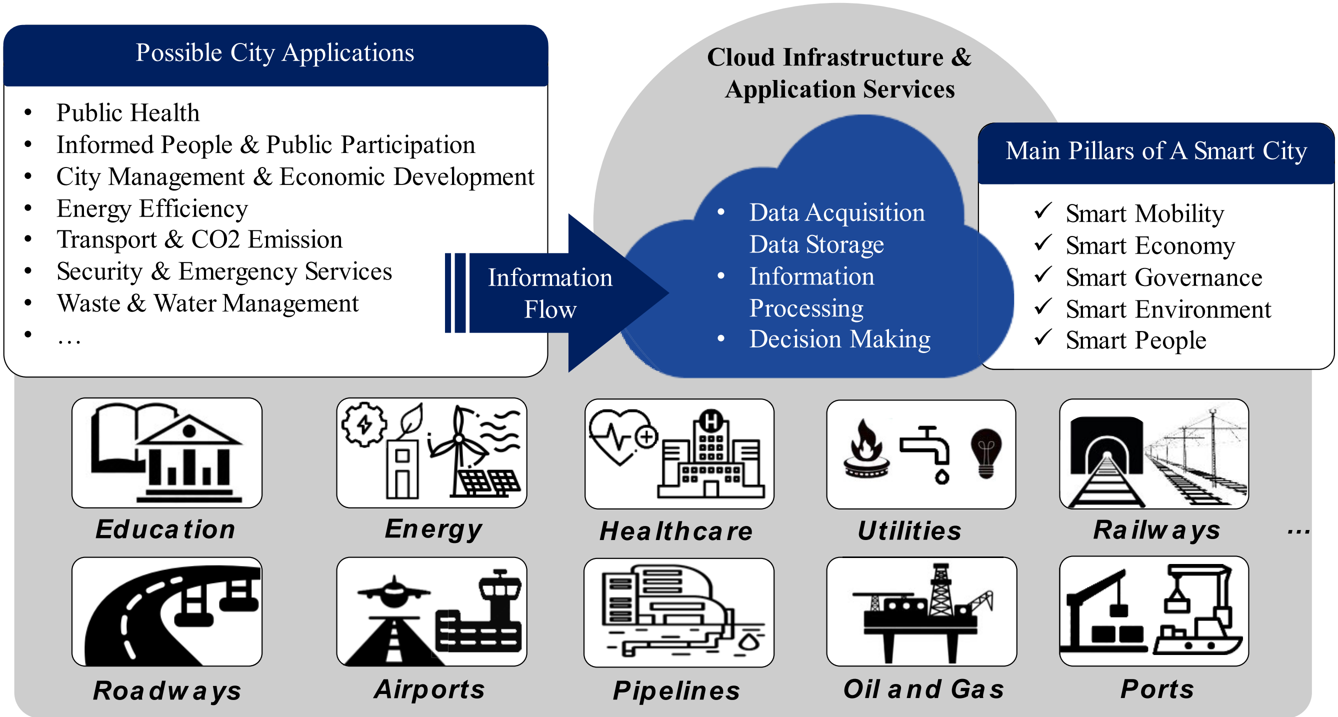

FIGURE 1. Application of a cloud to store data generated from different components of a city.

(intellectual endowment, and engagement), lifestyle, building, transportation, and public space aspects [1], [14].

SCs involve IoT applications. The concept of SCs has been largely accepted although no official definition can be found for an SC. An SC is aimed to enhance the use of public resources and Quality of Services (QoS) at reduced costs. Cities play a decisive role in the economy and society and can greatly affect the environment [21], [22]. Cities seek for efficient solutions to resolve transportation and land use problems to improve the quality of public services while offering economic benefits. For instance, an efficient and high-quality transportation system capable of responding to economic needs plays a key role in a city. Numerous novel approaches for municipal services are based on the latest technologies (i.e. information and communication technologies).

The implementation of the idea of the SC exponentially enhances the magnitude of involved data. Thus, such enormous data (i.e. big data) are at the heart of the IoT-rendered services [23]. The phenomenon of BD is known by volume, velocity, and a variety of data types created at ever-increasing rates [24]. BD presents promising opportunities to attain insights from a huge deal of data collected from diverse sources. Such data mainly encompass non-structured features [25]. Fig. 2 shows how BD and CC are combined with smart technologies. Various smart applications communicate with one another via embedded sensors and other devices that are integrated with CC infrastructures, which results in the production of a lot of unstructured data.

A. MOTIVATION AND CONTRIBUTION

The advancement of technology is one of the main factors to make daily life easier and more innovative. SCs use ICT and IoT to boost operational effectiveness, enhance the caliber of public services and citizen welfare, develop, implement, and promote sustainable development methods, and respond to the ever-expanding needs of their residents. Better management of available resources is possible by advances in information and communication. Additionally, research on upcoming technology aims to enhance the effectiveness of existing solutions. Wherever it was used, the digital revolution has proven to be essential for success—especially when considering how simple and effective it is to use. In contrast, the most important interacting aspect in any subject can be making an informed decision based on evidence and facts. The SC arena is the same. As a result, improving future technology to improve human lives becomes a primary research objective. A new study can be initiated by imagining the future features of SCs through the current circumstances. The interaction feature highlighted in this paper aids in the development of a whole new framework for SCs.

It is necessary to compile the works in these domains and build a uniform repository as the idea of SCs is evolving with the development of complementing technologies. This essay thus aims to provide a comprehensive examination and analysis of potential SC technologies. The following is a summary of our contributions:

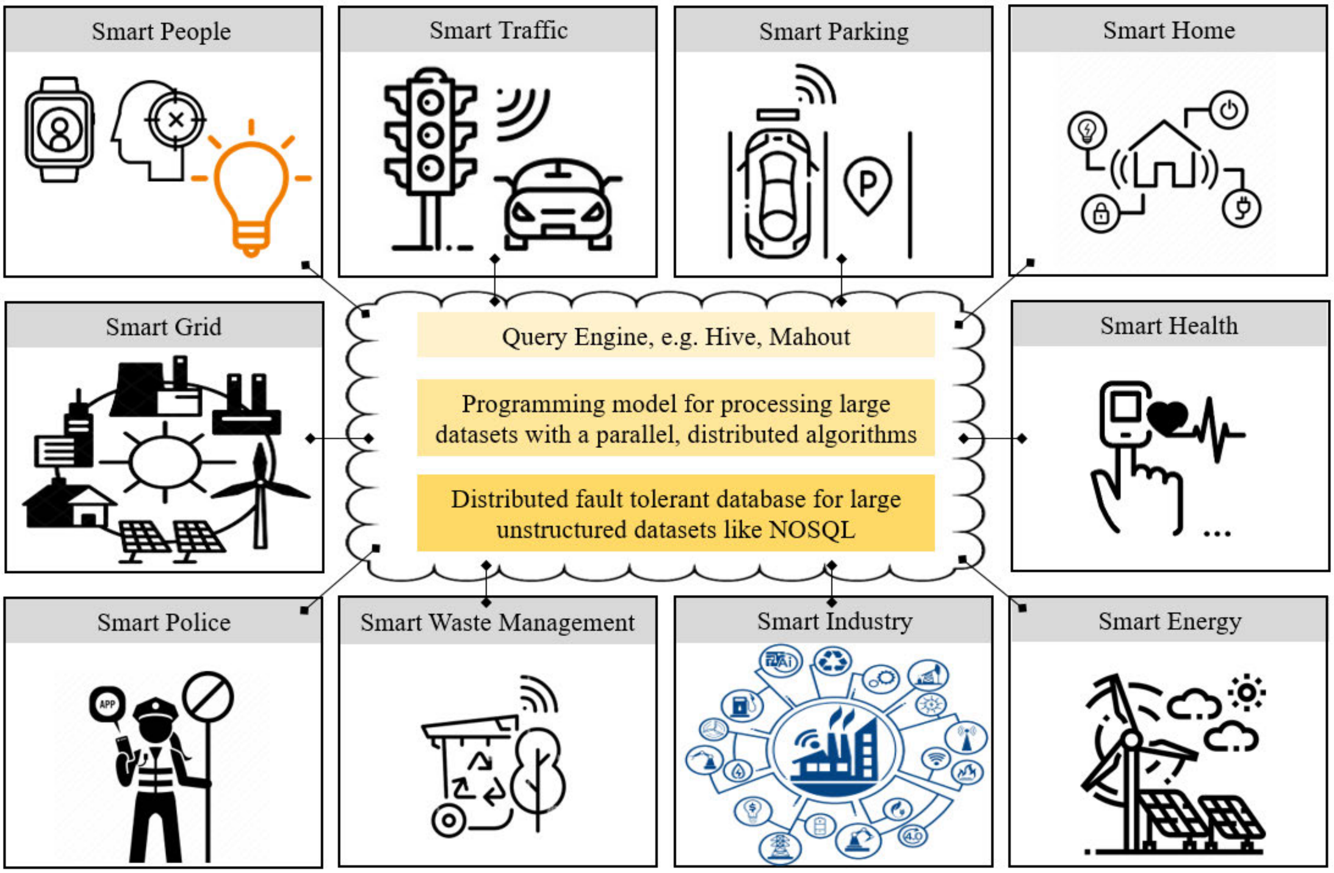

FIGURE 2. An overview of the SC and BD technology.

- Overview of the potential of future SCs to address numerous issues through intelligent solutions to improve the daily life.
- A survey of systematic SC-suitable BD tools, focusing on their effectiveness in data collection, preprocessing, and analysis.
- The main contribution of this thorough research is the identification and examination of recent technological developments, such as BD, CC, and IoT, that form the foundation of a new era.
- This research presents an innovative BDA framework for SCs that enhances data model management and aggregation. The framework introduces new functions that improve BDA capabilities.
- Next, we discuss technical challenges identified in the literature and propose a conceptual framework that serves as a roadmap for developing an integrated, cloud-based architecture for user services. However, realizing the full potential of this technology requires addressing various issues and obstacles related to CC.
- Finally, the future directions for developing SCs are presented.

The methods used in this paper are displayed in Fig. 3.

In continue, Section II highlights the key fundamental background of SC, BD, CC, and IoT by elaborating on ''Industry 4.0''. Section III defines the SC along with its applications and limitations. The concept of BD, its types, applications, and tools are addressed in Section IV. In addition, an intelligence framework for analysis of BD in SCs and its challenges are proposed in this section. Section V summarizes the classification and barriers of CC strategies in SCs. Likewise, Section VI presents the challenges of IoT in SCs. On top of that, the role of IoT in CC as well as BD is also discussed in Section VI. Section VII presents the architecture and implementation of SCs through BD, CC, and IoT. Section VIII reviews the findings of this paper from different perspectives. Then, a new framework analytics is proposed for SC data in Section IX. Future works on SC management are suggested in Section X. Finally, the important conclusions are drawn in Section XI.

II. RECENT ADVANCES IN IoT, CLOUD COMPUTING, BIG DATA AND ANALYTICS

The term ''industry'' refers to the creation of products, services, and facilities within an economy. Our world has experienced four steps of industrialization whose most important

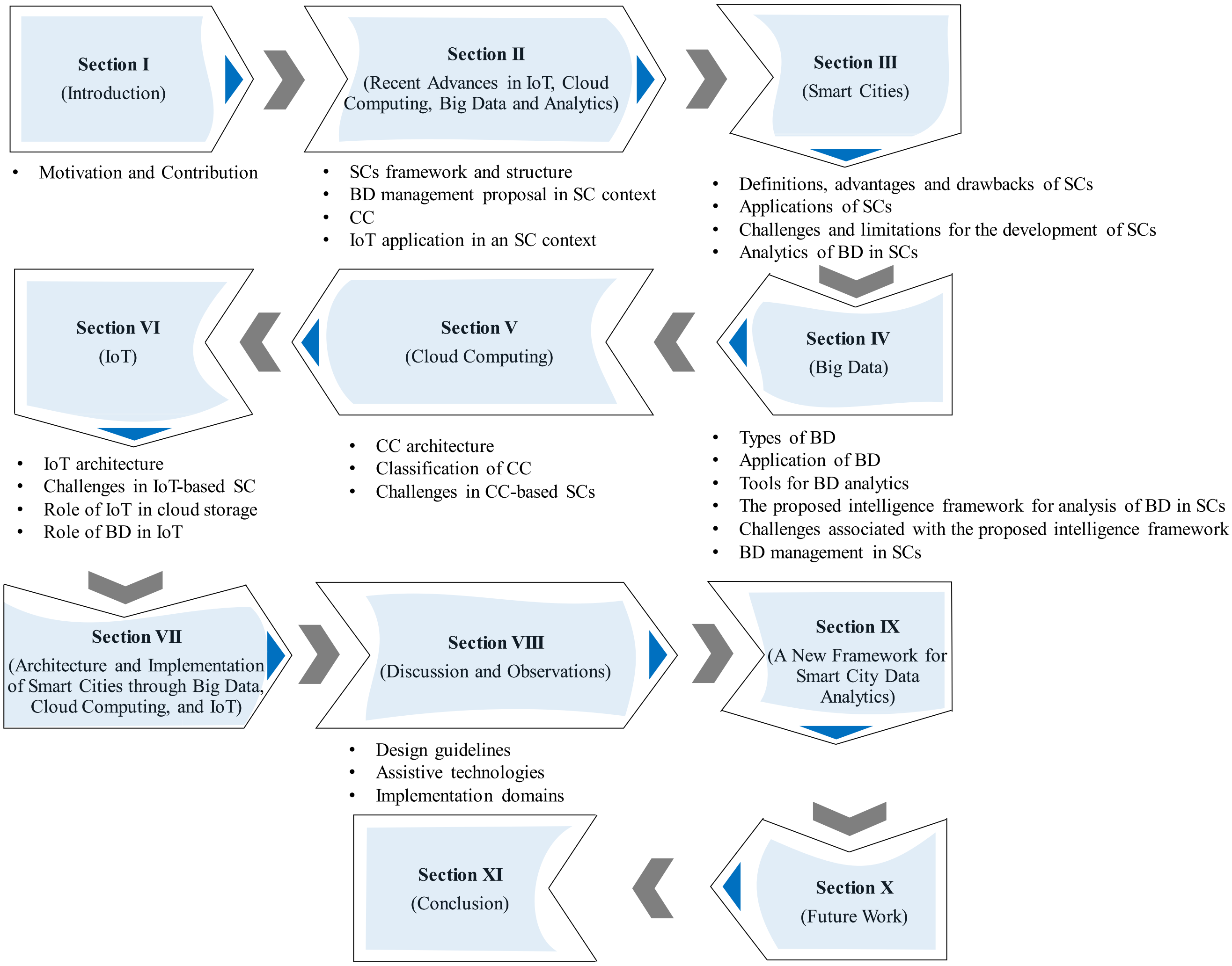

FIGURE 3. Outline of the survey.

contributions are listed in Table 1 [26], [27], [28], [29], [30], [31], [32]. As can be seen, IoT, BD, Artificial Intelligence (AI), CC, and SC are several key contributors of Industry 4.0.

An SC can be defined as the convergence of emerging IoT, BD, and AI techniques [33]. It offers a new paradigm in ICT to provide citizens with versatile infrastructures to easily access services. This concept also enables the governing bodies to manage and control the urban resources. SCs employ ICT for sensing, analyzing, and integrating the information of the cities [34], [35]. The concept of the SC has attracted the attention of local governments due to the population growth [36] and expansion of the boundaries of cities. This concept can be regarded as a versatile approach in the transformation of the traditional cities to boost the economy, technologies, and sustainability. The growth of IoT and BD are key factors in the creation of SCs. Large and complicated data sets are referred to as "BD", and they cannot be processed or managed using conventional data management technologies.

In the case of large volumes of data, local or enterprise storage and processing are not efficient. To find the right choice, consider the technologies for BD processing, IoT, and CC; then, explore the issues of these processes in the management of SCs. A comprehensive research paper was conducted considering journal articles and conference papers. Table 2 lists the studies in this area.

IoT provides a framework for seamless communication between sensors and actuators, enabling easier sharing of information across platforms within the context of SCs. With adoption of various wireless technologies, IoT's are positioned as the next breakthrough technology that fully leverages the internet-offered prospects. IoT has been used in the development of SC infrastructure, including smart grids [37], [38], smart energy [39], smart homes [40], [41], smart retail [42], smart water [43], [44], smart transportation [45], [46], and smart healthcare [47]. By connecting people on a scale and pace previously unheard of, the internet has revolutionized the communication. The development of

TABLE 1. Global industrial revolutions and their contributions.

Industrialization Steps	Key Contributions
Fourth Industrial Revolution (IR 4.0)	• Internet of Things (IoT) • Smart factories / Smart manufacturing / Robotics • Circular economy / Product – Lifecycle – Management (PLM) • Data Mining (DM) / Big Data Analytics (BDA) / Deep Learning (DL) • Artificial Intelligence (AI) / Machine learning (ML) • Smart sensors / Remote sensing / Wireless sensor network / Online monitoring • Cloud Computing (CC) / Cognitive computing / Mobile computing • Cybersecurity / Blockchain • Digital twin / Smart tasks and diagnostics / Smartification • Virtual Reality / Augmented reality (AR) / Building Information Modelling (BIM) • Unmanned Aerial Vehicles (UAVs) / Internet of Drone / Smart Cities (SCs) • Smart environment / Sustainable development / Renewable energy
Third Industrial Revolution (IR 3.0)	• Production Automation / Computer and Automation • Information and technology / Telecommunication • Linear economy • Leveraging electrical mechanization • First Programmable Logic Controller (PLC) • Industrial robotics / Electronic and nuclear industries • Business computers / Supercomputers / Business software • Internet / World wide web
Second Industrial Revolution (IR 2.0)	• Electrical energy / Steam power and petroleum • Skyscrapers • Large scale iron and steel production • Telephones and telegraphs, typewriter, phonograph, motion pictures • Widespread use of machinery in manufacturing • Automobile, Airplanes, Diesel engines, Bicycles, Railroads • Chemical, Rubber, Paper mills, Fertilizers • Applied science • New forms of business organizations
First Industrial Revolution (IR 1.0)	• Mechanical production • New energy resources / Water and steam power • New raw materials / Iron, Coal, Textile, Steam industries • New machines / Spinning jenny and the power loom • Factory system, Division of labor, Specialization • The locomotive • Expansion of world trade

an SC through object connectivity will be the next revolution, emphasizing the interconnection of sensor and actuator devices to enable information sharing using a single framework. CC serves as the foundation for this sharing, enabling seamless, omnipresent sensing, data analytics, and information representation.

This paper reviews the literature on the use of BD, CC, and IoT in smart environments to identify promising solutions and novel perspectives. It also aims to uncover potential obstacles and areas for future research. This review study also aims to promote a better understanding and application of these concepts among academics and industry professionals.

A. SMART CITIES FRAMEWORK AND STRUCTURE

The idea of a "SC" is covered in [23], with data being used to address problems and thereby improving the citizens' Quality of Life (QoL). BD technologies have attracted significant attention for developing new mechanisms for a higher QoL due to the enormous number, diversity, and speed of living requirements and city data.

The authors of [48] explained how SC environments are developing to enable the development and integration of SC services, by studying the principles of data processing and software platforms. In the same direction, this paper derives a reference architecture that will guide the advancement of future platforms and improvements for the SCs.

Petrolo et al. [49] proposed an SC idea which is crucial for research and industry in the field of SC technologies. As a result of the growth of numerous middleware platforms and IoT infrastructures, distinct techniques have arisen for building IoT eco-system fragments that are notable for their interconnected communications. In addition, the concept of an SC outlines key needs as well as the benefits of cloud and IoT service integration. Moreover, the pertinent difficulties are covered in this work.

Alavi et al. [50] provided a comprehensive literature review of the essential characteristics of the IoT paradigm to promote the creation of sustainable SCs. Sharifi [51] argued that assessment tools should take advantage of developments in smart solutions and BDA to develop more effective approaches for these criteria. The findings of this study can be used by interested target groups, including SC developers, planners, and policy-makers, to select tools that best suit their objectives in addition to revealing shortcomings that need to be addressed in the future. Another report [52] took the first step in revealing the significance of incorporating sustainability into the development of SCs. The recommended conceptual framework might serve as a practice manual for professionals as they get ready to build sustainable SCs.

The concept of Edge Computing (EC) was introduced by in [53], through several illustrative examples; ranging from cloud offloading to EC in smart homes and collaborative EC. An overview of SCs was presented in [7], followed by descriptions of their traits, general composition, architecture, and applications in real-world settings. The authors of [7] also discussed several issues and prospects of SCs.

The key privacy and security concerns in the architecture of SCs were described [54], highlighting the most important SC applications. Some of the most recent approaches to the security and privacy issues that arise in applications for information-centric SCs were also addressed in addition to introducing upcoming research difficulties that must still be taken into account for performance enhancement.

B. BIG DATA MANAGEMENT PROPOSAL IN SMART CITY CONTEXT

Several recent research studies have demonstrated, at least conceptually, the possibility of utilizing and managing BD to develop and create new SC services [55], [56], [57]. One of the most challenging and prevalent scenarios in an SC setting is the domain of transportation systems, which was covered in details by Neilson and colleagues [56] who discussed numerous approaches in this area. Honarvar and Sami [57] examined whether and how to economically monitor air pollution without the use of expensive sensors and equipment. A predictive model for particle matter prediction was created to achieve this objective. The research presented extremely encouraging experiments on an air pollution forecast scenario while not addressing data collection and management specifically. There are many architectures for offline-only data analysis for cloud data management and analytics, while others are solely focused on real-time data processing. The first type of work research [55] was exemplified through the development of data processing platforms for large GPS datasets to assess transportation policies. The study presented a promising platform for efficient data analysis, enabling policymakers to make informed decisions based on the results. The data used for analysis were pre-collected, and real-time storage and updating were not the primary focus. Gu et al. [58] conducted an exploratory study on the potential of integrating BD and IoT technologies to address the challenges of managing Waste Electrical and Electronic Equipment (WEEE).

These findings demonstrate that while some surveys cover BD and IoT, they tend to concentrate more on specific applications rather than general usage or overall benefits. In other words, these surveys tend to explore the potential of these technologies in specific contexts rather than examining their broader implications and advantages.

Shah et al. [59] highlighted the expanding role of BDA and IoT in disaster management. An analysis of recent studies has resulted in the development of a thematic taxonomy classification, a comprehensive review of widely used solutions, and a conceptual framework for the application of IoT and BDA in disaster management. The thematic taxonomy classification provides a systematic approach for categorizing various methods of using IoT and BDA in disaster management, while the review of solutions evaluates the effectiveness of the existing approaches. The conceptual framework outlines the key components necessary for successful implementation of IoT and BDA in disaster management, providing a roadmap for future research and development in this area.

The authors of [60] developed an interpretive structural model that illustrates the connections between the identified challenges related to BD in SCs development. This model helps stakeholders gain a better understanding of how these challenges interact with each other. The framework presented in their paper is useful for stakeholders to comprehend and analyze the difficulties associated with implementing BD in SC development. According to their report, the utilization of fuzzy DEMATEL analysis can provide a comprehensive understanding of the interdependencies among these difficulties, enabling the effective mitigation of obstacles and promoting the efficient and rapid development of SCs.

Soomro et al. [81] thoroughly assessed the BDA studies on SCs. The authors used specified keywords to search a variety of repositories. They chose the content for the review using a structured Data Mining (DM) process. Moreover, the authors conducted a comprehensive technological and thematic study of the literature, which involved narrowing down the selected papers based on specific criteria and identifying different DM and ML techniques that have been used in the context of the study. They presented a detailed analysis of the various techniques, their strengths and limitations, and potential applications in the field.

The application of IoT and DL for the creation of SCs has been reviewed in [15]. The authors of this study started by outlining the IoT and describing the traits of BD produced by the IoT. Next, they discussed various computing platforms, including cloud, fog, and EC, for IoT big data analytics. They also examined current research that integrates IoT with DL to create smart applications and services for SCs, as well as popular DL models. They finally concluded by outlining the present difficulties and problems in the creation of services for SCs. Pramanik et al [82] provided a critical review of the most recent privacy-preserving BDA methodology and a systematic evaluation of various privacy-preservation strategies. They recommended five suggestions for proper implementation of BDA for organizations while protecting customer privacy. In [83], the authors described the role of BD in understanding cities and provided a timeline of

TABLE 2. Papers addressing the concept of SCS, BD, and IoT.

Paper Title	Description	Reference
Big Data Technologies and Cloud Computing	The rise of BD, the current data-burst situation, the connection between BD and CC, and BD technologies are the main topics of this book, which also provides a thorough definition of BD from a variety of angles.	[61]
Integration of Cloud Computing and Internet of Things: a Survey	Integration of Cloud and IoT, i.e. the CloudIoT paradigm	[62]
Smart cyber society: Integration of capillary devices with high usability based on Cyber–Physical System	Introduction of the concept of Smart Cyber Society; through the idea of smart home	[63]
The role of big data in smart city	The state-of-the-art communication technologies and smart-based applications for SCs.	[23]
Smart cities – enabling services and applications	A review on the Infrastructure, Services and Performance for SCs.	[64]
Smart Health: Big Data Enabled Health Paradigm within Smart Cities	Organized evaluation of several BD and smart system technologies in the context of healthcare	[65]
Smart City Designing and Planning based on Big Data Analytics	A review of IoT, Communication systems that are energy-conscious for IoT environments and Energy-efficient communication systems for IoT environments	[66]
Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges	This study looked into cutting-edge research projects focused on massive IoT data analytics. BDA and IoT are related, as is explained.	[67]
Big Data in Smart-Cities: Current Research and Challenges	Reviewing big-data applications and presenting ten examples of smart-cities across the world.	[68]
Smart city with Chinese characteristics against the background of big data: Idea, action and risk	Proposing a development framework for Chinese SC considering BD.	[69]
Efficient IoT-based Sensor BIG Data collection–Processing and Analysis in Smart Buildings	BD, CC, IoT, and Monitoring technologies presented, with focus on common operations.	[70]
Cloud Computing and Big Data: Technologies, Applications and Security	A book addressing cloud and BD technologies, architecture and applications. The book is mainly devoted to the security aspects of CC and BD.	[71]
On big data, artificial intelligence and smart cities	This study examines AI's urban potential and provides a new paradigm for connecting AI and cities while assuring the integration of crucial cultural, metabolic, and governance components.	[72]
IoT big data analytics for smart homes with fog and cloud computing	Examining a knowledge base solution for IoT Data Analytics in Smart Homes, a complete data gathering, processing, and real-time recognition framework.	[3]
The Construction of Smart City Information System Based on the Internet of Things and Cloud Computing	This research initially looks at the evolution of the Internet of Things, CC-related technologies, and SCs, before concentrating on the important Internet of Things and CC technologies in terms of structure and application.	[73]
Future Trends and Current State of Smart City Concepts: A Survey	This article presented an overview of SC programs, as well as an analysis of their major principles and data handling methodologies, used a comprehensive literature matrix to conduct a thorough literature search and evaluation, which included terms like smart people, smart economy, smart governance, smart mobility, smart environment, and smart living.	[74]
IoT and Big Data Analytics for Smart Buildings: A Survey	Proposed an overview of connected works to IoT, BDA, and smart buildings in this paper.	[75]
IoT and Big Data Applications in Smart Cities: Recent Advances, Challenges, and Critical Issues	The notion of SCs is briefly discussed initially, followed by descriptions of its attributes and standards, as well as generic design, compositions, and real-world applications. In addition, the potential problems and opportunities in the subject of SCs are discussed. This study addresses a number of concerns and challenges, including analytics and the use of BD in SCs, which will aid in the development of applications for the aforementioned technologies.	[76]
Role of BIC (Big Data, IoT, and Cloud) for Smart Cities	This article discussed important ideas and plans for the intelligent community, in which the BIC (BD, IoT technologies and CC) increases the efficiency of innovative city applications. Using BD, CCs, and IoT, a variety of SC issues were solved.	[77]
Fundamentals of Smart Cities	The fundamentals of SCs, verticals in SCs, and BDA methodologies are covered.	[78]
Bootstrapping Urban Planning: Addressing Big Data Issues in Smart Cities	The author investigates the potential for BD to play a role in the efficient management of SCs. Representative BD applications, as well as their benefits and drawbacks, are examined.	[79]
Applications of Big Data and Green IoT-Enabling Technologies for Smart Cities	Reviewing- of BD, SCs, IoT , green-IoT concepts, technology and approaches, and applications around the world	[8]
Big data analysis of the Internet of Things in the digital twins of smart city based on deep learning	Utilization of BD applications for SCs.	[80]

its development. The study also discussed the potential of BD to facilitate data-driven approaches for urban planning, as well as its potential drawbacks, which were summarized in the final section.

C. CLOUD COMPUTING

Rittinghouse and Ransome [84] provided a comprehensive guide to CC, which included precise examples and practical strategies to help readers grasp its potential risks and design an effective cloud strategy. Additionally, the authors presented a historical overview of CC and discussed the influence of emerging technologies such as virtualization on its development and acceptance. By doing so, they aimed to dispel misconceptions that may impede the adoption of CC and facilitate the understanding of this critical technology.

Botta and colleagues [61] examined the emerging paradigm of CloudIoT, which combines cloud and IoT technologies [62], [62]. They surveyed CloudIoT applications in the literature, emphasizing the research challenges arisen from this integration. They also evaluated existing CloudIoT platforms, including both open-source and proprietary solutions. In conclusion, the study identified unresolved issues and potential directions for CloudIoT, which the authors believe will have a transformative impact on the future of the internet.

The CPSS (Cyber-Physical-Social System) framework was proposed for SCs [85] to provide residents and policymakers with lower-latency, in real-time, more effective, and proactive services. This framework moves some duties from the cloud center to network edge devices and brings the services and resources closer to the users.

Alam [86] explored cloud-based IoT applications and their functions in SCs. Furthermore, IoT and cloud convergence, cloud-based IoT solutions, and cloud-based IoT applications for SCs were all topics discussed by the author.

The benefits of CC for the development of SCs were also evaluated [87]. Additionally, it was investigated how CC fosters the growth of an SC and what social settings are suitable for full utilization of the infrastructure required to construct an SC using the cloud. Researchers worldwide are particularly interested in IoT, AI, and CC. They have proposed useful guidelines to support cutting-edge technologies and SCs. The authors investigate how new technologies fit into SCs.

Jamsa [88] discussed the nature of CC and the criteria used to evaluate its effectiveness. The author focused on browser-based Software as-a-Service (SaaS) offerings and their benefits for users. Furthermore, the study explored the use of cloud-based hardware and software platforms, which can help businesses of all sizes to rapidly and cost-effectively migrate their applications to the cloud. Finally, the author highlighted the potential of existing applications to be easily moved to the cloud by developers, and discussed the implications of this shift in the future CC.

D. IoT APPLICATION IN A SMART CITY CONTEXT

Mourtzis et al. [89] argued that the manufacturing industry can modernize its operations by adopting IoT technologies. Such a transformation leads to the generation of vast industrial data, which can be transformed into meaningful insights using advanced analytics. By leveraging these insights, enterprises can develop new data-driven strategies to effectively respond to competitive pressures. The authors illustrated their point using a hypothetical business with around 100 machines to demonstrate how the IoT paradigm can be applied in practice.

A literature survey on data fusion techniques was presented for smart ubiquitous environments, with a particular focus on mathematically-derived methodologies and their applications in various IoT contexts (e.g., distributed, heterogeneous, etc.) [90]. The study provided a comprehensive overview of each mathematical approach and its potential, as well as the challenges that arise when applying these techniques in different environments.

Talavera et al. [91] examined the agro-industrial and environmental sectors from an IoT perspective to identify the trends, application areas, architectural frameworks, and unresolved difficulties in these two distinct sectors.

Al Mamun and Yuce [92] provided an overview of recent studies and advancements in sensors, systems, and wearable technology for IoT applications in the environmental sector.

Quy et al. [93] presented IoT solutions and illustrated how IoT can be incorporated into the field of smart agriculture. To accomplish this goal, they assess the architecture (IoT devices, communication technologies, BD storage, and processing), applications, and research schedule of IoT-enabled smart agriculture ecosystems. Moreover, they also go through the prospects and trends for IoT applications in smart agriculture, as well as the problems and difficulties that still need to be resolved.

Syed and colleagues [94] presented a comprehensive review of IoT in SCs. They started by going over the basic elements of the IoT-based SC landscape, then moved on to the technologies that support these domains' existence in terms of the architectures employed, the networking technologies used, and the artificial algorithms deployed in IoT-based SC systems.

New concerns and potential directions were analyzed for the IoT business model, identifying elements that affect and obstruct an organization's ability to utilize IoT [95]. The survey identified several motivators and obstacles for IoT implementation. An innovative architecture was suggested in [96] to enhance the level of resilience of the IoT infrastructure. Furthermore, the implementation of the architecture's components using various technologies was recommended.

III. SMART CITIES

The paradigm of the SC has recently gained increasing popularity. Regarding the technical advancements in computing devices and wireless, mobile, and wearable sensing, its rapid development is expectable in the years to come. Efficient application of the real-world data input to the IT infrastructures is a vital component in improving the QoL of citizens through smart computing.

The idea of an SC has been originated from urbanization. Currently, more than 54% of people live in urban regions. This figure is estimated to rise to 66% by 2050 [97], accommodating an additional 2.5 billion people [98]. Cities are the main consumers of natural resources (75%) and are responsible for 60-80% of Greenhouse Gas (GHG) emissions [99] while occupying only 3% of the earth's land. Population growth has posed serious hazards to the environment due to enhancing energy, water, food, and transportation demands. The ever-increasing consumption rate has depleted natural resources while promoting global warming. Therefore, efficient measures have to be taken to establish economic, environmental, and social sustainability. In this context, the concept of "SC" has found increasing popularity. Cities are a key player in sustainable development due to their decisive role in social and economic aspects, as well as undeniable environmental impact [100]. Based on Fig. 4, the SC has six main contributions.

A. DEFINITIONS, ADVANTAGES AND DRAWBACKS OF SMART CITIES

According to IBM's documentation, SCs encompass smart equipment and devices. Table 3 lists various definitions of the concept of "SCs". As seen, the concept of the SC is no longer limited to the diffusion of ICT, as it considers the needs of people and the community. This aspect was further explained in [101] with emphasis on the urge to improve the use of ICT in cities to enhance the QoL.

"SC" is the common term referring to information and the digital city. SCs try to alter living and work by Information Technologies (IT) [102]. People are the leading actors in an SC by forming it through their lifelong interactions. Therefore, other terms are mainly associated with the concept of an SC. For instance, creativity is the main driving force of an SC, introducing education, learning, and knowledge as the key elements of an SC [100]. An SC is at the heart of higher education to recruit more educated experts and skillful human resources. SCs serve as magnets for the poor to make them smarter. Therefore, an SC offers numerous opportunities to its community members. Cities with a higher contribution of the educated workforce will achieve the maximal growth rate. The concepts of smart, skilled, capable, connected, and competitive are now the major elements in the development of cities. Cloud technologies have facilitated the development of knowledge-based environments through enhancing urban monitoring systems [103]. Industry, education, contribution, and infrastructure are the key components of an SC. The University of Vienna extended them to six items including economy, mobility, environment, people, living, and governance [104].

Table 4 examines the advantages and disadvantages of deploying SC infrastructure.

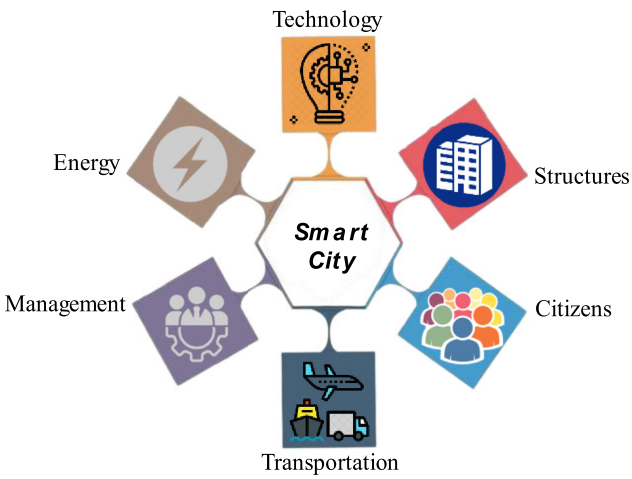

FIGURE 4. Significant components of an SC.

B. APPLICATIONS OF SMART CITIES

SCs can provide efficient and intelligent services for the public and different authorities through various sensor technologies and platforms that allow them to manage, share, and store the acquired data. For instance, positional, pressure, air movement, motion, and sound sensors are the most popular sensors that have been utilized in SCs. Those sensors are connected to the IoT and serve as the foundation of data collection in SCs for monitoring and managing urban activities as well as making real-time choices.

Wearable sensors are also becoming trendier these days. This may be due to the fact that the SC can gain significant information about the activities and behavior of its users by combining data obtained from wearable sensors such as smartphones. Furthermore, sensor systems can be utilized to provide information about traffic and available parking spots, minimizing driving stress, fuel consumption, and pollution, while enhancing economic output. In the same direction, smart public lighting with built-in motion sensors and energy-saving LED bulbs is another widely used improvement in urban infrastructure. Besides, smart buildings [118], utilizing sensors to optimize energy consumption are becoming more common. In addition to the aforementioned applications, sensor technology can be employed in various ways to enhance the functionality of SCs [119].

SCs contain different applications such as Smart Healthcare, Smart Transportation, Smart Governance, Smart Grid, and Smart Home. Table 5 summarizes the mentioned SC applications.

C. CHALLENGES AND LIMITATIONS FOR THE DEVELOPMENT OF SMART CITIES

BD has opened new horizons to the paradigm of SC [127]. Such innovations are, albeit, accompanied by numerous multi-dimensional challenges which can be discussed from diverse perspectives. This section presents the key research challenges. The business challenges such as planning, connection and sustainability, source of markets and customer, cost of acquiring SC, and CC integration are first addressed. Whereas the second part deals with technological challenges such as privacy, data analysis, data integration, QoS, and BDA using computational intelligence methods for SCs [23].

The challenges and limitations in Table 6 can be an interesting topic for future studies.

D. ANALYTICS OF BIG DATA IN SMART CITIES

BD play a significant role in the development of SCs [7]. Likewise, BDA comprise a wide range of applications in different aspects of SCs, e.g. traffic control and transportation [128], [129], Planning [79], [130], energy [20], criminal analysis [19], [131], and the environment [21]. However, specific non-functional and functional needs related to the type of data sources and applications in SCs should be explicitly established before designing software platforms and architectures for SC objectives. In this regard, business decisions can be improved through statistical analysis, predictive modeling, and other techniques. Additionally, analytics can aid in the development of the efficiency of critical processes, operations, and roles. In other words, by converting data into intelligence, analytics can expand the business performance. The diagram below depicts the journey of BD from its inception in the raw form to the extraction of useful information and insights for the benefit of decision-makers and citizens (Fig. 5).

Any city that has the following characteristics is an SC:

- Smart economy to support city
- Smart people and businesses will take advantage of all of the resources and services of the SC.
- Smart mobility to drive
- Smart governance allows for effective monitoring
- Smart environment will benefit the public.
- Smart living is required for such a large population.

The growth of SCs is driven by institutional, human, and technological factors. All around the world, people are migrating from villages to cities, which has led to the development of smarter cities. Data overload in SCs has forced the establishment of a management department for overseeing and control purposes. This department, however, is ineffective in terms of time and money. IoT devices made for traffic control are different from those used to collect data for healthcare (i.e. that is why we need to categorize the data into groups of domains to create certain technical BD for each domain).

TABLE 3. Definitions of SCs.

Explanation	Reference
A city is considered smart upon investments in (i) human and social capitals, (ii) conventional infrastructures, and (iii) fuel sustainable for economic growth and achieving high quality of life through rational management of natural resources.	[81]
Sustainable, innovative, connected and socially cohesive places that enhance the quality of urban life.	[105]
A city to be smart when investments in human and social capital and traditional (transport) and modern (ICT) communication infrastructure fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance.	[106]
A smart sustainable city is an innovative city that uses information and communication technologies (ICTs) and other means to improve quality of life, efficiency of urban operations, and services, and competitiveness, while ensuring that it meets the needs of present and future generations with respect to economic, social, and environmental aspects.	[107]
Cities around the world are entering a new era in which residents and their surrounding environments are increasingly connected through rapidly-changing intelligent technologies, sometimes called, smart technologies.	[108]
A Smart City is a city well performing in a forward-looking way in the following fundamental components (i.e., Smart Economy, Smart Mobility, Smart Environment, Smart People, Smart Living, and Smart Governance), built on the 'smart' combination of endowments and activities of self-decisive, independent and aware citizens.	[109]
A smart city is understood as a certain intellectual ability that addresses several innovative socio-technical and socio-economic aspects of growth. These aspects lead to smart city conceptions as "green" referring to urban infrastructure for environment protection and reduction of CO2 emission, "interconnected" related to revolution of broadband economy, "intelligent" declaring the capacity to produce added value information from the processing of city's real-time data from sensors and activators, whereas the terms "innovating", "knowledge" cities interchangeably refer to the ability of the city to raise innovation based on knowledgeable and creative human capital.	[110]
The application of information and communications technology (ICT) with their effects on human capital/education, social and relational capital, and environmental issues is often indicated by the notion of smart city.	[111]
Smart cities have high productivity as they have a relatively high share of highly educated people, knowledge-intensive jobs, output-oriented planning systems, creative activities and sustainability-oriented initiatives.	[112]
Smart cities are the result of knowledge-intensive and creative strategies aiming at enhancing the socioeconomic, ecological, logistic and competitive performance of cities. Such smart cities are based on a promising mix of human capital (e.g. skilled labor force), infrastructural capital (e.g. high-tech communication facilities), social capital (e.g. intense and open network linkages) and entrepreneurial capital (e.g. creative and risk-taking business activities).	[113]
Smart cities initiatives try to improve urban performance by using data, information and information technologies (IT) to provide more efficient services to citizens and monitor and optimize existing infrastructure to increase collaboration among different economic actors, and encourage innovative business models in both the private and public sectors.	[114]
Smart cities have a high capacity for learning and innovation coupled with creativity in people, knowledge-based institutions, and digital infrastructure for communication and knowledge management.	[115]
The main features of the smart city are smart economy, smart mobility, smart environment, smart people, smart living, and smart governance.	[97]
A city in which the traditional services and networks based on digital technologies are made more efficient for the benefit of its businesses, services, and inhabitants.	[116]
The concept of SCs was suggested as one of the most effective solutions to increase the quality of life and social welfare by considering human and environmental properties.	[117]

TABLE 4. Comparison table for advantages and drawbacks of SCs.

Advantages	Disadvantages
Improved interconnection	Inadequate social control
Lowering the carbon footprint	The project is still under progress
Efficient public services	Difficulty during the pre-commerce stage
Less crimes	Concerns about data security and privacy
Infrastructure improvements	The business case for implementation is difficult to release
There are more job opportunities	Overconfidence in the network
More secure communication	Restricted privacy

Fig. 5 displays six different sensors and IoT devices, including smart phones, smart cameras, smart thermometers, smart users with sensors, smart cars, and smart homes. A data center may collect data from items of various sorts, ranges, and values because sensors are available in a variety of sizes and shapes. Here, we outline the challenges that already exist and will emerge, along with the related solutions.

Challenges: Due to the shortage of equipment in every site where we wish to monitor and make judgements for that region, such as transmitting traffic information, the data collection process in SCs has improved but still remains challenging. Regardless of our focus whether it be on smart healthcare mobility or other areas of interest, IoT devices and other sensors may fail to precisely capture all data or may entirely miss data due to their limited storage and inefficient time-scales. Furthermore, data have greatly multiplied and are derived from a variety of sources.

Therefore, different approaches must be used to convert all of the data into single unit measurements, ranging from video and images to digits or strings. With the help of these metrics, we can swiftly apply ML and other DL algorithms to the data to get the most accurate conclusions.

Solution: Pre-processing is essential for handling values and missing data in the generated dataset. There are both simple and complex methodologies [25] for handling missing values, as well as tools and techniques for selecting important and pertinent features. For example, IFAB [26] uses an Artificial Bee colony to remove pointless features and ensure the validity and reproducibility of the results.

Data Engineering [21] is in charge of organizing and evaluating the input data, as well as applying labels to data that hasn't been labeled. This necessitates the use of professionals and time; hence it is not cost- or time-effective. As a result,

TABLE 5. Application of BD in SCs.

Application	Specific Use	IoT	Possible Communication Technology	Advantage	Limitation
Smart Healthcare [120], [121]	Health monitoring	Sensor, Smart wearable devices	Bluetooth, ZigBee	• Remote monitoring: Early diagnose the disease and save lives in case of a medical emergency • Prevention • Reduction of healthcare costs • Medical information is easily accessible • Improved treatment management	• Lack of precision • Security and privacy • Risk of failure • Uncertain network
Smart Transportation [45], [122]	Effective route management	Smart cars, Camera, RFID carts	RFID, 3G, and 4G	• Automatic traffic management • Efficient route management • Less congestion	• Network disconnectivity can result in serious accidents • inflexible to use in mixed traffic • Equipment of ITS is costly • High maintenance cost • Not dedicated for traffic data collection
Smart Governance [123]	Making smart policies to manage citizens	Smartphones, Camera, Sensors	WiFi, LTE, LTE-A, WiMax, Bluetooth, LoRaWAN	• Awareness of citizens' needs • clear policy • Enhanced participation of citizens • Access to curtail information • Suitable future • Involvement of the private sector	• Difficulties in data collection and analysis • Cybercrime/Leakage of Personal Information
Smart Grid [38], [124], [125]	Power supply management	Smart meters and smart readers	WiFi, Zigbee, Z-Wave	• Efficient power supply • Estimate future supply requirements • enhance the efficiency of electric power systems	• Expensive • Difficult to operate
Smart Home [126]	conserving energy	A collection of various sensors	WiFi, Zigbee, Bluetooth, RFID, Z-Wave	• Energy-conserving • Hands-free convenience • Enhanced Security • Save Time with Automated Tasks • improved management	• Internet Reliance • Configuration and Setup • Threats to Technical Security • High cost

utilizing BD techniques aids in right tuning and analysis of data.

IV. BIG DATA

BD is a hot topic in recent studies. Science, business, industry, government, and society are believed to be revolutionized under the influence of BD (Fig. 6). A huge deal of data are currently generated by various devices including smartphones, social media, Global Positioning System (GPS), computers, sensors, commercials, and games [147].

An ocean of data generated by these devices should be processed by BDA to produce relevant information. [148]. A BD Research and Development Initiative was founded in March 2012 by the Obama administration [149].

The SC can store and process data using BD technology, which will improve a variety of services. Moreover, BD can also assist decision-makers in increasing the services and resources in SCs. BD can accomplish these objectives by using the right tools for effective data analysis. These technologies and techniques encourage cooperation between organizations, provide services to numerous industries, enhance client experiences, and open up new business opportunities.

In terms of data analytics, BD can be defined by five V's — Volume, Velocity, Veracity, Variety, and Value (Fig. 7) —. Either all or any of them should be met to classify a problem as a BD problem. The volume deals with the size of data that is too large to be manageable by the currently available algorithms. As the most compelling of all V's, variety denotes different types and modalities of a given object [76].

The value is influenced by the usage, sample frequency, and age of the data. For instance, in a collision avoidance application, data of more than a few minutes old may not be useful, whereas route planning applications might make use of non-real-time data. Thus, the value can be viewed as a gauge of the capacity to extract useful data [150].

In terms of velocity, data should be streamed at rates faster than what can be handled by conventional algorithms and systems. Sensors rapidly read and communicate data streams. We are approaching the world of the quantified self, presenting data that have not been available hitherto [20].

TABLE 6. Constraints, challenges, and solution in the development of SCs.

Challenges Type	Limitation	Constraint Solving Method	Reference
Planning	The SC involves integration of a great deal of data from various sources which is one of the main challenges.	Various technologies have been developed for integrating data into the SC to reduce the technical issues in addressing data. Data quality is one of the challenging issues in any data integration mechanism.	[132]
Connection and Sustainable	The real-time interaction with IoT technologies and BD has led to new challenges	Implementation of sustainable SC applications requires the enhancement of IoT with new approaches for more reliability, resilience, autonomousness and intelligence. An SC can leverage the power of the IoT and BD to improve the services.	[133]
Source of Markets and Customer	The diversity of smart devices and applications in a city may confuse businesses in finding the right market and customer resources.	New technologies can extend the reach of organizations, improve management decisions, and accelerate the development of new products and services. Many social media apps are daily used by customers	[134]
Cost of Acquiring Smart City	The SC has to integrate diverse components which might be costly for the governments due to insufficient natural and human resources.	Standard technology frameworks can reduce the costs.	[105] [135] [136]
Cloud Computing Integration	CC technologies offer flexibility and low cost for BD hosts. Integration of these technologies with IoT is a huge challenge.	CC services have significantly improved the SCs as they offered efficient solutions regarding the security, management and open platforms.	[63] [137]
Privacy	The users' information may be misused, raising serious concerns for the citizens.	The challenge of safeguarding the personal data gathered by SC technology against thieves and hackers requires more research.	[81] [138] [139]
Data Analysis	In an SC, data are collected from different objects. The process of decision-making requires novel algorithms and visualization approaches, which may affect the activities of a centralized SC.	On-the-fly processing has gained increasing importance as the conventional storage and processing may be no longer applicable. Real-time data storage and processing are required.	[46] [140]
Data Integrating	The SC involves integrating a great deal of data from various sources which is a great challenge.	Various technologies have been developed for integrating data into the SC to reduce the technical issues in addressing data. Data quality is one of the challenging issues in any data integration mechanism.	[141] [142] [95]
QoS	Several technologies must be integrated to form an SC. The QoS provided by various technologies is a key challenge in the adoption of SC.	Satisfactory QoS should be provided by these technologies before full integration of SC application. The frameworks and methods to determine and apply QoS parameters in an SC is of crucial significance.	[143] [144]
Computational Intelligence Algorithms for Smart City Big Data Analytics	Computational intelligence algorithms, such as neural network, genetic algorithm, artificial bee colony, particle swarm optimization, cuckoo search algorithm, flower pollination algorithm, chicken swarm optimization, and bat algorithm can effectively engineer our knowledge. These algorithms encompass soft computing, ML, and DM. Nonetheless, the effectiveness, efficiency, and robustness of computational intelligence algorithms are limited to small datasets, making them inapplicable in BDA of an SC. The effectiveness, efficiency, and robustness of the computational_intelligence algorithms normally diminish upon enhancing the capacity of the dataset, making them inappropriate for investigating knowledge in BD of the SCs.	These solutions must be applied to hardware and software, data control, offline and online, and processing requirements, after which they can be further tailored to domain-specific dynamics and restrictions. As a result, these approaches can be used to provide efficient multi-purpose intelligent data analysis as well as decision support systems for a variety of commercial or industrial applications involving large amounts of complex or ambiguous data that require analysis in order to make operational and cost-effective decisions.	[145] [146]

Veracity suggests the serious concern on the data quality despite its availability. This implies that BD does not necessarily result in higher quality. Quality concerns are intensified in larger data which should be either tackled at the data pre-processing stage or by the learning algorithm.

The significance of the term "BD", which was coined a few years ago, has been debated by different scientific organizations. In this context, there are still a variety of technological challenges and business implications associated with BD due to the lack of global agreement on a uniform and widely acknowledged definition. To the best of our knowledge and according to the literature review, BD definitions can be divided into three categories based on their corresponding principal factors (i.e. "data quantity", "challenges", and "BD management complexities") which are utilized to formally elaborate the definition. Table 7 lists relevant literature sources for each of the aforementioned factors.

A. TYPES OF BIG DATA

Data of SCs can be generated by multiple sources in various formats. Conventional software may fail in storage and processing these data. BD can be classified into four classes, comprising structured, semi-structured, unstructured, and

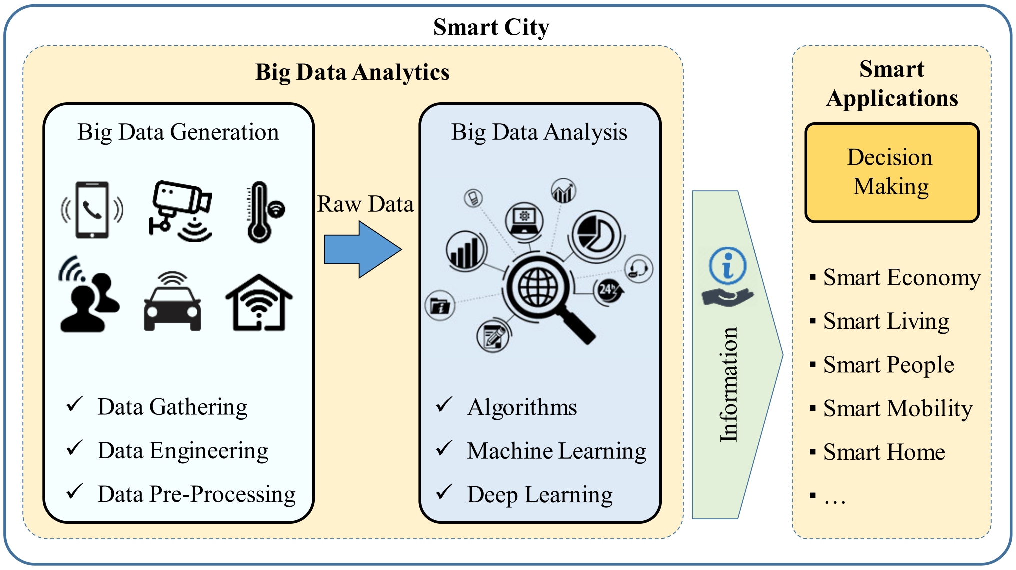

FIGURE 5. BDA in SCs.

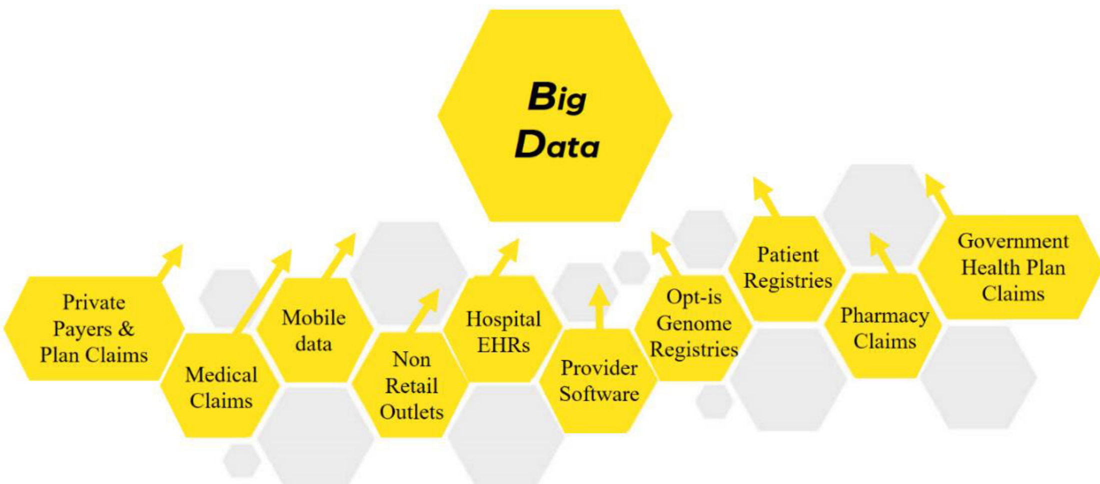

FIGURE 6. Major sources of BD in SCs.

mixed data, as shown in Fig. 8. Structured data possess a dedicated data model with a well-defined structure that follows a consistent order. Its design facilitates its access and usage either by a person or a computer. Structured data are often stored in well-defined columns and databases. Example could be Database Management Systems (DBMS). Semi-structured data refer to another form of structured data. Despite possessing a few properties of structured data, this type of data does not have a definite structure. Semi-structured data do not follow the formal structures of data models such as an RDBMS. Example: Comma Separated Values (CSV) File. Unstructured data neither have a definite structure nor exhibit the structural rules of data models. Their format is inconsistent and continuously varies. In rare cases, they might contain time-dependent information. Example: Audio Files and Images. Despite the fact that are not considered BD, mixed data are subtypes of data that are relevant to the field of analytics. Mixed data can be generated from different types of information, e.g. (1) Geographic Information Systems (GISs) which are used for capturing and analyzing geographically referenced data referring to infrastructure, roads, buildings, lakes, addresses, people, workplaces, and transit routes, (2) real-time media which consist of live or recorded media data streaming applications such as YouTube, Flicker, and Vimeo that produce a huge amount of videos, pictures, and audios, (3) natural language data that create vast amounts of text-like communication between objects, e.g. speech capture devices, land phones, cell phones, and the IoT, (4) time series analysis which is a set of data points (or observations) including multiple measurements taken over a period of time,

TABLE 7. Contributions of BD definitions based on their corresponding principal factors.

Principal Factors	Key Contributions of BD Definitions
Data Quantity	• Scientific data visualization for BD [151]; • Advanced data analytics due to the lack of conventional tools in the collection, preparation, analysis, and storage of BD [152]; • Complex analysis process for growing real-world information and big databases [153].
Challenges	• The 3Vs model comprising Velocity, Volume, and Variety is the most important issue in BD. Subsequently, another "V" so-called Veracity has been added to the 3Vs model and formed the 4Vs model [154]–[157]. • Value as a new "V" has been replaced with Veracity in the 4Vs model [158], [159]. • Then, the 5Vs model (i.e. Velocity, Volume, Variety, Value, and Veracity) was a further reproduction of V models [160], [161]. • Once again, Veracity has been removed and instead Variability as a new "V" has been replaced in the 4Vs model [83], [105]. • This works added variability (replacing veracity), so defining 5Vs as Volume, Velocity, Variety, Value and Variability [162]. • Last but not least, the 7Vs model (i.e. Volume, Variety, Velocity, Veracity, Value, Variability and Visualization) has been proposed by the new school of thought [76], [163].
Big Data Management Complexities	• Massive-scale datasets, which cannot be managed using conventional tools, can be defined by BD [161], [164]. • BD is recognized as a new data-intensive technology by various stakeholders such as industry, academia and research communities [165], [166]. • Complex large datasets with hidden relationships need to be integrated with BD technology [167]. • The vital steps of BD implementation are data collection, storage, and transportation in order to conduct the main contribution of the BD processing, i.e. exploitation [158], [168]. • Conventional approaches are challenged by the acquisition speed as well as deep processing of massive amounts of data which is one of the most significant assumptions of BD [75], [169]. • Definite complicated datasets, which cannot be handled by traditional data processing systems, are manageable using BD [23], [170]. • Large structured and unstructured data is defined as BD that cannot be managed through conventional software algorithms [171], [172].

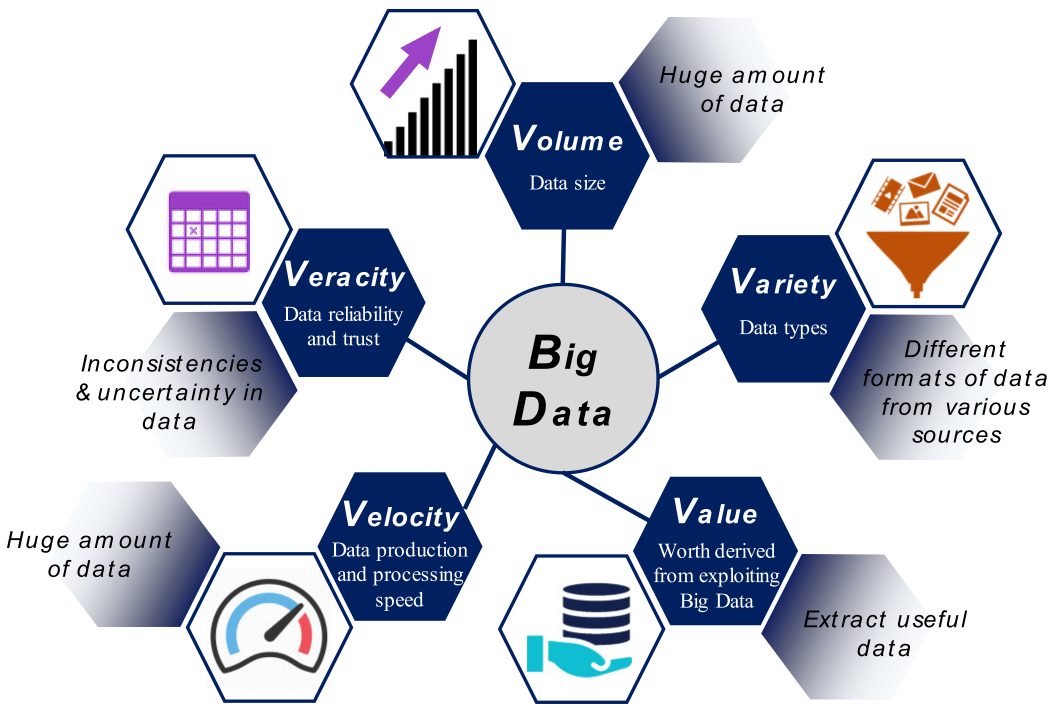

FIGURE 7. Characteristics of BD.

(5) network data that is concerned with vast networks comprising social networks, i.e. Facebook and Twitter, information networks, i.e. the World Wide Web, biological networks, i.e. biochemical, ecological, and neurological networks, and technological networks, e.g. the Internet, telephone and transportation networks, and (6) link data that uses standard Web technologies (e.g. HTTP, RDF, SPARQL, and URIs) for information communication through computers rather than human.

Here is a summary of various sources of BD to help in understanding its dimensions. Industry-specific content digitization has been provided as a new data source.

TABLE 8. Different BD Sources.

Sector	Data Generated	Usage
Astronomy	The motion of satellites and stars	Observing the asteroid bodies' actions
Financial	Videos, audio, tweets, and news reports are all examples of news content	Decision-making in trading
Healthcare	Digital photos and medical records	Both long-term epidemiology research and short-term public health monitoring
Internet of Things (IoT)	Sensor data	Keeping track of various events in SCs
Life Sciences	Gene sequences	Examination of genetic differences and potential remedies
Media/Entertainment	Content and user viewing behavior	To increase the audience size
Social Media	Tweets, blog entries, social networking sites, and log information	Analysis of the customer behavior pattern
Telecommunications	Call Detail Records (CDR)	Customer churn management
Transportation, Logistics, Retail, Utilities	Sensor information produced by smart meters, RFID tag readers, and fleet transceivers	Optimization of operations
Video Surveillance	Recordings from CCTV to IPTV cameras and recording system	Analysis of behavioral patterns for improvement of services and security

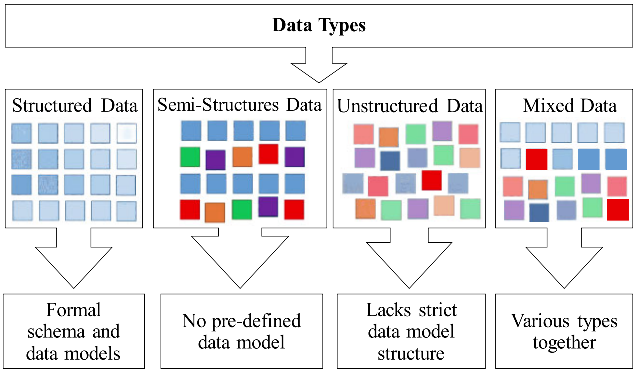

FIGURE 8. Types of BD.

Data generation has been accelerated by technological developments. Table 8 lists different types of data which can be generated by different sectors.

B. APPLICATION OF BIG DATA

BD technology has offered several benefits such as [173]:

- BD allows for predictive analysis to protect organizations from operational risks.
- Predictive analysis can help organizations to expand their business by analyzing the needs of their customers.
- Thanks to BD, multimedia platforms such as YouTube and Instagram have the opportunity to share their data.
- The medical and healthcare sectors are benefited from the opportunity of constant monitoring of the patients.
- BD can revolutionize customer-based companies and the global market.

BD is an invaluable and powerful fuel capable of running the massive IT industries in the 21st century. Soon, BD will be the most common technology in almost all business sectors [174]. We have compiled a list of some of the most important industries that use BD including transportation, insurance, energy and utilities, manufacturing, media and entertainment, healthcare, banking and securities, and education. (See Fig. 9).

- **Transportation:** One of the industries using BD technology most extensively is transportation. BD can help in predicting the requirements of travel facilities and improve the business by dynamic pricing.
- **Insurance:** BD has been used in the insurance industry to provide customer insights for simpler products through analyzing data from GPS-enabled devices, social media, CCTV footage, as well as forecasting customer behavior. BD can also be employed for claim administration, fraud detection, risk management, and real-time claim monitoring in the insurance sector.
- **Energy & Utilities:** Smart technologies such as CC, sensors, and power planning are continually being implemented in the energy and utility industries. All of these technologies generate massive amounts of data over time. Businesses in the energy and utility industries have difficulty extracting useful information from generated precious data such as dynamic energy management and smart utilities to use smart data.
- **Manufacturing:** BDA aid predictive modeling in the natural resources industry, allowing for better decision-making. They can be used to integrate massive amounts of data from geospatial, text, and temporal sources. This is due to the fact that the usage of BD technologies in manufacturing helps to improve product quality, track manufacturing faults, and plan supplies. In other words, in manufacturing sectors, BD solutions can be applied to enhance energy efficiency, predict output, test new manufacturing processes, and design new manufacturing processes.

- **Government and the Military:** also extensively utilize BD technology. High rates of data are created by the government and the military; for instance, a normal flight of a fighter jet plane requires processing petabytes of data [175].
- **Telecom:** by connecting people throughout the globe, telecom has eradicated the distance barrier. Some barriers have remained which require prompt solutions. BD can resolve the remaining issues of the telecom industry.
- **Retail:** The retail industry is a highly competitive field. In this context, retailers always seek novel approaches equipping them with a competitive superiority over competitors. Customers are the only rulers of the retail industry. Thus, a proper understanding of their behavior is a key factor in the survival of the retailers in such a competitive field. To this end, they should know their customers' demands and fulfill them in the best way [56].
- **Media & Entertainment:** are among the major users of Big Data. Zettabytes of data are daily generated whose handling requires BD technologies 1ę1.
- **Healthcare:** BD has already revolutionized healthcare [176]. Thanks to predictive analytics, personalized healthcare services can be provided for patients.
- **Banking & Securities:** Sectors widely employ BD technology to understand customer behavior according to their investment and shopping patterns, investment motivation, and personal or financial histories.
- **Education:** The data collected by an educational institute are of crucial significance. BD is a key player in the future of the people as it can improve the education system [13]. Prestigious universities employ BD for the renovation of their academic curriculum.

BD can be utilized to track the dropout rates of the students to apply the measures for declining this rate, if necessary.

C. TOOLS FOR BIG DATA ANALYTICS

The use of Spark, ETL (Extract, Transform, and Load) tools, and Hadoop components (Pig, HiveQL, etc.) is used for analytical processing such as aggregations and transformations (Fig. 10). DM is a labor-intensive procedure that can be accomplished by Spark (MLlib or custom code), making use of distributed processing, or with Weka and R. In situations where distributed processing is not necessary, the latter has demonstrated adequate performance. Furthermore, through the use of external packages (plyrmr, rmr, SparkR, distributedWekaHadoop, distributedWekaSpark), the two tools listed above can utilize Hadoop or Spark when appropriate. The expense of development and the acquisition of commercial tools must be balanced while choosing the data visualization tool (e.g. Tableau).

BASIS offers open-access BDA for external applications using Representational State Transfer (REST) web services. The proposed architecture aims to not only provide data processing, mining, and analysis methods in the Open (Big) Data portal but also to promote richer third-party applications that use these cutting-edge techniques to present insights within the context of SCs. This is analogous to the public availability of data in the Open (Big) Data portal. Based on [145], the community can be introduced to these analytical capacities, opening up intriguing possibilities for the production and sharing of public information and encouraging the co-creation of city services. As a result, the BDA plays a critical role in the proposed architecture, as in addition to data scientists and analysts working for the SC government, citizens and organizations can utilize these publicly available data.

Eventually, the administration, monitoring, and security components supervise the entire operation from another architectural component, to make sure of the sufficient functioning and safety of the system. The majority of technologies that are relevant to this component are built into Hadoop (Ambari, Knox, Sentry, Ranger, etc.); however, some of these technologies, such Kerberos authentication and Access Control Lists (ACLs), can also be used with Hadoop or other technologies. Although no monitoring tool has been specifically mentioned for non-relational storage, regarding its direct dependency on the selected technology, the relevance of its adequate utilization was assumed to guarantee greater efficiency. The mentioned technologies of this component should be only considered as a guideline for future techniques. Sufficient experience in infrastructure and security is of crucial significance.

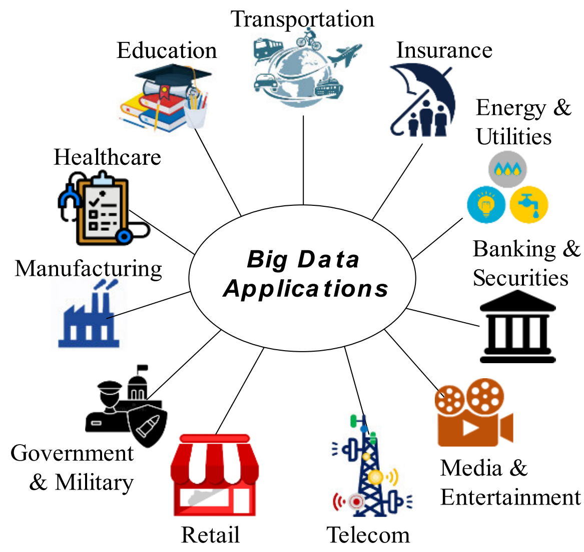

FIGURE 9. BD applications.

1) BIG DATA COLLECTION, PREPROCESSING, AND ANALYSIS TOOLS

Advancements in computing architecture are required to handle the data storage needs as well as the demanding server processing to economically analyze large volumes and varieties of data [159]. An overview of the technologies utilized in in-depth data analysis is given in this section.

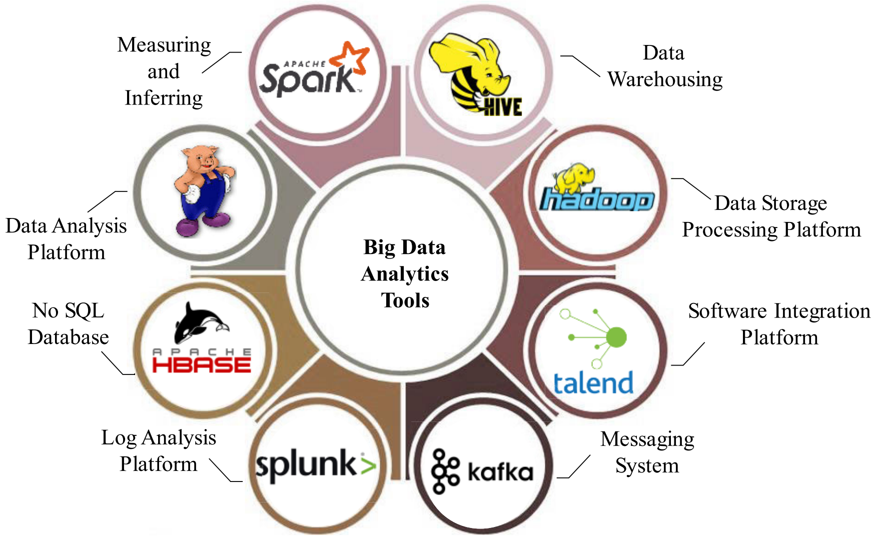

FIGURE 10. BDA tools.

a) **Not only Structured Query Languages (NoSQL):** Relational Database Management System (RDBMS) is a conventional technique for the management of structured data. This technique employs a relational database and schema to store and retrieve data. Storage and retrieval of large datasets are achieved by a Data warehouse. The most widely used database query language is Structured Query Language (SQL). Dimensional and normalized techniques are used to store data in a data warehouse. Data are separated into a fact table and a supporting dimension table when using the dimensional approach. However, when using the normalized approach, data is separated into entities, which results in the development of numerous tables in a relational database.
Scaling a significant amount of data is impossible due to the ACID constraint (atomicity, consistency, isolation, and durability). Additionally, RDBMS struggles to handle unstructured and semi-structured data. The concept of NoSQL was developed in response to the aforementioned RDBMS limitations.
NoSQL can be used in document storage, key-value storage, BigTable, and graph database. Compared to conventional methods, NoSQL utilizes looser consistency models. In the NoSQL database, data management and storage are divided. It enables data scalability. HBase, MongoDB, and Dynamo are among the examples of NoSQL databases.

b) **Hadoop:** Based on the Google File System and the Map Reduce programming paradigm, the open-source Apache Hadoop project was stablished and implemented in 2005. As open-source software, Hadoop can be employed for distributed storage and processing of ultra-large datasets. Some of its prominent features are summarized below: As open-source software, Hadoop can be employed for distributed storage and processing of ultra-large datasets. Some of its prominent features are summarized below:
- As Apache Hadoop is an **open-source** project, the user can modify its code depending on business requirements.
- **High data availability** of Hadoop despite hardware failure because of multiple copies of data. In the case of machine or hardware failure, data could be accessed from another path.
- **High scalability** of Hadoop makes it possible to easily incorporate the new hardware into the node. Hadoop also offers horizontal scalability (i.e. possibility to add nodes on the fly with no downtime).
- The **fault tolerance of** Hadoop can be assigned to the storage of 3 replicas of each block across the cluster. In the case of any node failure, data can be recovered from the other nodes.
- **Reliable data storage** on the cluster due to the replication of data on the other clusters.
- **Cost-effectiveness a** s Hadoop runs on a low-cost cluster of commodity hardware.
- **Easy use** of Hadoop as the client does not have to cope with distributed computing; the framework addresses everything.

c) **Hadoop Distributed File System (HDFS):** The distributed storage system for a Hadoop cluster that is fault-tolerant, scalable, and highly programmable is called HDFS. The data in the Hadoop cluster is divided up using HDFS. The Hadoop cluster's multiple servers are then used to spread the data. The dataset as a whole is only partially kept on the server.

d) **Hadoop MapReduce:** A software framework called MapReduce is fault-tolerant and reliable at distributing

TABLE 9. Capabilities of BD and their primary technologies.

Big Data Capability	Primary Technology	Features
Storage and management	Hadoop Distributed File System (HDFS)	Open-source distributed file system, Running on high-performance commodity hardware with high storage scalability and automatic data replication
Database capability	Oracle NoSQL	Dynamic and flexible schema design with high scalability, multi-node, multiple data centers, fault tolerance, ACID operations, operating based on high-performance key-value pair databases
	Apache HBase	Automatic failover support between Region servers, Automatic and configurable table sharding
	Apache Cassandra	Fault-tolerant for every node and column indices with log-structured updates and built-in caching
	Apache Hive	Query execution via MapReduce, Application of SQL-like language HiveQL, Feasible ETL process either from HDFS or Apache HBase
Processing capability	MapReduce	Distribution of data across thousands of nodes, Breaking down the problem into smaller subsections
	Apache Hadoop	Highly tunable infrastructure, Great scalability of parallel batch processing, and Fault tolerance
Data integration capability	Oracle BD connectors, Oracle data integrator	Exporting MapReduce outputs to RDBMS, Hadoop, and other targets, Supporting a Graphical User Interface
Statistical analysis capability	R and Oracle R Enterprise	Programming language for statistical analyses

TABLE 10. Shortcomings of HADOOP in BDA.

Limitation	Solution
Problems with Small Files	• Merging the small files to form a bigger one and copying that to HDFS. • File storage in HBase. We do not store millions of small files in HBase, as we rather add the binary contents of the file to a cell. • The use of sequence files in which the file name is taken as the key while its contents are regarded as the value. Using a program (100 KB), the files can be placed in a single Sequence file which can be processed in a streaming manner. MapReduce can break the Sequence file down into chunks and process them independently as the Sequence file is separable.
Slow Processing	• Spark can overcome this problem by in-memory processing of data. In-memory processing has a higher speed as no time is required to move the data/processes in and out of the disk. Spark has shown a 100-fold enhancement in processing speed compared to MapReduce. It also processes everything in memory. Flink is even faster than spark due to its streaming architecture. It can be programmed to only process a specific part of data with actual change, significantly incrementing the job performance.
Only supporting Batch Processing	• Spark can be employed to resolve this restriction and improve performance. The stream processing of Spark is not, however, as efficient as Flink since it applies micro-batch processing. By providing single run-time for the streaming as well as batch processing, Flink can enhance the overall performance. Flink uses native closed-loop iteration operators which make ML and graph processing faster.
Lack of Real-time Data Processing	• Apache Spark supports stream processing which involves a continuous data stream. It puts emphasis on the data speed and quick data. • Apache Flink offers single run-time for streaming and batch processing. A common run-time is employed for data streaming and batch processing applications. Flink is capable of real-time row-by-row processing.
Lack of Delta Iteration	• Apache Spark can resolve this issue and significantly enhance the performance of iterative algorithms by accessing data from RAM rather than disk. Spark can iterate its data in batches. Each iteration is separately scheduled and executed.
Latency	• Spark can be applied to solve this problem. Apache Spark is another batch system with a relatively higher pace as it can cache most of the input data on memory by RDD (Resilient Distributed Dataset) and maintain intermediate data in memory. Data streaming provided by Flink has shown low latency and high throughput.
No Caching	• By caching data in memory for further iterations, Spark and Flink improve the system performance.
Security	• Spark offers a security advantage in this regard. Spark uses HDFS ACLs and file-level permissions in HDFS. Moreover, Spark can run on YARN providing the possibility of using Kerberos authentication.
Infeasibility	• Spark offers an interactive mode thus developers and users can present intermediate feedback for queries and other activities. Spark is easily programmable thanks to its numerous high-level operators.
Long Code Lines	• Despite being written in scala and java, Spark and Flink are implemented in Scala, shortening the code lines compared to Hadoop. Therefore, the program can be executed faster.

the processing of enormous volumes of data. Map and Reduce phases are two separate parts. Map Phase is the place where the burden is divided into more manageable chunks. The Mapper, which processes unit blocks of data to create a sorted list of (key, value) pairs, is then given the jobs. The list, which represents the mapper's output, is handed on to the following stage. Another name for this action is shuffling. Data analysis and merging are done during the reduce phase to produce the final result that will be written to the cluster's HDFS. The capabilities of the current primary technologies and BD are listed in Table 9.

Although Hadoop is known as the strongest tool in BD, it suffers from several drawbacks. Hadoop does not suite for small files and fails to handle the live data. It also exhibits a slow processing rate and low efficiency for iterative processing and caching. The solutions to these limitations are listed in Table 10.

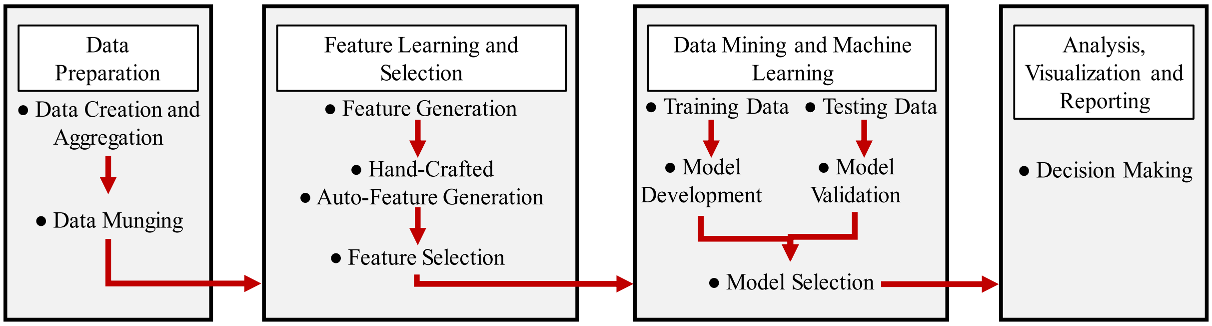

FIGURE 11. BD processes illustration.

2) TOOLS FOR HANDLING BIG DATA IN SOFTWARE

- **a) Python:** Python is a well-known open-source programming language that is supported by the Windows, Linux, and Mac operating systems. It contains a large number of packages from third-party or communitycontributed modules. Among the widely used packages for ML and DM are NumPy, Scikit, and Pandas support. The majority of scientific computations use NumPy. Python is able to support large, multi-dimensional arrays and matrices. Scikit can be used for preprocessing, model selection, feature selection, dimensionality reduction, clustering, regression, classification, and clustering.
  Pandas helps with data munging and preparing data for modeling and analysis. Thanks to its NetworkX library and nltk for text analytics and Natural Language Processing (NLP), it strongly supports graph analysis. Python is a user-friendly language that works well for rapid and dirty problem investigation. The pyspark library enables it to be seamlessly linked with Spark as well.
- **b) Scala:** As an object-oriented language, scala is an acronym for "Scalable Language". Similar to other object-oriented languages, its objects and operations are method-call. A Java virtual machine environment is necessary for Scala. Scala is used to create Spark, an inmemory cluster computing framework. As a solution to BD issues, Scala is becoming more and more popular.
- **c) R:** As an open-source statistical computing language, R offers diverse statistical and graphical techniques. R presents vector operations, efficient data handling, and storage capabilities, while enabling quicker processing. R provides conditional arguments, loops, and user-defined functions in addition to having all the properties of a typical programming language. R is supported by several packages and the Comprehensive R Archive Network (CRAN). R has good documentation and is accessible on Windows, Linux, and Mac platforms. It also robustly supports data munging, DM, and ML algorithms, in addition, to properly supporting reading and writing in the distributed environments, further promoting its suitability for BD handling. Nonetheless, speed, memory management, and efficiency are the major challenges in the application of R. R Studio has been built for programming in the R language. All browsers can access it, and it is released for standalone desktop computers and supports client-server architecture.
- **d) Apache Spark:** refers to a Scala-implemented inmemory cluster computing technique for fast computation. Apache Spark includes 80 high-level operators for interactive queries. Its Resilient Distributed Data (RDD) framework supports in-memory computation, which accelerates computation by breaking the data into smaller parts on different machines. In addition to providing SQL, data streaming, graph processing, and ML methods, it also offers Map and Reduce for data processing. Apache Spark is accessible through Python, Java, and R, while Scala offers stronger stability and support.
- **e) Apache Hive:** As an open-source platform, Apache Hive facilitates query and management of large datasets with distributed storage (e.g., HDFS). Due to its similarity to SQL, it is also known as HiveQL. It processes the queries by Map Reduce. Moreover, it enables programmers to insert their own custom mapper and reducer routines where HiveQL is unable to adequately represent the desired logic.
- **f) Apache Pig:** Is capable of analyzing large datasets. As a high-level programming language, it creates MapReduce programs that require Hadoop for data storage. The Pig Latin code can be extended by User-Defined Functions written in Java and Python. It is open to extensive parallelization, enabling the handling of very large datasets.
- **g) Amazon Elastic Compute Cloud (EC2):** Is a web service that provides access to CC resources. Because it offers total control over the computing resources, developers can carry out their work in any computing environment. It uses a pay-as-you-go business model and is one of the most popular CC platforms.

BD is also supported by other frameworks as BlinkDB, CouchDB, Clojure, MongoDB, Tachyon, Cassandra, Tableau, and Splunk. The following diagram illustrates the phases in BD processing. (Fig. 11).

D. THE PROPOSED INTELLIGENCE FRAMEWORK FOR ANALYSIS OF BIG DATA IN SMART CITIES

The developed framework is depicted in Fig. 12. As seen, this framework involved several iterative steps which formed the formulation of the statement of the research problem, followed by data entry and processing, attaining prominent visualizable insights to enhance the competitiveness of the companies in the analysis of BD, and develop smart applications and techniques. The applied results are evaluated in the final step, which led to the statement of a novel research problem. The research problem is first formulated to decide on the type of collected data and required analysis. In the next step (i.e. data entry), the potential social media platforms (e.g. Facebook, Twitter, and Instagram), and the data selection and sharing modes are determined. Diverse types of data including text, image, video, or sound, might be structured, semi-structured, or unstructured. Various tools can be employed for data collection from social media among which, direct databases access to the company's servers, questionnaires, Web crawlers, sanctioned API (application programming interface), and custom applications based on social media platforms can be mentioned [177], [178]. The data processing step involves the preparation, storage, and analysis of the data. The preparation process includes data extraction, purification, transformation, and loading [179]. This process leads to homogeneously structured data that can be readily stored in BD warehouses.

BD warehouses are utilized to store and process the BD and support the data analysis at various levels of detail. Their difference from traditional data warehouses lies in their support of online analytical processing. Moreover, BD warehouses are utilized by organizations for the collection of daily data and decision-making [180]. Hive is a BD warehouse that has facilitated querying and managing a huge deal of distributed data. It required no rapid change nor fast responses since it is located in a static data case. Moreover, NoSQL databases can be also regarded as a BD warehouse upon transforming traditional data models into data models of the NoSQL databases [181]. The analysis process is carried out after data storage. The social media contents are analyzed by data frameworks like Spark and Apache Hadoop [182]. This process uses algorithms and methods to obtain new insights from the data entry.

The textual data, in particular, can be better analyzed by ML and NLP [183] using diverse types of analysis including descriptive, predictive, diagnostic, and prescriptive analyses, simulation, and optimization. Furthermore, sentiment, trend, and network analyses can be employed in addition to market, financial, and crime analyses.

In this stage, the obtained insights can be visualized to further help the decision-makers. Visualization is an essential step. It can be employed as a presentation tool to illustrate and explain the results in diverse representations including charts and bars. Visual analytic techniques can offer a prompt and dynamic understanding of a huge deal of information in real-time. The findings of the previous stages will result in business development, enhancing the competitiveness of the cooperation in smart technologies. Finally, all findings should be evaluated, for better assessment of the decisions and discovery of emerging issues.

According to Fig. 12, the structure of BD in an SC can be grouped into multiple layers, i.e. data collection, data processing, data analysis, and data visualization to promote integrated BD management and SC technologies.

i) **Data Collection:** This stage refers to the collection of data from all available sources mostly by web crawlers, sensors, log files, and network monitoring software.
ii) **Data processing:** These collected data [184] might be inconsistent and noisy, requiring a pre-processing phase to enhance the quality of data to improve the analysis precision and decline the storage costs. The following steps can be adopted for pre-processing:
    - **Integration:** Through two conventional techniques, data federation and data warehousing, the data gathered from many sources can be merged in a more uniform framework. Data warehousing performs several processes including ETL. In the former process, the data are selected, gathered, processed, and analyzed; while their conversion into a standard format is accomplished through the transformation step. The transformed data are finally imported to a storage infrastructure in the Loading process. A dynamic data integration can be achieved by aggregating data from diverse sources by a virtual database that contains no data but the details of the original data (i.e. metadata).
    - **Cleaning:** In this phase, data are tested in terms of precision, comprehensiveness, and consistency. The data might be omitted or modified to enhance their quality. The cleaning process generally involves five processes: determination of the error types, identification of errors from data, correction of errors, documentation of error types and corresponding examples, and modification of data entry procedure for avoiding future errors.
    - **Redundant Data Elimination:** Most datasets contain surplus data or data repetitions, i.e. data redundancy. Redundance data enhance the storage cost, leading to inconsistent and poor-quality data. This issue can be resolved by diverse data reduction techniques like filtering and compression. These data reduction methods are, however, limited due to their high computational cost, requiring a cost-effective analysis prior to implementing data reduction.

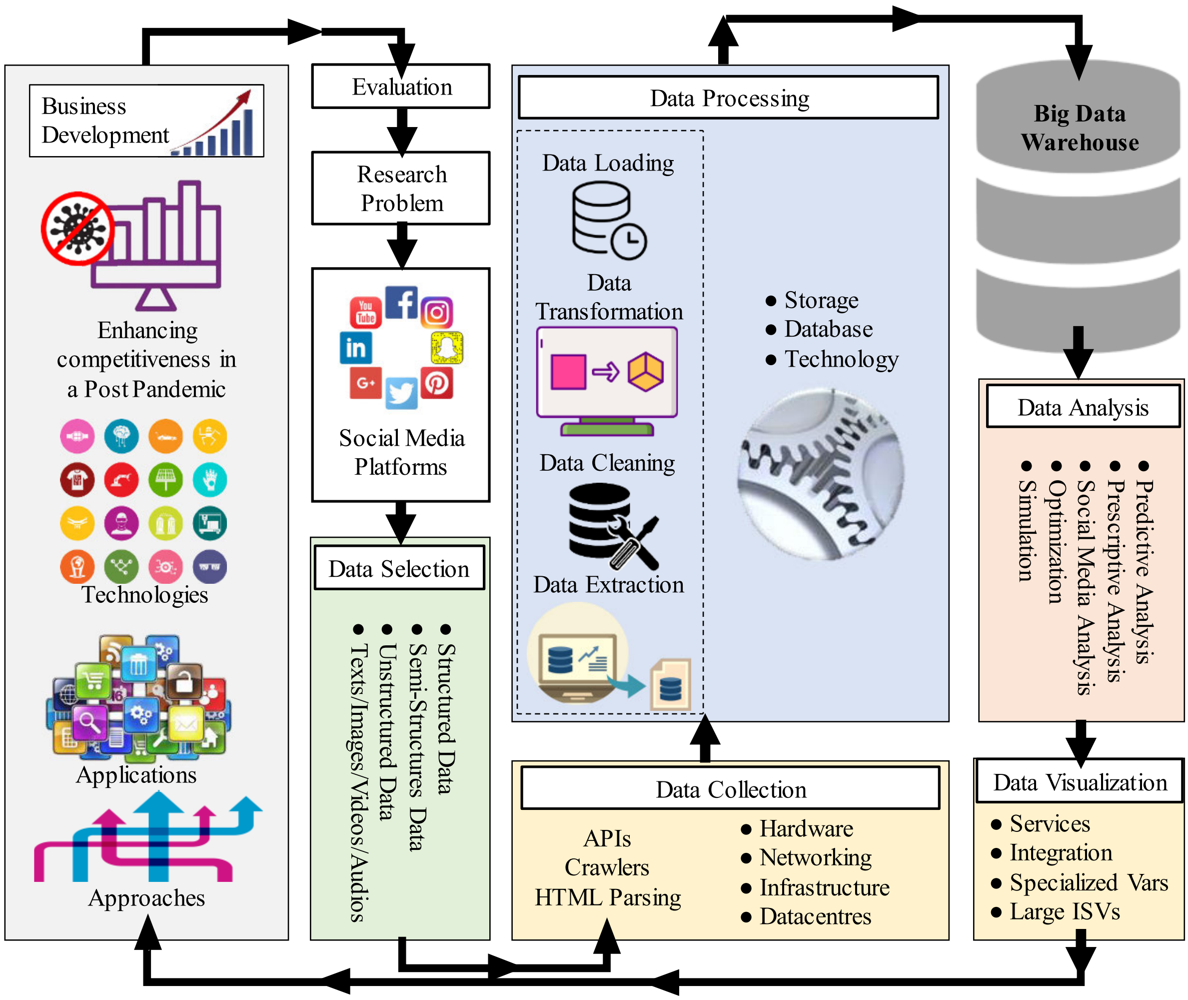

FIGURE 12. Advanced intelligence framework.

iii) **Data Analysis:** After data collection, transformation, and storage, the data are exploited or analyzed through the following stages [137]:

- **Metrics Definition:** A series of metrics can be defined regarding the problem specifications. For example, for identifying a potential customer who is going to churn out, the frequency of his/her contact (a voice call, tweets, or complaints on Facebook pages) can be taken into account.
- **Architecture selection based on the type of analysis:** An appropriate architecture can be chosen based on the timeliness of conducted analyses. Real-time analysis is suitable in cases with continuously-changing data which does not require rapid analysis for taking prompt actions. Among the architectures that are currently accessible are memory-based calculations and parallel processing. Real-time analysis examples include telecom fraud and fraud detection in the retail sector. Applications that don't need quick responses can be run using offline analysis. Additionally, the data might be taken, saved, and examined later. The Hadoop platform is frequently used.
- **Algorithms:** An appropriate technique can dramatically improve data analysis. BDA has shown promising outcomes for several traditional data analysis techniques like cluster and regression analyses and DM algorithms. As an unsupervised technique, cluster analysis classifies objects in terms of their specifications. On the other hand, DM can draw out unknown, concealed, and beneficial information from a sizable dataset. Various tools such as open-source and commercial software have been developed for data analysis. R, Weka/Pentaho, and RapidMiner are among the examples of open-source software developed for DM and visualization, ML, and predictive analyses, respectively.

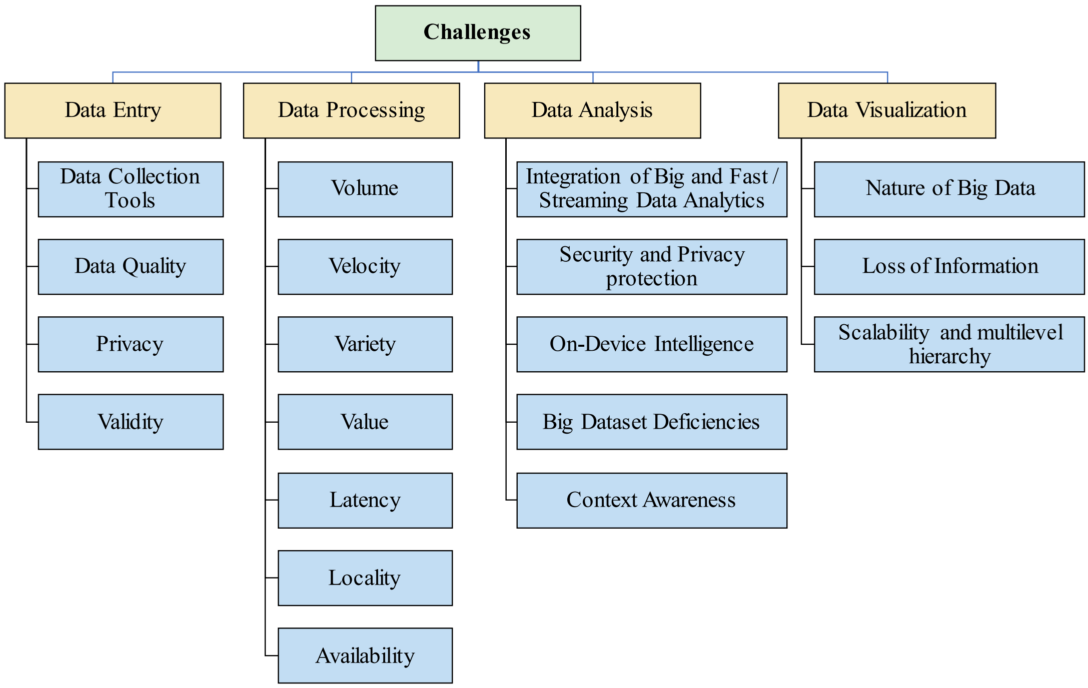

FIGURE 13. Challenges in the intelligence frameworks.

iv) Data Visualization: The data visualization originated from the need for detailed inspection at various scales. The temporal variation of the patterns of BD can be identified by the visual interface combined with statistical analyses and relevant context. Visual Analytics (VA) refers to analytical reasoning through a visual interactive interface such as Centrifuge, Tableau, VA, QlikView, Spotfire, JMP, Jaspersoft, Visual Mining, and Board.

E. CHALLENGES ASSOCIATED WITH THE PROPOSED INTELLIGENCE FRAMEWORK

Regardless of the sector, the industry must take three questions into account prior to BDA implementation, as follows: (1) Are valuable data gathered in addition to those from the current systems? (2) Does the use of BDA improve the accuracy of the information? And ultimately, (3) Does BDA promote timely response?

Nowadays, social media can be simply accessed throughout the globe in real-time. Over a billion users of various social media are continuously producing a large deal of data through their posts, likes, and comments. Such a huge deal of data has brought several opportunities and challenges. These data may be of various structures (unstructured, semi-structured, structured) and types (text, photo, video, and audio) [185]. Some of these challenges are listed in Fig. 13. These challenges can be classified into four groups: data collection, processing, analysis, and execution Challenges.

i) Challenges in Data collection: Several issues have to be addressed in data entry:

- **Data collection tools:** Regarding the wide diversity of data structures and types on social media, the development of effective data collection tools could be a huge challenge. For instance, API handling, the flexibility of the scaling process, real-time report, generalization, and reliability are among the major issues, requiring an effective solution [186] that can be properly addressed by constructing an efficient data collection tool for each social media platform. For instance, API could be a versatile solution for Twitter. In the case of Facebook, however, the privacy of users, long messages, and streaming data still challenge the implementation of an effective data collection tool.
- **Data quality:** The unstructured nature of data gathered from various users may decrement the data quality. Several issues like the difficulties of data integration as a result of the diversity of data origins, deciding on data quality during a reasonable time, and short timeliness of data because of their rapid changes, and advanced technologies required for data processing are some of these issues [187] which could be addressed by data cleaning and filtering [188].
- **Privacy:** The privacy of users is one of the prominent issues of social media platforms. Collection of the users' data ignoring the privacy concerns will result in an untrusted situation. Thus, a balance should be established between the use of personal data and privacy protection issues based on the

service type, personal data, customers, and regulatory rules. Several challenges like social identification and assessment of privacy-preserving services with real data are still open [105] which can be resolved by "k-anonymity" and "differential privacy" [22].

Wiretapping behavior is possible during data collection which may lead to information leakage. Numerous studies have been devoted to the confidentiality of wireless networks. Encryption schemes are applicable to wireless networks [76]. The energy and computational restrictions of smart objects have limited the application of these methods in IoT, requiring the development of novel lightweight protective schemes for IoT [189].

- **Validity:** Problems such as misinformation, fake accounts, and fake news threat the validity of social media data. The adverse effects of these issues on analytical processes may result in biased insights. Fake news can be classified by ensemble ML approaches based on effective and precise feature extraction [82].

ii) **Challenges in Data Processing:** Some of the challenges of the data processing stages can be assigned to the characteristics of BD. Data preparation, storage, and analysis may encounter several issues such as:

- **Volume:** Social media platforms generate a huge deal of data every day, for instance, over 500 terabytes of data (more than 2.5 petabytes/h) are added to Facebook just through Walmart customer transactions. Retrieval and processing of such a huge amount of data will challenge DM [179]. Furthermore, dimensionality reduction techniques like Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) can effectively decline the complexity of high-dimensional datasets in BD [1].
- **Velocity:** Social media platforms generate such a huge amount of data at high speeds, which may be challenging. Real-time analysis could be a proper alternative to batch analysis. The interconnected devices can share data, hence, increasing the velocity [190].
- **Variety:** Various structures of data (structured, semi-structured, and unstructured) with diverse formats are involved in BDA. These data are dissimilar following no specific formats (e.g., emails, tweets, images/videos/audios). Such a huge diversity and heterogeneity decrement the data quality, challenging their management and comprehension [15].
- **Value:** extraction of value from terabytes and petabytes of data through cost-effective and reliable methods could be a demanding challenge. It can promote the competitiveness of the firms and organizations on the global platforms [161].
- **Latency:** Latency plays an essential role in the performance of the system. Lower latencies reflect the prompt system response, which could be a serious challenge in large-scale data. It is dependent on the storage and organization of large-scale data, direct processing of the streaming data, and performing effective tasks of various priorities utilizing highly capable hardware and software [191]. Declined data complexity can lower the latency. Load-balancing techniques could be also helpful [192].
- **Locality:** BD systems involve data storage and distribution in different locations. The majority of ML algorithms assume that the datasets are placed in the memory, posing a serious challenge to the data retrieval process [15].
- **Availability:** ML algorithms are highly dependent on the availability of the dataset. In streaming data analysis, however, the data are not available as they arrive in a continuous manner [15].

iii) **Challenges in Data Analysis:** Data analytics offers an approach for presenting processed outputs of BD through visualization. In this regard, data analytics collects real-world data for digitization and analysis of the processes and environments, thus helping managers in the transformation of field data into meaningful information and supporting the decision-making process. As a versatile management tool, it can optimize the process and enhance the efficiency of diverse departments and industries.

The development of BDA-supported SC applications faces serious challenges which should be resolved to realize a reliable and precise system:

- **Integration of Big and Fast/Streaming Data Analytics:** The context of an SC involves numerous time-sensitive applications requiring real-time or near-real-time analytics of the data stream. The mentioned applications highlight the need for the development of novel analytic frameworks capable of supporting BDA in combination with fast/streaming data analytics.
- **Security and Privacy protection:** Data-driven ML techniques (e.g., DL) [193] are prone to attacks from False Data Injection (FDI), endangering the system reliability and trustworthiness. Thus, ML algorithms should be sufficiently resilient against these issues. On the other hand, privacy protection is a prominent issue as SC data are largely composed of users' information which must be protected against public availability. ML algorithms [1] can resolve these concerns resulting in extensive acceptance of SC systems among organizations and individuals.
- **On-Device Intelligence:** Light ML algorithms capable of being deployed on resource-constrained devices are highly welcome in hard real-time intelligence. It also coincides with the security and privacy

concerns since data will no longer be transferred to the fog or cloud.

- **Big Dataset Deficiencies:** Development and evaluation of SC applications require real-world datasets; these datasets are not, however, accessible for diverse application domains; necessitating result confirmation based on simulated BD.
- **Context Awareness:** The contextual information should be integrated with raw data to achieve more valuable, faster, and more precise reasoning and action. For instance, the detection of a sleepy face by a human pose detection system leads to fundamentally different actions in driving a car or relaxing at home.

iv) **Challenges in Data Visualization:** In the age of BD, data execution is critical for detecting hidden patterns and trends in massive amounts of data and converting it from useless to valuable and intelligible information. The following are some of the issues that large data visualization faces:

- **Nature of big data**: As discussed in Section VI, the visualization process could be a challenging task due to the large volume of diverse, heterogeneous, complex, and high-dimensional data. The majority of the current data visualization tools have shown poor functionality, scalability, and response time, necessitating the development of an efficient real-time interactive visualization tool to promote process flexibility.
- **Loss of information**: The approaches utilized to reduce the dimensionality of data may result in the loss of information, posing adverse impacts on data accuracy.
- **Scalability and multilevel hierarchy:** multilevel hierarchy method has been recognized as the main solution for various visual scalability problems. The major challenges of scalability analysis include scroll around a deep multilevel hierarchy and finding an optimal resolution [194].

F. BIG DATA MANAGEMENT IN SMART CITIES

The process of Big Data Management (BDM) is illustrated in Fig. 14. The challenges in the application of BD can be resolved by smart decisions. Diverse data from various sources (e.g. antenna, RFID, sensors, and social media) are collected, retrieved, and stored in clouds to decline the storage and operational cost in SCs. The privacy protection and security of BD should be also guaranteed. Nonetheless, the conventional approaches for the collection of massive datasets suffer from security deficiencies. Insufficient compliance policies often endanger data security and privacy intentionally or unintentionally. Thus, data protection is a must to decrease the risk of data theft and identity fraud to decline the possibility of financial or structural losses.

The massive collected data may lack basic structure which might be useful with efficient and timely processing. These problems can be resolved by diagnosing the data beforehand and converting them into viable data to promote effective decision-making.

A great deal of data is produced from transactions and stored as events in numerous files and databases. Unique data structures should be drawn from these events according to the proximity of data and users' demands. Thus, BD can be presented as accessible, smaller, and manageable data records with proper scalability and coherence. Filtering, sampling, binned aggregation, and model fitting are among the widely employed data reduction approaches.

The variety of BD may cause data inconsistency and non-aligned data structures, giving rise to analytic sprawl. Comprehensive monitoring and investigation of human societies result in heterogeneous data. This issue can be addressed by drawing structures from BD using tagcloud visualization methods to achieve meaningful data and take real-time actions. High-speed BD generation challenges data handling as new and/or updated data are constantly added to the datasets. Ineffective data processing approaches also challenge the real-time visualization of BD which can be resolved by employing human analysts to extract the relationships between the data.

BD veracity involves two aspects of data consistency and data trustworthiness. Ambiguity, incompleteness, and inconsistency can decrement the data quality. Uncertainty can be managed by several reliable sources, evaluating data credibility through diverse quantitative tools and identification frameworks. High-quality features enhance the value of BD by facilitating its utilization with improved precision and insights.

Maximal benefits can be achieved without compromising data authenticity by identifying the registered users capable of filtering the data to extract valuable information. The filtered data have to be structured in terms of geographic place and pattern homogeneity. In this regard, visualization techniques can help streamline the data and draw a structure to obtain meaningful information for more analyses. Diverse computational devices across the city collect timely information on traffic and weather which can be used in taking timely decisions to save time and costs. Devices such as safety gadgets, energy-harvesting wireless technological devices, and IoT sensors can enhance the QoL of the citizens of SCs.

V. CLOUD COMPUTING

SCs rely on various IoT devices and sensors to collect and analyze data about their infrastructure, environment, and residents. This generates large amounts of data that need to be processed, stored, and analyzed in real-time to derive actionable insights. CC provides necessary infrastructure to handle such massive data processing and storage requirement. Utilizing cloud services, SCs can access virtually unlimited computing power and storage capacity to handle their data requirements. In the context of SCs, CC plays a vital role in processing and analyzing the large volumes of data generated by various urban systems and devices, such as sensors,

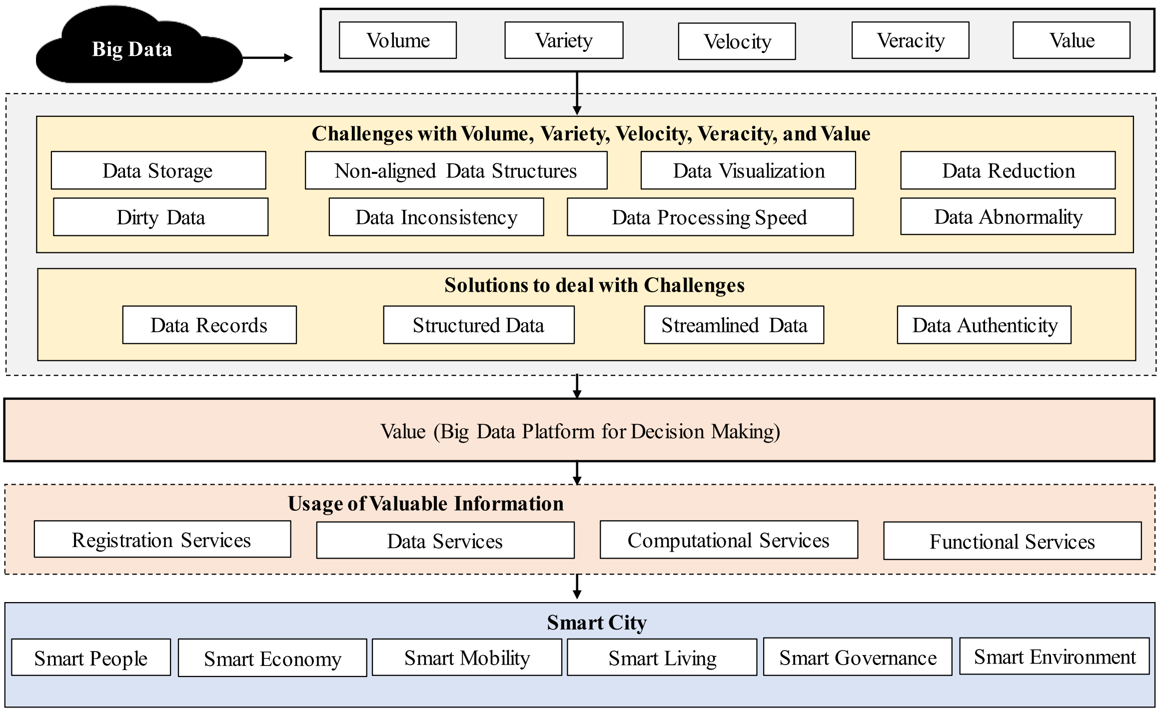

FIGURE 14. BDM in SCs.

cameras, and smart meters. By leveraging CC, SCs can optimize their operations, enhance their services, and improve the QoL of citizens.

CC refers to a computing model which provides on-demand access to a huge pool of network, storage, and computing resources through the internet [195]. It also declines the cost as customers do not have to apply hardware; they can rather register in computing resources from the Cloud Service Provider (CSP). In this way, they can access what they need and be charged only for services they used [196].

The underlying engine is also provided by CC through BD technologies such as the Hadoop framework [67] to establish a platform and programming model to process large datasets distributed over various clusters [197]. Hadoop includes two major and tightly-related components: HDFS and MapReduce [3]. Despite considering the real-time requirements of data storage and processing in SCs, streaming architecture guarantees efficient and intact communication between sensing devices. This technology has been lately implemented upon the utilization of diverse stream processing platforms, like Apache S4 [133], Storm [198], and Spark streaming, enabling data storage and processing through different interconnected nodes [23], [154].

A. CLOUD COMPUTING ARCHITECTURE

A carrier management environment is provided by CC infrastructure in addition to the integrated system of physical computing and its processes for performance monitoring. Secure operations are provided by CC architecture in an effort to safeguard sensitive data [71].

Khan et al. [199] described the capabilities necessary in a cloud environment to collect integrated intelligence for urban management systems. By creating a context-aware component in a cloud environment, these capabilities offer a solid foundation for meeting the demands of SCs.

The proposed architecture, as shown in Fig. 15, can be utilized to create a Platform-as-a-Service (PaaS) or a software-as-a-service solution. For instance, when it is used to create modelling services, it transforms into a PaaS that additional services can be built on top of that. It turns into a software-as-a-service solution when used to provide a service that directly engages users, as through visualization. The architecture shown in the figure is principally composed of two vertical layers and five horizontal layers. The Platform Integration, Thematic, and Data Acquisition and Analysis levels of our bottom-up strategy produce general data that may be adapted to particular application demands connected to SCs in the top three layers. The goal of the following architecture walkthrough is to demonstrate how each layer contributes to giving end users contextual information for a variety of objectives.

- The platform integration layer illustrates a hybrid cloud environment-based cyber infrastructure that ensures data is accessible across platforms.
- The data acquisition and analysis layer are utilized to acquire environmental data from a variety of sources, such as remote database repositories, sensor nets, and

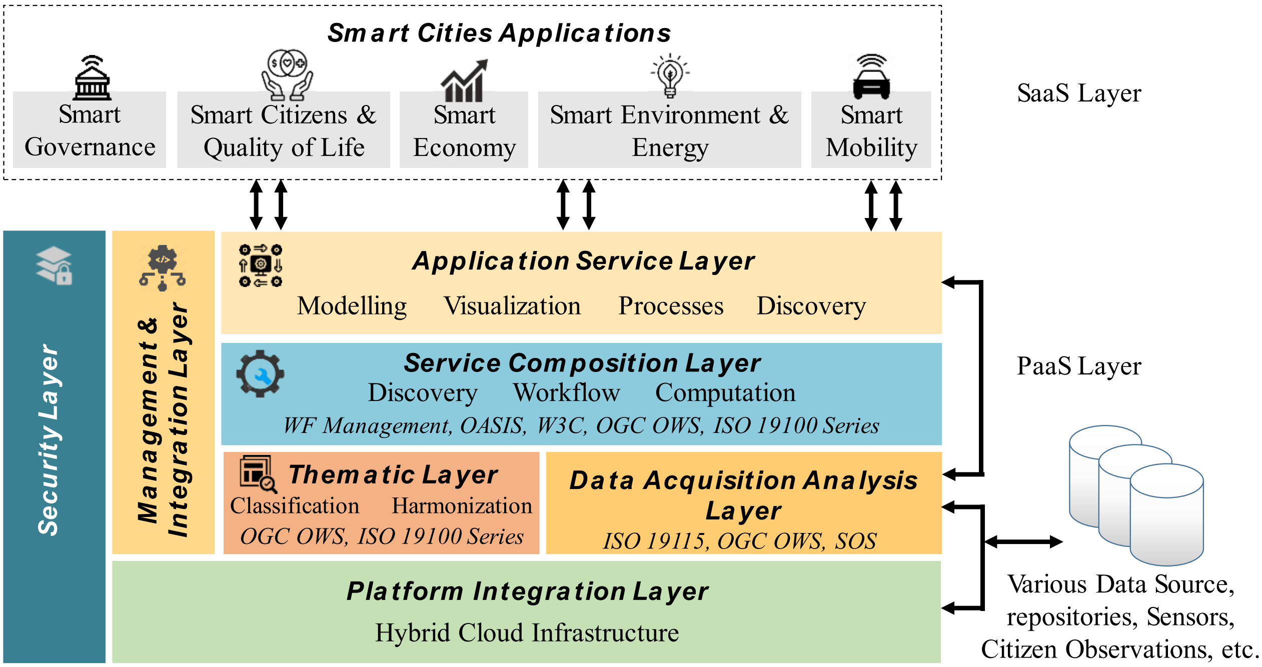

FIGURE 15. Architecture of CC services for SCs Infrastructure.

citizen observations, e.g., using smart phones in the cloud environment. Additionally, this layer guarantees the accuracy of the acquired data and reveals the need for data harmonization and cleaning.

- The thematic layer organizes the collected data into application-specific thematic categories, harmonizes the data, and updates the data/service catalogues to enable further data utilization.
- The service composition layer is required to create work-flows, identify data sources, and link relevant processing components for implementing the workflows. Furthermore, this layer can be used to perform any necessary analytical analyses of the outputs of the work-flow. Moreover, this layer makes sure that the provenance of data and certain processes is preserved so that they can be used by various expert systems at the application layer for analysis.
- The application service layer performs contextual analysis for decision-making by using the results from the service composition layer in application domain-specific tools like simulations and visual maps. Moreover, this layer offers stakeholders the ability to utilize current tools and create new application domain-specific components and services (at the SaaS level) to meet the contextual needs of the end users.
- The management and integration layer is implemented to automate the transfer of filtered data and information between the horizontal layers. It makes sure that the processed outputs from one layer to the next are sound both grammatically and contextually. Additionally, it seeks to lower the amount of administrative overhead required by the layered design and handle change management that happens at various tiers.
- The security layer guarantees that the appropriate auditing, authentication, and authorization are carried out before data and services are used by authorized users. Additionally, it guarantees that end users' personalization and profiling are secured to analyze and retrieve contextual data from a cloud environment.

The suggested architecture promotes processing at many stages while ensuring the delivery of contextual information to end users.

B. CLASSIFICATION OF CLOUD STORAGE

Cloud storage is an Internet-based storage system that involves sending data to a distant storage system, where the IoT-generated data are kept, preserved, controlled, backed up, and accessible via the Internet. Customers are frequently billed monthly based on their consumption. Database services, computer services, and storage services are some of the main cloud-based services [88].

CC can be categorized into four deployment models: Public, private, hybrid, and Community data storage schemes. Table 11 compares the benefits and drawbacks. As inferred by its title, an individual organization possesses and manages a private cloud; while cloud infrastructures of a community cloud model can be owned and shared by different organizations with similar operations. The hybrid cloud refers to a combination of two or more clouds either public, private,

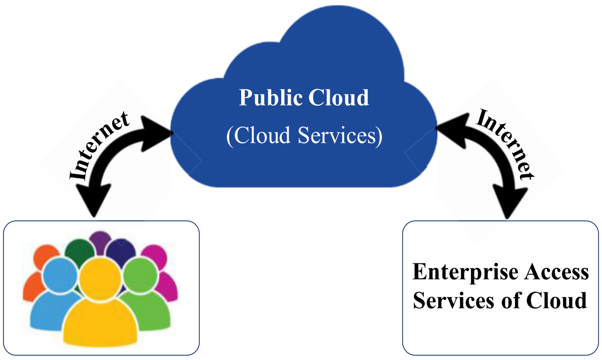

FIGURE 16. Public cloud storage.

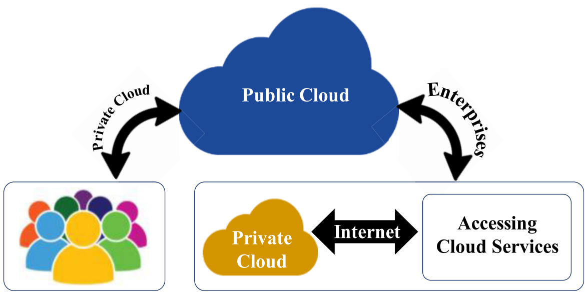

FIGURE 17. Private cloud storage.

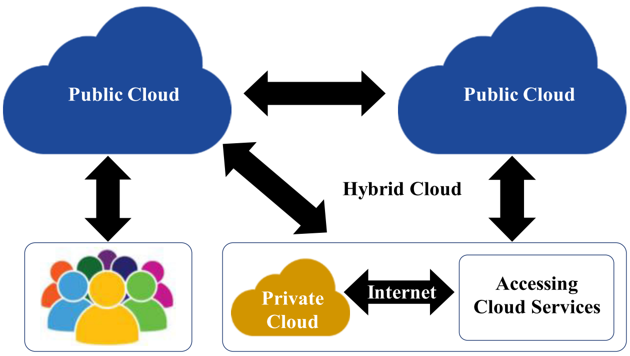

FIGURE 18. Hybrid cloud storage.

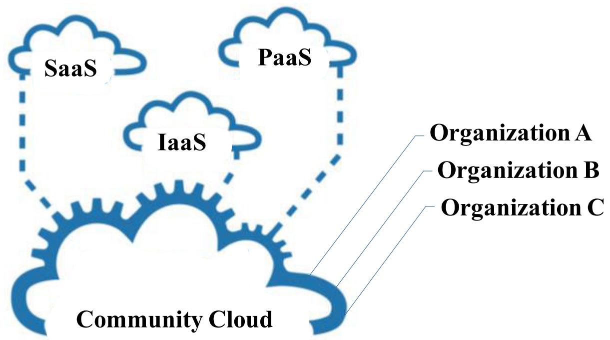

FIGURE 19. Community cloud storage.

or community models which remain distinct but may be joined through uniform technologies to facilitate the mobility of data and applications. The cloud infrastructures of the public cloud model are accessible for the public based on pay-as-you-use principles using the CSP [86], [88].

- **Public Cloud:** The public cloud, also known as SaaS, is created for open use by the entire public. This type of cloud storage can be owned, maintained, and operated by corporation, municipal, or academic organizations. Pay-per-use data storage is also offered by it. According to Fig. 16, it is possible to implement the communication linkages via the open Internet. Through the open Internet, any user, including customers, individuals, or companies, can use cloud services. Microsoft Windows Azure, Google App Engine, IBM Smart Cloud, and Amazon Elastic Compute Cloud (EC2) can be mentioned as examples of public cloud storage [200].
- **Private cloud:** An individual user's data are saved on distant servers under this expandable and independent storage arrangement, as shown in Fig. 17. Private cloud storage is safer than public cloud storage since it can be located on business property or in the data center. Two sub-categories of private clouds are distinguishable: (1) On-Site Private Cloud which is implemented at the location of the user. (2) Outsourced Private Cloud which is utilized at the server-side by a hosting business.
- **Hybrid Cloud:** Hybrid cloud method is one of today's most popular infrastructure configurations, which allows to use both your on-premises servers and public cloud services like Google Cloud. A combined computing, storage, and service environment known as a hybrid cloud consists of on-premises infrastructure, private cloud services, and a public cloud (such as Amazon Web Services (AWS) or Microsoft Azure) with orchestration between the various platforms as represented in Fig. 18.
- **Community Cloud:** Community CC helps users better understand and analyze their business requirements. Community Cloud may be hosted on-site or off-site in a data center that is operated by one of the tenants or by a different cloud services provider (Fig. 19). The U.S.-based dedicated IBM SoftLayer cloud for federal organizations is a notable example. In essence, this platform can only be used by governmental institutions that share common traits like security, auditability, and privacy concerns or requirements. Such a strategy greatly boosts the platform's credibility; which cloud users will employ to deploy their delicate workloads.

C. CHALLENGES IN CLOUD COMPUTING-BASED SMART CITIES

CC has several advantages as it allows the establishment of a virtual office; hence the user can access business files anywhere, anytime.

Overall, challenges related to the use of the cloud have been divided into eight primary categories: data

TABLE 11. Type of cloud storage and their advantages and drawbacks.

Cloud Type	Advantages	Drawback
Public Cloud	• The service provider is in charge of setting up and maintaining the infrastructure for engineering support and storage. • The cloud environment is under provider control. As a result, the workload or data of the subscriber could be moved at any time. • The workload might be displaced to computer servers with reduced operating costs. • Public clouds are benefited from high flexibility. • An appropriate option for small and medium-sized businesses.	• Increased security threats as a result of the public domain and providers' full control.
Private Cloud	• The user has full control over his/her data. • Great security, scalability, and reliability. • Appropriate for large enterprises or organizations.	• More costly than public cloud storage.
Hybrid Cloud	• Ease of customization. • Safe and reliable. • Increase workload efficiency.	• Excessively difficult security. • Visibility issues. • Hardware costs.
Community Cloud	• The ability to change the properties to meet the needs of the user. • Businesses can use a variety of heterogeneous devices and communicate with their distant workers. • Offering block facilities, which allows service providers to prevent users from adding, removing, changing, and downloading particular data sets and services.	• Security issues since several entities can use the infrastructure and exert control over it. • The availability of bandwidth and storage space might be an issue when numerous firms share the same resources. • Community cloud deployment is still less common than other deployment models.

management and resource allocation, load-balancing, migration to clouds and compatibility, security and privacy, reliability and availability, scalability and elasticity, virtualization,interoperability and communication between clouds. Regarding conceptual underpinnings, each of these problems can affect the dependability and effectiveness of cloud-based platforms.

- **Data management and resource allocation:** Resource allocation is one of the key ideas in CC data centers. The significance of this concept can be defined regarding the abundance of resources in cloud-based contexts. Consequently, the process of resource allocation should satisfy the needs of the network QoS, remove performance issues without significant rise of the cost of the service provider, and control energy usage [201]. Data center network resources, data center processing resources, and energy-efficient data center resource allocation are the three main categories under which resource allocation challenges are categorized.
  The first section is concerned with allocating resources in accordance with user requirements. As a result, the processing throughput in cloud-based data centers should be as low as feasible [202], and the distance between the locations of virtual machines in cloud data center should be reduced by dispersing virtual machines [203].
  Optimizing virtual network provisioning while maximizing revenue is the most difficult problem in data center network resources. In addition, choosing the best virtual network using IP over a cloud-based network is a significant issue that has been taken into account in relation to propagation, delay, and flow conversion constant [204].
- **Load balancing:** Load balancing is a crucial problem in CC context which affects both download speed and storage utilization. The major goal in this situation is to create an algorithm for effective allocation of work to the cloud nodes in accordance with current constraints for instance, heterogeneity and high communication delays [205]. The challenges and issues associated with load balancing are typically divided into four main categories [205], [206]: spatial distribution of the cloud nodes, data replication, performance, and point of failure. The management of the load balancing process among all spatially dispersed cloud nodes is the initial problem. This management should consider the delays arisen from the speed of network connectivity between cloud nodes, the distance between service nodes, and the distance between clients and task processing nodes [207].
  The second issue is the method of data replication (i.e. full and partial) across several nodes, which is complicated by the load-balancing algorithm across various cloud nodes for partial replication and more storage requirements for complete replication. The performance of the load-balancing algorithm is the next problem.
  The complexity of the designed algorithm should be as low as feasible due to the nature of load-balancing procedures to prevent delays and faults during

complex processes. The capacity of load-balancing algorithms to resist regular or unanticipated failures is the final obstacle. In cloud-based environments, managing and minimising failures caused by the complexity of load-balancing algorithms is a challenging issue that has attracted attention of academic communities.

- **Migration to clouds and compatibility:** Traditional IT suppliers have been urged to shift and adapt their products (such as traditional applications, operating systems, and middleware) to cloud-based environments due to the rapid expansion and popularity of CC among users and businesses. However, the likelihood of success in this migration process is low issue due to the current restrictions on traditional IT solutions. An IT product should meet five key requirements to increase the rate of success in the migration process [208]:
  1) **Modularity:** The ability to increase or decrease computing power based on identical and replicated components on several virtual or physical nodes.
  2) **Backward Compatibility**: The ability for ongoing interaction between components when they are accessible as representational state transfer URLs.
  3) **Portability**: The ability to integrate components into different Infrastructure-as-a-Service (IaaS) infrastructure.
  4) **Scalability:** the ability to coordinate multiple tasks.
  5) **Changeability**: the ability to properly modify an IT product in the cloud environment.

  Migration to cloud-based services provides many advantages for both users and businesses, including increasing job satisfaction, appropriate modification of organisational growth and the development of new talents, and lowering monotonous labour. However, some significant issues need to be addressed during the migration process, such as the growing reliance of businesses on external third parties, departmental reductions, the lack of knowledge of cloud capabilities and structure, the lack of supporting resources, and concerns on new technologies [209]. Therefore, businesses and service providers should take into account all potential outcomes and salient features both during and after the transfer process.
- **Security and privacy:** In CC environments, protecting the security and privacy of stored resources is one of the most difficult problems which can affect the reliability and efficiency.The main factor obstructing the development of CC is security, which has emerged as a key problem in industry and academic research [210]. Security concerns in cloud-based environments can be divided into three categories: attack vulnerability, best practises for security, and compliance with local, state, or federal regulations governing the storage of personal data or laws governing record-keeping. These problems raised significant concerns on surface at many levels, including those of service providers, infrastructure, and end users. Two popular methods for improving the dependability of authentication processes are the application of agent-based authentication systems [211] and multifactor authentication processes [212]. Additionally, researchers and service providers consider how to ensure the security of resources during transmission operations between providers and tenants.

  The ability of cloud-based servers to resist potential attacks or unforeseen occurrences is the other crucial security concern. A significant resistance could raise the rate of reliability in CC environments and improve the data security process of the service providers. Real-time cryptography algorithms are the best way to improve defences against potential assaults or unforeseen circumstances [213].
- **Reliability and availability:** The strength of any technology is determined by its reliability and accessibility. The frequency of failures and interruptions (such as data loss or code resets during execution) is referred to as reliability. Downtime is a significant factor that gravely undermines the dependability of CC. Reliability can be attained in part through the use of redundant resources. Considering the time for providing these resources, availability can be defined as the potential of accessing the resources whenever they are needed. The services in CC are susceptible to denial-of-service attacks, performance slowdowns, equipment failures, and natural disasters even when they use architectures with excellent reliability and availability characteristics [88].
- **Scalability and Elasticity:** A difficult problem in providing cloud-based services is the ability to adjust cloud capacity to on-demand services under diverse workloads such as static, periodic, once-in-a-lifetime, unpredictable, and continuously changing ones. The absence of this capability may result in performance reduction (during the peak of workload) or even oversizing (during the bottom of workload) when on-demand services are provided [214].

  Elasticity and scalability are among the most impressive and distinctive aspects of CC. These capabilities enable customers to employ cloud resources that are provided according to their needs in an unlimited number of instances as needed. Scalability can be defined as the ability of the system to properly function even when the resources have been scaled up. On the other hand, elasticity refers to the capacity to adjust resource levels as needed. However, elasticity goes a step further and permits the dynamic integration of the infrastructure and extraction of physical resources. The elastic nature of CC allows resource allocation to change in size in response to demand [215].
- **Virtualization:** The process of creating a virtual copy of a server, an operating system, a storage device, or other network resources is known as virtualization. To make it easier for other systems, applications, or end

users to engage with those resources, the virtualization process divides the resource into many execution contexts and conceals the physical properties of computing resources. Virtualization is one of the essential technologies in the CC implementation. Customers can often relocate their computation and data to a remote site using virtualization, with different effects on performance. Virtualization offers many advantages such as elasticity and scalability, cost efficiency, infrastructure independence, customizability, and streamlined access interface. However, while virtualization offers many advantages to users, it also presents numerous difficulties for CC. Infrastructure management becomes increasingly difficult to manage as a result of virtualization. Enormous automation is needed to meet important needs like automation, on-demand, and elasticity requirements [88], [216].

- **Interoperability and communication between clouds:** A significant problem brought on by the different approaches and organisational structures of different cloud vendors is the lack of interoperability.
  Inter-cloud interoperability can be established at four different levels [217]:
  1) **Agreement**: Employing a specified data format, communication protocols, or a standard reference model for information sharing.
  2) **Deployment**: Establishing interoperability for operations at the same or various deployment levels, whether horizontally or vertically.
  3) **Interaction:** Real-time services that require quick responses may use synchronous patterns, or employ asynchronous patterns with less interaction between them.
  4) **Adaptation:** Utilizing the same standards to produce goods that reduce service providers' ability to compete.

Achieving interoperable CC environments is more challenging as each level of interoperability has some limitations.

VI. INTERNET OF THINGS (IoT)

IoT has dramatically altered our world. IoT offers the possibility of collecting data on the surrounding physical environment which can result in invaluable knowledge on the underlying mechanisms of the world. These data can be employed for the development of new real-world applications, facilitating smart decisions for the improvement of the QoL. The concept of an SC could be a proper instance to illustrate the use of this valuable information. The complexity of the IoT paradigm can be mainly assigned to the huge deal of data collected. These data mostly require processing to be used as useful knowledge [25].

IoT can be defined as an internet connection between objects through sensors and actuators, hence, representing a bridge from technology to the physical environment. The surrounding objects will become a member of a network [1]. IoT is a global network of interconnected objects with unique addresses which operated according to standard communication protocols. Its convergence can be considered the Internet. The objects serve as sensors and can achieve a common goal through interaction with each other. IoT can drastically affect the daily lives of its users.

IoT can significantly influence occupational and domestic scenarios as it will soon play a vital role in assisting people's lives, health, and smart transportation. IoT can also contribute to logistics, industrial automation, freight transportation, and security. IoT encompasses the tiny real world with limited storage and processing capacity, raising serious concerns about performance, security, and privacy. Some of the major aspects of IoT are discussed in the following.

A. IoT ARCHITECTURE

No generally agreed consensus can be found on IoT architecture as different researchers have proposed various architectures.

IoT has offered a prominent and innovative platform for communication between digital devices, as well as their connections with the external environment [218] and users. IoT architecture integrates and unifies data sensing/actuating from/to devices, transmitting/receiving messages, data storage, processing, analysis as well as final exploitation through the use of cloud, fog, and edge computation, services, and applications. Various technologies contribute to the development of IoT frameworks. Nonetheless, several broad functional architectures have been suggested based on a condensed version of the Open Systems Interconnection (OSI) concept. A typical IoT framework has been defined which comprised three layers [93], [219]: the application layer; the network layer and the perception/sensing layer. The earlier approach was expanded in certain other research [220], [221], [222], [223] into five layers of architecture, including the perception/sensing layer, transportation/network layer, middleware/processing layer, application layer, and business layer. Both the three-layer and five-layer IoT designs are depicted in Fig. 20.

Various computing paradigms can be defined based on the considered level of the stack including EC, Fog Computing (FC), and CC. The following list summarizes these various levels while taking into account their uses and necessary technology. This article considers the five-layer architecture, which can be seen as an expanded version of the three-layer design.

- **Perception/Sensing Layer:** The physical level, which consists of sensors and actuators, interacts with the outside environment through data sending and receiving across wireless networks. This is referred to as the perception or sensing layer [15]. This layer requires mobile devices, sensors, and actuators technologies. Diverse commercial devices are available to evaluate various physical quantities such as temperature, humidity, pressure, speed, acceleration, distance, geographical

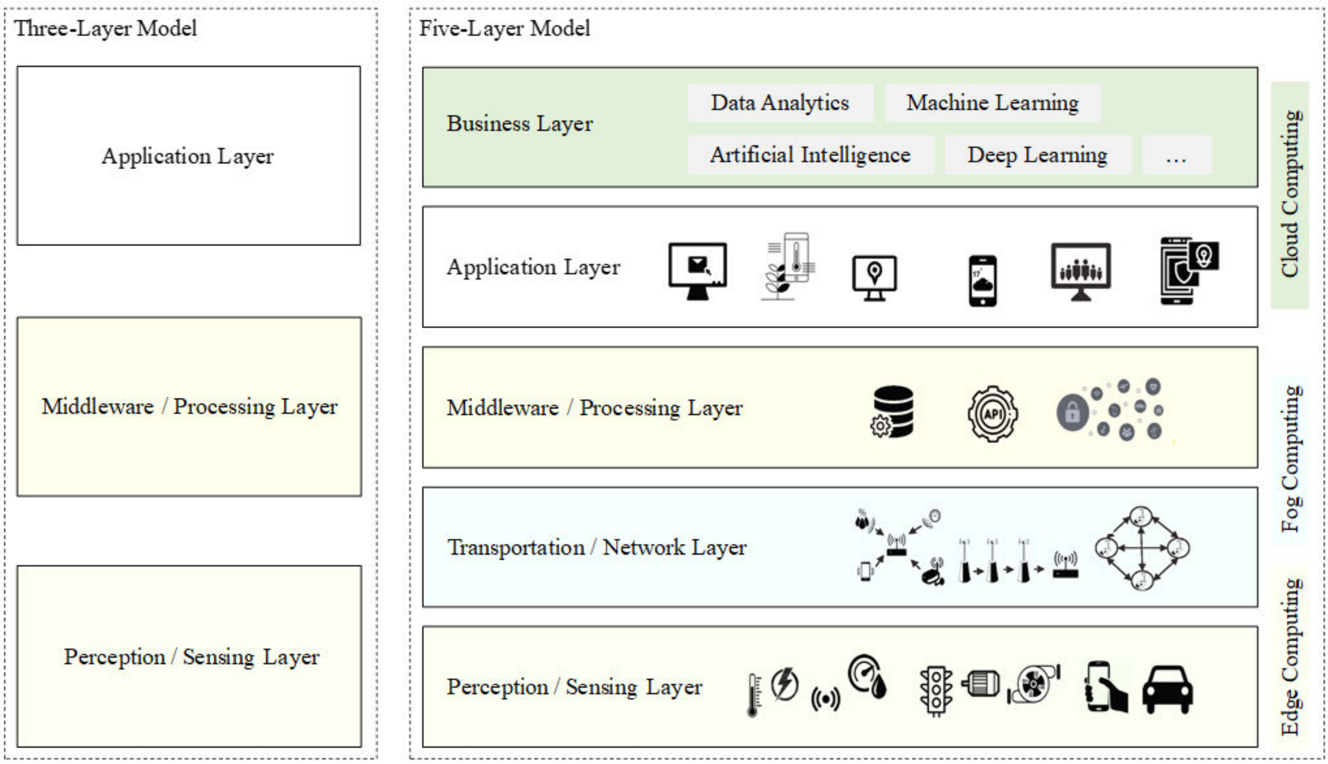

FIGURE 20. An overview of IoT–based architecture for SCs.

coordinates, weight, pollution particles, and brightness level. Actuators are responsible for physically or virtually controlling or displacing other devices [224].

- **Transportation/Network Layer:** Data transmission and routing take place at the transportation or network layer. Network gateways and brokers can act as intermediaries to connect various IoT nodes, allowing data to be sent to and received from sensors for M2M communication. For data transfer, numerous methods and protocols have been adopted. As examples of proximity communication protocols, we can name Bluetooth, radio-frequency identification RFID tag technology, and Near-Field Communication (NFC). Networks covering larger ranges use wireless technologies, such as WiFi, Zigbee, Long-Range Wide-Area Network (LoRaWAN), Sigfox, and 5G.
Personal Area Networks (PANs) have been created for low-power wireless communication technologies like Bluetooth and Bluetooth Low Energy [225] which are suitable for low-bandwidth data transmission among mobile devices across a short range of the order of 10 m. To connect personal gadgets, these technologies are frequently used in intelligent environments.
RFID utilizes radiofrequency to uniquely detect objects, vehicles, and people. The reader-tag device communication could be passive or active. Despite its similarity to RFID, NFC [226] utilizes even shorter ranges for communication (several centimeters). NFC is useful in two-way communications and does not build a reader/tag hierarchy, in contrast to RFID. NFC is often employed for mobile payment and checking access.
Wi-Fi is based on the IEEE 802.11 standard and makes use of wireless frequencies (at 2.4, 5, and 60 GHz bands) to provide high-speed internet connectivity with a coverage area of up to 100 meters [152]. For Wireless Local Area Networks (WLAN), there is Wi-Fi (WLAN).
An open international standard known as ZigBee (IEEE 802.15.4) is used for low-cost, Machine-to-Machine (M2M) and IoT enabled wireless network communication (WNS). The ZigBee protocol was developed to transport data in noisy radio frequency (RF) environments, such as those seen in commercial and industrial settings [227]. The range of its data transmission is similar to that of Wi-Fi (from 10 to 100 m).
LoRaWAN is a Low-Power Wide-Area Network (LPWAN) that plays a pivotal role in IoT applications, owing to its capability to meet the key IoT requirements (e.g., long range transmission (about 10 km), low cost, small data volumes, massive device numbers, and low energy consumption) [228].
As a narrowband communication system, Sigfox is used for long-range data transmission (up to 40 km [229]). Despite using narrowband signals, Sigfox suits for diverse applications including geolocation services.

5G refers to the fifth generation of cellular networks. It will provide unprecedented levels of connectivity compared to earlier generations of mobile communication technology in terms of higher data rates, traffic density, reliability, and user speed, as well as lower latency, connectivity for a large number of devices, and lower power, enabling IoT applications [137].

The last network layer stack ensures that data is formatted correctly. The client/server architectures offer a variety of message/data delivery methods, including push and pull. Web services, FTP (File Transfer Protocol), HTTP/HTTPS, and REST calls are frequently used in pull protocols.

The most popular push protocols for receiving data via data-driven subscriptions include WebSocket (WS), Constrained Application Protocol (CoAP), Message Queue Telemetry Transport (MQTT), Advanced Message Queuing Protocol (AMQP), FIWARE NGSI, and NGSI-V2On [230]. Moreover, data exchange among multiple devices and applications can be achieved by distributed publish-subscribe messaging system (like Apache Kafka, RabbitMQ) to efficiently and scalably handle multiple data streams. The data persistence of queue (like Apache Kafka) gains increasing significance in fault tolerance during handling of data that should not be lost in case of failure.

- **Middleware/Processing Layer:** This layer offers various functionalities. Since the data may be obtained from heterogeneous devices using different protocols with no contact or interaction, it may, for instance, function as a data aggregator module. In this regard, the middleware layer should promote interoperability among the connected devices and offer required programming and/or model abstractions. Additionally, interoperability must be ensured at several levels:
  - **i)** Technical level [231], for efficient and reliable end-to-end communication between devices, gateways, and brokers
  - **ii)** A level of syntax to accommodate many protocols and formats;
  - **iii)** Semantic level, to exploit SemanticWeb technologies, such as XML, Resource Description Framework (RDF), Web Ontology Language (OWL), and Linked Data (LD), to obtain clear data representation and data semantic enrichment, and improve the expressiveness of the system.

  IoT middleware can also provide scalability and reliability, enabling the systems to cope with an increasing number of IoT connections at the scale of BD and supply stable and fault-tolerant services. Additionally, to integrate the data from different protocols and enable data processing through IoT, database-level data persistence is also necessary. This goal can be achieved at the middleware layer. As a result, middleware encompasses diverse data storage modalities and thus covers other functionalities, such as context identification, information extraction, and the reconciliation of collected data.
- **Application Layer:** The application layer, which provides the user with application-specific services, is where user interaction occurs. The formatting and presentation of data fall under the purview of the application layer. Typically, HTTP serves as the foundation for the Internet's application layer. However, due to its relatively verbose nature and high parsing overhead, HTTP is not appropriate in contexts with limited resources. Besides, various frameworks and ecosystems have been utilized in the design and implementation of event-driven IoT applications such as Visual Programming Language (VPL) tools [101]. Node-RED is one of the most applied applications relying on the Node.js engine which permits application flow in a graphic environment by a combination of visual nodes or blocks.
- **Business Layer:** The business layer is also in charge of service management for IoT devices as well as the overall management of IoT applications. The information processed at the application layer and the analysis of the operations' results are used by the business layer to develop a business model [232] For business intelligence, the business layer often uses reports and real-time dashboards. Through additional integrations, data gathered from the application layer can be further enhanced. For instance, by comparing electricity consumption before and after the installation of smart lighting sensors, business intelligence analysts can determine cost savings [194].

B. CHALLENGES IN IoT-BASED SMART CITIES

This section addresses the advances, usage, and technologies in the field of IoT-based SCs. Accordingly, several applications of IoT-based SCs may raise various limitations such as security, privacy, heterogeneity, reliability, large scale, BD, and sensor networks as illustrated in Fig. 21.

- **Data Privacy and Security:** Security and privacy issues often emerge in the case of an assortment of data in a general platform that might be prone to attacks. IoT depends on the interconnectedness of diverse devices connected across wired or wireless networks to send data to service providers. Security is a major concern with regard to its wireless data gathering and transmission on both the server and client sides. Upon real-time data transmission, attacks may be also activated, necessitating additional care to prevent stealing of data and restrict unauthorized access. Users of some smart systems are continuously generating data (for instance, healthcare systems); in other words, all the information of the users is present in such systems. Therefore, user identity and data privacy should be protected against violation. Thus, only doctors and individuals should have access to modify information. In this regard, the security of the infrastructure is a key factor in the development of

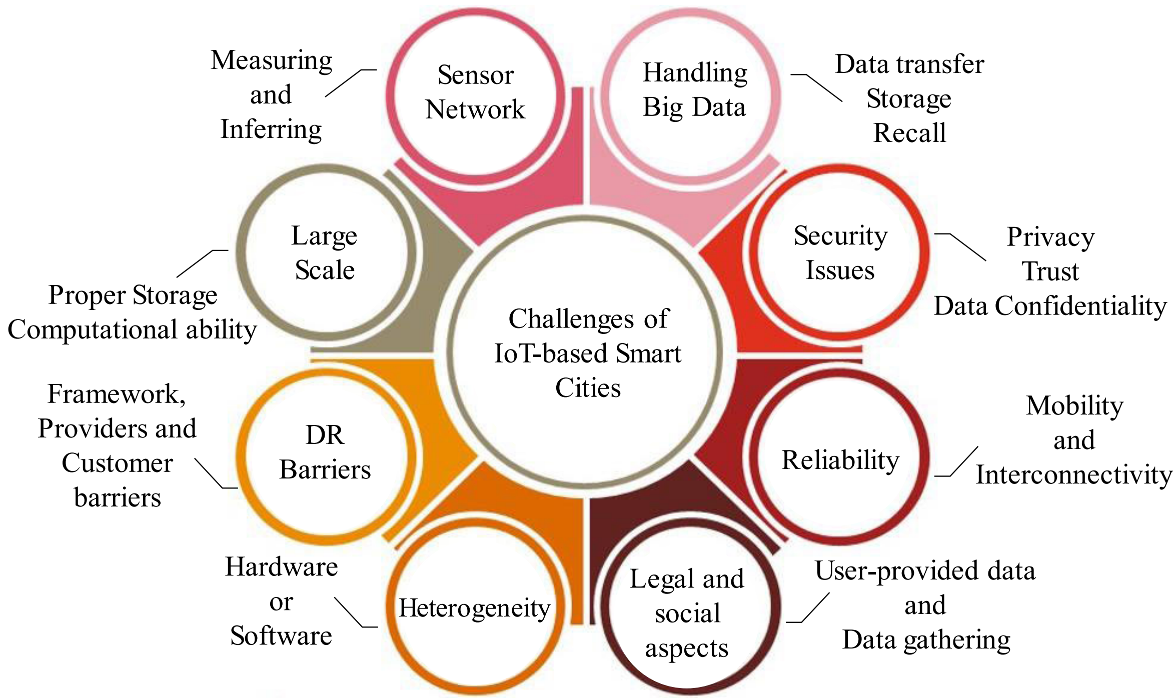

FIGURE 21. Challenges of IoT-based SCs.

the IoT-based SC infrastructure. Moreover, individuals' information must be confidential letting the authorized person access the information and make changes.

- **Reliability:** As one of the prominent issues in the IoT-based SCs paradigm, reliability takes care of mobility and interconnection. The IoT infrastructure relies on sensors or devices with pre-built sensors. These devices wirelessly send their data to the cloud and other linked devices. Thus, the reliability of information should be ensured. Moving vehicles are the best instance of a reliability issue whose interconnections are not consistent or reliable.
- **Legal and Social Aspects:** In IoT, heterogeneous sensors generate; as these sensors are used by the users, the users are data providers. The service provider receives the user's data to assess or use in the decision-making process. Strict rules must be established to ensure that the sensitive data of the customers are not shared to guarantee data security and integrity. Healthcare and transportation are of crucial significance as they utilize users' data. In smart healthcare, users' information must be tightly protected. In smart transportation, the real-time location of the users should not be public. These issues can be addressed by establishing and following proper rules and policies. An agreement must be also set between the user and service provider.
- **Heterogeneity:** The issue of heterogeneity can be resolved by smart solutions which involve joining the system to a specific application context. The development of an IoT-based SC could be highly complicated as it includes several sub-tasks. IoT can be defined as the interconnection of heterogeneous sensor-enabled devices connected by a wireless medium. Therefore, coping with various devices is one of these tasks which could be more difficult in a larger infrastructure such as an SC in which diverse applications are integrated and operated collectively. A large infrastructure such as an SC, thus, requires aggregating individual applications. An IoT-based SC involves heterogeneous devices and diverse applications. By connecting all of these smart applications, an SC infrastructure can be constructed. Cooperation and integration are two crucial challenges which have to be addressed.
- **DR Barriers:** IoT can help the system's responsive demand contribute more effectively. There are still a variety of impediments that can prevent people from participating in DR programs. These barriers associated with IoT-based SCs can be categorized into three sections: Framework Barrier, Provider Barrier, and Customer Barrier:

1) **Framework Barriers:** The framework perspective encompasses communication constraints, guidelines, and strategies.
   a) Communication constraints: The communication channel is a prominent barrier. Millions of connected devices are generating data in real-time, while there is a limited number of channels for communication. Hence, the communication channel should be quick and effective since a delay in communication may result in significant costs. Another hurdle that needs

to be taken into account while designing the SCs infrastructure is data transfer speed.

b) Guidelines and strategies: The framework must comply with the guidelines and strategies ruled by the government. Confidentiality of the data is a must and the system must be tightly secured.

2) **Providers Barriers:** The constraints of the providers can be classified into two sections: market trends and revenue.

a) Market Trends: Market analysis is a complex task involving following the latest trends to target the customers. Hence, the customer's perception and essential requirements must be considered to fulfill their demands cost-effectively.

b) Revenue: The final goal of the service providers is to obtain revenue. They desire twice as much as their investment, i.e. they only want profit, not a loss. This barrier limits the service providers to pay for any new IoT system. Thus, the creation of such a vast system will take longer than expected.

3) **Customer Barriers:** Concerning constraints from the customer's point of view, they can be characterized as knowledge and financial issues

a) Knowledge constraints: Ordinary people rarely know about current or changing technologies. The unfamiliarity with current technology is the main obstacle in the way of developing an SC. The insufficient technical knowledge to deal with the problems has always hindered the users.

b) Financial constraints: Not all users afford a smart device. Ordinary people prefer to save money for the future rather than spend on fancy devices.

- **Large scale:** Large-scale data require proper storage and computational capabilities [233]. IoT service is generally based on user-supplied information. Candidates have enough motivation to focus on an individual scenario and data gathering as a major local and social challenge, respectively [234].
- **Sensor Networks:** The IoT system can be regarded as a lattice of different sensors to collect data from the environment. The data are then passed to data storage vendors or service providers. The complexity of the system rises by incrementing the number of sensors, challenging the real-time handling of the events. In addition to data transfer, sensor devices can also interact with other smart devices. In this regard, the reasonable application of proper sensors is another challenge. The slow pace of data sensing, transmission, receiving, or computing is a serious concern that may lead to system breakdown. Other prominent constraints including energy efficiency and network capacity should be also addressed.
- **Handling Big Data:** Handling BD in a real-time system like SC is highly complex. In the BD paradigm, computation depends on data, thus, it is vital to consider the influence of data storage, transfer, recollection, and examination [3]. Sensor network was considered to accept the IoT as a novel technology [76] enabling data computation and collection, as well as broadening our insight into the environmental problems. Thus, the data from these devices are big and real-time. In this context, five "v"s have to be considered.

IoT is expected to be widely employed for connecting billions of devices in the future. For the comprehension of the concept of a SC, IoT technologies play a key role in the development of a SC strategy. The significance and benefits of innovative technologies such as IoT are undeniable in SCs. The growth of IoT will experience further technological advancements in SCs, enabling the users to pursue a simple yet interactive lifestyle.

C. ROLE OF IoT IN CLOUD STORAGE

More generally, CC can be classified into three service models: PaaS, SaaS, and IaaS. The SaaS model allows cloud clients to use the provider's applications to operate on a cloud infrastructure, in contrast to PaaS, which gives users the option to rent virtualized platforms to host own applications or services [137]. In IaaS, however, customers can have CPU, network bandwidth, storage, and other computing resources, which can be reconfigured based on their needs. Cloud services are presented to customers and users based on virtualization and distributed computing concepts. Fig. 22 illustrates the tiered architecture and standard service models as well as cloud advantages.

Many industries such as health care [235], agriculture [236], transportation [45], telecommunication [237], and real estate [238] have practically adopted IoT. This is because IoT can give "things" the finest connected environment possible and creates enormous amounts of data that must be handled, examined, and transferred to the cloud. By allowing third-party access to the business, public cloud storage aids IoT services. Therefore, by integrating IoT and cloud, IoT data or computational components operating over IoT devices can be provided.

Fig. 23 depicts three elements of cloud-based IoT storage. The first component presents IoT-based infrastructure in which devices are connected to the clouds through a second component, referred to as middleware, which covers communication technologies including 5G, Internet, Satellite Network, and Wi-Fi. The cloud infrastructure is the third element.

The architecture of IoT-based infrastructure consists of three layers: perception, network, and application layers [239].

- The physical or sensor layer, also known as the perception layer, is where sensors and actuators collect and transmit data for further processing [240]. This layer detects particular physical characteristics or other intelligent objects.

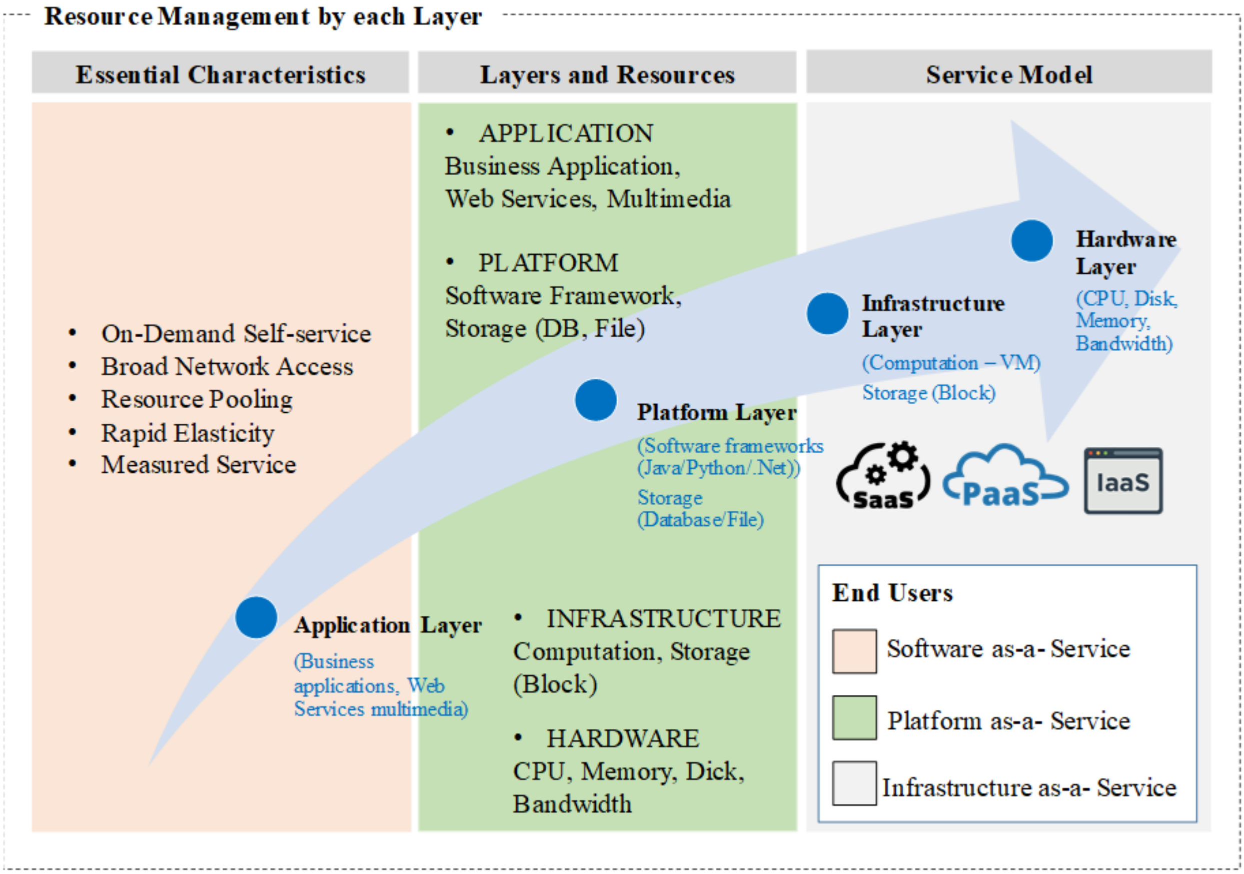

FIGURE 22. Type of CC.

- The network layer is in charge of connecting to servers, devices, and objects (things).
- As the uppermost layer, the application layer deploys IoT. This layer conveys application-specific facilities to the users.

The main advantage of hosting the IoT system on the cloud is that connected applications get more flexibility, scalability, and reliability.

D. ROLE OF BIG DATA IN IoT

IoT serves as a crucial hotspot for those who collect massive amount of data for analysis. That is why IoT has become quite popular in BD. This is because BDA is emerging as a critical component for deciphering IoT-generated data from "connected devices," which aid in the essential activity update. In this direction, the role of BD in the IoT is to handle as well as store large amount of data in a reliable way using a variety of high-capacity technologies. The preparation of IoT-based BD can be divided into four steps, as follows:

1) IoT devices generate a massive volume of unstructured data in the context of BD, which has five variables, i.e. volume, velocity, variety, value, and variability [162]. These variables are commonly referred to as 5V's.
2) Large amounts of data are stored in BD documents within a widely diffused database, so-called, the BD framework.
3) Logical tools such as Hadoop, MapReduce or Spark are utilized for analyzing the stored IoT BD.
4) The last step is the construction of reports based on the stored data which need to be analyzed, accordingly.

It is worth noting that the unstructured IoT data can be collected by the Web. In order to quickly identify key pieces of knowledge from data and make judgments, BD for IoT hence requires rapid inquiry with more inquiries. As a result, the importance of BD in IoT is undeniable.

The fact that the Web can gather unstructured IoT data is meaningless. In order to quickly identify critical pieces of knowledge from data and make judgments, BD for IoT hence requires quick inquiry with more inquiries. BD's significance in IoT is therefore undeniable [19]. BD's usability can be also discovered to a large part in the realm of IoT. In recent years, people have been progressively moving to cities. Cities are important sources of BD because they are home to large populations of people, technical equipment, and vehicles, all of which produce data every day. Green Internet of Things (G-IoT), as applied by Chithaluru et al. [241] for SCs, plays one of the most crucial roles in the creation of a green and

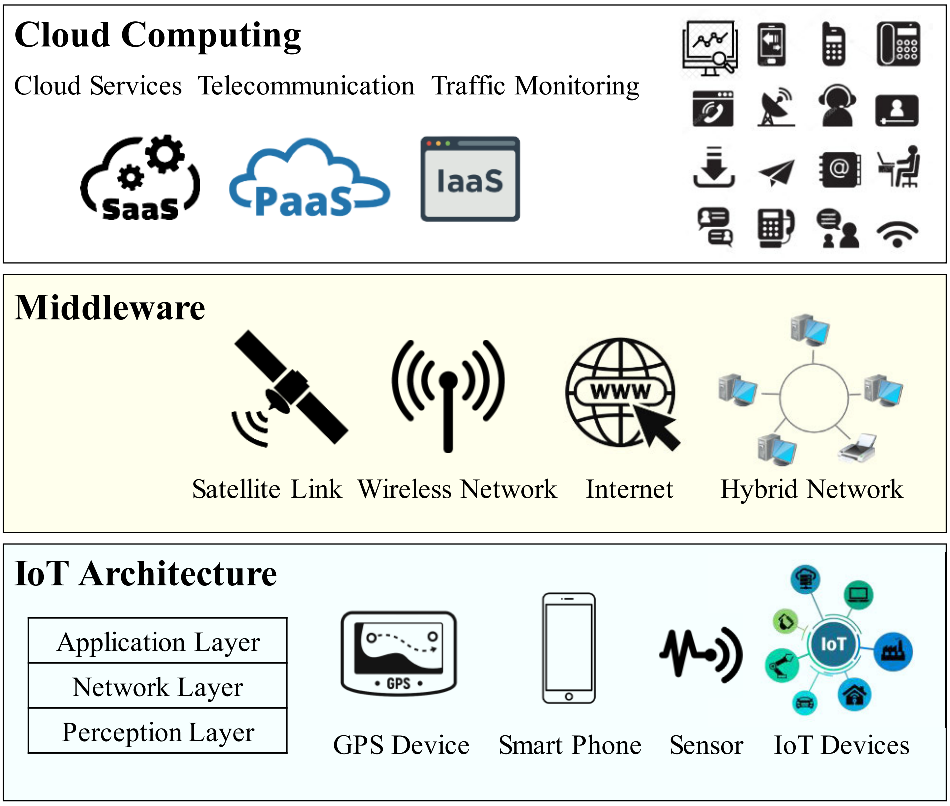

FIGURE 23. Cloud and IoT-based environments.

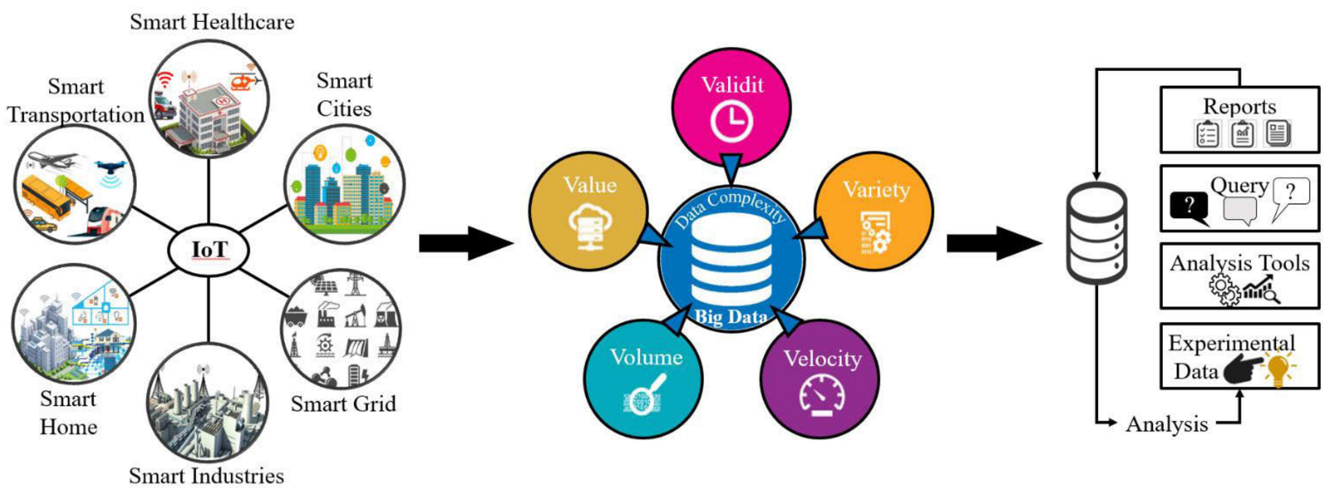

FIGURE 24. Environmental based on BD and IoT.

safe living environment. As mentioned before, the ability to extract usable information from enormous amounts of data generated by IoT necessitates BDA.

Cities are important sources of BD because they are home to large populations of people, technical equipment, and vehicles, all of which produce data every day. The goal of an IoT integrated with BDA system is to overcome the obstacles of storing and examining a large amount of data originating from smart structures. The link between BD and IoT is depicted in Fig. 24.

VII. ARCHITECTURE AND IMPLEMENTATION OF SMART CITIES THROUGH BIG DATA, CLOUD COMPUTING, AND IoT

SCs [54] can be smarter through the use of digital technologies. These cities are equipped with electronic devices and

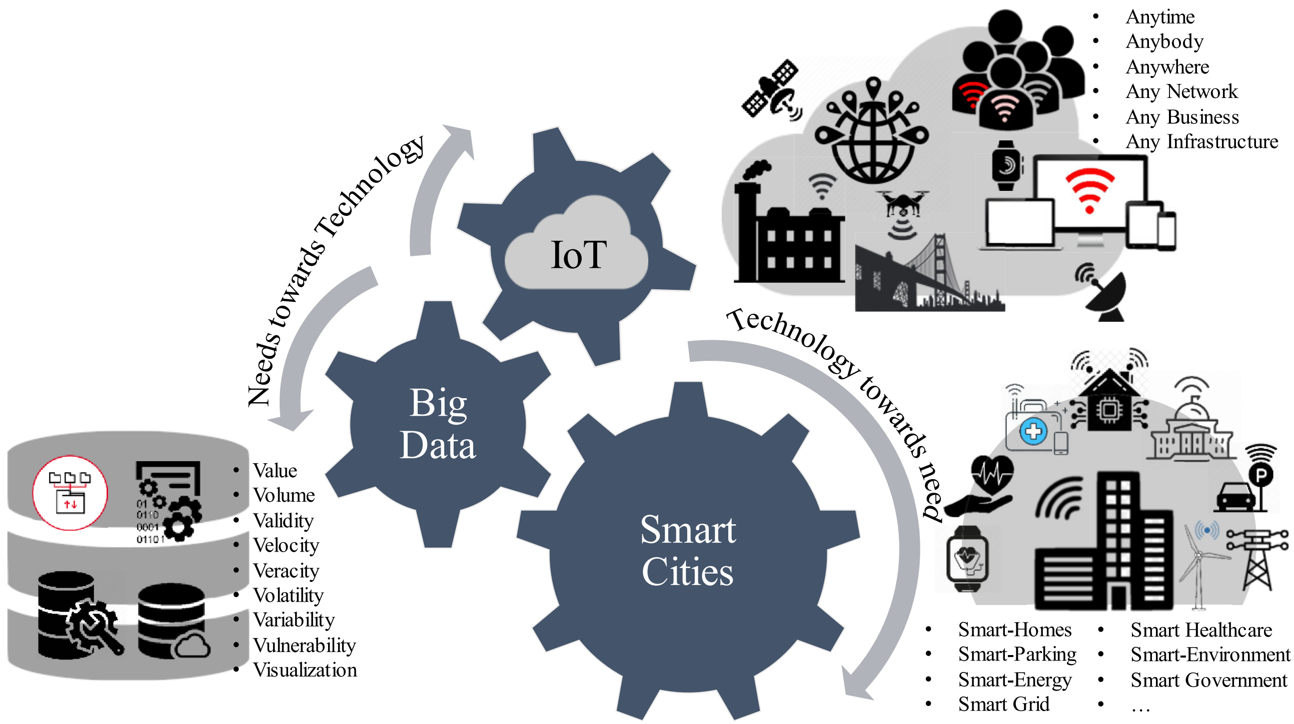

FIGURE 25. Relationship between IoT, BD, and SCs.

smart applications such as street-level cameras and sensors for surveillance and transportation. Some initiatives use these elements to establish valuable services such as Google Street View and GPS.

Urban planning applications can be benefited from the capabilities of the IoT in SCs [242], affecting local regions. Diverse projects are employed in these cities for cycling, automobiles, and public monitoring through sensors for data collection. Fig. 25 depicted the relationship between IoT, BD, and SCs.

SCs and urban planning drastically affect national development by enhancing the decision power of the communities through smart and effective decision-making at the right time [63].

A complete system is developed by Hadoop technologies with Spark to attain real-time processing. The simple IoT-based SC data sets like vehicular networks, smart parking, smart home, weather, pollution, and surveillance are assessed for the establishment of SCs and making urban planning decisions. BD is employed for the analyses of diverse aspects of the SC using the previous information [94].

The solution in the SC can be applied and evaluated by the Spark and Hadoop frameworks to consider the effect of online decision-making. The proposed system was assessed by the Hadoop framework with Spark to attain the real-time impacts upon real-time SC decisions. Furthermore, Hadoop and MapReduce can be utilized on large historical data for urban planning [243].

A 4-tier architecture is proposed to analyze IoT BD for the establishment of SCs according to the requirements of an SC and urban planning as depicted in Fig. 26. The Data section is the first layer; while digitalization and intelligence are the second and third layers, respectively whose functionalities will be explained as follows:

- **Bottom Layer:** This layer is responsible for data generation through diverse IoT sources as well as data collection. As numerous IoT sensors are engaged in data generation, lots of heterogeneous data will be produced with different formats, origins, and periodicities. Additionally, these data have security, privacy, and quality requirements. In the case of sensor data, Metadata has been always larger than the actual measures. Thus, early registration and filtration are required in this layer to filter unnecessary Metadata, as well as duplicate ones. The sensor data are then sent to the nodes via ZigBee technology, allowing communications on the internet through diverse technologies, such as Wi-Fi, WiMAX, LTE, 4G, and 5G [244]. Ethernet can be employed for the analysis of various analyzing servers.
- **Intermediate Layer:** As the major layer of the entire analytical system, it is in charge of data processing. All the data stored on the physical level are processed at this level. Statistical calculations, computations, and graph analyses are also conducted in this layer. After data collection from the IoT and the Internet, they are transferred to the cloud. A third-party real-time tool is required to

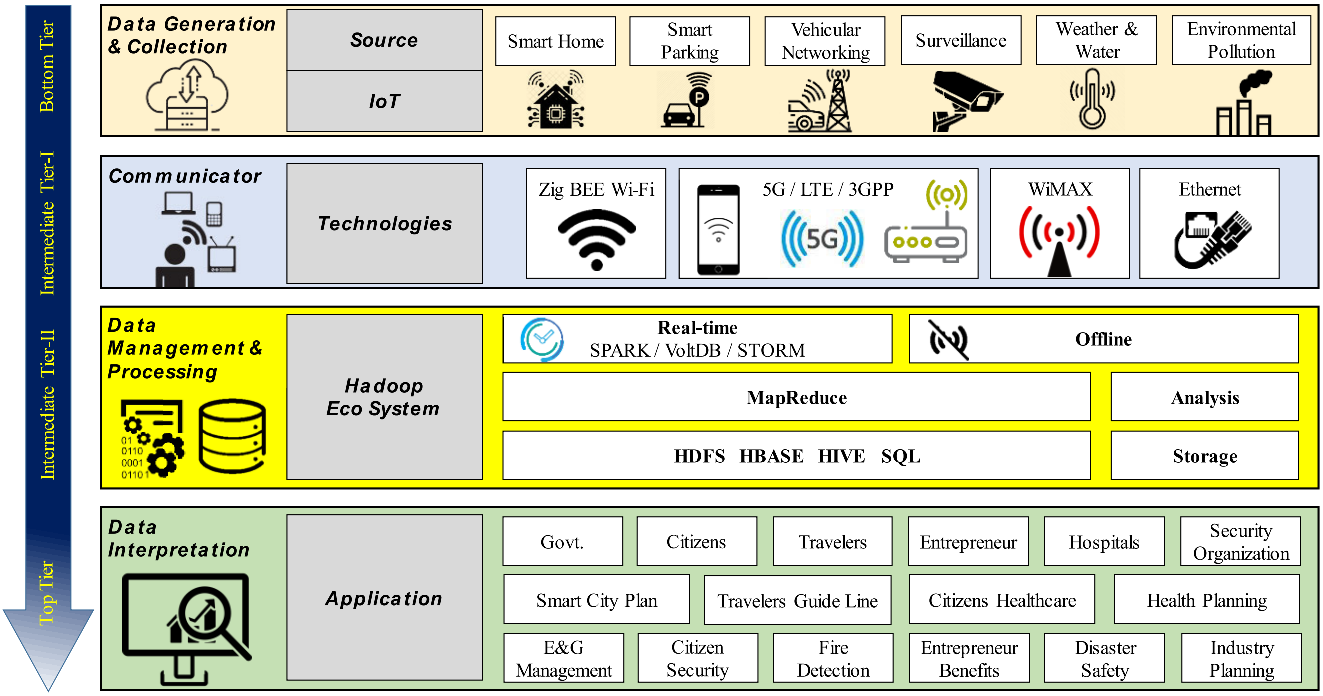

FIGURE 26. Architecture and implementation of SCs based on IoT, BD, Cloud.

combine with Hadoop and achieve a real-time implementation due to the demand for real-time analysis in smart systems. The Hadoop [245] ecosystem possesses HDFS file storage to divide the data into equal chunks and store them on diverse data nodes. Then, parallel processing is carried out on these chunks by the MapReduce [246] system. The results of the processing calculations are conducted at the Hadoop ecosystem. Finally, the decisions are made based on the results of the Hadoop ecosystem utilizing ML, pattern recognition, soft computing, and decision models. Real-time implementation can be achieved by Strom [247], Spark [198], and VoltDb. However, the system was implemented by Spark. At the lower layer of Hadoop, the same structures of MapReduce and HDFS were employed. HIVE, HBASE, and SQL can be also applied for managing Databases (in-memory or Offline) to store historical information. In urban planning, the real-time results are not much important, thus, Hadoop was utilized with MapReduce programming. As a demand-based usage model, the cloud offers unlimited virtualization. Cloud technology also contributes to overcoming issues such as privacy, security, and data reliability through the improvement of the security of available data and services. Owing to the efficient management of sensor data, auxiliary services are timely provided to individuals, governments, centers, and organizations for improving urban planning management.

- **Top Layer:** The upper level, also known as the decision level, encompasses the urban planning decisions. This level involves various modules depending on the type of planning (e.g. road planning and building planning). The number written under the planning module signifies the number of datasets used by the module as the source of input data. Once BD and IoT information are gathered, processed, analyzed, integrated, and secured, the data are brought to the required level of intelligence to make decisions based on the results of the system. Decision-making involves ML, recognition patterns, calculations, and decision models. This layer deals with software services for the virtual interaction of individuals, organizations, and centers for optimal planning, analysis, and decision-making, with the help of the data obtained from the previous. This level can result in the successful implementation of smart water resource management, sewer, security, tourism industry, home, and education. This process can also enhance the quality while improving the living standards. In the case of real-time traffic management, Packet Capture (Pcap) can be generated by Wireshark libraries and retransmitted to the developed system. Hadoop-pcap-lib, Hadoop-pcap-serde, and Hadoop Pcap Input libraries are utilized to process network packets and generate Hadoop Readable (sequence file) at the collection and aggregation for being processed by Spark. MapReduce programming can be also employed for offline analysis in urban planning. The Dataset mentioned in section IV was employed to assess the efficiency of the system.

Vehicular traffic information is one of the most significant sources in an SC which can offer useful real-time analysis for citizens and the government. The citizens can decide on their destination according to the traffic load and the mean speed of the vehicles. This opportunity can decline fuel consumption and level of contamination due to the crowded traffic. Government can also obtain real-time data about the road blockage due to accidents. Thus, they can take necessary measures for timely traffic management. In the proposed SC, the traffic data can be collected by GPRS, vehicular sensors, and sensors installed on the front screen of the cars. These data include the location of each car and the number of vehicles between two pairs of sensors. Moreover, upon accident and damage of the front screen, the sensor will send an alert to the police, traffic authorities, and hospital.

VIII. DISCUSSION AND OBSERVATIONS

In order to identify the features of BDA frameworks in SCs and address the research objectives, the findings of this paper can be viewed from three various perspectives (i.e., design guidelines, assistive technologies, and Implementation domains).

A. DESIGN GUIDELINES

- Eventually, the layered design concept is employed. Each layer has a distinct job and interacts with the above and lower layers in ways that are clearly defined [162],
- It is important to standardize data collecting from the outside world (interoperability and mobility). Likewise, to enable fact-based applications, consistent access to output data (e.g., via API or standard message formats) is also required [248],
- It is crucial to provide both historical and real-time data analysis,
- Both iterative and sequential data processing are provided appropriately to suit the needs of real-time and historical data analytics,
- Scalability to handle the rising number of data-gathering devices and the data volumes that these devices created.

B. ASSISTIVE TECHNOLOGIES

The key technologies of BDA in SCs are discussed in this section. The elements directly related to applications for SCs are the main focus.

- Horizontally scalable platforms: The scalability demands of BDA frameworks in SCs can potentially be met by Hadoop. They offer durable platforms that are affordable and able to handle massive amounts of BD [138], [249].
- FC and IoT go hand in hand [204]. Computing and storage are offered via FC, sometimes referred to as EC, in end devices and clouds.
- Cloud systems provide robust, scalable, and accessible hosting environments for data storage and computing, In keeping with the massive amount of BD and intricate dynamic features of SCs [3], [218].
- DM and ML are critical technologies for a data-centric SC. While DM is employed for the extraction of hidden and potentially valuable information, DM involves a broad field encompassing diverse algorithms and statistical techniques [193], [250], [251]. Moreover, ML can be regarded as an application of AI enabling computers to learn and improve from experience with no programming. ML algorithms can be grouped into two categories of supervised and unsupervised learning methods. The former can learn from previous data with labeled examples (training dataset). With sufficient training iteration, the system can predict future events. On the other hand, unsupervised ML algorithms apply to the non-labeled training data. Unsupervised learning studies are focused on how systems can infer a function for describing a hidden structure from unlabeled data [252].
  Numerous SC applications using DM and ML have been reported in the literature [253]. However, the main challenge is to find suitable datasets meeting the DM requirements. Resolving the problems of dataset selection and data combination are highly crucial in the DM process of SCs [145].
- In-memory databases for high-performance analytics: Relational Databases Systems (RDBMS) are not designed to resolve BD scalability and agility issues, nor are they set to exploit the current commodity storage and processing power. On the other hand, SQL databases (NoSQL) are designed to meet BD requirements. NoSQL databases can be classified into key-value, document-based, column-based, and graph-based types. Apache Hadoop suite includes HBase which is a NoSQL database running on top of HDFS, supporting both key-value and column-based types. HBase also offers real-time read/write access to large datasets. The distinct advantage of HBase lies in its high linear scalability in coping with huge datasets containing many rows and columns. Moreover, HBase is capable of combining data sources with diverse structures and schemas.
- Data visualization can be defined as presenting data and \ or information in an easily understanding illustrative format intended for the rapid identification of interesting patterns. In the era of BD, visualization is becoming more challenging due to its V characteristics. For instance, in BD applications of SCs, decision maker-specific ones increase the complexity of data visualization due to various origins of data [254]. To overcome such bottlenecks, the development of cutting-edge technologies (e.g. Augmented Reality (AR), Virtual Reality (VR), Mixed Reality (MR), and Google Maps) have paved the way for the advancement of efficient SC applications. Points of Interest (POI) [255] and traffic management [18], [256] are two examples of data visualization applications in SCs.

C. IMPLEMENTATION DOMAINS

Application domains are tackled in two ways in the scope of reviewed publications. Some research papers presented analytics solutions based on a specific application area (e.g., environment, traffic, etc.), while others were motivated by the need of identifying technical solutions for analytical functions (e.g., interoperability, mobility, etc.). Therefore, the application domain is employed as a proof of concept for the proposed idea in the latter situation.

Based on the article analysis and review in the previous section and the above conclusions, the variables influencing the design of the suggested frameworks can be driven by either finding solutions for specific SC domains or answering to some technological issues.

Another noteworthy remark is that the basic design concept of the evaluated architecture fails to meet the holistic perspective of SCs as complex systems in the interrelationships between different SC domains which must be considered accordingly.

The interrelationship between different SC domains requires comprehensive analytics based on datasets generated from different domains which may be repeated for a broader scope of extracted information and insights (for example, strategic objectives are defined during the planning phase of the SC) [257], [258]. There are two options for meeting this difficult criterion. The initial option is to save datasets from various domains, combine selected datasets for analysis and start the analytics process. The second strategy involves saving the results of the analytics procedures for future study and allowing the extracted data models to be recovered and integrated for comprehensive analytics. The applicability of integrating the models is certainly constrained by this option. However, compared to the first strategy, the second approach will save time as it is not required to reanalyze the same dataset several times. The second strategy does not obviously preclude by the first option. However, in order to optimize execution time, both mechanisms are complementary to each other.

We recommend that two essential design components, i.e. model management, and aggregation can be included in BDA for SCs. The first part allows storing the extracted data and insights (data models) for future study without repeating the analysis procedure. In other words, the first component will help in creating a library for the extracted models. On the other hand, the second part will allow for more extensive multi-domain analytics through model ensembles.

Diverse parts of an SC such as living, governance, public, economy, mobility, and environment sectors are in continuous interaction with each other through feedforward intelligent information and feedback promoting the competitiveness and effectiveness of the city. The nonstop human-machine interactions lead to develop emerging technologies and serve as a platform in the combat against the challenges.

IX. A NEW FRAMEWORK FOR SMART CITY DATA ANALYTICS

BDA refers to the process of database exploring, DM and analysis to enhance business performance to better serve customers and assist in the decision-making procedures. Furthermore, the datasets collected from multiple city domains must be merged and evaluated to get useful insights on creating smart information services at the city level. As a consequence, the fundamental goal of BDA is to help business associations having a better data knowledge, as a result, making effective and well-informed decisions. Data scientists and miners can analyze a massive number of data using BDA that may not be possible with conventional technologies.

Regarding the mentioned discussion on analysis and the challenges of BD technology, and in light of the results and debate in Section VIII, as well as an awareness of an SC's holistic perspective, a novel architecture of Smart City Data Analytics (SCDA) is proposed in this section. BDA frameworks represented in data model management and aggregation are given new features by the suggested framework. The following research queries serve as the foundation for the SCDAP design:

- What traits do BDA frameworks use in SCs in the literature,
- and what are the fundamental design concepts that ought to direct the development of BDA frameworks to fulfil the needs of SCs?

We conducted a thorough literature study on BDA frameworks in SCs to respond to these queries. In addition to managing and aggregating data models, the suggested framework adds new functions to BDA frameworks. As depicted in Fig. 27, this architecture involves three layers: operational platform, security, and data processing layers.

a) **Operational Platform Layer:** As a horizontally scalable platform, this layer includes operating systems, hardware clusters, and communication protocols. Additional computing nodes could be included if necessary.

b) **Security Layer:** Despite the complete vision for the functionality of the security layer in the physical implementation of Smart City Data Analytics Dashboard (SCDAD), several security measures should be considered in physical design, in particular for critical analytics:
   - Restricted sign-on access to the framework for critical and sensitive data,
   - Multi-level user authentication,
   - Complete audit log for prominent operations.

c) **Data processing layer:** As the core data processing engine, this layer encompasses all the data processing functionalities from data acquisition to knowledge extraction. Online and batch data processing are supported for real-time and historical data analytics, respectively. Moreover, it offers two important functionalities

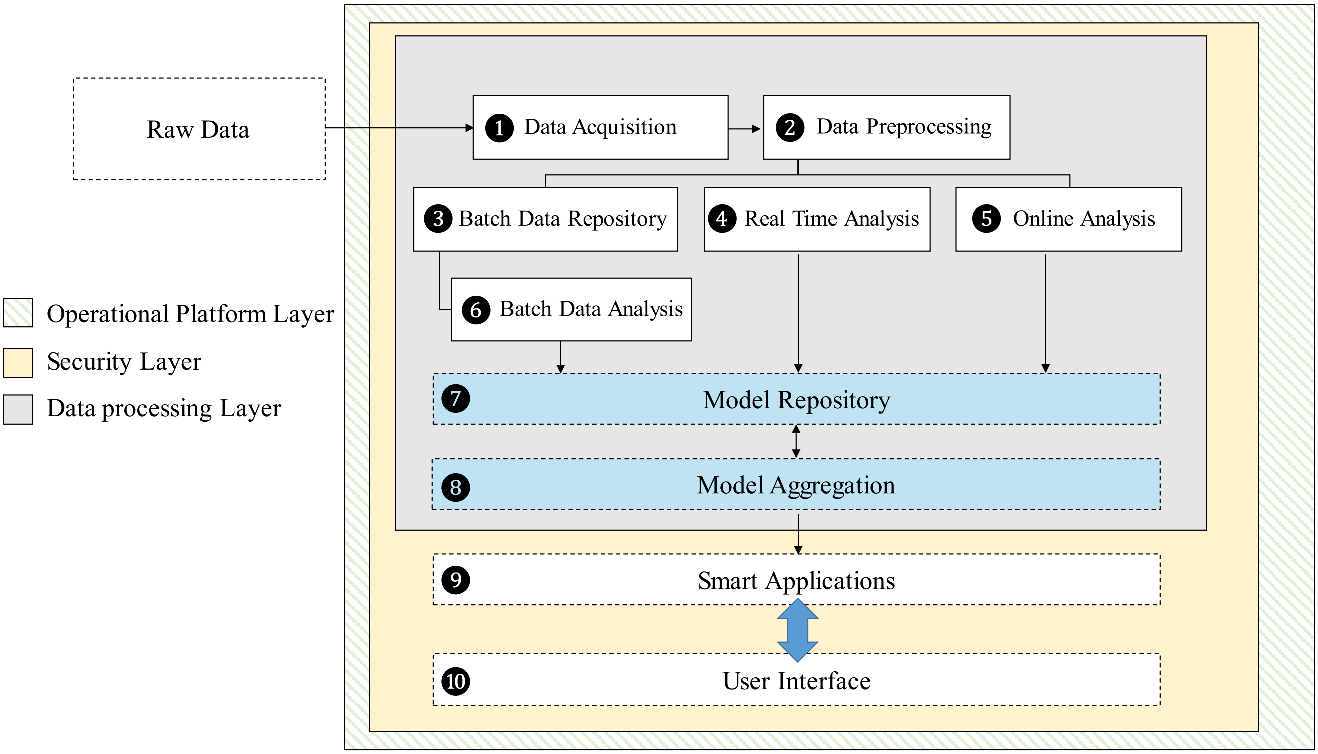

FIGURE 27. SCDAD architecture.

to distinguish SCDAD: Model manager and model aggregation. While the former manages the extract data models, the latter is for data aggregation. The components of this layer can be listed as follows:

1) **Data acquisition:** Data acquisition plays a vital role in capturing data from the external world. This component is characterized by scalability and interoperability which respectively refer to the ability of dynamic accommodation with the increase of the number of data-generating artifacts and the ability to interact with various types of artifacts (Wireless Sensors Networks (WSN), IoT, and Social Media Network (SMN)).
2) **Data preprocessing:** this component is responsible for cleansing input data, transforming, and integrating functionalities. It also transforms input data into analysis-ready formats.
3) **Online analytics:** Capability of stream data processing for interactive applications with acceptable latency.
4) **Real-Time Analytics:** Capability of stream data processing in real-time applications with sufficiently low latency.
5) **Batch data repository:** Data storage management systems (e.g. Hadoop HDFS, NoSQL data management system).
6) **Batch Data analytics:** Batch data analytics for applications affording latency.
7) **Model management:** Extracted data model management system, in which resultant data analysis models are persisted and retrieved (or deleted) by relevant metadata for further inquiries or reuse. Furthermore, static (semi-static data) of the city can be persisted in this repository.
8) **Model aggregation:** Extract data model ensemble functionality for more advanced and complicated analytics and inquiries.
9) **Smart application:** Smart application developed based on resultant data analysis models.
10) **User interface:** End user interface which provides efficient and flexible tools for access, report, and ad hoc inquiries for persisted and/or aggregated models.

X. FUTURE WORK

Concerning SC management, examination and understanding the significance of decision-making analytical systems in BD, IoT, and CC may lead to questions whose solutions can be the topic of future investigations.

a) How to employ old sensor information combined with BDA for urban management and planning?
b) How to enhance the intelligence level of current equipment?
c) How to enable the equipment to cope with the current status?
d) How to manage the cost of data collection?
e) How to provide real-time BDA?

Accordingly, the SC employs ICT to help the citizens use minimum resources. Recently-developed technologies offer advanced sensor and information storage capabilities for a large amount of information for controlling great volumes of data. A proper system should be capable of knowing and anticipating all the latest developments in ICT. Such systems require environmental sensing abilities and metric data analysis.

Several studies have addressed the development of intelligent systems for SCs. However, significant questions have remained unanswered or have not been fully explored. This section presents a glimpse into the mostly unexplored field of intelligent system designs for SCs. Additionally, a brief description of the following areas is provided for improvement.

- **Data management:** Data play a crucial role in the development of SCs. SCs generate a lot of data whose management and comprehension could be a difficult task. Furthermore, such enormous deal of data requires a large storage capacity. Greater synchronisation between the application viewpoint and implementation is necessary to address these issues
- **IoT integration into a broader perspective:** Although IoT is essential for smooth operation of SCs, much effort is required to fully utilize IoT in these cities. IoT is an essential component in extracting useful information from real-world data. The IoT services cover a broad range of industries. The role of IoT in SCs can be investigated in more detail.
- **Interoperability protection:** Data can move freely among several technologies which use advanced safety measures in city services like lighting, and transit. Therefore, intelligent communication between such devices can be ensured. Thus, it is essential to ensure compatibility among such systems, which can be addressed as a potential research project.
- **Cloud platform:** As cloud services might further enhance living in SCs, the use of CC platform could be one of the interesting research areas. To ensure public safety and security, this can be done by leveraging BD-driven dispatch of police or traffic details, coordinated public works scheduling, and municipal maintenance. As a result, researchers are very interested in the problem of BD management in SCs using cloud services.

XI. CONCLUSION

The drastic rise of connected devices in urban cities has promoted the rapid growth of data, attracting the attention of researchers from various fields. SCs and urban planning have shown a great and direct influence on the development of communities by making smart, efficient, and timely decisions. This article provides a comprehensive review of SCs, BD, IoT, and CC. The limitations and advantages of BD applications were also discussed. IoT offers a framework for secure communication of devices and sensors and allows for information sharing across the SC. CC improves resource integration, information sharing, and software collaboration. Generally speaking, this article is aimed to comprehensively describe the application of IoT, cloud, and BD technologies in the management and planning of SCs. A three-layer system was also introduced for SCs and urban planning based on IoT, BDA, and CC. The proposed system features aggregation, communication, processing, interpretation, and translation. The system was further developed by Hadoop technology to approach real-time processing cases. Finally, BD and IoT with CC can play a key role in extracting valuable information for better decision-making. The studies on the application of BD in an SC are at the beginning of a long journey. Resolving the mentioned challenges can promote its applicability. This review article could serve as a benchmark for researchers and industry to further develop SCs.

REFERENCES

[1] S. Mishra and A. K. Tyagi, "The role of machine learning techniques in Internet of Things-based cloud applications," in *Artificial Intelligence-based Internet of Things Systems*. Cham, Switzerland: Springer, 2022, pp. 105–135, doi: 10.1007/978-3-030-87059-1_4.

[2] P. Brous, M. Janssen, and P. Herder, "The dual effects of the Internet of Things (IoT): A systematic review of the benefits and risks of IoT adoption by organizations," *Int. J. Inf. Manage.*, vol. 51, Apr. 2020, Art. no. 101952, doi: 10.1016/j.ijinfomgt.2019.05.008.

[3] A. Yassine, S. Singh, M. S. Hossain, and G. Muhammad, "IoT big data analytics for smart homes with fog and cloud computing," *Future Gener. Comput. Syst.*, vol. 91, pp. 563–573, Feb. 2019, doi: 10.1016/j.future.2018.08.040.

[4] X. Lu, K. Zhou, X. Zhang, and S. Yang, "A systematic review of supply and demand side optimal load scheduling in a smart grid environment," *J. Cleaner Prod.*, vol. 203, pp. 757–768, Dec. 2018, doi: 10.1016/j.jclepro.2018.08.301.

[5] A. Pundir, S. Singh, M. Kumar, A. Bafila, and G. J. Saxena, "Cyber-physical systems enabled transport networks in smart cities: Challenges and enabling technologies of the new mobility era," *IEEE Access*, vol. 10, pp. 16350–16364, 2022, doi: 10.1109/ACCESS.2022.3147323.

[6] D. Pal, S. Funilkul, N. Charoenkitkarn, and P. Kanthamanon, "Internet-of-Things and smart homes for elderly healthcare: An end user perspective," *IEEE Access*, vol. 6, pp. 10483–10496, 2018, doi: 10.1109/ACCESS.2018.2808472.

[7] B. N. Silva, M. Khan, and K. Han, "Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities," *Sustain. Cities Soc.*, vol. 38, pp. 697–713, Apr. 2018, doi: 10.1016/j.scs.2018.01.053.

[8] O. Dogan and O. F. Gurcan, *Applications of Big Data and Green IoT-Enabling Technologies for Smart Cities*. Hershey, PA, USA: IGI Global, 2022, doi: 10.4018/978-1-5225-7432-3.ch002.

[9] B. Esmaeilian, B. Wang, K. Lewis, F. Duarte, C. Ratti, and S. Behdad, "The future of waste management in smart and sustainable cities: A review and concept paper," *Waste Manage.*, vol. 81, pp. 177–195, Nov. 2018, doi: 10.1016/j.wasman.2018.09.047.

[10] M. D. Lytras, P. Papadopoulou, and A. Sarirete, "Smart healthcare: Emerging technologies, best practices, and sustainable policies," in *Innovation in Health Informatics*. New York, NY, USA: Academic Press, 2020, pp. 3–38, doi: 10.1016/b978-0-12-819043-2.00001-0.

[11] T. S. J. Darwish and K. Abu Bakar, "Fog based intelligent transportation big data analytics in the Internet of Vehicles environment: Motivations, architecture, challenges, and critical issues," *IEEE Access*, vol. 6, pp. 15679–15701, 2018, doi: 10.1109/ACCESS.2018.2815989.

[12] K. Ding, Y. Zhang, F. T. S. Chan, C. Zhang, J. Lv, Q. Liu, J. Leng, and H. Fu, "A cyber-physical production monitoring service system for energy-aware collaborative production monitoring in a smart shop floor," *J. Cleaner Prod.*, vol. 297, May 2021, Art. no. 126599, doi: 10.1016/j.jclepro.2021.126599.

[13] S. K. Podder and D. Samanta, "Impact of business analytics for smart education system and management functions," in *Advances in Computing and Intelligent Systems*. Cham, Switzerland: Springer, 2020, pp. 479–488.

[14] S. Praharaj and S. Bandyopadhyay, "Understanding the open data challenge for building smart cities in India," in *Open Cities | Open Data*. Cham, Switzerland: Springer, 2020, pp. 359–382, doi: 10.1007/978-981-13-6605-5_16.

[15] S. B. Atitallah, M. Driss, W. Boulila, and H. B. Ghézala, "Leveraging deep learning and IoT big data analytics to support the smart cities development: Review and future directions," *Comput. Sci. Rev.*, vol. 38, Nov. 2020, Art. no. 100303.

[16] Z. S. Ageed, R. K. Ibrahim, and M. A. M. Sadeeq, "Unified ontology implementation of cloud computing for distributed systems," *Current J. Appl. Sci. Technol.*, pp. 82–97, Nov. 2020, doi: 10.9734/CJAST/2020/v39i3431039.

[17] B. B. Sinha and R. Dhanalakshmi, "Recent advancements and challenges of Internet of Things in smart agriculture: A survey," *Future Gener. Comput. Syst.*, vol. 126, pp. 169–184, Jan. 2022, doi: 10.1016/j.future.2021.08.006.

[18] J. Zeng, L. T. Yang, M. Lin, H. Ning, and J. Ma, "A survey: Cyber-physical-social systems and their system-level design methodology," *Future Gener. Comput. Syst.*, vol. 105, pp. 1028–1042, Apr. 2020, doi: 10.1016/j.future.2016.06.034.

[19] X. Xu, R. Gu, F. Dai, L. Qi, and S. Wan, "Multi-objective computation offloading for Internet of Vehicles in cloud-edge computing," *Wireless Netw.*, vol. 26, pp. 1611–1629, Sep. 2019, doi: 10.1007/s11276-019-02127-y.

[20] J. Laufs, H. Borrion, and B. Bradford, "New security technologies and their functions in smart cities: A systematic review," *Sustain. Cities Soc.*, vol. 55, Apr. 2020, Art. no. 102023, doi: 10.1016/j.neubiorev.2019.07.019.

[21] S. E. Bibri and J. Krogstie, "Smart sustainable cities of the future: An extensive interdisciplinary literature review," *Sustain. Cities Soc.*, vol. 31, pp. 183–212, May 2017, doi: 10.1016/j.scs.2017.02.016.

[22] E. P. Trindade, M. P. F. Hinnig, E. M. da Costa, J. S. Marques, R. C. Bastos, and T. Yigitcanlar, "Sustainable development of smart cities: A systematic review of the literature," *J. Open Innov., Technol., Market, Complex.*, vol. 3, no. 3, p. 11, 2017, doi: 10.1186/s40852-017-0063-2.

[23] I. A. T. Hashem, V. Chang, N. B. Anuar, K. Adewole, I. Yaqoob, A. Gani, E. Ahmed, and H. Chiroma, "The role of big data in smart city," *Int. J. Inf. Manage.*, vol. 36, no. 5, pp. 748–758, Oct. 2016, doi: 10.1016/j.ijinfomgt.2016.05.002.

[24] I. A. T. Hashem, N. B. Anuar, A. Gani, I. Yaqoob, F. Xia, and S. U. Khan, "MapReduce: Review and open challenges," *Scientometrics*, vol. 109, no. 1, pp. 389–422, Oct. 2016, doi: 10.1007/s11192-016-1945-y.

[25] D. Gil, M. Johnsson, H. Mora, and J. Szymański, "Review of the complexity of managing big data of the Internet of Things," *Complexity*, vol. 2019, pp. 1–12, Feb. 2019.

[26] A. Chauhan, S. K. Jakhar, and C. Chauhan, "The interplay of circular economy with industry 4.0 enabled smart city drivers of healthcare waste disposal," *J. Cleaner Prod.*, vol. 279, Jan. 2021, Art. no. 123854, doi: 10.1016/j.jclepro.2020.123854.

[27] K. Douaioui, M. Fri, C. Mabroukki, and E. A. Semma, "The interaction between industry 4.0 and smart logistics: Concepts and perspectives," in *Proc. Int. Colloq. Logistics Supply Chain Manage. (LOGISTIQUA)*, Apr. 2018, pp. 128–132.

[28] M. Gordan, S.-R. Sabbagh-Yazdi, Z. Ismail, K. Ghaedi, P. Carroll, D. McCrum, and B. Samali, "State-of-the-art review on advancements of data mining in structural health monitoring," *Measurement*, vol. 193, Apr. 2022, Art. no. 110939, doi: 10.1016/j.measurement.2022.110939.

[29] C.-C. Kuo, J. Z. Shyu, and K. Ding, "Industrial revitalization via industry 4.0—A comparative policy analysis among China, Germany and the USA," *Global Transitions*, vol. 1, pp. 3–14, 2019, doi: 10.1016/j.glt.2018.12.001.

[30] A. Nayyar, A. Kumar, and D. Gupta, *A Roadmap to Industry 4.0: Smart Production, Sharp Bus. and Sustainable Development*. Berlin, Germany: Springer, 2020.

[31] H. Oliff and Y. Liu, "Towards industry 4.0 utilizing data-mining techniques: A case study on quality improvement," *Proc. CIRP*, vol. 63, pp. 167–172, 2017, doi: 10.1016/j.procir.2017.03.311.

[32] M. Gordan, K. Ghaedi, and V. Saleh, *Industry 4.0-Perspectives and Applications*. London, U.K.: IntechOpen, 2023.

[33] M. Talebkhah, A. Sali, M. Marjani, M. Gordan, S. J. Hashim, and F. Z. Rokhani, "Edge computing: Architecture, applications and future perspectives," in *Proc. IEEE 2nd Int. Conf. Artif. Intell. Eng. Technol. (IICAIET)*, Sep. 2020, pp. 1–6.

[34] C. Esposito, M. Ficco, and B. B. Gupta, "Blockchain-based authentication and authorization for smart city applications," *Inf. Process. Manage.*, vol. 58, no. 2, Mar. 2021, Art. no. 102468, doi: 10.1016/j.ipm.2020.102468.

[35] E. Ismagilova, L. Hughes, N. P. Rana, and Y. K. Dwivedi, "Security, privacy and risks within smart cities: Literature review and development of a smart city interaction framework," *Inf. Syst. Front.*, vol. 24, pp. 393–414, Jul. 2020, doi: 10.1007/s10796-020-10044-1.

[36] F. Chiariotti, M. Condoluci, T. Mahmoodi, and A. Zanella, "SymbioCity: Smart cities for smarter networks," *Trans. Emerg. Telecommun. Technol.*, vol. 29, no. 1, p. e3206, Jan. 2018, doi: 10.1002/ett.3206.

[37] M. Hussain and M. M. Beg, "Fog computing for Internet of Things (IoT)-aided smart grid architectures," *Big Data Cognit. Comput.*, vol. 3, no. 1, p. 8, Jan. 2019, doi: 10.3390/bdcc3010008.

[38] F. Al-Turjman and M. Abujubbeh, "IoT-enabled smart grid via SM: An overview," *Future Gener. Comput. Syst.*, vol. 96, pp. 579–590, Jul. 2019, doi: 10.1016/j.future.2019.02.012.

[39] H. Lund, P. A. Østergaard, D. Connolly, and B. V. Mathiesen, "Smart energy and smart energy systems," *Energy*, vol. 137, pp. 556–565, Oct. 2017, doi: 10.1016/j.energy.2017.05.123.

[40] I. Malik, A. Bhardwaj, H. Bhardwaj, and A. Sakalle, "IoT-enabled smart homes: Architecture, challenges, and issues," in *Revolutionizing Industrial Automation Through the Convergence of Artificial Intelligence and the Internet of Things*. Hershey, PA, USA: IGI International, 2023, pp. 160–176, doi: 10.4018/978-1-6684-4991-2.ch008.

[41] W. Lee, S. Cho, P. Chu, H. Vu, S. Helal, W. Song, Y.-S. Jeong, and K. Cho, "Automatic agent generation for IoT-based smart house simulator," *Neurocomputing*, vol. 209, pp. 14–24, Oct. 2016, doi: 10.1016/j.neucom.2015.04.130.

[42] S. G. Dacko, "Enabling smart retail settings via mobile augmented reality shopping apps," *Technological Forecasting Social Change*, vol. 124, pp. 243–256, Nov. 2017, doi: 10.1016/j.techfore.2016.09.032.

[43] A. D. Gupta, P. Pandey, A. Feijóo, Z. M. Yaseen, and N. D. Bokde, "Smart water technology for efficient water resource management: A review," *Energies*, vol. 13, no. 23, p. 6268, Nov. 2020, doi: 10.3390/en13236268.

[44] J. Dong, G. Wang, H. Yan, J. Xu, and X. Zhang, "A survey of smart water quality monitoring system," *Environ. Sci. Pollut. Res.*, vol. 22, no. 7, pp. 4893–4906, Apr. 2015, doi: 10.1007/s11356-014-4026-x.

[45] F. Ghofrani, Q. He, R. M. P. Goverde, and X. Liu, "Recent applications of big data analytics in railway transportation systems: A survey," *Transp. Res. C, Emerg. Technol.*, vol. 90, pp. 226–246, May 2018, doi: 10.1016/j.trc.2018.03.010.

[46] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, "Big data analytics in intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 1, pp. 383–398, Jan. 2019, doi: 10.1109/TITS.2018.2815678.

[47] R. Yadav, W. Zhang, I. A. Elgendy, G. Dong, M. Shafiq, A. A. Laghari, and S. Prakash, "Smart healthcare: RL-based task offloading scheme for edge-enable sensor networks," *IEEE Sensors J.*, vol. 21, no. 22, pp. 24910–24918, Nov. 2021.

[48] E. F. Z. Santana, A. P. Chaves, M. A. Gerosa, F. Kon, and D. S. Milojicic, "Software platforms for smart cities: Concepts, requirements, challenges, and a unified reference architecture," *ACM Comput. Surveys*, vol. 50, no. 6, pp. 1–37, Nov. 2018, doi: 10.1145/3124391.

[49] R. Petrolo, V. Loscrì, and N. Mitton, "Towards a smart city based on cloud of things, a survey on the smart city vision and paradigms," *Trans. Emerg. Telecommun. Technol.*, vol. 28, no. 1, p. e2931, Jan. 2017, doi: 10.1002/ett.2931.

[50] A. H. Alavi, P. Jiao, W. G. Buttlar, and N. Lajnef, "Internet of Things-enabled smart cities: State-of-the-art and future trends," *Measurement*, vol. 129, pp. 589–606, Dec. 2018, doi: 10.1016/j.measurement.2018.07.067.

[51] A. Sharifi, "A critical review of selected smart city assessment tools and indicator sets," *J. Cleaner Prod.*, vol. 233, pp. 1269–1283, Oct. 2019, doi: 10.1016/j.jclepro.2019.06.172.

[52] H. H. Khan, M. N. Malik, R. Zafar, F. A. Goni, A. G. Chofreh, J. J. Klemeš, and Y. Alotaibi, "Challenges for sustainable smart city development: A conceptual framework," *Sustain. Develop.*, vol. 28, no. 5, pp. 1507–1518, Sep. 2020, doi: 10.1002/sd.2090.

[53] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, Oct. 2016, doi: 10.1109/JIOT.2016.2579198.

[54] F. Al-Turjman, H. Zahmatkesh, and R. Shahroze, "An overview of security and privacy in smart cities' IoT communications," *Trans. Emerg. Telecommun. Technol.*, vol. 33, no. 3, p. e3677, Mar. 2022, doi: 10.1002/ett.3677.

[55] M. De Gennaro, E. Paffumi, and G. Martini, "Big data for supporting low-carbon road transport policies in Europe: Applications, challenges and opportunities," *Big Data Res.*, vol. 6, pp. 11–25, Dec. 2016, doi: 10.1016/j.bdr.2016.04.003.

[56] A. Neilson, B. Daniel, and S. Tjandra, "Systematic review of the literature on big data in the transportation domain: Concepts and applications," *Big Data Res.*, vol. 17, pp. 35–44, Sep. 2019, doi: 10.1016/j.bdr.2019.03.001.

[57] A. R. Honarvar and A. Sami, "Towards sustainable smart city by particulate matter prediction using urban big data, excluding expensive air pollution infrastructures," *Big Data Res.*, vol. 17, pp. 56–65, Sep. 2019, doi: 10.1016/j.bdr.2018.05.006.

[58] F. Gu, B. Ma, J. Guo, P. A. Summers, and P. Hall, "Internet of Things and big data as potential solutions to the problems in waste electrical and electronic equipment management: An exploratory study," *Waste Manage.*, vol. 68, pp. 434–448, Oct. 2017, doi: 10.1016/j.wasman.2017.07.037.

[59] S. A. Shah, D. Z. Seker, S. Hameed, and D. Draheim, "The rising role of big data analytics and IoT in disaster management: Recent advances, taxonomy and prospects," *IEEE Access*, vol. 7, pp. 54595–54614, 2019, doi: 10.1109/ACCESS.2019.2913340.

[60] M. I. Khan, S. Khan, U. Khan, and A. Haleem, "Modeling the big data challenges in context of smart cities—An integrated fuzzy ISM-DEMATEL approach," *Int. J. Building Pathol. Adaptation*, vol. 41, no. 2, pp. 422–453, Apr. 2023, doi: 10.1108/IJBPA-02-2021-0027.

[61] W. Tian and Y. Zhao, "Big data technologies and cloud computing," in *Optimized Cloud Resource Management and Scheduling*. Amsterdam, The Netherlands: Elsevier, 2015, pp. 17–49, doi: 10.1016/b978-0-12-801476-9.00002-1.

[62] A. Botta, W. de Donato, V. Persico, and A. Pescapé, "Integration of cloud computing and Internet of Things: A survey," *Future Gener. Comput. Syst.*, vol. 56, pp. 684–700, Mar. 2016, doi: 10.1016/j.future.2015.09.021.

[63] A. Ahmad, A. Paul, M. M. Rathore, and H. Chang, "Smart cyber society: Integration of capillary devices with high usability based on cyber–physical system," *Future Gener. Comput. Syst.*, vol. 56, pp. 493–503, Mar. 2016, doi: 10.1016/j.future.2015.08.004.

[64] E. Curry, S. Dustdar, Q. Z. Sheng, and A. Sheth, "Smart cities–enabling services and applications," *J. Internet Services Appl.*, vol. 7, no. 1, pp. 1–3, Dec. 2016, doi: 10.1186/s13174-016-0048-6.

[65] M. I. Pramanik, R. Y. K. Lau, H. Demirkan, and M. A. K. Azad, "Smart health: Big data enabled health paradigm within smart cities," *Exp. Syst. Appl.*, vol. 87, pp. 370–383, Nov. 2017, doi: 10.1016/j.eswa.2017.06.027.

[66] M. Khan, M. Babar, S. H. Ahmed, S. C. Shah, and K. Han, "Smart city designing and planning based on big data analytics," *Sustain. Cities Soc.*, vol. 35, pp. 271–279, Nov. 2017, doi: 10.1016/j.scs.2017.07.012.

[67] M. Marjani, F. Nasaruddin, A. Gani, A. Karim, I. A. T. Hashem, A. Siddiqa, and I. Yaqoob, "Big IoT data analytics: Architecture, opportunities, and open research challenges," *IEEE Access*, vol. 5, pp. 5247–5261, 2017, doi: 10.1109/ACCESS.2017.2689040.

[68] D. Pal, T. Triyason, and P. Padungweang, "Big data in smart-cities: Current research and challenges," *Indonesian J. Electr. Eng. Informat.*, vol. 6, no. 4, pp. 351–360, Dec. 2018, doi: 10.11591/ijeei.v6i4.543.

[69] Y. Wu, W. Zhang, J. Shen, Z. Mo, and Y. Peng, "Smart city with Chinese characteristics against the background of big data: Idea, action and risk," *J. Cleaner Prod.*, vol. 173, pp. 60–66, Feb. 2018, doi: 10.1016/j.jclepro.2017.01.047.

[70] A. P. Plageras, K. E. Psannis, C. Stergiou, H. Wang, and B. B. Gupta, "Efficient IoT-based sensor BIG data collection–processing and analysis in smart buildings," *Future Gener. Comput. Syst.*, vol. 82, pp. 349–357, May 2018, doi: 10.1016/j.future.2017.09.082.

[71] C. Zbakh, M. Essaaidi, M. Manneback, and P. Rong, *Cloud Computing and Big Data: Technologies, Applications and Security*. Cham, Switzerland: Springer, 2019, doi: 10.1007/978-3-319-97719-5.

[72] Z. Allam and Z. A. Dhunny, "On big data, artificial intelligence and smart cities," *Cities*, vol. 89, pp. 80–91, Jun. 2019, doi: 10.1016/j.cities.2019.01.032.

[73] D. Jiang, "The construction of smart city information system based on the Internet of Things and cloud computing," *Comput. Commun.*, vol. 150, pp. 158–166, Jan. 2020, doi: 10.1016/j.comcom.2019.10.035.

[74] A. Kirimtat, O. Krejcar, A. Kertesz, and M. F. Tasgetiren, "Future trends and current state of smart city concepts: A survey," *IEEE Access*, vol. 8, pp. 86448–86467, 2020, doi: 10.1109/ACCESS.2020.2992441.

[75] A. Daissaoui, A. Boulmakoul, L. Karim, and A. Lbath, "IoT and big data analytics for smart buildings: A survey," *Proc. Comput. Sci.*, vol. 170, pp. 161–168, 2020, doi: 10.1016/j.procs.2020.03.021.

[76] M. Talebkhah, A. Sali, M. Marjani, M. Gordan, S. J. Hashim, and F. Z. Rokhani, "IoT and big data applications in smart cities: Recent advances, challenges, and critical issues," *IEEE Access*, vol. 9, pp. 55465–55484, 2021, doi: 10.1109/ACCESS.2021.3070905.

[77] N. Chai, C. Mao, M. Ren, W. Zhang, P. Poovendran, and P. Balamurugan, "Role of BIC (big data, IoT, and cloud) for smart cities," *Arabian J. Sci. Eng.*, vol. 48, p. 4115, Jul. 2021, doi: 10.1007/s13369-021-05888-x.

[78] M. T. Quasim, M. A. Khan, F. Algarni, and M. M. Alshahrani, "Fundamentals of smart cities," in *Smart Cities: A Data Analytics Perspective*. Cham, Switzerland: Springer, 2021, pp. 3–16, doi: 10.1007/978-3-030-60922-1_1.

[79] A. Lohachab, "Bootstrapping urban planning: Addressing big data issues in smart cities," in *Research Anthology on Big Data Analytics, Architectures, and Applications*. Hershey, PA, USA: IGI Global, 2022, pp. 1329–1358.

[80] X. Li, H. Liu, W. Wang, Y. Zheng, H. Lv, and Z. Lv, "Big data analysis of the Internet of Things in the digital twins of smart city based on deep learning," *Future Gener. Comput. Syst.*, vol. 128, pp. 167–177, Mar. 2022, doi: 10.1016/j.future.2021.10.006.

[81] K. Soomro, M. N. M. Bhutta, Z. Khan, and M. A. Tahir, "Smart city big data analytics: An advanced review," *WIREs Data Mining Knowl. Discovery*, vol. 9, no. 5, p. e1319, Sep. 2019, doi: 10.1002/widm.1319.

[82] M. I. Pramanik, R. Y. K. Lau, Md S. Hossain, Md M. Rahoman, S. K. Debnath, Md G. Rashed, and Md Z. Uddin, "Privacy preserving big data analytics: A critical analysis of state-of-the-art," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 11, no. 1, p. e1387, 2021, doi: 10.1002/widm.1387.

[83] C. J. Pettit, S. Zarpelon Leao, O. Lock, M. Ng, and J. Reades, "Big data: The engine to future Cities—A reflective case study in urban transport," *Sustainability*, vol. 14, no. 3, p. 1727, Feb. 2022, doi: 10.3390/su14031727.

[84] J. W. Rittinghouse and J. F. Ransome, *Cloud Computing: Implementation, Management, and Security*. Boca Raton, FL, USA: CRC Press, 2016.

[85] P. Wang, L. T. Yang, and J. Li, "An edge cloud-assisted CPSS framework for smart city," *IEEE Cloud Comput.*, vol. 5, no. 5, pp. 37–46, Sep. 2018, doi: 10.1109/MCC.2018.053711665.

[86] T. Alam, "Cloud-based IoT applications and their roles in smart cities," *Smart Cities*, vol. 4, no. 3, pp. 1196–1219, Sep. 2021, doi: 10.3390/smartcities4030064.

[87] T. Alam, M. Tajammul, and R. Gupta, "Towards the sustainable development of smart cities through cloud computing," in *AI and IoT for Smart City Applications*. Singapore: Springer, 2022, pp. 199–222, doi: 10.1007/978-981-16-7498-3_13.

[88] K. Jamsa, *Cloud Computing*. Burlington, MA, USA: Jones & Bartlett Learning, 2022.

[89] D. Mourtzis, E. Vlachou, and N. Milas, "Industrial big data as a result of IoT adoption in manufacturing," *Proc. CIRP*, vol. 55, pp. 290–295, 2016, doi: 10.1016/j.procir.2016.07.038.

[90] F. Alam, R. Mehmood, I. Katib, N. N. Albogami, and A. Albeshri, "Data fusion and IoT for smart ubiquitous environments: A survey," *IEEE Access*, vol. 5, pp. 9533–9554, 2017, doi: 10.1109/ACCESS.2017.2697839.

[91] J. M. Talavera, L. E. Tobón, J. A. Gómez, M. A. Culman, J. M. Aranda, D. T. Parra, L. A. Quiroz, A. Hoyos, and L. E. Garreta, "Review of IoT applications in agro-industrial and environmental fields," *Comput. Electron. Agricult.*, vol. 142, pp. 283–297, Nov. 2017, doi: 10.1016/j.compag.2017.09.015.

[92] M. A. A. Mamun and M. R. Yuce, "Sensors and systems for wearable environmental monitoring toward IoT-enabled applications: A review," *IEEE Sensors J.*, vol. 19, no. 18, pp. 7771–7788, Sep. 2019, doi: 10.1109/JSEN.2019.2919352.

[93] V. K. Quy, N. V. Hau, D. V. Anh, N. M. Quy, N. T. Ban, S. Lanza, G. Randazzo, and A. Muzirafuti, "IoT-enabled smart agriculture: Architecture, applications, and challenges," *Appl. Sci.*, vol. 12, no. 7, p. 3396, Mar. 2022, doi: 10.3390/app12073396.

[94] A. S. Syed, D. Sierra-Sosa, A. Kumar, and A. Elmaghraby, "IoT in smart cities: A survey of technologies, practices and challenges," *Smart Cities*, vol. 4, no. 2, pp. 429–475, Mar. 2021, doi: 10.3390/smartcities4020024.

[95] E. E. Cranmer, M. Papalexi, M. C. T. Dieck, and D. Bamford, "Internet of Things: Aspiration, implementation and contribution," *J. Bus. Res.*, vol. 139, pp. 69–80, Feb. 2022, doi: 10.1016/j.jbusres.2021.09.025.

[96] D. P. Abreu, K. Velasquez, M. Curado, and E. Monteiro, "A resilient Internet of Things architecture for smart cities," *Ann. Telecommun.*, vol. 72, nos. 1–2, pp. 19–30, Feb. 2017, doi: 10.1007/s12243-016-0530-y.

[97] S. Chang, N. Saha, D. Castro-Lacouture, and P. P.-J. Yang, "Multivariate relationships between campus design parameters and energy performance using reinforcement learning and parametric modeling," *Appl. Energy*, vol. 249, pp. 253–264, Sep. 2019, doi: 10.1016/j.apenergy.2019.04.109.

[98] K. Wang, H. Yin, W. Quan, and G. Min, "Enabling collaborative edge computing for software defined vehicular networks," *IEEE Netw.*, vol. 32, no. 5, pp. 112–117, Sep. 2018, doi: 10.1109/MNET.2018.1700364.

[99] P. Chamoso, A. González-Briones, S. Rodríguez, and J. M. Corchado, "Tendencies of technologies and platforms in smart cities: A State-of-the-Art review," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–17, Aug. 2018, doi: 10.1155/2018/3086854.

[100] V. Albino, U. Berardi, and R. M. Dangelico, "Smart cities: Definitions, dimensions, performance, and initiatives," *J. Urban Technol.*, vol. 22, no. 1, pp. 3–21, Jan. 2015, doi: 10.1080/10630732.2014.942092.

[101] R. Sánchez-Corcuera and A. Nuñez-Marcos, "Smart cities survey: Technologies, application domains and challenges for the cities of the future," *Int. J. Distrib. Sens. Netw.*, vol. 15, no. 6, 2019, Art. no. 1550147719853984, doi: 10.1177/1550147719853984.

[102] E. Costales, "Identifying sources of innovation: Building a conceptual framework of the smart city through a social innovation perspective," *Cities*, vol. 120, Jan. 2022, Art. no. 103459, doi: 10.1016/j.cities.2021.103459.

[103] H. K. Channi and R. Kumar, "The role of smart sensors in smart city," in *Smart Sensor Networks*. Cham, Switzerland: Springer, 2022, pp. 27–48, doi: 10.1007/978-3-030-77214-7_2.

[104] D. J. Langley, J. van Doorn, I. C. L. Ng, S. Stieglitz, A. Lazovik, and A. Boonstra, "The Internet of Everything: Smart things and their impact on business models," *J. Bus. Res.*, vol. 122, pp. 853–863, Jan. 2021, doi: 10.1016/j.jbusres.2019.12.035.

[105] S. P. Caird and S. H. Hallett, "Towards evaluation design for smart city development," *J. Urban Design*, vol. 24, no. 2, pp. 188–209, Mar. 2019, doi: 10.1080/13574809.2018.1469402.

[106] A. Caragliu, C. D. Bo, and P. Nijkamp, "Smart cities in Europe," *J. Urban Technol.*, vol. 18, no. 2, pp. 65–82, Apr. 2011, doi: 10.1080/10630732.2011.601117.

[107] C. K. Toh, J. A. Sanguesa, J. C. Cano, and F. J. Martinez, "Advances in smart roads for future smart cities," *Proc. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 476, no. 2233, Jan. 2020, Art. no. 20190439, doi: 10.1098/rspa.2019.0439.

[108] S. Bhattacharyya, S. Banerjee, I. Bose, and A. Kankanhalli, "Temporal effects of repeated recognition and lack of recognition on online community contributions," *J. Manage. Inf. Syst.*, vol. 37, no. 2, pp. 536–562, Apr. 2020, doi: 10.1080/07421222.2020.1759341.

[109] T. Anagnostopoulos, C. Anagnostopoulos, S. Behdad, and I. Salmon, "IoT-enabled tip and swap waste management models for smart cities," *Int. J. Environ. Waste Manag.*, vol. 28, no. 4, pp. 521–539, 2020.

[110] S. Zygiaris, "Smart city reference model: Assisting planners to conceptualize the building of smart city innovation ecosystems," *J. Knowl. Economy*, vol. 4, no. 2, pp. 217–231, Jun. 2013, doi: 10.1007/s13132-012-0089-4.

[111] A. Herdiyanti, P. S. Hapsari, and T. D. Susanto, "Modelling the smart governance performance to support smart city program in Indonesia," *Proc. Comput. Sci.*, vol. 161, pp. 367–377, 2019, doi: 10.1016/j.procs.2019.11.135.

[112] M. J. P. L. dos Santos, "Smart cities and urban areas—Aquaponics as innovative urban agriculture," *Urban Forestry Urban Greening*, vol. 20, pp. 402–406, Dec. 2016, doi: 10.1016/j.ufug.2016.10.004.

[113] K. Kourtit and P. Nijkamp, "Smart cities in the innovation age," *Innovation, Eur. J. Social Sci. Res.*, vol. 25, no. 2, pp. 93–95, Jun. 2012, doi: 10.1080/13511610.2012.660331.

[114] M.-L. Marsal-Llacuna, J. Colomer-Llinàs, and J. Meléndez-Frigola, "Lessons in urban monitoring taken from sustainable and livable cities to better address the smart cities initiative," *Technological Forecasting Social Change*, vol. 90, pp. 611–622, Jan. 2015, doi: 10.1016/j.techfore.2014.01.012.

[115] M. Peris-Ortiz and D. R. Bennett, *Sustainable Smart Cities*. Hoboken, NJ, USA: Wiley, 2015.

[116] T. Yigitcanlar, "Smart cities: An effective urban development and management model?" *Austral. Planner*, vol. 52, no. 1, pp. 27–34, Jan. 2015, doi: 10.1080/07293682.2015.1019752.

[117] M. S. H. Abdi, "Smart city: A review on concepts, definitions, standards, experiments, and challenges," *J. Energy Manag. Technol.*, vol. 4, no. 3, pp. 1–6, 2020, doi: dx.doi.org/10.22109/jemt.2020.206444.1205.

[118] F. Mofidi and H. Akbari, "Intelligent buildings: An overview," *Energy Buildings*, vol. 223, Sep. 2020, Art. no. 110192, doi: 10.1016/j.enbuild.2020.110192.

[119] D. Kim, Y. Yoon, J. Lee, P. J. Mago, K. Lee, and H. Cho, "Design and implementation of smart buildings: A review of current research trend," *Energies*, vol. 15, no. 12, p. 4278, Jun. 2022, doi: 10.3390/en15124278.

[120] J. B. Awotunde, R. G. Jimoh, R. O. Ogundokun, S. Misra, and O. C. Abikoye, "Big data analytics of IoT-based cloud system framework: Smart healthcare monitoring systems," in *Artificial Intelligence for Cloud and Edge Computing*. Cham, Switzerland: Springer, 2022, pp. 181–208, doi: 10.1007/978-3-030-80821-1_9.

[121] W. Li, Y. Chai, F. Khan, S. R. U. Jan, S. Verma, V. G. Menon, and X. Li, "A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system," *Mobile Netw. Appl.*, vol. 26, no. 1, pp. 234–252, Feb. 2021, doi: 10.1007/s11036-020-01700-6.

[122] Y. Liu, J. J. Q. Yu, J. Kang, D. Niyato, and S. Zhang, "Privacy-preserving traffic flow prediction: A federated learning approach," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7751–7763, Aug. 2020, doi: 10.1109/JIOT.2020.2991401.

[123] S. R. Chohan and G. Hu, "Success factors influencing citizens' adoption of IoT service orchestration for public value creation in smart government," *IEEE Access*, vol. 8, pp. 208427–208448, 2020, doi: 10.1109/ACCESS.2020.3036054.

[124] M. Sheha, K. Mohammadi, and K. Powell, "Techno-economic analysis of the impact of dynamic electricity prices on solar penetration in a smart grid environment with distributed energy storage," *Appl. Energy*, vol. 282, Jan. 2021, Art. no. 116168, doi: 10.1016/j.apenergy.2020.116168.

[125] E. Sarker, P. Halder, M. Seyedmahmoudian, E. Jamei, B. Horan, S. Mekhilef, and A. Stojcevski, "Progress on the demand side management in smart grid and optimization approaches," *Int. J. Energy Res.*, vol. 45, no. 1, pp. 36–64, Jan. 2021, doi: 10.1002/er.5631.

[126] D. Mocrii, Y. Chen, and P. Musilek, "IoT-based smart homes: A review of system architecture, software, communications, privacy and security," *Internet Things*, vols. 1–2, pp. 81–98, Sep. 2018, doi: 10.1016/j.iot.2018.08.009.

[127] V. M. Schönberger and T. Ramge, *Reinventing Capitalism in the Age of Big Data*. Paris, France: Hachette, 2018.

[128] M. Salman, S. Ozdemir, and F. Celebi, "Fuzzy traffic control with vehicle-to-everything communication," *Sensors*, vol. 18, no. 2, p. 368, Jan. 2018, doi: 10.3390/s18020368.

[129] R. Xu, H. Xiang, Z. Tu, X. Xia, M.-H. Yang, and J. Ma, "V2X-ViT: Vehicle-to-everything cooperative perception with vision transformer," 2022, *arXiv:2203.10638*.

[130] C. You, J. Lu, D. Filev, and P. Tsiotras, "Advanced planning for autonomous vehicles using reinforcement learning and deep inverse reinforcement learning," *Robot. Auto. Syst.*, vol. 114, pp. 1–18, Apr. 2019, doi: 10.1016/j.robot.2019.01.003.

[131] G. C. Oatley, "Themes in data mining, big data, and crime analytics," *WIREs Data Mining Knowl. Discovery*, vol. 12, no. 2, p. e1432, Mar. 2022.

[132] A. Karvonen, M. Cook, and H. Haarstad, "Urban planning and the smart city: Projects, practices and politics," *Urban Planning*, vol. 5, no. 1, pp. 65–68, Mar. 2020, doi: 10.17645/UP.V5I1.2936.

[133] S. E. Bibri, "The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability," *Sustain. Cities Soc.*, vol. 38, pp. 230–253, Apr. 2018, doi: 10.1016/j.scs.2017.12.034.

[134] B. P. Bhattarai, S. Paudyal, Y. Luo, M. Mohanpurkar, K. Cheung, R. Tonkoski, R. Hovsapian, K. S. Myers, R. Zhang, P. Zhao, M. Manic, S. Zhang, and X. Zhang, "Big data analytics in smart grids: State-of-the-art, challenges, opportunities, and future directions," *IET Smart Grid*, vol. 2, no. 2, pp. 141–154, Jun. 2019, doi: 10.1049/iet-stg.2018.0261.

[135] R. G. Hollands, "Will the real smart city please stand up?" in *The Routledge Companion to Smart Cities*. London, U.K.: Routledge, 2020.

[136] S. Sharma, G. Saini, K. Kumar, and K. Saini, *Futuristic Approach to Energy in Smart Cities*. Boca Raton, FL, USA: CRC Press, 2022, doi: 10.1201/9781003287186-14.

[137] Z. Ullah, F. Al-Turjman, L. Mostarda, and R. Gagliardi, "Applications of artificial intelligence and machine learning in smart cities," *Comput. Commun.*, vol. 154, pp. 313–323, Mar. 2020, doi: 10.1016/j.comcom.2020.02.069.

[138] A. Gharaibeh, M. A. Salahuddin, S. J. Hussini, A. Khreishah, I. Khalil, M. Guizani, and A. Al-Fuqaha, "Smart cities: A survey on data management, security, and enabling technologies," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2456–2501, 4th Quart., 2017, doi: 10.1109/COMST.2017.2736886.

[139] C. Badii, P. Bellini, A. Difino, and P. Nesi, "Smart city IoT platform respecting GDPR privacy and security aspects," *IEEE Access*, vol. 8, pp. 23601–23623, 2020, doi: 10.1109/ACCESS.2020.2968741.

[140] P. R. Krishnan and J. Jacob, "An IoT based efficient energy management in smart grid using DHOCSA technique," *Sustain. Cities Soc.*, vol. 79, Apr. 2022, Art. no. 103727, doi: 10.1016/j.scs.2022.103727.

[141] A. Sharifi, S. Kawakubo, and A. Milovidona, "Urban sustainability assessment tools: Toward integrating smart city indicators," in *Urban Systems Design*. Amsterdam, The Netherlands: Elsevier, 2020, pp. 345–372, doi: doi.org/10.1016/B978-0-12-816055-8.00011-7.

[142] L. Anthopoulos and V. Kazantzi, "Urban energy efficiency assessment models from an AI and big data perspective: Tools for policy makers," *Sustain. Cities Soc.*, vol. 76, Jan. 2022, Art. no. 103492, doi: 10.1016/j.scs.2021.103492.

[143] M. H. Kashani, A. M. Rahmani, and N. J. Navimipour, "Quality of service-aware approaches in fog computing," *Int. J. Commun. Syst.*, vol. 33, no. 8, p. e4340, May 2020, doi: 10.1002/dac.4340.

[144] W. Chen, B. Liu, H. Huang, S. Guo, and Z. Zheng, "When UAV swarm meets edge-cloud computing: The QoS perspective," *IEEE Netw.*, vol. 33, no. 2, pp. 36–43, Mar. 2019, doi: 10.1109/MNET.2019.1800222.

[145] R. Iqbal, F. Doctor, B. More, S. Mahmud, and U. Yousuf, "Big data analytics: Computational intelligence techniques and application areas," *Technological Forecasting Social Change*, vol. 153, Apr. 2020, Art. no. 119253, doi: 10.1016/j.techfore.2018.03.024.

[146] A. Kumar, M. Payal, P. Dixit, and J. M. Chatterjee, "Framework for realization of green smart cities through the Internet of Things (IoT)," in *Trends in Cloud-based IoT*. 2020, pp. 85–111, doi: 10.1007/978-3-030-40037-8_6.

[147] M. A. Javed, S. Zeadally, and E. B. Hamida, "Data analytics for cooperative intelligent transport systems," *Veh. Commun.*, vol. 15, pp. 63–72, Jan. 2019, doi: 10.1016/j.vehcom.2018.10.004.

[148] R. K. R. Kummitha, "Smart cities and entrepreneurship: An agenda for future research," *Technological Forecasting Social Change*, vol. 149, Dec. 2019, Art. no. 119763, doi: 10.1016/j.techfore.2019.119763.

[149] D. Li, J. Cao, and Y. Yao, "Big data in smart cities," *Sci. China Inf. Sci.*, vol. 58, no. 10, pp. 1–12, Oct. 2015, doi: 10.1007/s11432-015-5396-5.

[150] R. Arghandeh and Y. Zhou, *Big Data Application in Power Systems*. Amsterdam, The Netherlands: Elsevier, 2017, doi: 10.1016/c2016-0-00194-8.

[151] J. D. Miller, *Big Data Visualization*. Birmingham, U.K.: Packt, 2017.

[152] B. Marr, *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results*. Hoboken, NJ, USA: Wiley, 2016.

[153] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: Management, analysis and future prospects," *J. Big Data*, vol. 6, no. 1, pp. 1–25, Dec. 2019, doi: 10.1186/s40537-019-0217-0.

[154] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "The rise of 'big data' on cloud computing: Review and open research issues," *Inf. Syst.*, vol. 47, pp. 98–115, Jan. 2015, doi: 10.1016/j.is.2014.07.006.

[155] J. J. Berman. (2013). *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information*. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780124045767000083

[156] D. E. O'Leary, "Artificial intelligence and big data," *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 96–99, Mar. 2013, doi: 10.1109/MIS.2013.39.

[157] M. S. Mahdavinejad, M. Rezvan, M. Barekatain, P. Adibi, P. Barnaghi, and A. P. Sheth, "Machine learning for Internet of Things data analysis: A survey," *Digit. Commun. Netw.*, vol. 4, no. 3, pp. 161–175, Aug. 2018, doi: 10.1016/j.dcan.2017.10.002.

[158] Z.-H. Zhou, N. V. Chawla, Y. Jin, and G. J. Williams, "Big data opportunities and challenges: Discussions from data analytics perspectives [discussion forum]," *IEEE Comput. Intell. Mag.*, vol. 9, no. 4, pp. 62–74, Nov. 2014.

[159] N. Zhang, P. Yang, J. Ren, D. Chen, L. Yu, and X. Shen, "Synergy of big data and 5G wireless networks: Opportunities, approaches, and challenges," *IEEE Wireless Commun.*, vol. 25, no. 1, pp. 12–18, Feb. 2018, doi: 10.1109/MWC.2018.1700193.

[160] S. Chauhan, N. Agarwal, and A. K. Kar, "Addressing big data challenges in smart cities: A systematic literature review," *Eletronic Library*, vol. 18, no. 4, pp. 73–90, Jun. 2016.

[161] H. N. Dai, R. C. W. Wong, H. Wang, Z. Zheng, and A. V. Vasilakos, "Big data analytics for large-scale wireless networks: Challenges and opportunities," *ACM Comput. Surv.*, vol. 52, no. 5, pp. 1–36, 2019, doi: 10.1145/3337065.

[162] D. Zhang, L. G. Pee, S. L. Pan, and L. Cui, "Big data analytics, resource orchestration, and digital sustainability: A case study of smart city development," *Government Inf. Quart.*, vol. 39, no. 1, Jan. 2022, Art. no. 101626, doi: 10.1016/j.giq.2021.101626.

[163] A. Alsaig, V. Alagar, Z. Chammaa, and N. Shiri, "Characterization and efficient management of big data in IoT-driven smart city development," *Sensors*, vol. 19, no. 11, p. 2430, May 2019, doi: 10.3390/s19112430.

[164] K. Abouelmehdi, A. Beni-Hssane, H. Khaloufi, and M. Saadi, "Big data security and privacy in healthcare: A review," *Proc. Comput. Sci.*, vol. 113, pp. 73–80, 2017, doi: 10.1016/j.procs.2017.08.292.

[165] L. A. Tawalbeh and G. Saldamli, "Reconsidering big data security and privacy in cloud and mobile cloud systems," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 33, no. 7, pp. 810–819, Sep. 2021, doi: 10.1016/j.jksuci.2019.05.007.

[166] X. Jin, B. W. Wah, X. Cheng, and Y. Wang, "Significance and challenges of big data research," *Big Data Res.*, vol. 2, no. 2, pp. 59–64, Jun. 2015, doi: 10.1016/j.bdr.2015.01.006.

[167] M. Naeem, T. Jamal, J. Diaz-Martinez, S. A. Butt, N. Montesano, and M. I. Tariq, "Trends and future perspective challenges in big data," in *Advances in Intelligent Data Analysis and Applications*, vol. 253. Arad, Romania: Springer, 2022, pp. 309–325, doi: 10.1007/978-981-16-5036-9_30.

[168] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of big data challenges and analytical methods," *J. Bus. Res.*, vol. 70, pp. 263–286, Jan. 2017, doi: 10.1016/j.jbusres.2016.08.001.

[169] M. I. Baig, L. Shuib, and E. Yadegaridehkordi, "Big data adoption: State of the art and research challenges," *Inf. Process. Manage.*, vol. 56, no. 6, Nov. 2019, Art. no. 102095, doi: 10.1016/j.ipm.2019.102095.

[170] R. Kune, P. K. Konugurthi, A. Agarwal, R. R. Chillarige, and R. Buyya, "The anatomy of big data computing," *Software: Pract. Exper.*, vol. 46, no. 1, pp. 79–105, Jan. 2016, doi: 10.1002/spe.2374.

[171] B. Diène, J. J. P. C. Rodrigues, O. Diallo, E. H. M. Ndoye, and V. V. Korotaev, "Data management techniques for Internet of Things," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106564, doi: 10.1016/j.ymssp.2019.106564.

[172] Y. Chen, H. Chen, A. Gorkhali, Y. Lu, Y. Ma, and L. Li, "Big data analytics and big data science: A survey," *J. Manage. Analytics*, vol. 3, no. 1, pp. 1–42, Jan. 2016, doi: 10.1080/23270012.2016.1141332.

[173] A. Siddiqa, I. A. T. Hashem, I. Yaqoob, M. Marjani, S. Shamshirband, A. Gani, and F. Nasaruddin, "A survey of big data management: Taxonomy and state-of-the-art," *J. Netw. Comput. Appl.*, vol. 71, pp. 151–166, Aug. 2016, doi: 10.1016/j.jnca.2016.04.008.

[174] C. Lim, K.-J. Kim, and P. P. Maglio, "Smart cities with big data: Reference models, challenges, and considerations," *Cities*, vol. 82, pp. 86–99, Dec. 2018, doi: 10.1016/j.cities.2018.04.011.

[175] A. A. Guenduez, S. Singler, T. Tomczak, K. Schedler, and M. Oberli, "Smart government success factors," *Jahrbuch der Schweizerischen Verwaltungswissenschaften*, vol. 9, no. 1, pp. 96–110, 2018, doi: 10.5334/ssas.124.

[176] S. Alotaibi, R. Mehmood, and I. Katib, "The role of big data and Twitter data analytics in healthcare supply chain management," in *Smart Infrastructure and Applications*. Cham, Switzerland: Springer, 2020, pp. 267–279, doi: 10.1007/978-3-030-13705-2_11.

[177] A. Goswami and A. Kumar, "Challenges in the analysis of online social networks: A data collection tool perspective," *Wireless Pers. Commun.*, vol. 97, no. 3, pp. 4015–4061, Dec. 2017, doi: 10.1007/s11277-017-4712-3.

[178] T. H. McCormick, H. Lee, N. Cesare, A. Shojaie, and E. S. Spiro, "Using Twitter for demographic and social science research: Tools for data collection and processing," *Sociol. Methods Res.*, vol. 46, no. 3, pp. 390–421, Aug. 2017, doi: 10.1177/0049124115605339.

[179] S. De, S. Dey, S. Bhatia, and S. Bhattacharyya, "An introduction to data mining in social networks," in *Advanced Data Mining Tools and Methods for Social Computing*. New York, NY, USA: Academic, 2022, pp. 1–25.

[180] J. Ranjan and C. Foropon, "Big data analytics in building the competitive intelligence of organizations," *Int. J. Inf. Manage.*, vol. 56, Feb. 2021, Art. no. 102231, doi: 10.1016/j.ijinfomgt.2020.102231.

[181] I. Kotenko, A. Krasov, I. Ushakov, and K. Izrailov, "An approach for Stego-insider detection based on a hybrid NoSQL database," *J. Sensor Actuator Netw.*, vol. 10, no. 2, p. 25, Mar. 2021.

[182] J. Schwaiger, T. Hammerl, J. Florian, and S. Leist, "UR: SMART—A tool for analyzing social media content," *Inf. Syst. e-Business Manage.*, vol. 19, no. 4, pp. 1275–1320, Dec. 2021, doi: 10.1007/s10257-021-00541-4.

[183] N. Ribelles, J. M. Jerez, P. Rodriguez-Brazzarola, B. Jimenez, T. Diaz-Redondo, H. Mesa, A. Marquez, A. Sanchez-Muñoz, B. Pajares, F. Carabantes, M. J. Bermejo, E. Villar, M. E. Dominguez-Recio, E. Saez, L. Galvez, A. Godoy, L. Franco, S. Ruiz-Medina, I. Lopez, and E. Alba, "Machine learning and natural language processing (NLP) approach to predict early progression to first-line treatment in real-world hormone receptor-positive (HR+)/HER2-negative advanced breast cancer patients," *Eur. J. Cancer*, vol. 144, pp. 224–231, Feb. 2021, doi: 10.1016/j.ejca.2020.11.030.

[184] G. R. Halegoua, *Smart Cities*. Cambridge, MA, USA: MIT Press, 2020, doi: 10.1201/9781003027799-3.

[185] J. A. M. Hayajneh, M. B. H. Elayan, M. A. M. Abdellatif, and A. M. Abubakar, "Impact of business analytics and π-shaped skills on innovative performance: Findings from PLS-SEM and fsQCA," *Technol. Soc.*, vol. 68, Feb. 2022, Art. no. 101914.

[186] H. Zhang, D. van Berkel, P. D. Howe, Z. D. Miller, and J. W. Smith, "Using social media to measure and map visitation to public lands in Utah," *Appl. Geography*, vol. 128, Mar. 2021, Art. no. 102389, doi: 10.1016/j.apgeog.2021.102389.

[187] D. McGilvray, *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information (TM)*. New York, NY, USA: Academic, 2021. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780123743695500052

[188] S. Stieglitz, M. Mirbabaie, B. Ross, and C. Neuberger, "Social media analytics–challenges in topic discovery, data collection, and data preparation," *Int. J. Inf. Manage.*, vol. 39, pp. 156–168, Apr. 2018, doi: 10.1016/j.ijinfomgt.2017.12.002.

[189] Y. Pan and L. Zhang, "Roles of artificial intelligence in construction engineering and management: A critical review and future trends," *Autom. Construction*, vol. 122, Feb. 2021, Art. no. 103517, doi: 10.1016/j.autcon.2020.103517.

[190] M. Majid, S. Habib, A. R. Javed, M. Rizwan, G. Srivastava, T. R. Gadekallu, and J. C.-W. Lin, "Applications of wireless sensor networks and Internet of Things frameworks in the industry revolution 4.0: A systematic literature review," *Sensors*, vol. 22, no. 6, p. 2087, Mar. 2022, doi: 10.3390/s22062087.

[191] M. Nahri, A. Boulmakoul, L. Karim, and A. Lbath, "IoV distributed architecture for real-time traffic data analytics," *Proc. Comput. Sci.*, vol. 130, pp. 480–487, 2018, doi: 10.1016/j.procs.2018.04.055.

[192] J. Wang, D. Feng, S. Zhang, J. Tang, and T. Q. S. Quek, "Computation offloading for mobile edge computing enabled vehicular networks," *IEEE Access*, vol. 7, pp. 62624–62632, 2019, doi: 10.1109/ACCESS.2019.2915959.

[193] J. Souza, A. Francisco, C. Piekarski, and G. Prado, "Data mining and machine learning to promote smart cities: A systematic review from 2000 to 2018," *Sustainability*, vol. 11, no. 4, p. 1077, Feb. 2019, doi: 10.3390/su11041077.

[194] W. Cui, "Visual analytics: A comprehensive overview," *IEEE Access*, vol. 7, pp. 81555–81573, 2019, doi: 10.1109/ACCESS.2019.2923736.

[195] D. Assante, M. Castro, I. Hamburg, and S. Martin, "The use of cloud computing in SMEs," *Proc. Comput. Sci.*, vol. 83, pp. 1207–1212, 2016, doi: 10.1016/j.procs.2016.04.250.

[196] S. Akintoye and A. Bagula, "Improving quality-of-service in cloud/fog computing through efficient resource allocation," *Sensors*, vol. 19, no. 6, p. 1267, Mar. 2019, doi: 10.3390/s19061267.

[197] A. Sharma and G. Singh, "A review of scheduling algorithms in Hadoop," in *Proc. ICRIC*, 2020, pp. 125–135.

[198] H. Nasiri, S. Nasehi, and M. Goudarzi, "Evaluation of distributed stream processing frameworks for IoT applications in smart cities," *J. Big Data*, vol. 6, no. 1, pp. 1–24, Dec. 2019, doi: 10.1186/s40537-019-0215-2.

[199] Z. Khan, S. L. Kiani, and K. Soomro, "A framework for cloud-based context-aware information services for citizens in smart cities," *J. Cloud Comput.*, vol. 3, pp. 1–17, Oct. 2014, doi: 10.1186/preaccept-1562408279114774.

[200] H. A. S. Ahmed, M. H. Ali, L. M. Kadhum, M. F. Bin Zolkipli, and Y. A. Alsariera, "A review of challenges and security risks of cloud computing," *J. Telecommun. Electron. Comput. Eng.*, vol. 9, nos. 1–2, pp. 87–91, 2017.

[201] M. Ghobaei-Arani, A. Souri, and A. A. Rahmanian, "Resource management approaches in fog computing: A comprehensive review," *J. Grid Comput.*, vol. 18, no. 1, pp. 1–42, Mar. 2020, doi: 10.1007/s10723-019-09491-1.

[202] N. Iqbal, A.-N. Khan, A. Rizwan, F. Qayyum, S. Malik, R. Ahmad, and D.-H. Kim, "Enhanced time-constraint aware tasks scheduling mechanism based on predictive optimization for efficient load balancing in smart manufacturing," *J. Manuf. Syst.*, vol. 64, pp. 19–39, Jul. 2022, doi: 10.1016/j.jmsy.2022.05.015.

[203] P. Wang, C. Yao, Z. Zheng, G. Sun, and L. Song, "Joint task assignment, transmission, and computing resource allocation in multilayer mobile edge computing systems," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2872–2884, Apr. 2019, doi: 10.1109/JIOT.2018.2876198.

[204] R. Basir, S. Qaisar, M. Ali, M. Aldwairi, M. I. Ashraf, A. Mahmood, and M. Gidlund, "Fog computing enabling industrial Internet of Things: State-of-the-art and research challenges," *Sensors*, vol. 19, no. 21, p. 4807, Nov. 2019, doi: 10.3390/s19214807.

[205] E. Jafarnejad Ghomi, A. Masoud Rahmani, and N. Nasih Qader, "Load-balancing algorithms in cloud computing: A survey," *J. Netw. Comput. Appl.*, vol. 88, pp. 50–71, Jun. 2017, doi: 10.1016/j.jnca.2017.04.007.

[206] D. A. Shafiq, N. Z. Jhanjhi, A. Abdullah, and M. A. Alzain, "A load balancing algorithm for the data centres to optimize cloud computing applications," *IEEE Access*, vol. 9, pp. 41731–41744, 2021, doi: 10.1109/ACCESS.2021.3065308.

[207] L. Zhang, L. Zhou, and A. Salah, "Efficient scientific workflow scheduling for deadline-constrained parallel tasks in cloud computing environments," *Inf. Sci.*, vol. 531, pp. 31–46, Aug. 2020, doi: 10.1016/j.ins.2020.04.039.

[208] H. Aydin, "A study of cloud computing adoption in universities as a guideline to cloud migration," *SAGE Open*, vol. 11, no. 3, 2021, Art. no. 21582440211030280, doi: 10.1177/21582440211030280.

[209] M. F. Gholami, F. Daneshgar, G. Low, and G. Beydoun, "Cloud migration process—A survey, evaluation framework, and open challenges," *J. Syst. Softw.*, vol. 120, pp. 31–69, Oct. 2016, doi: 10.1016/j.jss.2016.06.068.

[210] S. Chenthara, K. Ahmed, H. Wang, and F. Whittaker, "Security and privacy-preserving challenges of e-health solutions in cloud computing," *IEEE Access*, vol. 7, pp. 74361–74382, 2019, doi: 10.1109/ACCESS.2019.2919982.

[211] N. Agrawal and S. Tapaswi, "A trustworthy agent-based encrypted access control method for mobile cloud computing environment," *Pervas. Mobile Comput.*, vol. 52, pp. 13–28, Jan. 2019, doi: 10.1016/j.pmcj.2018.11.003.

[212] D. Prabakaran and S. Ramachandran, "Multi-factor authentication for secured financial transactions in cloud environment," *Comput., Mater. Continua*, vol. 70, no. 1, pp. 1781–1798, 2022, doi: 10.32604/cmc.2022.019591.

[213] K. R. Sajay, S. S. Babu, and Y. Vijayalakshmi, "Enhancing the security of cloud data using hybrid encryption algorithm," *J. Ambient Intell. Hum. Comput.*, no. 2018, pp. 1–10, Jul. 2019, doi: 10.1007/s12652-019-01403-1.

[214] S. A. Bello, L. O. Oyedele, O. O. Akinade, M. Bilal, J. M. Davila Delgado, L. A. Akanbi, A. O. Ajayi, and H. A. Owolabi, "Cloud computing in construction industry: Use cases, benefits and challenges," *Autom. Construction*, vol. 122, Feb. 2021, Art. no. 103441, doi: 10.1016/j.autcon.2020.103441.

[215] C. Li, J. Bai, Y. Ge, and Y. Luo, "Heterogeneity-aware elastic provisioning in cloud-assisted edge computing systems," *Future Gener. Comput. Syst.*, vol. 112, pp. 1106–1121, Nov. 2020, doi: 10.1016/j.future.2020.06.022.

[216] B. Asvija, R. Eswari, and M. B. Bijoy, "Security in hardware assisted virtualization for cloud computing—State of the art issues and challenges," *Comput. Netw.*, vol. 151, pp. 68–92, Mar. 2019, doi: 10.1016/j.comnet.2019.01.013.

[217] J. Moura and D. Hutchison, "Review and analysis of networking challenges in cloud computing," *J. Netw. Comput. Appl.*, vol. 60, pp. 113–129, Jan. 2016, doi: 10.1016/j.jnca.2015.11.015.

[218] J. Dizdarević, F. Carpio, A. Jukan, and X. Masip-Bruin, "A survey of communication protocols for Internet of Things and related challenges of fog and cloud computing integration," *ACM Comput. Surveys*, vol. 51, no. 6, pp. 1–29, Nov. 2019, doi: 10.1145/3292674.

[219] I. Mashal, O. Alsaryrah, T.-Y. Chung, C.-Z. Yang, W.-H. Kuo, and D. P. Agrawal, "Choices for interaction with things on Internet and underlying issues," *Ad Hoc Netw.*, vol. 28, pp. 68–90, May 2015, doi: 10.1016/j.adhoc.2014.12.006.

[220] J. Jamali, M. Ali, B. Bahrami, A. Heidari, P. Allahverdizadeh, and F. Norouzi, "IoT architecture," in *Towards the Internet of Things*. Cham, Switzerland: Springer, 2020, pp. 9–31, doi: 10.1007/978-3-030-18468-1_2.

[221] K. K. Goyal, "A literature survey on Internet of Things (IoT)," *Int. J. Adv. Netw. Appl.*, vol. 9, no. 6, pp. 3663–3668, 2018.

[222] D. Goyal, A. Kumar, V. Piuri, and M. Ganzha, "Lecture notes in networks and IoT architectures and its security: A review," in *Proc. 2nd Int. Conf. Inf. Manag. Mach. Intell.* Singapore: Springer, 2020, pp. 87–94.

[223] M. Mohammadi and A. Al-Fuqaha, "Enabling cognitive smart cities using big data and machine learning: Approaches and challenges," *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 94–101, Feb. 2018.

[224] C. C. Sobin, "A survey on architecture, protocols and challenges in IoT," *Wireless Pers. Commun.*, vol. 112, no. 3, pp. 1383–1429, Jun. 2020, doi: 10.1007/s11277-020-07108-5.

[225] M. S. BenSaleh, R. Saida, Y. H. Kacem, and M. Abid, "Wireless sensor network design methodologies: A survey," *J. Sensors*, vol. 2020, pp. 1–13, Jan. 2020, doi: 10.1155/2020/9592836.

[226] M. Laroui, B. Nour, H. Moungla, M. A. Cherif, H. Afifi, and M. Guizani, "Edge and fog computing for IoT: A survey on current research activities & future directions," *Comput. Commun.*, vol. 180, pp. 210–231, Dec. 2021, doi: 10.1016/j.comcom.2021.09.003.

[227] D. S. Pereira, M. R. De Morais, L. B. P. Nascimento, P. J. Alsina, Vitor. G. Santos, D. H. S. Fernandes, and M. R. Silva, "Zigbee protocol-based communication network for multi-unmanned aerial vehicle networks," *IEEE Access*, vol. 8, pp. 57762–57771, 2020, doi: 10.1109/ACCESS.2020.2982402.

[228] E. S. A. Ahmed and M. E. Yousef, "Internet of Things in smart environment: Concept, applications, challenges, and future directions," *World Sci. News*, vol. 134, no. 1, pp. 1–51, 2019.

[229] K. Mekki, E. Bajic, F. Chaxel, and F. Meyer, "A comparative study of LPWAN technologies for large-scale IoT deployment," *ICT Exp.*, vol. 5, no. 1, pp. 1–7, Mar. 2019, doi: 10.1016/j.icte.2017.12.005.

[230] F. Cirillo, D. Gómez, L. Diez, I. Elicegui Maestro, T. B. J. Gilbert, and R. Akhavan, "Smart city IoT services creation through large-scale collaboration," *IEEE Internet Things J.*, vol. 7, no. 6, pp. 5267–5275, Jun. 2020, doi: 10.1109/JIOT.2020.2978770.

[231] C. S. Lai, Y. Jia, Z. Dong, D. Wang, Y. Tao, Q. H. Lai, R. T. K. Wong, A. F. Zobaa, R. Wu, and L. L. Lai, "A review of technical standards for smart cities," *Clean Technol.*, vol. 2, no. 3, pp. 290–310, Aug. 2020.

[232] L. Mendiboure, M.-A. Chalouf, and F. Krief, "Edge computing based applications in vehicular environments: Comparative study and main issues," *J. Comput. Sci. Technol.*, vol. 34, no. 4, pp. 869–886, Jul. 2019, doi: 10.1007/s11390-019-1947-3.

[233] D. Li, L. Deng, and Z. Cai, "Intelligent vehicle network system and smart city management based on genetic algorithms and image perception," *Mech. Syst. Signal Process.*, vol. 141, Jul. 2020, Art. no. 106623.

[234] B. N. Silva, M. Khan, and K. Han, "Integration of big data analytics embedded smart city architecture with RESTful web of things for efficient service provision and energy management," *Future Gener. Comput. Syst.*, vol. 107, pp. 975–987, Jun. 2020, doi: 10.1016/j.future.2017.06.024.

[235] K. Abouelmehdi, A. Beni-Hessane, and H. Khaloufi, "Big healthcare data: Preserving security and privacy," *J. Big Data*, vol. 5, no. 1, pp. 1–18, Dec. 2018, doi: 10.1186/s40537-017-0110-7.

[236] M. Babar, F. Khan, W. Iqbal, A. Yahya, F. Arif, Z. Tan, and J. M. Chuma, "A secured data management scheme for smart societies in industrial Internet of Things environment," *IEEE Access*, vol. 6, pp. 43088–43099, 2018, doi: 10.1109/ACCESS.2018.2861421.

[237] P. Singh and R. Agrawal, "A customer centric best connected channel model for heterogeneous and IoT networks," *J. Organizational End User Comput.*, vol. 30, no. 4, pp. 32–50, Oct. 2018, doi: 10.4018/JOEUC.2018100103.

[238] K. Zhang, Y. Mao, S. Leng, Y. He, and Y. Zhang, "Predictive offloading in cloud-driven vehicles: Using mobile-edge computing for a promising network paradigm," *IEEE Veh. Technol. Mag.*, vol. 12, no. 2, pp. 36–44, Apr. 2017.

[239] M. Burhan, R. Rehman, B. Khan, and B.-S. Kim, "IoT elements, layered architectures and security issues: A comprehensive survey," *Sensors*, vol. 18, no. 9, p. 2796, Aug. 2018, doi: 10.3390/s18092796.

[240] T. Stephan, F. Al-Turjman, K. Suresh Joseph, B. Balusamy, and S. Srivastava, "Artificial intelligence inspired energy and spectrum aware cluster based routing protocol for cognitive radio sensor networks," *J. Parallel Distrib. Comput.*, vol. 142, pp. 90–105, Aug. 2020, doi: 10.1016/j.jpdc.2020.04.007.

[241] P. Chithaluru, F. Al-Turjman, M. Kumar, and T. Stephan, "I-AREOR: An energy-balanced clustering protocol for implementing green IoT in smart cities," *Sustain. Cities Soc.*, vol. 61, Oct. 2020, Art. no. 102254, doi: 10.1016/j.scs.2020.102254.

[242] U. Majeed, L. U. Khan, I. Yaqoob, S. M. A. Kazmi, K. Salah, and C. S. Hong, "Blockchain for IoT-based smart cities: Recent advances, requirements, and future challenges," *J. Netw. Comput. Appl.*, vol. 181, May 2021, Art. no. 103007, doi: 10.1016/j.jnca.2021.103007.

[243] M. M. Rathore, A. Ahmad, A. Paul, and S. Rho, "Urban planning and building smart cities based on the Internet of Things using big data analytics," *Comput. Netw.*, vol. 101, pp. 63–80, Jun. 2016, doi: 10.1016/j.comnet.2015.12.023.

[244] I. Jawhar, N. Mohamed, and J. Al-Jaroodi, "Networking architectures and protocols for smart city systems," *J. Internet Services Appl.*, vol. 9, no. 1, pp. 1–16, Dec. 2018, doi: 10.1186/s13174-018-0097-0.

[245] M. Babar, F. Arif, M. A. Jan, Z. Tan, and F. Khan, "Urban data management system: Towards big data analytics for Internet of Things based smart urban environment using customized Hadoop," *Future Gener. Comput. Syst.*, vol. 96, pp. 398–409, Jul. 2019, doi: 10.1016/j.future.2019.02.035.

[246] P. Sowkuntla and P. S. V. S. S. Prasad, "MapReduce based parallel fuzzy-rough attribute reduction using discernibility matrix," *Appl. Intell.*, vol. 52, no. 1, pp. 154–173, 2022, doi: 10.1007/s10489-021-02253-1.

[247] M. Ghorbanian, S. H. Dolatabadi, and P. Siano, "Big data issues in smart grids: A survey," *IEEE Syst. J.*, vol. 13, no. 4, pp. 4158–4168, Dec. 2019, doi: 10.1109/JSYST.2019.2931879.

[248] H. Samih, "Smart cities and Internet of Things," *J. Inf. Technol. Case Appl. Res.*, vol. 21, no. 1, pp. 3–12, Jan. 2019, doi: 10.1080/15228053.2019.1587572.

[249] H. S. Munawar, M. Mojtahedi, A. W. A. Hammad, A. Kouzani, and M. A. P. Mahmud, "Disruptive technologies as a solution for disaster risk management: A review," *Sci. Total Environ.*, vol. 806, Feb. 2022, Art. no. 151351, doi: 10.1016/j.scitotenv.2021.151351.

[250] M. Gordan, H. A. Razak, Z. Ismail, K. Ghaedi, Z. X. Tan, and H. H. Ghayeb, "A hybrid ANN-based imperial competitive algorithm methodology for structural damage identification of slab-on-girder bridge using data mining," *Appl. Soft Comput.*, vol. 88, Mar. 2020, Art. no. 106013, doi: 10.1016/j.asoc.2019.106013.

[251] M. Gordan, Z. Ismail, H. A. Razak, K. Ghaedi, Z. Ibrahim, Z. X. Tan, and H. H. Ghayeb, "Data mining-based damage identification of a slab-on-girder bridge using inverse analysis," *Measurement*, vol. 151, Feb. 2020, Art. no. 107175, doi: 10.1016/j.measurement.2019.107175.

[252] G. Nguyen, S. Dlugolinsky, M. Bobák, V. Tran, Á. L. García, I. Heredia, P. Malík, and L. Hluchý, "Machine learning and deep learning frameworks and libraries for large-scale data mining: A survey," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 77–124, Jun. 2019, doi: 10.1007/s10462-018-09679-z.

[253] A. Dogan and D. Birant, "Machine learning and data mining in manufacturing," *Exp. Syst. Appl.*, vol. 166, Mar. 2021, Art. no. 114060, doi: 10.1016/j.eswa.2020.114060.

[254] Y. Hajjaji, W. Boulila, I. R. Farah, I. Romdhani, and A. Hussain, "Big data and IoT-based applications in smart environments: A systematic review," *Comput. Sci. Rev.*, vol. 39, Feb. 2021, Art. no. 100318, doi: 10.1016/j.cosrev.2020.100318.

[255] Y. Sun, H. Song, A. J. Jara, and R. Bie, "Internet of Things and big data analytics for smart and connected communities," *IEEE Access*, vol. 4, pp. 766–773, 2016, doi: 10.1109/ACCESS.2016.2529723.
[256] J. Barthélemy, N. Verstaevel, H. Forehead, and P. Perez, "Edge-computing video analytics for real-time traffic monitoring in a smart city," *Sensors*, vol. 19, no. 9, p. 2048, May 2019, doi: 10.3390/s19092048.
[257] G. Phillips-Wren, M. Daly, and F. Burstein, "Reconciling business intelligence, analytics and decision support systems: More data, deeper insight," *Decis. Support Syst.*, vol. 146, Jul. 2021, Art. no. 113560, doi: 10.1016/j.dss.2021.113560.
[258] G. Perboli and M. Rosano, "A taxonomic analysis of smart city projects in North America and Europe," *Sustainability*, vol. 12, no. 18, p. 7813, Sep. 2020, doi: 10.3390/su12187813.

MARIEH TALEBKHAH (Member, IEEE) received the B.Sc. and M.Sc. degrees in computer software engineering from Iran, in 2009 and 2017, respectively. She is currently pursuing the Ph.D. degree with Universiti Putra Malaysia (UPM), Kuala Lumpur. Her current research interests include vehicular communication, intelligent transportation systems, wireless ad hoc networks, cloud and Fog computing, and big data analytics.

ADUWATI SALI (Senior Member, IEEE) received the B.Eng. degree in electrical and electronics engineering from The University of Edinburgh, U.K., in 1999, the M.Sc. degree in communications and network engineering from Universiti Putra Malaysia, in 2002, and the Ph.D. degree in mobile and satellite communications from the University of Surrey, U.K., in 2009. From 2016 to 2019, she was the Deputy Director of the Research Management Centre (RMC), Universiti Putra Malaysia. She is currently a Professor with the Department of Computer and Communication System Engineering and a Researcher with the Wireless and Photonic Network Research Center of Excellence (WiPNET), Universiti Putra Malaysia. Her research interest includes mobile and satellite communication systems.

MEISAM GORDAN received the Ph.D. degree in structural engineering and materials from the University of Malaya, in 2020, with a focus on structural health monitoring, data mining, and artificial intelligence. He is currently on a European Union's Horizon 2020-funded project, named PRECINCT as a Postdoctoral Research Fellow with University College Dublin (UCD), Ireland. He has published more than 30 articles in scientific international journals and conferences in different fields, such as structural health monitoring, signal processing, data mining, artificial intelligence, soft computing, vibration control systems, the IoT, big data, blockchain, remote sensing, smart cities, unmanned aerial systems, finite element analysis, and materials. His research interest includes the connection between structural engineering and computer science using industry 4.0 emerging technologies. He is a member of renowned institutes, such as the American Society of Civil Engineers (ASCE), the Board of Engineers Malaysia (BEM), and the Institution of Engineers Malaysia (IEM). He has been recognized by several peer-reviewed journals as a reliable referee with more than 40 peer-reviewed articles and being an editorial board member of several peer-reviewed journals.

SHAIFUL JAHARI HASHIM (Member, IEEE) received the B.Eng. degree in electrical and electronics engineering from the University of Birmingham, U.K., in 1998, the M.Sc. degree from the National University of Malaysia, in 2003, and the Ph.D. degree from Cardiff University, U.K., in 2011. He is currently an Associate Professor with the Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia (UPM). His research interests include cloud computing, the Internet of Things (IoT), network security, and non-linear wireless measurement systems.

FAKHRUL ZAMAN ROKHANI (Member, IEEE) is with the Department of Computer and Communication Systems Engineering, Universiti Putra Malaysia, where he is currently a Lecturer. He has published numerous publications in various national and international peer-reviewed journals and presented scientific articles across the world. His clinical and scientific research interest includes embedded systems. Because of the active association with different societies, academies, and contributions, he is recognized by the subject experts around the world. His contributions are appreciated by various reputed awards.