

Received March 4, 2021, accepted March 20, 2021, date of publication April 5, 2021, date of current version April 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3070905

IoT and Big Data Applications in Smart Cities: Recent Advances, Challenges, and Critical Issues

MARIEH TALEBKHAH¹, ADUWATI SALI¹, (Senior Member, IEEE), MOHSEN MARJANI², MEISAM GORDAN³, SHAIFUL JAHARI HASHIM¹, AND FAKHRUL ZAMAN ROKHANI¹, (Member, IEEE)

¹Wireless and Photonics Networks Research Centre (WiPNET), Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Seri Kembangan 43400, Malaysia

²Faculty of Innovation and Technology, School of Computer Science and Engineering, Subang Jaya 47500, Malaysia

³Department of Civil Engineering, Faculty of Engineering, University of Malaya, Kuala Lumpur 50603, 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 BIDANET: Parametric Big Data Analytics over Wireless Networks (Vote No: 9696300), and in part by the EMOSEN-Energy Efficient MIMO-Based Wireless Transmission for SWIPT-Enabled Network (UPM/800-3/3/1/GPB/2019/9671600, Vote No: 9671600).

ABSTRACT The notion of smart cities has remained under evolution as its global implementations are challenged by numerous technological, economic, and governmental obstacles. Moreover, the synergy of the Internet of Things (IoT) and big data technologies could result in promising horizons in terms of smart city development which has not been explored yet. Thus, the current research aims to address the essence of smart cities. To this end, first, the concept of smart cities is briefly overviewed; then, their properties and specifications as well as generic architecture, compositions, and real-world implementations are addressed. Furthermore, possible challenges and opportunities in the field of smart cities are described. Numerous issues and challenges such as analytics and using big data in smart cities introduced in this study offers an enhancement in developing applications of the above-mentioned technologies. Hence, this study paves the way for future research on the issues and challenges of big data applications in smart cities.

INDEX TERMS Big data, Internet of Things (IoT), smart city.

I. INTRODUCTION

Based on the latest estimation of the United Nations, about 68% of people will reside in urban regions in 2050. The report indicated a dramatic rise in urban living from 751 million in 1950 to 4.2 billion people in 2018 [1]. Bibri and Krogstie [2] argued that urban zones account for 70% of total natural resource consumptions which had led to environmental contamination, ecosystems destruction, and energy shortage. Limited access to resources is a prominent challenge in the cities development since these cities are invented to decline the cost and unemployment rate with special emphasis on climate changes and potable water supplies. Thus, there is an urgent need to adopt smart methods to help the citizens to address all the mentioned aspects [3] in this context, smart cities could be a solution to such problems. “Smart city” can be employed to improve the environmental, economic, alleviate mobility, safety, governance, and living standards of their residents [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Seyedali Mirjalili¹.

Marsal-Llacuna *et al.* [5] described a smart city as an urban medium utilizing information and communication technologies (ICT) as well as other relevant technologies to promote more efficient ordinary city operations and improve the quality of services (QoS) received by the citizens [6]. The extreme objective of the first smart cities was to alleviate the quality of life (QoL) of their citizens by modulating the demand-supply contradictions in different functionalities [7]. Concerning QoL demands, modern smart cities specifically emphasize sustainability and efficient energy management, transportation, health care, and governance to comply with the utmost requirement of urbanization [8].

The implementation of the notion of smart city will drastically raise the data. Such a huge deal of data (i.e. big data) will be at the center of the services rendering by the Internet of Things (IoT). The big data phenomenon has been explored by volume, velocity, and a variety of data types at ever-rising rates [9], [10]. Big data allows the city to attain precise insight into a huge deal of data collected from different sources. Such data mainly encompass unstructured specifications in comparison with big data collected by other methods [11].

Such an enormous deal of unstructured data could be gathered and saved in the clouds or data centers through the use of distributed fault-tolerant databases like Not Only SQL, improving a single service or application [12]. Therefore, the programming models for large datasets processing using parallel algorithms could be implemented for data analytics to attain values from the saved data.

Nonetheless, the application of big data analytics (BDA) in smart media is still at the initial stages. In this regard, the big data analytics could be promising in the enhancement of smart city services [13]. Currently, a huge deal of data is continuously generated in various sources, including smartphones, PCs, sensors, cameras, global positioning systems, social networking platforms, commercial transactions, and games. Regarding the ever-growing data generation in our digitized era, high-efficiency data storage and processing have brought up new issues for the conventional data mining and analytics platform. Big data analytics is capable of extracting valuable information from an ocean of data generated by various sensors.

This article is beneficial for the individual with a specific attention to the concept of smart cities as it is mainly concentrated on the analysis of big data applications as a major issue in the field of smart cities. The research gaps in big data studies of the smart cities are identified in various sectors. Ultimately, the open challenges and trends will be presented to elucidate the direction of future studies in the field of big data in smart cities. The current paper will also address various usages of big data in smart cities.

The important aims of this work are outlined as follows.

- A brief review on the main characteristics of smart cities (e.g., architecture, challenges, and opportunities) and function of IoT in smart cities,
- Discussion the needs and challenges of big data systems in smart cities,
- A survey on big data systematic tools representing for smart cities in terms of data collection, pre-processing of data and analytical process of data,
- Summary of potential upcoming research paths.

The rest of this article comprises the following sections: Section 2 presents state-of-the-art smart cities architecture, applications and the concept of IoT. The issues and opportunities of smart cities will be described in Section 3. As big data is the most prominent issue of smart cities, the obstacles in the path of its implementation in the smart city will be addressed in Section 4 considering various case studies in this field. Unsolved issues and future study topics will be also discussed. Section 5 introduces an innovative big data framework for the smart cities; while, Section 6 highlights the future research opportunities. Last but not least, Section 7 presents the conclusions.

Based on our literature review, this study is the first research addressing the issues of big data in a smart city. In this regard, the present article can pave the way for future

studies aimed to develop smart cities based on the concept of big data.

II. OVERVIEW OF SMART CITIES AND IoT

A. SMART CITIES

Based on Albino *et al.* [14], a smart city refers to a series of paradigms extended over diverse areas: economy, public, government, mobility, environment, and lifestyle. This concept involves various fields including environment monitoring, traffic analysis, utility monitoring, public transportation, and incident reporting. Data collection from the mentioned areas helps the city authorities to improve the infrastructure and optimize their resources [15]. A smart city has to encompass some prominent factors to centralize the data. Such factors could be found in various shapes and forms from a plain webpage to a complicated and context-aware mobile application or specialized hardware [16], [17]. Furthermore, the data accessibility must be guaranteed so that the system can freely access the citizens as well as allowing them to modify and correct. Upon the citizens' contribution, broader information angles could be achieved as it will be easier to collect more data from citizens [18].

A high-level concept of a smart city and its various functionalities are depicted in Fig. 1. As it can be observed from this figure, Unmanned Aerial Vehicles (UAVs) and Internet of Drone (IoD) resources have been employed extensively for conducting the expected services and applications [19].

This prompt urbanization growth has brought up new infrastructural challenges. The city management gets more complicated upon city expansion and its service extension. Therefore, cities have to evolve to cope with the newly-emerged social, economic, engineering, and environmental issues. In other words, the cities should develop smart features to be able to properly encounter these issues.

B. ARCHITECTURE OF SMART CITIES

A smart city is intended to improve the smartness of its systems and applications. Here, some of its requirements and features are listed [2], [20], [21]:

- A rigid and extensive framework offering safe and open access;
- A citizen-oriented architecture;
- A considerable deal of mobile and wearable public and private data which can be stored, found, shared, and tagged providing the citizens with access to information from anywhere, if necessary;
- An application with analytical and integrated features; and finally
- A smart physical and network infrastructure offers the possibility to transfer large amounts of heterogeneous data and support complicated and distributed services and applications [6], [22].

The studied smart city possesses a layered configuration including four levels as depicted in Fig. 2.

- The physical components level serves as a host for all devices, sensors, equipment, apparatus, and components

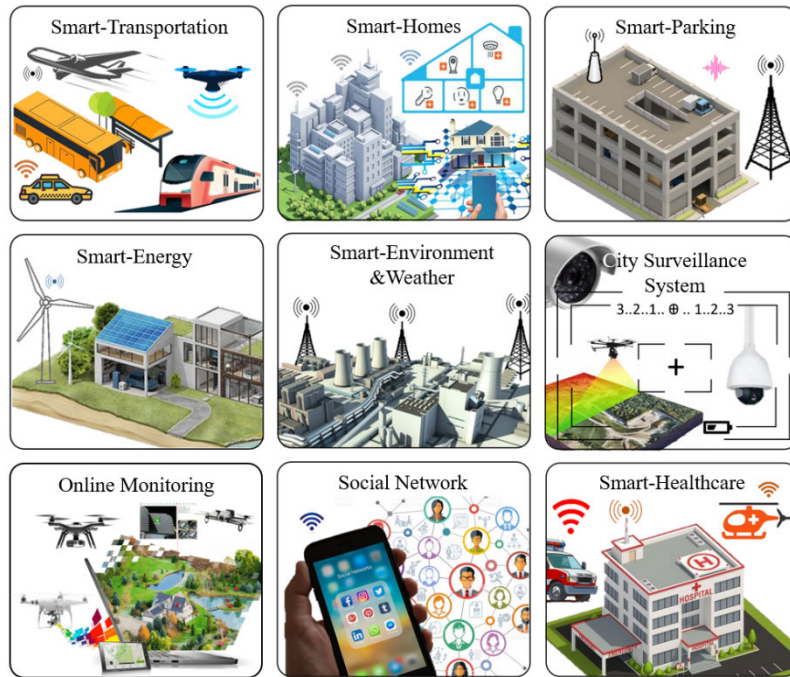


FIGURE 1. Smart-City functionalities.

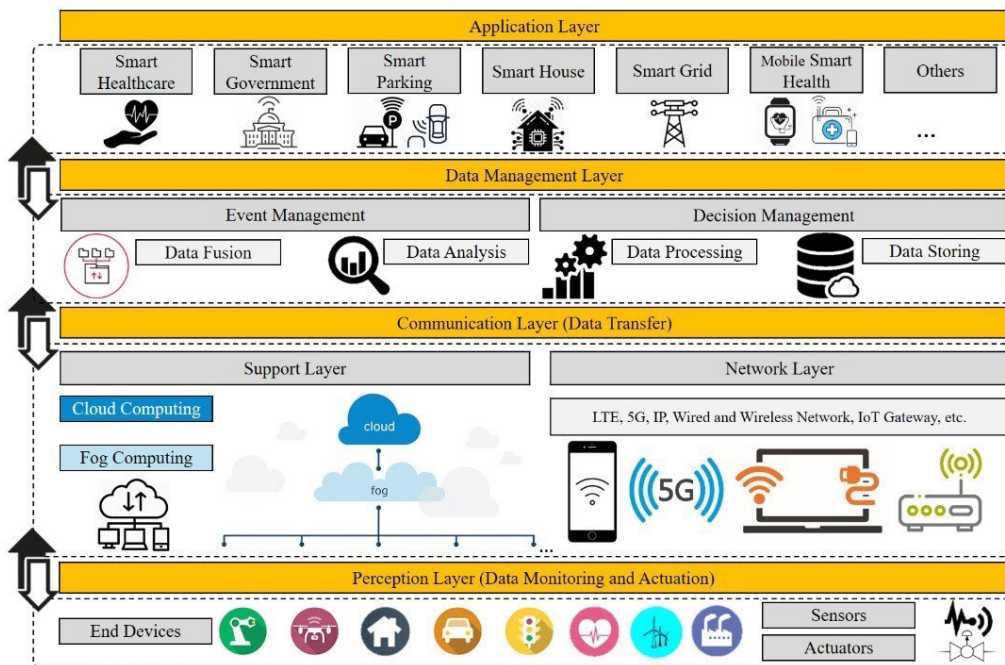


FIGURE 2. Proposed architecture for smart city deployment framework.

of different smart systems, services, and applications. Besides that, data collection plays the most decisive role due to controlling the rest of the operations in a smart city. Moreover, substantial data heterogeneity has increased the challenges of data collection.

The major challenges at the physical level can be summarized as:

- the large volume of devices;
- heterogeneity; and

- low-processibility and restricted sources of energy.

Based on Fig. 2, the bottom layer is responsible for sensing/data collection. This layer is composed of wireless sensor networks (WSNs), intelligent components, and data collection tools. This layer also collects the entire data categories from various sensors and devices [23]. This layer exploits diverse techniques to present more efficient data collection. Its sensor network captures a wide variety of parameters such as humidity, temperature, pressure, and light. In this regard,

various sensing devices (e.g. radio frequency identification (RFID) sensors, global positioning system (GPS) terminals, Bluetooth, actuators, Zigbee, and cameras [24]) have been employed. Its maximum network coverage offers excellent accessibility and smartness as well as extended data capturing capabilities [6].

- The communication level: it encompasses the entire network technological alternatives to establish communication between physical elements and applications. By connecting data sources to the management, the communication (transmission) layer is the most prominent component of the smart city architectures. This layer converges diverse communication networks. Thus, innumerable devices can be linked to a single network satisfying routing through its distinctive addressability. This layer includes different types of technologies (i.e. wired, wireless, and satellite). Concerning its coverage, the transmission layer can be categorized into two sub-layers: access and network transmissions. Technologies offering short-range coverage (e.g. near field communication (NFC), Zwave, Bluetooth, M2M, Zigbee, and RFID) are considered as access network whereas those supplied broader coverage (like low-power wide-area networks (LP-WAN)), 5G, 4G long-term evolution (LTE), and 3G are classified in transmission class.
- The Management Level: The data management layer can be regarded as the brain of the smart city as it is placed between the acquisition and application layers. The management layer conducts a wide diversity of data manipulation, organization, analysis, storing, and decision-making actions. An efficient data management layer is a prerequisite for a maintainable smart city as its functions depends on data management. The primary performance of the data layer is to preserve the strength of data by focusing on cleaning of data, development, involvement, and protection. This layer could be divided into several subunits including integration, processing, analysis, storing of data, and finally decision making supervision [25].
- The application layer: As the top layer of a smart city, the application layer is a mediator between the citizens and the data management layer. Its performance has a drastic influence on the users' perspective and satisfaction from the smart city, as it is in direct interaction with the citizens. Including diverse applications applied by citizens, municipal staff, and administrators are connected to different departments of the smart city. This layer encompasses the entire smart systems and developments. These systems can be fed and process huge data used by the communication level to flow the entities in the application processes. The smart applications are liable for decision implementations received from the data management layer.

Each level exhibits a series of features and requirements. In this regard, various innovative computing paradigms have been utilized and explored.

As big data offers the possibility to solve the smart city challenges, the relevant challenges should be mentioned first. A limited number of studies have dealt with the aforesaid issues. However, scarce studies have consolidated the existing works and established them as a framework to offer an efficient approach for the management of big data in a smart city [26]. Here, we tried to present the recommended effective methods to cope with big data issues in a deep systematic review article. Considering the current importance of this topic, the present article can offer major insights for practitioners and researchers in this field [27].

C. IoT IN SMART CITIES

The emergence of the Internet of Things (IoT) approached the abstract concept of the smart city into reality. The smart cities are intended to provide their citizens with superior healthcare, safety, convenience, and wisdom. In other words, they are aimed at improving the quality of life. IoT could be definitely regarded as the most effective technology giving rise to a significant mutation in the world [28], [29]. IoT is the developed version of conventional networks aimed at connecting a myriad of interconnected devices. The concept of IoT has been further reinforced with advanced technologies such as wireless sensor networks (WSNs) and machine-to-machine (M2M) communication [6].

The resources of a smart city are managed and controlled by intelligent information systems that require considering the food, energy, and water nexus. Optimized supply and high-efficiency of resources require the development of IoT-based systems to meet the mentioned issues and the big data originating from such systems [26]. Huge efforts have been devoted to developing smart houses [30], smart transportation [31], traffic-managements [32], waste disposal system [33], energy management systems [34], and healthcare [35], as well as various facilities, synergizing towards the establishment of a smart city.

Fig. 3 illustrates the idea of IoT which is one of the most effective approaches for enhancing the quality of life and social welfare taking into account human and environmental issues.

As a newly-emerged concept, IoT uses mobile devices, public services, transportation resources, and home appliances as its data collection tools. In such an environment, the entire daily-life electronic devices (e.g. emergency alarms, wristwatches, garage doors, and vending machines), and also household appliances like freezers, water heaters, washing machines, microwave ovens, refrigerators, dishwashers, stoves, and air conditioners are connected to an IoT network and could be remotely controlled.

Citizens are continuously and increasingly demanding for novel services supporting a high-quality lifestyle. These opportunities have led to the cities evolving; however, numerous challenges could influence the daily lives of the citizens [8]. As the hearth of this evolution, technology has remarkably altered our world and life. A digitalized world with interconnected objects, people, and devices has

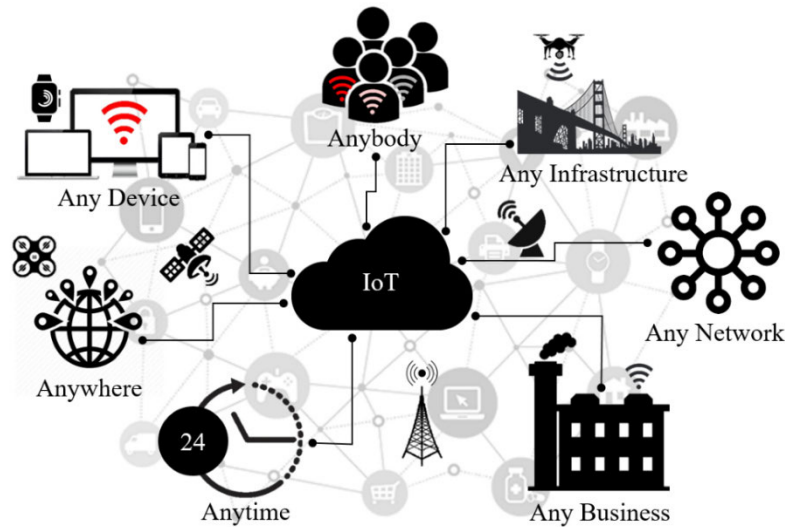


FIGURE 3. IoT concept.

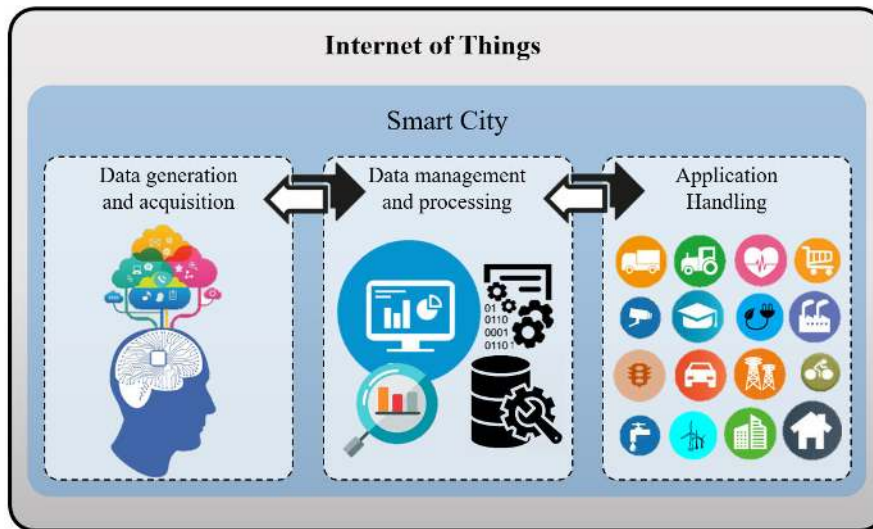


FIGURE 4. The role of IoT to the establishment of smart city.

drastically influenced our jobs, travels, societies, and interactions with our environment; profoundly affecting various fields including urban systems, environmental monitoring, management platforms, and healthcare [36].

The smart city has found increasing popularity over other urban development models (information city, digital city, and telicity) as it manifests the concept of the entire mentioned approaches [37]. A smart city is an implementation of the IoT [38], thus, inheriting its underlying operational mechanisms. According to Fig. 4, IoT offers vital building blocks of a smart city.

III. SMART CITIES OPPORTUNITIES AND CHALLENGES

Despite the extensive application of the concept of a smart city, its further development requires resolving its remaining

challenges. In this section, the issues and potential opportunities for the real application of a smart city will be discussed. These challenges are recognized by extensive literature review on the latest studies about the smart cities while the opportunities are presented by surveying the works and experiences in this area.

A. OPPORTUNITES FOR SMART CITIES

The concept of a smart city is yet developing and experiments and implementations are restricted inside the developed countries. Nonetheless, the promising advantages of smart cities can be applied to any urban region as shown in Fig. 5.

Thus, more research works on cost-effective design and implementation can enhance the popularity of smart cities around the globe. The incorporation of renewable energy

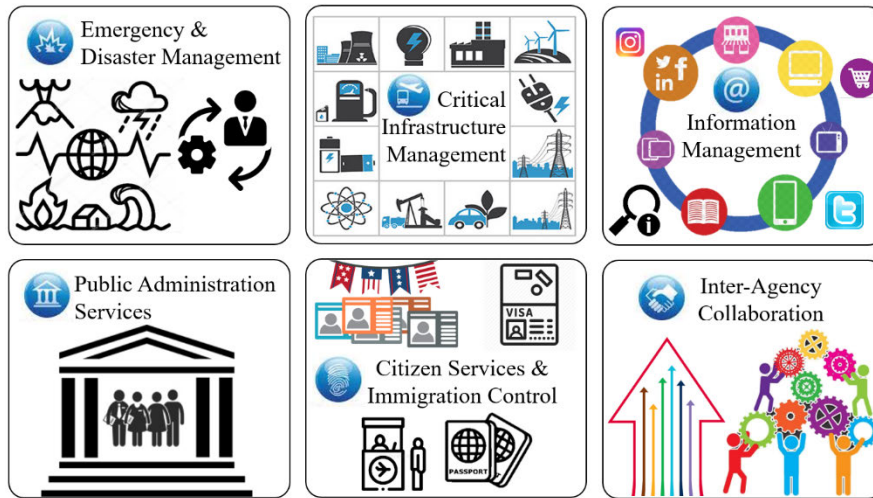


FIGURE 5. Opportunities for smart city deployment.

resources is a promising method to guarantee sustainable city functions and successful management of scarce non-renewable energy sources.

Smart devices are producing an immense deal of data that needs enormous data storage facilities. Upon Big Data generation, the traditional data processing approaches have lost their application in the architecture of modern smart cities. Thus, Big Data analytics should be integrated into a smart city environment. This issue has been addressed in a limited number of studies, however, the majority of these works have been performed in proposal forms instead of practical implementation. In this context, the development and application of Big Data analytics [39] in real smart cities could be an interesting research topic. The secure processing of sensitive data is an urgent need in connected media. Doubtful security of sensitive data will result in the citizens' avoidance to use ICT plan of smart cities hindering the robustness of city operations. Therefore, the introduction of global security measures for smart cities is of high priority requiring deeper investigations. The heterogeneous devices are another interesting research field. A smart city integrates diverse sub-systems in its application layer to present prompt and consistent assistance. The issue of aggregation at the application layer is another potential research area [6].

B. CHALLENGES FOR SMART CITIES

The realization of a smart city will face diverse issues during design, arrangement, construction and operation steps. Design and operational costs, device heterogeneity, collection, and analysis of huge data, data security, and sustainability are just to name a few alarming challenges. Fig. 6 shows some of the major issues in the design of a smart city. Furthermore, some approaches are proposed to resolve the above-mentioned challenges that are discussed in the followings.



FIGURE 6. Issues of smart city implementation.

Most of the cities are challenged by urban waste collection and they continuously try to develop a more efficient waste management system. Therefore, a smart waste management system is of crucial significance in a smart city [40].

The ability of smart cities to resolve environmental problems (particularly waste management) is a prominent question which should be explored academically [41]. The transformation of a smart city into zero-waste robust one necessitates four inter-related major approaches: waste prevention, upstream waste separation, timely waste collection, and suitable recycling collected wastes. The objective is to design and develop an IoT-enabled waste management system with an emphasis on associating waste management measures in the whole product life-cycle.

Heterogeneity [42] is another major issue in the development of a smart city. Smart cities encompass multi-vendor and multi-functional sensors, tools, and appliances. The implementation of the concept of smart city requires integrating the aforesaid heterogeneous items at the

application layer. However, the platform incompatibility due to the heterogeneity hinders the integration and inter-operations at the application layer. Despite the challenges of facilitating universal access, smart cities are focused on the design, identification, and purchasing hardware and software enabling the integration of such heterogeneous subsystems.

Design and maintenance costs [43] are a key challenge in realizing the notion of a smart city. Design cost can be defined as the financial capital spent to deploy a smart city. Thus, the lower the design rate, the higher the probability of its realization. Operational expenses are related to regular procedures and supports. Minimum operational rates guarantee the sustainability of service provision with no extra costs on the municipality. Nonetheless, the rate of optimization in the lifespan of a smart city has remained an unsolved challenge.

Smart technologies could offer interesting solutions for various urban challenges including transportation, waste and environmental management. Security and crime prevention, however, are often unaddressed in most cases. Furthermore, when researchers propose a new smart security technology [44], they seldom address the implementation or possible impacts on conventional policing and the urban planning process [45].

Failure management is a major aspect in any smart city as well. Failures are predictable after natural disasters and system collapses (e.g. infrastructure breakdown and network inaccessibility). Sustainable design offers prompt recovery strategies to cope with the failure and retain the normal urban functionalities. However, the detection and application of such recovery strategies will augment design and operational costs. The objective is the implementation of failure recovery measures with the lowest cost and the highest efficiency.

With rising greenhouse gas emission, trashes in seas, jungles, and cities, smart cities are in critical situation to decline the environmental damages [46]. Energy-efficient structures, i.e. buildings, air quality monitoring, and renewable energy sources can help the cities to reduce their adverse ecological effects. Protection of the environment and resources for our children through lowering the carbon emission [47] and efficient utilization of the resources are among the main issues of modern smart cities with a decisive role in policies intended to mitigate the direct and indirect consequences of the city economy on the sustainability. In this regard, modern cities are concentrated on renewable energy supplies to decline the carbon emission and ensure the sustainability of the city.

Air quality monitoring throughout the city [48] can be used to find times of low air quality, detect pollution roots, and help the data analysts in developing preventive measures. Air quality sensors could contribute to the establishment of a foundation aimed at declining air pollution in even the cities with high population; while will surely save many lives as pollution-related diseases are annually threatening millions of lives.

A smart city connects the social-physical information-technology, and business infrastructures for leveraging the

cooperative understanding power of the city. The connectivity issue [49] can be divided into three main stakeholders: operators who should offer capable coverage solutions for various sites throughout the city in a reasonable limit; venue owners who support coverage capacity ensuring that they will not lose due to poor mobile connection. Last but not least, local governments who have to control the mentioned groups to guarantee the supply of coverage for developing sustainable and competitive smart city infrastructures.

The urban spaces may seriously challenge the coverage. First, materials applied in the construction of highly-populated smart cities are often reflective thus prohibiting the radio frequency (RF) propagation. Basements and parking lots could be another example since these reinforced concrete structures can block the RF signals. Some buildings may support RF propagation but as they are located in a dense urban region, neighboring buildings could serve as a shield leading to poor coverage.

In this context, an efficient, inexpensive, and scalable solution capable of supporting M2M and IoT usages, smart city patterns, and mobile users is highly required. Moreover, networks have to offer solid and secure public communications for emergency services to further promote the safety of the users of such novel 'smart' media.

Due to their high amounts of data generation, the application of smart devices requires large data storage facilities. In this context, Big Data generation has replaced the conventional data processing approaches. In this regard, the examination of Big Data analytical systems and their integration into smart cities sounds crucial. A limited number of studies have addressed this issue most of which have remained as a proposal and not implemented in real-world situations. Thus, development and testing of Big Data analytical systems in real smart cities could be an interesting research topic. The security of sensitive data is an urgent requirement in interconnected media.

Several newly-emerged technologies like IoT, Radio-frequency identification (RFID), and upcoming Internet technologies can widely assist to the development of smarter cities [50]–[53]. The rising presence of the mentioned technologies has led to the generation of a huge amount of data [54]. As mentioned() [55], about 2.5 quintillion bytes of data are daily created. Moreover, 90% of these data have been generated in the last two years. Upon proper management and analysis, such a huge amount of data (i.e. big data) could drastically influence the operations in the smart cities [39]. The worthy information from big data analytical systems can rapidly transform cities into the artificial ecosystems of mutually dependent, interrelated, and smart digital entities [56], [57].

IV. BIG DATA

A. APPLICATIONS OF BIG DATA IN THE SMART CITY

Thanks to the implementation of big data technology in smart cities, the data can be efficiently stored and

TABLE 1. Common big data use-cases in smart cities.

Smart-city use-case	IoT	Advantages	Limitations
Smart Home [35], [58], [59]	A combination of different sensors	<ul style="list-style-type: none"> Enhanced energy efficiency Better management Better appliance functionalities 	<ul style="list-style-type: none"> High cost Internet Dependence Professional dependence
Smart Government [60]	Smart phone, Camera, sensor	<ul style="list-style-type: none"> Help in decreasing crimes as a result of city-wide surveillances Providing a platform to hear citizens' comments on governmental policies and approaches Promotion of online business which can save time and costs. Businesses could gather real-time data from government More efficient city management 	<ul style="list-style-type: none"> Difficulties in collecting and analyzing data
Smart Transportation [61], [62]	Smart automobiles, Cameras, RFID carts	<ul style="list-style-type: none"> Road safety improvement Reduced cost of transportation supplying various demands Improved environmental friendliness Capable of optimizing routes and schedules using big data 	<ul style="list-style-type: none"> Network disturbance may result in serious car crashes
Smart Environmental monitoring [63]–[65]	Temperature & humidity, sensors, gas sensors	<ul style="list-style-type: none"> Real-time monitoring of pollutants Improved weather forecasting Alert services upon citizens' requirements 	<ul style="list-style-type: none"> Heterogeneity of sensors and devices Difficult management
Energy management [34], [66], [67]	Smart meters, energy sensors,	<ul style="list-style-type: none"> A more efficient power supply as a consequence of better demand-supply predictions Establishment of a dynamic tariff system in favor of the customers 	<ul style="list-style-type: none"> Low data quality Data integration and sharing
Healthcare [35]	Sensors, Smart wearable device	<ul style="list-style-type: none"> Better healthcare through improved preventive measures Early diagnoses and treatments Better handling of healthcare records and patients' care Access to citizen's health data 	<ul style="list-style-type: none"> Imprecise network
Smart Grid [68], [69]	Smart meters and Smart readers	<ul style="list-style-type: none"> Improved performance of electric power systems More efficient, reliable, economic, and sustainable electric power generation and distribution Sufficient supply of electricity Prediction of future demands 	<ul style="list-style-type: none"> High cost Management difficulties

processed into knowledge, (i.e. information) in order to develop smart city services. Furthermore, big data can help in service and resource expansion. In this regard, advanced tools and approaches can result in high-efficiency and effective data analysis. Such tools and routes can promote cooperation and communication between entities offering services to various sections of a smart city in addition to improving customer satisfaction and business opportunities. Table 1 summarizes common big data usages in a smart city.

B. RECENT ADVANCES ON BIG DATA

Ongoing studies in this area have emphasized the significance of big data in the development of smart cities. Moreover, some of these works have addressed the challenges of adopting big data in the notion of the smart city [70].

Q. Zhang *et al.* [71] proposed a processing model, so-called firework as a novel computational paradigm allowing assigned processing of data and distributing in a shared edge medium established based on IoT. The developed firework combined distributed data through the provision of virtual data views for the end-users applying pre-determined connections which are represented by a series of functions and databases. This firework is aimed at minimizing latency access of data through approaching processing techniques to the data producer in the edge networks. Firework possesses several participants who have to enroll their databases and corresponding tasks as data views. These data views could be

accessed by any stakeholder in order to combine various data views into a particular one to conduct in depth systematic data analysis.

B. Ahlgren and colleagues [72] addressed the importance of IoT in services delivery to improve the citizens' lives (e.g. transportation, and energy efficiency). They emphasized that the IoT-based system has to rely on accessible data and benchmarks (i.e. interfaces and set of rules) to promote third party innovation through improving creator lock-ins. Accordingly, they planned and established a GreenIoT scheme in Sweden for determining the benefits of open platforms and accessible data in developing smart cities. Some guidelines should be designed to deal with the procurements of a user-friendly accessible IoT system (such as conventional data layouts and open application programming interfaces (APIs)).

B. Cheng and coworkers [73] developed an edge analytics platform (GeeLytics) capable of conducting synchronous processing of data at the network edges as well as the clouds. This approach coped with the geo-distributed and low-latency analytics due to the huge deal of IoT data. GeeLytics is aimed at supporting active stream preparation strategies in respect to the features of heterogeneous edge or cloud nodes as well as the capacity of the system.

E. Ahmed and M. Rehmani [74] analyzed human behaviors through the use of big data and analytical outlines within a social IoT model. The authors proposed an approach comprising three operational areas. Moreover, the ecosystem of smart cities and big data was explored in this research. It was

concluded that, collaborative filtering schemes could help in accurate analysis of human behavior.

H. R. Arkian and colleagues [75] introduced a fog-based data analysis methodology with a cost-effective resource optimization technique applicable in IoT crowd sensing purposes. This scheme (known as MIST) was aimed at reducing the service latency supply in conventional cloud computing. According to the empirical outcomes, the MIST fog-based method outperformed the conventional cloud computing by increasing the number of real-time applications. This realm of research can be further expanded through the following approaches: (a) addition of optional sensing modules to the fog layer, (b) Architectural reinforcement using privacy protection data analytics, and (c) Taking into account the data creators' and users' mobility in the store supply section.

M. M. Rathore and coworkers [76] proposed a system capable of resolving several issues of a smart city environment for instance enabling objects reaction to the context, minimization of data collection cost, and attaining insight on the real-time collected and processed data. Their developed system possessed a four-tier configuration, in which the bottom tier had the duty of collecting the data. The intermediate tier 1 enabled communications between base stations, sensors, Internet, and relays; while intermediate tier 2 was in charge of data supervision and preparation by means of the Hadoop framework. At last, the top tier applied data analysis approaches to generate results. According to their findings, the developed system exhibited higher scalability and efficiency concerning throughput and processing time as compared to the currently-used systems. Their system did not have an intelligent decision-making approach to handle big data in an IoT medium.

F. Alam *et al.* [77] reported the use of a number of data mining algorithms, i.e. deep learning artificial neural network (DLANN), pre-developed ANN, support vector machine (SVM), Bayesian, K-nearest neighbors (KNN), linear discriminant analysis (LDA), and decision tree algorithms (C4.5 and C5.0) in IoT-based data. The execution time, classification robustness, and confusion matrix of the aforesaid algorithms were compared in this work. Concerning classification robustness, DLANN, ANN, C4.5, and C5.0 outperformed SVM, Bayesian, KNN, and LDA. Nonetheless, C4.5, C5.0, and ANN showed high similarity in classification precision. Bayesian and LDA exhibited the best execution time, while LDA showed a slight superiority over Bayesian in terms of processing time. The authors were intended to examine extensive IoT datasets in more detail. In overall, neural networks could provide the best performance amongst all algorithms. This is due to the fact that neural networks are able to achieve the highest accuracy due to the nature of artificial intelligence technology [78], [79].

D. Mourtzis *et al.* [80] revealed that the use of IoT in the manufacturing industries can modernize the outdated mechanisms which may lead to the data production process turning industrial data into industrial big data. Such big data would be useless in the absence of analytics power. The

implementation of data analytics empowers companies to adopt novel data-driven approaches to manage competitive pressure. Moreover, they also illustrated the simple implementation of an IoT paradigm in a company possessing ~100 machines.

Alotaibi and colleagues [81] reviewed the diverse relevant concepts such as big data analytical systems and big data in addition to its role in healthcare and healthcare supply chain management (SCM). The influence of Twitter data on SCM was examined as well. They addressed opportunities and challenges of big data-enabled healthcare supply chain in addition to recommending future study topics. Based on their conclusion, big data has substantial promises in healthcare supply chains which requires deeper investigations.

Das and Griffin [82] systematically reviewed bias sources as well as their resolving strategies by overviewing the published papers and interviewing the experts. Their paper analyzed the topic frequency and evaluated the concept reliability by two independent trained coders. The trend of the unstructured textual contents was explored by a text mining pipeline which determined patterns and biases. The findings suggest maintaining the central locations of transportation experts and the public for determining the suitable objectives and metrics for assessing transportation safety, developing novel approaches to correlate big data to the total population, solving difficult problems with the help of big data, and progressing toward new trends and techniques.

N.Abdul Ghani *et al.* [83] reviewed the latest projects to broaden our insight on the big social media analytics. In this context, the authors classified the literature according to some prominent features. They also compared potential big data analytical technologies. Furthermore, they discussed the applicability of social big media data analytical systems through the emphasis on the presented state-of-the-art. Unsolved research challenges were also described.

C.Lim and coworkers presented the results obtained in diverse cases of big data application in cities as well as their projects in collaboration with the governmental organizations to develop smart cities. In particular, this review article classified the usage of urban data in 4 reference models and identified 6 major issues of data transformation into knowledge. According to the relevant studies, they also proposed five considerations in coping with issues arising in the implementation of the benchmark patterns in real-world cases. The benchmark patterns and assumptions formed a framework to apply data in smart cities. This article contributes to urban planning and policies toward a modern data-rich economy [84].

Abaker Targio Hashem *et al.* [85] described the latest communication techniques and smart-based applications which can be applied in smart cities. The concept of big data analytics in supporting smart cities was addressed with emphasis on the fundamental role of big data in altering urban populations at various levels. Furthermore, a business model was proposed for a smart city to identify the business and technological issues.

C. BIG DATA CHALLENGES IN SMART CITY

Challenges of Big Data in Smart Cities include the challenges that Big Data by its nature facing different challenges such as Volume, Velocity, Veracity, Value, etc. Big Data brings these challenges into any domain that deals with. This study refers to these challenges as Big Data Challenges in Smart Cities. Besides this group of challenges, implementing big data and relevant technologies in Smart Cities are also facing another group of challenges for instance Variety, Variability, Validity, Vulnerability, etc. This research categorizes them as Big Data Implementation Challenges in Smart Cities. Following sections describe these two categories of challenges in details.

1) INTEGRATION CHALLENGES OF BIG DATA IN SMART CITIES

Data generated from various sources will result in a concept known as big data. Data sources surround us as depicted in Fig.7. Diverse applications such as social media, digital pictures and videos, commercial transactions, advertisements, and games have recently accelerated data generation.



FIGURE 7. Big Data (BD) in smart cities.

Big data (BD) is a normal consequence of the developed digital artifacts along with applications. Modern digital technologies such as sensors, mobiles, and social media have penetrated our lives leading to an unprecedented huge deal of data (i.e. Big Data). “Big Data” can be defined as an ever-rising dataset whose management with conventional relational management systems is difficult. BD is often described by ten Vs: Volume, Velocity, Veracity, Value, Variety, Variability, Validity, Vulnerability, Volatility, and Visualization [86]–[88] as illustrated in Fig.8.

- The first “V” refers to the exponentially-growing volume of intelligent transportation systems (ITS) data. The volume of BD grows supersedes the conventional operational databases or data warehouses which mostly



FIGURE 8. Major attributes of big data.

rise to the order of GB or TB. Such huge amounts of BD could be quantified by new units such as Petabyte (10¹⁵) and Exabyte (10¹⁸).

- The second “V” stands for the velocity of ITS data, which has a wide variation range. Data could be continuously generated and collected in real-time at regularly-set intervals. The velocity of BD indicates the rate of data streamed into the host platforms. Besides the high incoming data rate, velocity causes significant questions on the issue of data aging for instance the validity duration of these data. Sometimes, streaming data have to be analyzed in real-time. As an instance, real-time analyses of video streams obtained from traffic surveillance cameras are crucial in predicting traffic jams and preventing bottlenecks.
- The third “V” refers to veracity also known as the ITS data trustworthiness level. In-time and reliable transportation data collection has been one of the serious challenges in the ITS community.
- The fourth “V” stands for the ITS data value that is highly dependent on the data age, their sampling frequency, and purposes. For example, consider a collision-prevention application where data aged few minutes old may be no longer valuable. In contrast, route planning applications can utilize non-real-time data. The value could be regarded as a measure evaluating the capability of extracting meaningful data and actionable business insight [89].
- As the fifth “V”, variety denotes the complexity of BD formats. Besides the conventional structured data, BD often includes semi-structured (such as text data and images stream) and unstructured (like geospatial data streams) data with the ratios of 20% and 80%, respectively. It is believed that the term “big” will lose its meaning over time and the self-evident concept of

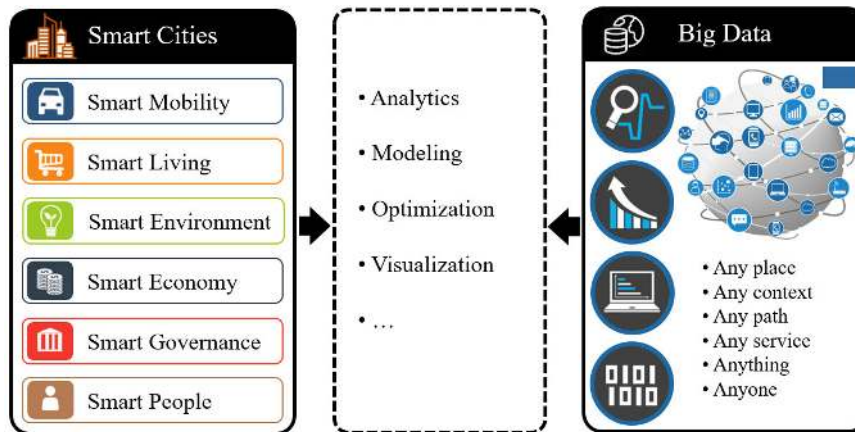


FIGURE 9. The relationship between sensing (as a service paradigm) and smart cities and big data.

“data” will be further extended to encompass the entire data types.

As the sixth “V” of ITS data, variability stands for a few different items including the lack of consistency in the data which have to be detected by anomaly and outlier detection routes to provide the possibility of any meaningful analytics. The variability of big data could be also assigned to the multitudes of data dimension as a result of various data types and sources. Variability could also indicate the inconsistency in big data loading speed onto the databases.

The seventh “V” of ITS data refers to the validity. Like veracity, validity deals with data reliability for the considered application. As reported by Forbes, scientists spend about 60% of their time to clean their data prior to any analyses. The advantages of big data analytical systems are as beneficial as their principal data; thus, proper data management practices should be adopted to guarantee reliable data characteristic, common definition, and metadata.

The eighth “V” of ITS data indicates the Vulnerability. Big data raise new security issues. Big data containing a breach can be considered as a big breach.

The ninth “V” of ITS data refers to the Volatility. Prior to the emergence of big data, organizations stored their data indefinitely since a few terabytes of data may not result in high storage costs; they could be even stored in the live database leading to no functioning issue. Classical data settings may not even possess a data archiving policy. Nonetheless, the high velocity and volume of big data require careful consideration of volatility. Some rules should be established for data currency and availability in addition to ensuring prompt information retrieval if necessary.

Finally, as the tenth “V” of ITS data, the visualization is a measure of challenges in visualizing. Currently-available big data visualization approaches encounter several challenges as a consequence of in-memory technological restrictions and weak scalability, response time, and capability. Conventional graphs are not applicable in plotting a billion data points this necessitates various data representation methods, i.e.

clustering of data or utilizing tree maps, sunbursts, parallel coordinates, round network plans, or cone trees. Combined with the variables due to the wide diversity of big data and velocity and their complicated associations, the development of a meaningful visualization would be a difficult task.

Storage, analysis, collection, and processing of the data are impossible through the current approaches as they include a huge amount of fast, dynamic, and various information.

Despite that, the comprehension of data usage in smart cities has remained limited. Some works have addressed the utilization of big data in a smart city to identify challenges in specific applications, including transportation, public safety, and sustainability [90]–[95]. Nonetheless, a limited number of investigations were focused on general knowledge about employing of big data in smart cities regardless of their application area [96].

As observed, BD is a result of technological advancements, not the applications or user’s requirements. On the other hand, SCs [96] are intended to resolve the issues of modern cities. The rural population migration and suburban concentration to the cities have posed serious challenges to citizens as well as city authorities. Waste management, energy, education, transportation, health, water resources, crime, and unemployment are among these serious problems [35]. SCs have been developed to resolve the mentioned challenges through the use of ICT. According to its definition, SC should possess six smart features: economy, people, governance, mobility, environment, and living [59]. According to Fig. 9, SC and Big Data, with different origins, are converged toward a common objective. Sensing can serve as a service paradigm between them with numerous academic and industrial patterns.

2) IMPLEMENTATION CHALLENGES OF BIG DATA IN SMART CITIES

Design, improvement, and operation of big data in smart cities could be highly challenging. Since a smart city is a

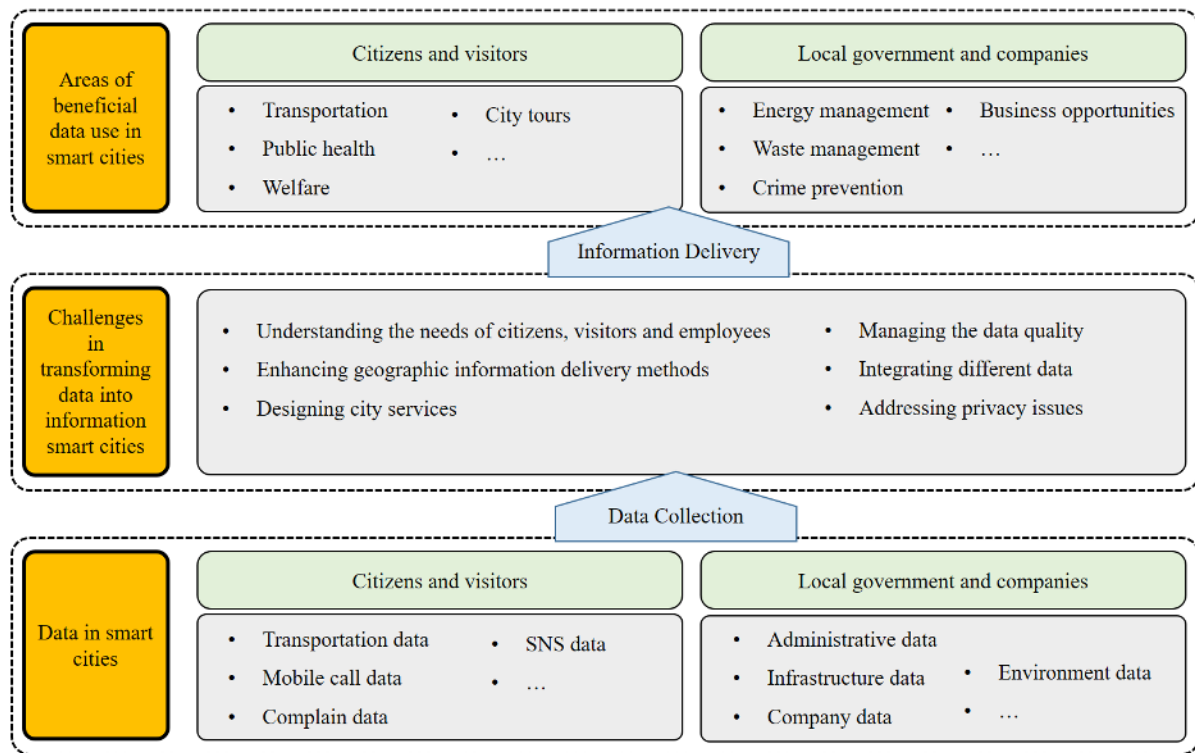


FIGURE 10. Challenges arising in transforming data into information for smart cities.

highly dynamic and evolving environment, these challenges should be eliminated or at least minimized [97], [98].

As mentioned before, the categories and amounts of data have immensely increased in modern cities offering versatile opportunities for the establishment of data-based smart cities. Similar to other large-scale transformations, moving toward a data-based smart city is not a simple task. In this context, issues arisen in the implementation of smart city projects due to the application of urban big data should be identified. This section is thus devoted to the problems emerging during the data-information transformation in a smart city [84], [99]. Fig. 10 shows the stages involved in collection of data to knowledge discovery in a smart city as well as the six major issues.

- Quality management: The proper quality of urban data is a prerequisite in the identification of reliable smart city information.
- Integration of diverse data: This process involves integrating data from various sources. According to Figure 10, different data could be gathered from various sources in a modern city. The important point is connecting these diverse data to provide the citizens and city authorities with reliable knowledge and excellent-quality information. As various organizations apply diverse data structures, this task could be highly demanding.
- Consideration of privacy issues: Privacy protection is a vital need in developing data-based smart cities to produce high-validity and sustainable value.

- Consideration of the citizens', visitors', and employees' demands: The determination of useful information plays an essential role in developing a data-based smart city as the usefulness of the delivered information has a direct association with the costumers' satisfaction from service.
- Improving the geographic information delivery method: A majority of big data are aimed at data analysis and delivering identified information based on a geographic element (e.g., district and building). In developing a data-based smart city, the knowledge content should be clearly visualized to increase information approval by the audiences.

Design of city services: This aspect involves designing smart city services to deliver knowledge from urban big data. Urban big data offers diverse information [84], [100]. In this regard, the proper design of data-based smart city service is vital as it can integrate all the data analytical systems, forming related ideas, and knowledge content design results.

Fig.10 illustrates the big data management in a smart city. This figure indicates the issues arisen upon implementing big data and the potential clarifications to resolve them. These smart decisions will positively affect the operations in a smart city.

Sometimes the recorded data might be short of the essential configuration which involves lost values. Then, these data could be useful after an effective and timely process [101].

Table 2 lists some of the major challenges of the implementation of big data in a smart city.

TABLE 2. Open issues in the application of big data in a smart city.

Open challenges and issues	Description, significance, and research directions to address the challenge
Data source and characteristics	Big data can be generated in diverse formats from various sources. Many of these data formats are unstructured (e.g. Audio, Video, Server logs, etc.) which can be handled into a structured format by highly-advanced database systems. Big data can generate numerous approaches and complicated models making its management even more difficult. Data mining tools fail in handling such large size and high complexity. Data collection is a complex task in multiple sources with different access policies and usage of various formats. The unstructured data are organized and categorized data which can be easily accessed and used.
Data and information sharing	Data and information sharing between various urban sectors is a serious concern. Urban authorities and sectors often possess private storage places for their confidential data. They are usually not willing to share their proprietary data. Moreover, some of this information may have special privacy requirements making their sharing even harder. The fine lines should not be crossed to guarantee the protection of citizens' privacy. Smart applications should devise a way to eliminate or decline the barriers in reaching seamless information exchange between various entities. Moreover, given the widely distributed data sources, some data types (e.g. spatio-temporal ones) could be rapidly updated. Thus, the establishment of a unified understanding of data semantics, and extraction of knowledge based on specific cycles and real-time data could be a difficult task. Therefore, the creation of a knowledge base for a smart city would be hard.
Data Quality	Big data provide data from various people and store them in unique databases. These data lack heterogeneity and consistency giving rise to disparity issues. Thus, there are no global approaches for transforming and retrieving the data source from useful analyses. For instance, sensor data are gathered from a third party with no central control.
Cost	Cost is a decisive issue affecting people's opinions on the application of ICT solutions. The use of energy reduction, components, and features for information recording are some of these examples.
Skill Gaps	Deficient data skills can hinder the efficient use of big data to manage the city. Management and analysis of large data sets and the development of proper policies require special skills, which can be rarely found especially in the public sector.
Smart city population	People influence smart city applications as the big data size depends on the city's population. The data size quickly grows by the rise in the population size. In cases with rapid population growth, traffic, social, economics, technical, problems, pollution, and environmental issues will proportionally increase. Big data prevent issues arisen from the smart city applications and efficiently manage such prompt growth. The goal is to develop and implement smart applications to manage such an ever-growing data.
Lack of Standards	Concerning SC development, relevant standards play a key role in complying with common platforms. We are now in shortage of standards related to data integration, policies, procedures, and formats.
Technology Advancement	Keeping pace with today's fast-growing technologies is highly demanding and costly. The smart city solutions have to progress in line with the advancements in the relevant technologies. Upgrading to the latest solution could be easier.

V. BIG DATA FRAMEWORK FOR SMART CITIES

Smart cities produce a huge deal of data that originated from various sources. Thus, the infrastructures should be able to store, process, and analyze such ever-growing data. Several prominent issues have to be considered in the design of a big-data framework for a smart city. Firstly, the big-data framework should ensure the effective storage of various data (structured, non-structured, and semi-structured) [56], [102]. Secondly, they should be capable of processing both real-time and historical data. Third, they should offer flexible data storage and processing (in cases with an unexpected load boost). Finally, they must show the ability to share the processed results across diverse applications/services ascendingly and scalable.

Considering the mentioned requirements, a conceptual big-data framework is presented in Fig.11. This framework can be categorized into four zones: Data acquisition (Zone 1), Data pre-processing (Zone 2), Data storage (Zone 3), and Data Analytics (Zone 4). All these zones are interconnected, in a way that the output of one of them is used as the input to the subsequent zone. As the first physical layer, Data acquisition encompasses various types of sensors and objects interlinked through diverse networking techniques. Sensors generate the intended data. The communication occurs by wires or through wireless routes (RFID, WiFi, Zigbee, and

Bluetooth). A gateway is also applied to connect Zones 1 and 2. Various gateways could be incorporated to obtain more robust frameworks [103].

Zone 2 pre-processes the raw naïve data from Zone 1 as these raw data have to be preprocessed prior to storage to modify their large volume and remove the duplicate data or uncertainty information. A typical preprocessing technique includes data cleaning, integration, and compression.

Data storage is accomplished in Zone 3 which stores and manages immense data sets. The data storage systems can be classified into two classes: storage infrastructure and data management software. The former encompasses the storage devices in addition to the networking devices to link the storage devices. Data management software is another essential requirement of data storage systems.

Zone 4 applies different data analytics approaches for extracting valuable information from large datasets. The data analytics techniques could be divided into three groups: (i) descriptive, (ii) predictive, and (iii) prescriptive as explained in Table 3.

VI. FUTURE RESEARCH OPPORTUNITIES

Table 3 lists the possible solutions for the issues arising in big data analytics of large wireless networks in terms of the mentioned zones. Many of these research challenges have

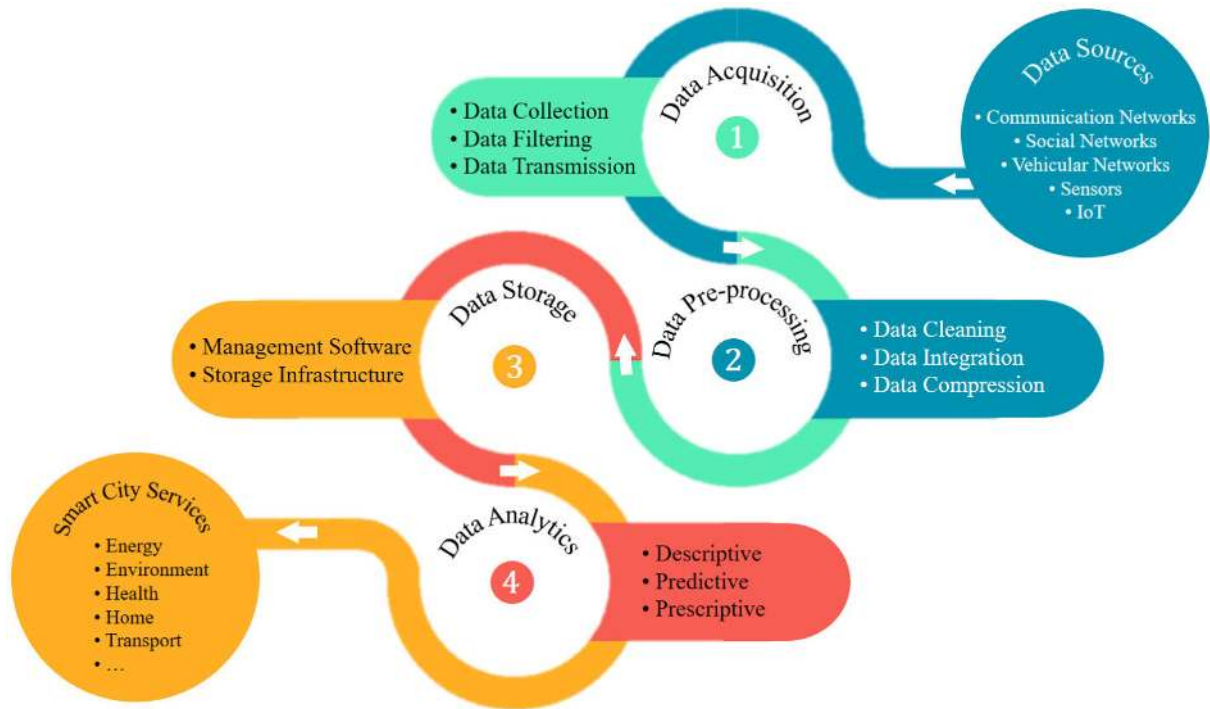


FIGURE 11. Big data framework for a smart city.

TABLE 3. Challenges and opportunities in analyzing big data.

Big data	Challenges	Opportunities
Data acquisition [104]–[107]	<ul style="list-style-type: none"> • Data representation • Efficient data collection • Effective data transmission 	<ul style="list-style-type: none"> • Heterogeneity of data • Secure data transmission • Energy harvesting within the data transmission
Data pre-processing [108]–[110]	<ul style="list-style-type: none"> • Integrating diverse data • Declining duplications • Data cleaning and compression 	<ul style="list-style-type: none"> • Privacy protection • Security • Energy-efficiency
Data storage [15], [86], [111]	<ul style="list-style-type: none"> • Reliable and consistent data storage • Scalability • Efficient data storage 	<ul style="list-style-type: none"> • Data storage and distribution of models used in heterogeneous networks • Computing models with low computational costs • Privacy and security
Data Analytics [112]–[114]	<ul style="list-style-type: none"> • Temporal and spatial correlation of data • Effective data mining • Privacy issues 	<ul style="list-style-type: none"> • Energy-efficiency and time-sensitiveness • Privacy • Security

been resolved, however, many have remained unsolved. In the next section, the future directions in big data analytics for large-scale wireless networks are described in Fig. 12.

A. DISTRIBUTED DATA PROCESSING MODELS

Despite the huge efforts devoted to the development of distributed data processing models to be used in large-scale wireless networks, numerous issues have remained unsolved.

- Stream data processing: data storage and processing in memory is practically impossible regarding the huge

amount of real-time data (e.g., sensor data from WSNs). Therefore, the conventional data analyzing algorithms requiring access to the entire data sets will not operate under this scenario [56].

- In-network processing: Given the tremendous wireless nodes distributed throughout the large-scale wireless networks, the integration of their data is a vital step in the data processing. Nonetheless, data fusion among distributed networks will significantly increase the communication cost. One approach to resolve this problem

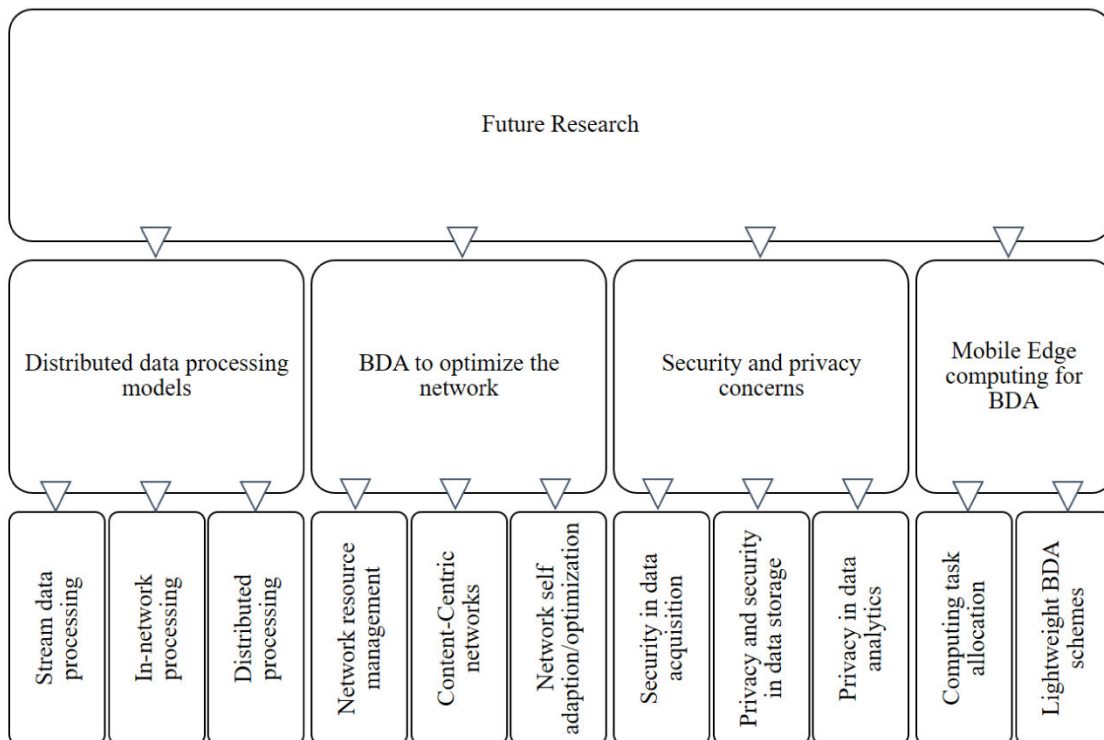


FIGURE 12. Future research opportunities for BDA of wireless network.

is in-network processing throughout the network which involves data processing at each individual node rather than the centralized servers. The computational costs can be further declined by clustering the nodes [115] which might result in new challenges: cluster selection to meet the big data requirement.

- **Distributed processing:** MapReduce [116] could offer several advantages, including simplicity, fault tolerance, and scalability. It, however, suffers from some limitations due to its lower efficiency compared to the other parallel processing models. Conventional parallel computing models (e.g. Message Passing Interface (MPI), OpenMP (Open Multi-Processing), FPGA-oriented programming, and GPU-based parallel processing (e.g., NVIDIA's CUDA)) exhibit better performances compared to the MapReduce and its analogous models. The integration of MapReduce-type models with parallel computing ones could enhance the performance. Moreover, the advantages of some processing platforms could ameliorate the available data analysis algorithms. For instance, FPGA-oriented Convolutional neural network (CNN) [96] and FPGA deep-learning have been explored in this regard.

- **Network resource management:** According to BDA of network data, the network administrator is capable of predicting the network resource demands. As mentioned in [117], it is easy to predict congestion in some parts of the city upon the incidence of some social events like the marathon.
- **Content-centric networks:** As suggested in many previous works [118], storage of some popular content (also known as caches) at the base station could remarkably decline the real-time traffic and hence enhance the network activity. The determination of the cache is a new issue. It is possible to obtain cache information by the analysis of the application data. The privacy concerns and the heterogeneity of data from diverse applications make it difficult to attain precise user information.
- **Network self-adaption/self-optimization:** BDA can remarkably contribute to network self-adaption or self-optimization in self-organizing networks (SONs) [119]. Fan *et al.* [120] developed a self-optimization approach through integrating fuzzy neural network and reinforcement learning which managed to meet the coverage and capacity criteria of SONs. However, further studies are required in this novel field of research.

B. BIG DATA ANALYTICS TO OPTIMISE THE NETWORK

Big data analytics (BDA) can be employed for the optimization of large-scale wireless networks. Here, several unsolved challenges in the use of BDA for network optimization will be discussed.

C. SECURITY AND PRIVACY CONCERNS

As two prominent challenges in BDA of wireless networks, security and privacy are tightly related whereas showing some differences [121]: Security involves data confidentiality, integrity, and availability while privacy guarantees the

appropriate data usage avoiding the disclosure of user private information without his/her consent. Possible research themes in BDA security and privacy within wireless networks are:

- Security in data acquisition: During data acquisition, wiretapping behavior may occur leading to information leakage. In this regard, huge studies have addressed the confidentiality of wireless networks. Encryption schemes could be applied in wireless networks [121]. Nonetheless, energy and computational limitations of smart objects have made these techniques impractical in IoT. Thus, novel lightweight protective schemes should be extended for IoT [122].
- Privacy and security in data storage: Personal confidential data could be leaked upon invasion to data storage highlighting the importance of data protection in this phase. Fortunately, the application of encryption algorithms in the data storage phase is more feasible compared to the data acquisition (transmission) phase. Albeit, the enforcement of privacy-preserved operations in data storage phase might be difficult, in particular when the service is provided by a third party entity [123]. Mobile Edge Computing (MEC) [124] is able to resolve the issue through data offloading from the unreliable third party to a trusted MEC server (near the user).
- Privacy in data analytics: The balancing privacy and efficiency of data analysis is a major challenge. The private documents could be protected by encryption and storage in a secure server (or a cloud). On the other hand, analysis of encrypted document is time-consuming, hence declining the data analytics efficiency [56]. There are numerous unsolved issues in data publishing [125], data mining output, and distributed data privacy [126] requiring deeper investigations.

D. MOBILE EDGE COMPUTING FOR BDA

Regarding the intrinsic limitations of wireless nodes like their power restrictions and lower computational ability, the computing tasks are deployed to a remote cloud server with higher computational ability with no resource limitations. Drawbacks such as high latency, performance bottleneck, context unawareness, and privacy exposure have also limited cloud computing [127]. MEC (or Fog computing) could be employed as a complementary part of cloud computing systems to overcome the mentioned drawbacks. The MEC relies on offloading the computational processes from a remote cloud to diverse MEC servers at base stations, IoT gateways, and WiFi APs near the end-users. Therefore, the extensive delay-tolerant jobs would be implemented at remote clouds whereas the less-intensive delay-critical, and context-aware ones are offloaded to an edge server.

- Computing task allocation: Different computing resources can be found in wireless networks. Therefore, the computation resources should be allocated to various computing devices which could be difficult in

cases of allocating and coordinating within distributed large-scale networks.

- Lightweight BDA schemes: As expressed in [127] AlexNet (i.e., a typical convolutional neural network) can lead to high server-edge node communication costs. The resource restrictions of edge servers and mobile devices have led to development of lightweight BDA scheme and compressing BDA model whose hardware architecture, optimization, data compression, distributed computing, and machine learning require further investigations [103].

VII. CONCLUSION

The ever-growing increase in connected devices in the cities has resulted in speeding data increment attracting the attention of numerous researchers from diverse fields of study. This article is aimed at presenting a comprehensive review of the undeniable obligation of big data in the smart cities. To this end, the enabling technologies were described. The architecture was also suggested to manage big data in a smart city. The essential role of big data analytics in the smart city's applications was also addressed. Different cases were explained. Several unsolved challenges were also described to suggest possible research directions for future studies. Finally, it was concluded that big data could play a decisive role in obtaining valuable information and decision-making processes. Nonetheless, research on the role of big data in smart cities is at the beginning of its journey, and resolving the mentioned challenges can facilitate its practical use.

REFERENCES

- [1] 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.
- [2] 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.
- [3] C. Liao and L. Nong, "Smart city sports tourism integration based on 5G network and Internet of Things," *Microprocessors Microsyst.*, Jan. 2021, Art. no. 103971.
- [4] A. Abella, M. Ortiz-de-Urbina-Criado, and C. De-Pablos-Heredero, "A model for the analysis of data-driven innovation and value generation in smart cities' ecosystems," *Cities*, vol. 64, pp. 47–53, Apr. 2017.
- [5] 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," *Technol. Forecasting Social Change*, vol. 90, pp. 611–622, Jan. 2015.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] H. M. Kim, S. Sabri, and A. Kent, "Smart cities as a platform for technological and social innovation in productivity, sustainability, and livability: A conceptual framework," in *Smart Cities for Technological and Social Innovation*. New York, NY, USA: Academic, 2021, pp. 9–28.
- [10] M. Lom and O. Pribyl, "Smart city model based on systems theory," *Int. J. Inf. Manage.*, vol. 56, Feb. 2021, Art. no. 102092.

- [11] M. I. Pramanik, R. Y. K. Lau, S. Hossain, M. Rahoman, S. K. Debnath, G. Rashed, and Z. Uddin, "Privacy preserving big data analytics: A critical analysis of state-of-the-art," *Wiley Interdiscipl. Rev.: Data Mining Knowl. Discovery*, vol. 11, no. 1, p. e1387, 2021.
- [12] L. Leenen and T. Meyer, "Artificial intelligence and big data analytics in support of cyber defense," in *Research Anthology on Artificial Intelligence Applications in Security*. Hershey, PA, USA: IGI Global, 2021, pp. 1738–1753.
- [13] 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.
- [14] 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.
- [15] L. Al-Awami and H. S. Hassanein, "Robust decentralized data storage and retrieval for wireless networks," *Comput. Netw.*, vol. 128, pp. 41–50, Dec. 2017.
- [16] 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.
- [17] J. Tian and L. Gao, "Using data monitoring algorithms to physiological indicators in motion based on Internet of Things in smart city," *Sustain. Cities Soc.*, vol. 67, Apr. 2021, Art. no. 102727.
- [18] D. Pal, T. Triyason, and P. Padungweang, "Big data in smart-cities: Current research and challenges," *Indonesian J. Electr. Eng. Informat. (IJEEI)*, vol. 6, no. 4, pp. 351–360, Dec. 2018.
- [19] M. Gordan, Z. Ismail, K. Ghaedi, Z. Ibrahim, H. Hashim, H. H. Ghayeb, and M. Talebkhah, "A brief overview and future perspective of unmanned aerial systems for in-service structural health monitoring," *Eng. Adv.*, vol. 1, no. 1, pp. 9–15, Feb. 2021.
- [20] S. Sengan, V. Subramaniaswamy, S. K. Nair, V. Indragandhi, J. Manikandan, and L. Ravi, "Enhancing cyber-physical systems with hybrid smart city cyber security architecture for secure public data-smart network," *Future Gener. Comput. Syst.*, vol. 112, pp. 724–737, Nov. 2020.
- [21] 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. Frontiers*, pp. 1–22, Jul. 2020.
- [22] J. S. B. Martins, "Towards smart city innovation under the perspective of software-defined networking, artificial intelligence and big data," *arXiv1810.11665*, vol. 8, no. 2, pp. 1–8, Oct. 2018.
- [23] 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.
- [24] X. He, J. Zhu, W. Su, and M. M. Tentzeris, "RFID based non-contact human activity detection exploiting cross polarization," *IEEE Access*, vol. 8, pp. 46585–46595, 2020.
- [25] J. Maria de Fuentes, L. Gonzalez-Manzano, A. Solanas, and F. Veseli, "Attribute-based credentials for privacy-aware smart health services in IoT-based smart cities," *Computer*, vol. 51, no. 7, pp. 44–53, Jul. 2018.
- [26] 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.
- [27] S. J. Nawaz, S. K. Sharma, S. Wyne, M. N. Patwary, and M. Asaduzzaman, "Quantum machine learning for 6G communication networks: State-of-the-art and vision for the future," *IEEE Access*, vol. 7, pp. 46317–46350, 2019.
- [28] 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. (HICAJET)*, Sep. 2020, pp. 1–6.
- [29] G. Javadzadeh and A. M. Rahmani, "Fog computing applications in smart cities: A systematic survey," *Wireless Netw.*, vol. 26, no. 2, pp. 1433–1457, Feb. 2020.
- [30] F. Fontes, R. Antão, A. Mota, and P. Pedreiras, "Improving the ambient temperature control performance in smart homes and buildings," *Sensors*, vol. 21, no. 2, p. 423, 2021.
- [31] E. Ismagilova, L. Hughes, Y. K. Dwivedi, and K. R. Raman, "Smart cities: Advances in research—An information systems perspective," *Int. J. Inf. Manage.*, vol. 47, pp. 88–100, Aug. 2019.
- [32] T. S. Gomides, R. E. De Grande, A. M. D. Souza, F. S. H. Souza, L. A. Villas, and D. L. Guidoni, "An adaptive and distributed traffic management system using vehicular ad-hoc networks," *Comput. Commun.*, vol. 159, pp. 317–330, Jun. 2020.
- [33] Y. A. Fatimah, K. Govindan, R. Murningsih, and A. Setiawan, "Industry 4.0 based sustainable circular economy approach for smart waste management system to achieve sustainable development goals: A case study of Indonesia," *J. Cleaner Prod.*, vol. 269, Oct. 2020, Art. no. 122263.
- [34] W. Ejaz, M. Naeem, A. Shahid, A. Anpalagan, and M. Jo, "Efficient energy management for the Internet of Things in smart cities," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 84–91, Jan. 2017.
- [35] 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.
- [36] R. K. R. Kummitha, "Smart cities and entrepreneurship: An agenda for future research," *Technological Forecasting Social Change*, vol. 149, Dec. 2019, Art. no. 119763.
- [37] B. Bhushan, A. Khamparia, K. M. Sagayam, S. K. Sharma, M. A. Ahad, and N. C. Debnath, "Blockchain for smart cities: A review of architectures, integration trends and future research directions," *Sustain. Cities Soc.*, vol. 61, Oct. 2020, Art. no. 102360.
- [38] 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.
- [39] A. M. S. Osman, "A novel big data analytics framework for smart cities," *Future Gener. Comput. Syst.*, vol. 91, pp. 620–633, Feb. 2019.
- [40] Z. Mingaleva, N. Vukovic, I. Volkova, and T. Salimova, "Waste management in green and smart cities: A case study of Russia," *Sustainability*, vol. 12, no. 1, p. 94, 2020.
- [41] 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.
- [42] 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.
- [43] S. Idwan, I. Mahmood, J. A. Zubairi, and I. Matar, "Optimal management of solid waste in smart cities using Internet of Things," *Wireless Pers. Commun.*, vol. 110, no. 1, pp. 485–501, Jan. 2020.
- [44] J. Laufs, H. Borrión, and B. Bradford, "New security technologies and their functions in smart cities: A systematic review," *Sustain. Cities Soc.*, vol. 55, Mar. 2020, Art. no. 102023.
- [45] 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.
- [46] K. D. Kang, H. Kang, I. M. S. K. Ilankoon, and C. Y. Chong, "Electronic waste collection systems using Internet of Things (IoT): Household electronic waste management in Malaysia," *J. Cleaner Prod.*, vol. 252, Apr. 2020, Art. no. 119801.
- [47] T. Wakiyama, M. Lenzen, A. Geschke, R. Bamba, and K. Nansai, "A flexible multiregional input-output database for city-level sustainability footprint analysis in Japan," *Resour. Conservation Recycling*, vol. 154, Mar. 2020, Art. no. 104588.
- [48] P. Partheeban, H. P. Raju, R. R. Hemamalini, and B. Shanthini, "Real-time vehicular air quality monitoring using sensing technology for Chennai," in *Transportation Research Singapore*: Springer, 2020, pp. 19–28.
- [49] 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.
- [50] R. Chauhan, H. Kaur, and V. Chang, "An optimized integrated framework of big data analytics managing security and privacy in healthcare data," *Wireless Pers. Commun.*, vol. 117, pp. 87–108, Feb. 2020.
- [51] 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.
- [52] C. Zhu, H. Zhou, V. C. M. Leung, K. Wang, Y. Zhang, and L. T. Yang, "Toward big data in green city," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 14–18, Nov. 2017.
- [53] C. Metallo, R. Agrifoglio, F. Schiavone, and J. Mueller, "Understanding business model in the Internet of Things industry," *Technol. Forecasting Social Change*, vol. 136, pp. 298–306, Nov. 2018.

- [54] F. Cirillo, D. Gomez, L. Diez, I. Elicegeui 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.
- [55] 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.
- [56] 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, p. 99, 2019.
- [57] 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.
- [58] J. Zhang, Z. Zhou, S. Li, L. Gan, X. Zhang, L. Qi, X. Xu, and W. Dou, "Hybrid computation offloading for smart home automation in mobile cloud computing," *Pers. Ubiquitous Comput.*, vol. 22, no. 1, pp. 121–134, Feb. 2018.
- [59] A. Muthanna, A. A. Ateya, A. Khakimov, I. Gudkova, A. Abuarqoub, K. Samouylov, and A. Koucheryavy, "Secure and reliable IoT networks using fog computing with software-defined networking and blockchain," *J. Sensor Actuator Netw.*, vol. 8, no. 1, p. 15, Feb. 2019.
- [60] A. A. Guenduez, S. Singler, T. Tomczak, K. Schedler, and M. Oberli, "Smart government success factors," *Swiss Yearbook Administ. Sci.*, vol. 9, no. 1, pp. 96–110, Dec. 2018.
- [61] 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.
- [62] 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.
- [63] N. Mohamed, J. Al-Jaroodi, I. Jawhar, A. Idries, and F. Mohammed, "Unmanned aerial vehicles applications in future smart cities," *Technol. Forecasting Social Change*, vol. 153, Apr. 2020, Art. no. 119293.
- [64] J. Trevanthan and R. Johnstone, "Smart environmental monitoring and assessment technologies (SEMAT)—A new paradigm for low-cost, remote aquatic environmental monitoring," *Sensors*, vol. 18, no. 7, p. 2248, 2018.
- [65] S. A. A. Elmustafa and E. Y. Mujtaba, "Internet of Things in smart environment: Concept, applications, challenges, and future directions," *World Sci. News*, vol. 134, no. 1, pp. 1–51, 2019.
- [66] A. Jiang, H. Yuan, D. Li, and J. Tian, "Key technologies of ubiquitous power Internet of Things-aided smart grid," *J. Renew. Sustain. Energy*, vol. 11, no. 6, Nov. 2019, Art. no. 062702.
- [67] A. A. Munshi and Y. A.-R.-I. Mohamed, "Big data framework for analytics in smart grids," *Electric Power Syst. Res.*, vol. 151, pp. 369–380, Oct. 2017.
- [68] 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.
- [69] 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, 2021.
- [70] S. E. Bibri, "Compact urbanism and the synergic potential of its integration with data-driven smart urbanism: An extensive interdisciplinary literature review," *Land Use Policy*, vol. 97, Sep. 2020, Art. no. 104703.
- [71] Q. Zhang, X. Zhang, Q. Zhang, W. Shi, and H. Zhong, "Firework: Big data sharing and processing in collaborative edge environment," in *Proc. 4th IEEE Workshop Hot Topics Web Syst. Technol. (HotWeb)*, Oct. 2016, pp. 20–25.
- [72] B. Ahlgren, M. Hidell, and E. C. H. Ngai, "Internet of Things for smart cities: Interoperability and open data," *IEEE Internet Comput.*, vol. 20, no. 6, pp. 52–56, Nov./Dec. 2016.
- [73] B. Cheng, A. Papageorgiou, F. Cirillo, and E. Kovacs, "Geelytics: Geo-distributed edge analytics for large scale IoT systems based on dynamic topology," in *Proc. IEEE 2nd World Forum Internet Things (WF-IoT)*, Dec. 2015, pp. 565–570.
- [74] E. Ahmed and M. H. Rehmani, "Introduction to the special section on social collaborative Internet of Things," *Comput. Electr. Eng.*, vol. 58, pp. 382–384, Feb. 2017.
- [75] H. R. Arkian, A. Diyanat, and A. Pourkhalili, "MIST: Fog-based data analytics scheme with cost-efficient resource provisioning for IoT crowd-sensing applications," *J. Netw. Comput. Appl.*, vol. 82, pp. 152–165, Mar. 2017.
- [76] 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.
- [77] F. Alam, R. Mehmood, I. Katib, and A. Albeshri, "Analysis of eight data mining algorithms for smarter Internet of Things (IoT)," *Procedia Comput. Sci.*, vol. 98, pp. 437–442, 2016.
- [78] 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.
- [79] M. Gordan, Z. Ismail, H. Abdul 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.
- [80] D. Mourtzis, E. Vlachou, and N. Milas, "Industrial big data as a result of IoT adoption in manufacturing," *Procedia CIRP*, vol. 55, pp. 290–295, 2016.
- [81] 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.
- [82] S. Das and G. P. Griffin, "Big data and transportation safety: Connecting the dots," in *Proc. Transp. Res. Board Annu. Meeting*, Washington, DC, USA, 2020, pp. 12–16.
- [83] N. A. Ghani, S. Hamid, I. A. T. Hashem, and E. Ahmed, "Social media big data analytics: A survey," *Comput. Hum. Behav.*, vol. 101, pp. 417–428, Dec. 2018.
- [84] C. Lim, K. Kim, and P. P. Maglio, "Smart cities with big data: Reference models, challenges, and considerations," *Cities*, vol. 82, pp. 86–99, Dec. 2018.
- [85] 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, 2016.
- [86] S. K. Sharma and X. Wang, "Live data analytics with collaborative edge and cloud processing in wireless IoT networks," *IEEE Access*, vol. 5, pp. 4621–4635, 2017.
- [87] 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.
- [88] 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, 2020, pp. 3–38.
- [89] Y. Cui, S. Kara, and K. C. Chan, "Manufacturing big data ecosystem: A systematic literature review," *Robot. Comput.-Integr. Manuf.*, vol. 62, Apr. 2020, Art. no. 101861.
- [90] 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, p. 100303, 2020.
- [91] F. Al-turjman and J. P. Lemayian, "Intelligence, security, and vehicular sensor networks in Internet of Things (IoT)-enabled smart-cities: An overview," *Comput. Electr. Eng.*, vol. 87, Art. no. 2020, Art. no. 106776.
- [92] R. Díaz-díaz, L. Muñoz, and D. Pérez-gonzález, "Business model analysis of public services operating in the smart city ecosystem: The case of SmartSantander," *Future Gener. Comput. Syst.*, vol. 76, pp. 198–214, Nov. 2017.
- [93] 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.
- [94] 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, 2020.
- [95] A. Kirimtat, O. Krejcar, A. Kertes, and M. F. Tasgetiren, "Future trends and current state of smart city concepts: A survey," *IEEE Access*, vol. 8, pp. 86448–86467, 2020.

- [96] K. Soomro, M. Nasir Mumtaz Bhutta, M. Bhutta, Z. Khan, and M. A. Tahir, "Smart city big data analytics: An advanced review," *Wiley Interdiscipl. Rev.: Data Mining Knowl. Discovery*, vol. 9, no. 5, p. e1319, 2019.
- [97] S. Singh and A. Yassine, "Big data mining of energy time series for behavioral analytics and energy consumption forecasting," *Energies*, vol. 11, no. 2, p. 452, Feb. 2018.
- [98] M. A. Amin, S. Hadouej, and T. S. J. Darwish, "Big data role in improving intelligent transportation systems safety: A survey," in *Proc. Int. Conf. Emerg. Internetworking, Data Web Technol.*, 2019, pp. 187–199.
- [99] 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, 2017.
- [100] A. Aliyu, A. H. Abdullah, O. Kaiwartya, Y. Cao, J. Lloret, N. Aslam, and U. M. Joda, "Towards video streaming in IoT environments: Vehicular communication perspective," *Comput. Commun.*, vol. 118, pp. 93–119, Mar. 2018.
- [101] I. Mistry, S. Tanwar, S. Tyagi, and N. Kumar, "Blockchain for 5G-enabled IoT for industrial automation: A systematic review, solutions, and challenges," *Mech. Syst. Signal Process.*, vol. 135, Jan. 2020, Art. no. 106382.
- [102] D. Gil, M. Johnsson, H. Mora, and J. Szymanski, "Review of the complexity of managing big data of the Internet of Things," *Complexity*, vol. 2019, pp. 1–12, Feb. 2019.
- [103] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "Model compression and acceleration for deep neural networks: The principles, progress, and challenges," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 126–136, Jan. 2018.
- [104] H. Zhou, H. Wang, X. Li, and V. C. M. Leung, "A survey on mobile data offloading technologies," *IEEE Access*, vol. 6, pp. 5101–5111, 2018.
- [105] J. Sahoo, S. Cherkaoui, A. Hafid, and P. K. Sahu, "Dynamic hierarchical aggregation for vehicular sensing," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2539–2556, Sep. 2017.
- [106] M. Hajarian, A. Bastanfard, J. Mohammadzadeh, and M. Khalilian, "Introducing fuzzy like in social networks and its effects on advertising profits and human behavior," *Comput. Hum. Behav.*, vol. 77, pp. 282–293, Dec. 2017.
- [107] 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.
- [108] A. Belhadi, Y. Djenouri, G. Srivastava, D. Djenouri, J. C.-W. Lin, and G. Fortino, "Deep learning for pedestrian collective behavior analysis in smart cities: A model of group trajectory outlier detection," *Inf. Fusion*, vol. 65, pp. 13–20, Jan. 2021.
- [109] G. Ertek, X. Chi, and A. N. Zhang, "A framework for mining RFID data from schedule-based systems," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 11, pp. 2967–2984, Nov. 2017.
- [110] C. Deng, R. Guo, C. Liu, R. Y. Zhong, and X. Xu, "Data cleansing for energy-saving: A case of cyber-physical machine tools health monitoring system," *Int. J. Prod. Res.*, vol. 56, nos. 1–2, pp. 1000–1015, Jan. 2018.
- [111] H. Hu, Y. Wen, and D. Niyato, "Public cloud storage-assisted mobile social video sharing: A supermodular game approach," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 3, pp. 545–556, Mar. 2017.
- [112] Q.-V. Pham, F. Fang, V. N. Ha, M. J. Piran, M. Le, L. B. Le, W.-J. Hwang, and Z. Ding, "A survey of multi-access edge computing in 5G and beyond: Fundamentals, technology integration, and state-of-the-art," *IEEE Access*, vol. 8, pp. 116974–117017, 2020.
- [113] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transp. Res. Part C: Emerg. Technol.*, vol. 79, pp. 1–17, Jun. 2017.
- [114] H. Saeedi Emadi and S. M. Mazinani, "A novel anomaly detection algorithm using DBSCAN and SVM in wireless sensor networks," *Wireless Pers. Commun.*, vol. 98, no. 2, pp. 2025–2035, Jan. 2018.
- [115] B. Diène, J. J. P. C. Rodrigues, O. Diallo, E. H. M. Ndoeye, and V. V. Korotaev, "Data management techniques for Internet of Things," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106564.
- [116] A. Yousefpour, C. Fung, T. Nguyen, K. Kadiyala, F. Jalali, A. Niakanlahiji, J. Kong, and J. P. Jue, "All one needs to know about fog computing and related edge computing paradigms: A complete survey," *J. Syst. Archit.*, vol. 98, pp. 289–330, Sep. 2019.
- [117] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou, "Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1606–1615, Apr. 2018.
- [118] Z. Su, Q. Xu, and Q. Qi, "Big data in mobile social networks: A QoE-oriented framework," *IEEE Netw.*, vol. 30, no. 1, pp. 52–57, Jan. 2016.
- [119] X. Wang, X. Li, and V. C. M. Leung, "Artificial intelligence-based techniques for emerging heterogeneous network: State of the arts, opportunities, and challenges," *IEEE Access*, vol. 3, pp. 1379–1391, 2015.
- [120] J. Tang, F. Gao, F. Liu, and X. Chen, "A denoising scheme-based traffic flow prediction model: Combination of ensemble empirical mode decomposition and fuzzy C-means neural network," *IEEE Access*, vol. 8, pp. 11546–11559, 2020.
- [121] 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.
- [122] Y. Pan and L. Zhang, "Roles of artificial intelligence in construction engineering and management: A critical review and future trends," *Autom. Construct.*, vol. 122, Feb. 2021, Art. no. 103517.
- [123] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A survey on Internet of Things: Architecture, enabling technologies, security and privacy, and applications," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1125–1142, Oct. 2017.
- [124] Q.-V. Pham, L. B. Le, S.-H. Chung, and W.-J. Hwang, "Mobile edge computing with wireless backhaul: Joint task offloading and resource allocation," *IEEE Access*, vol. 7, pp. 16444–16459, 2019.
- [125] J. Zhang, G. Cormode, C. M. Procopiuc, D. Srivastava, and X. Xiao, "PrivBayes: Private data release via Bayesian networks," *ACM Trans. Database Syst.*, vol. 42, no. 4, p. 25, 2017.
- [126] R. Mendes and J. P. Vilela, "Privacy-preserving data mining: Methods, metrics, and applications," *IEEE Access*, vol. 5, pp. 10562–10582, 2017.
- [127] G. Zou, Z. Qin, S. Deng, K.-C. Li, Y. Gan, and B. Zhang, "Towards the optimality of service instance selection in mobile edge computing," *Knowl.-Based Syst.*, vol. 217, Apr. 2021, Art. no. 106831.



MARIEH TALEBKHAH has received the B.Sc. degree in computer software engineering from the University of Birjand, Iran, in 2009, and the M.Sc. degree in computer software engineering from the University of Damghan, Iran, in 2017. 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 computing, and big data analytics.



ADUWATI SALI (Senior Member, IEEE) received the B.Eng. degree in electrical electronics engineering (communications) from The University of Edinburgh, U.K., in 1999, the M.Sc. degree in communications and network engineering from the Universiti Putra Malaysia (UPM), Malaysia, in April 2002, and the Ph.D. degree in mobile and satellite communications from the University of Surrey, U.K., in July 2009.

She worked as an Assistant Manager with Telekom Malaysia Berhad, from 1999 to 2000. She was a Deputy Director with the UPM Research Management Centre (RMC) responsible for Research Planning and Knowledge Management, from 2016 to 2019. She has been a Professor with the Department of Computer and Communication Systems, Faculty of Engineering, UPM, since February 2019. She is also a Chartered Engineer (C.Eng) registered under U.K. Engineering Council and a Professional Engineer (P.Eng.) under Board of Engineers Malaysia (BEM). She gave consultations to Malaysian Ministry of Information and Multimedia, Malaysian Ministry of Higher Education, National Space Agency (ANGKASA), ATSB Bhd and Petronas Bhd. on projects related to mobile and satellite communications. In 2014, the fateful event of missing MH370 has requested her to be in printed and broadcasting media, specifically Astro Awani, RTM, TV Al-Hijrah, BERNAMA, City-Plus FM, Harian Metro, and Metro Ahad, regarding analysis on satellite communication in tracking the aircraft. She was involved with EU-IST Satellite Network of Excellence (SatNEx) I and II, from 2004 to 2009. She is the Principle Investigator and a Collaborator for projects under the local and international funding bodies; namely Malaysian Ministry of Science, Technology and Innovation (MOSTI), Malaysian Ministry of Higher Education (MoHE), Malaysian Communications and Multimedia Commission (MCMC), Research University Grant Scheme (RUGS) (now known as Putra Initiative Grant) UPM, The Academy of Sciences for the Developing World (TWAS-COMSTTECH) Joint Grants, EU Horizon2020 Research and Innovation Staff Exchange (H2020-RISE), EU ERASMUS+ Capacity Building for Higher Education (CBHE), and NICT Japan-ASEAN IVO. Her research interests include radio resource management, MAC layer protocols, satellite communications, satellite-assisted emergency communications, the IoT systems for environmental monitoring, and 3D video transmission over wireless networks.

Dr. Sali was a recipient of the 2018 Top Research Scientists Malaysia (TRSM) Award from the Academy of Sciences Malaysia (ASM). She was involved with IEEE as the Chair of ComSoc/VTS Malaysia from 2017 to 2018 and a Young Professionals (YP) in 2015. She has been with the Young Scientists Network-Academy of Sciences Malaysia (YSN-ASM) as an Honorary Member, since 2020. She was also the Chair in 2018 and a Co-Chair in 2017 for Science Policy.



MOHSEN MARJANI received the master's degree in information (multimedia computing) from Multimedia University (MMU), Malaysia, in 2011, and the Ph.D. degree in computer science from the University of Malaya (UM), Malaysia. He is currently a Lecturer with the School of Computer Science and Engineering, Taylor's University, Malaysia. Since mid-1999, he has been started teaching Mathematics and IT-based subjects in different public, and private institutions,

including universities, pre-universities, high schools, and primary schools. He has experience of collaborating with and working for a number of IT companies as a IT Consultant, Senior Web and Mobile Developer, Tech Lead, Project Manager, and CTO. He has published a number of research articles in refereed international journals. His research interests include big data, the IoT, data analytics, AI, machine learning, deep learning, and image processing.



MEISAM GORDAN received the M.Sc. degree in civil engineering from the Department of Structure and Materials, University Technology of Malaysia (UTM), in 2014, and the Ph.D. degree from the University of Malaya (UM), Kuala Lumpur, in 2020. His doctoral thesis entitled "Data Mining for Structural Damage Identification Using Hybrid Artificial Neural Network Based Algorithm for Beam and Slab Girder" was supervised by Prof. Hashim Abdul Razak and Prof. Zubaidah Ismail.

His research interests include civil engineering and computer science using emerging technologies.



SHAIFUL JAHARI HASHIM received the B.Eng. degree from the University of Birmingham, the M.Eng. degree from Universiti Kebangsaan Malaysia, and the Ph.D. degree from Cardiff University, U.K. He is affiliated with the Department of Computer and Communication Systems Engineering, University Putra Malaysia, where he is currently working as an Associate Professor. He has authored and coauthored several national and international publications. His major research

interest includes network security. He is having an active association with different societies and academies around the world. He made his mark in the scientific community with the contributions and widely recognition from honorable subject experts around the world. He has received several awards for the contributions to the scientific community. He is also working as a reviewer for reputed professional journals.



FAKHRUL ZAMAN ROKHANI (Member, IEEE) is affiliated with the Department of Computer and Communication Systems Engineering, University Putra Malaysia, where he is currently working as 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.

...