# Smart Cities Unleashed: Leveraging IoT and Big Data for Urban Advancement

**Article history**

Received:
25 Oct 2024

Received in revised form:
15 Nov 2024

Accepted:
15 Dec 2024

Published online:
27 Dec 2024

Chiam Jia Ying [1], Suriani Mohd Sam[2], Norliza Mohamed[3], Hafiza Abas[4]

[1,2,3,4] Faculty of Artificial Intelligence,
Universiti Teknologi Malaysia, Malaysia
chiamjiaying@graduate.utm.my, suriani.kl@utm.my, norlizam.kl@utm.my, hafiza.kl@utm.my

*Corresponding author
suriani.kl@utm.my

***Abstract***

*Abstract— The implementation of Internet of Things (IoT) and Big Data Technology has facilitated the development of Smart Cities. IoT collects data through sensors and actuators from different sources, such as smart devices, wearables, infrastructure, vehicle and so on, are tightly coupled with the usage of big data analysis to generate significant insights. This paper first introduces the reason for the emergence of smart cities, following with the concept overview of IoT, big data and smart cities. In addition, the following reviewing different frameworks proposed to implement IoT and big data in various domains.*

***Keywords: — Smart City, Internet of Things, IoT, Big Data, Urbanization***

## 1. Introduction

Cities are often highly populated due to the various opportunities that they offer to the citizens. They are usually being said as the centre of industry, business, and education, which offers a wide range of opportunities in the employment market [1]. In addition, high standard amenities, advanced healthcare, convenient transportation systems and the huge variety of resources that they offer also contributes to a higher quality of life. However, the increasing of density in city populations also leads to the increasing in demand for resources, such as the infrastructure, services, transportation and so on. The United Nations has estimated that there will be around 68% of people living in urban areas by 2050 [2]. While the urbanization provides more opportunities, it also leads to negative impacts such as traffic congestion, air pollution and the increasing needs for utility resources and management of waste and energy [3].

Thus, as urbanization in cities will be a continuous phenomenon, it is one of the utmost priorities to maintain the quality standard in the urban setting with these huge population of citizens by providing adequate provision. In such consequences, with various research and development process being carried out, the emerging of

* *Corresponding author. suriani.kl@utm.my*

smart city would be the solution needed to tackle the challenges faced due to the rapid growth of population in the city.

To achieve full potential of smart cities, there's a need to effectively collect and manage the extensive data obtained, therefore the success of smart cities hinges on the combination of technology solutions, IoT and big data analytics to empower its capabilities by providing reliable methods for data collection, communication, integration, and analysis of real-time and historical data, along with the storage of data [4]. IoT and big data in smart city works simultaneously, as the data used in big data analytics requires the environmental data collected from IoT devices from various domains in the smart city. The combination of IoT and big data improves the decision-making process as it provides statistics and analytic data to make informed decisions.

## 2. Overview of IoT, Big Data and Smart Cities

Internet of Things (IoT) has been widely discussed since the past decade. IoT does not have a standard definition [5], however it is usually defined as a system which allows a massive number of physical devices to be connected to the Internet, which each of the devices are uniquely identifiable and can be interacted through the network [6]. The "things" in IoT can communicate, and does computation among themselves, without the need of human interrupt in the middle of the process. Its underlying technology includes cyber physical systems, ubiquitous computing, internet protocols (IPs) and applications. The computation sometimes involves artificial intelligence and machine learning to create a better algorithm model for data analysis.

By embedding sensors, actuators along with communication capabilities, physical devices which are connected to the internet can communicate with each other intelligently, transmit data to a central back end or cloud platform for analysis, and respond to changes. This promotes the creation of smart environment, and leads to improvements in efficiency, convenience, and productivity.

Big data usually refers to the retrieval, processing and management of a massive data sets, that are collected within a specific timeframe with the usage of software tools. The data sets can be categorised into structured data, which are often numeric and easily to be stored in a relational database; semi-structured data, which does not strictly follow the structure of traditional relational database, but still contain some organization level, for example XML or JSON; and unstructured data, which does not have a consistent structure, for example, videos, images or raw data from IoT sensors.

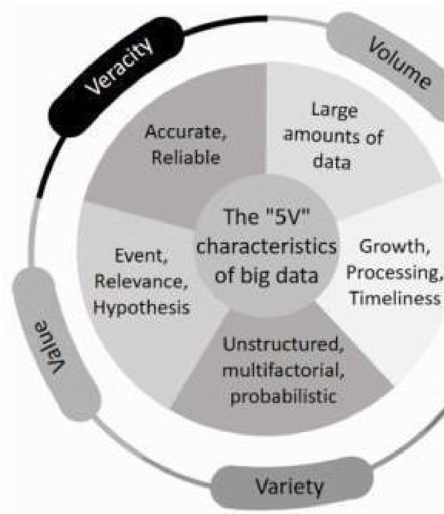


Figure 1. Big data's "5Vs" [5]

The characteristics of big data are categorized into "5Vs" as shown in Figure 1, which are as follow:

i. Volume: Big data consists of a vast number of data, which comes from various sources like sensors, social media and manually uploaded data. Large storage and powerful computation is needed to process the data.
ii. Velocity: The growth rate of data for big data is fast. It involves continuous real-time data streams, such as data from IoT devices. Processing of those data are time-sensitive and requires accurate insights.
iii. Variety: Big data consists of 3 different types of data, structured, semi-structured and unstructured. Data storage would need to be able to manage these diverse data types.
iv. Value: Valuable data would need to be processed from the massive amount of raw data, which there might only be below 10% of value ratio [7].
v. Veracity: Data obtained would need to perform data filtering and validation to prevent involvement of inaccurate or incomplete data, and ensure quality, accuracy and reliability of data.

Apache Hadoop is a one of the powerful open-source technology used in big data, is scalable and suitable to process and store large amounts of data with its cluster of powerful hardware backend. The core services of Hadoop - Hadoop Distributed File System (HDFS) service is offered to store massive data, and MapReduce is used to provide parallel computing services. It provides benefits like high availability, high scalability, fault tolerance and cost-effectiveness, as it efficiently manages large amount data disregard their types by distributing them among all available compute nodes. The application of Hadoop and its services will be further discussed in the literature reviews, where [15]-[18] have all integrated Hadoop services into their implementation. Big data in smart cities contain rich and diverse data resources, where various fields and departments are included. For example, government, healthcare, transportation, and economy data. The variant of data is rich, and it's usually processed real-time and has a fast growing speed.

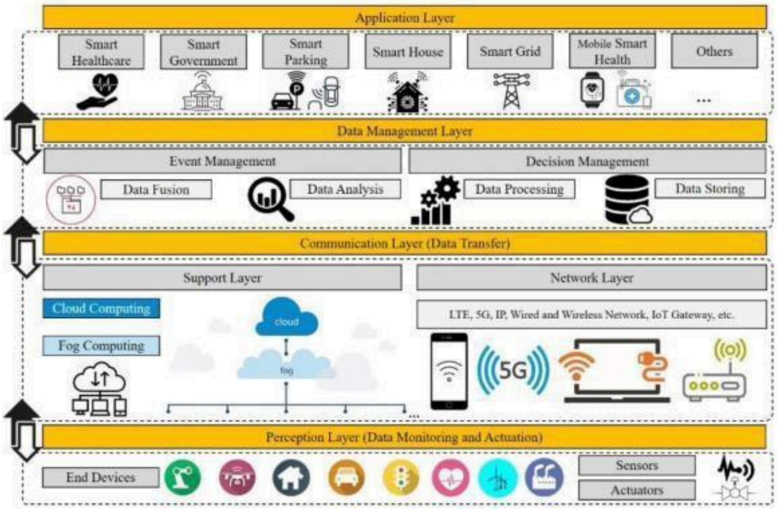

Figure 2. Architecture of Smart City [2]

The concept of smart cities utilizes the information and communication technology (ICT) to promote and enhance the quality of services and life to the citizens, and foster sustainability, energy efficiency, and a green environment. Smart city involves multiple smart systems across different domains, which involves smart homes, smart traffic, smart transportation and so on.

Based on Figure 2, the architecture of a smart city consists of 4 layers, which is based on the architecture of IoT. It includes the perception layer, communication layer, data management layer and the topmost layer is the application layer. The perception layer will play the role of data gathering. This layer is built from a large fleet of sensors from multiple domains to collect the temperature, humidity, air quality, location and much more. The communication layer is where the sensors and smart devices are being connected to the network wired or wirelessly and through different network protocols that covers short-range and wide range of data transmission. This enables the sensors to be able to communicate with other sensors, and to the backend computation in the fog or cloud. The data management layer is where the computation and big data analysis takes place, where raw data are being processed, analysed, stored, and perform decision making. The application layer is where most of the citizens have direct interaction with. The applications are fed by the data processed in data management layer and will be visualised in different applications according to the government departments or the service vendor.

## 3. Problem Definition

The motivation which leads to the need of this review mainly to understand how IoT and Big Data can seamlessly integrate into different smart city applications to provide quality of services and quality of live.

There are a few challenges led by the rapid urbanization and high population of cities, which includes air pollution, traffic congestions, outdated or limited infrastructure and so on. IoT which consists of interconnected sensors and actuators can collect real-time data from diverse aspects of the city, for example like traffic patterns and air quality. However, this causes another issue as the sensors collects raw and unorganized data which might not contain 100% valuable data. Therefore, big data is integrated in smart cities to analyse all the data collected from IoT devices by processing, storing and provides valuable insights.

As the IoT and Big Data is always being discussed that it brings a lot of benefits theoretically, there are needs to have an in-depth review of how the implementation of the IoT and big data framework really works and how are they being integrated in smart city initiatives. Through a comprehensive review of journals and case studies regarding IoT and big data frameworks in multiple domains, this paper aims to:

- Give an overview of the concept of IoT, big data, and smart city
- Explore how existing frameworks implements IoT and big data in smart cities

These provide valuable insights for researchers and urban planners to have a clearer understanding of the involvement of IoT and big data in smart city, which enables them to provide innovative solutions to solve existing and emerging issues in current cities.

## 4. Review : IoT in Smart City

The integration of the Internet of Things (IoT) into urban infrastructures is revolutionizing the concept of smart cities, offering advanced solutions for improving everyday life. This review focuses on three critical areas of IoT applications: smart homes, smart traffic and transportation, and the smart environment. In smart homes, IoT enables enhanced energy efficiency, automation, and security through interconnected devices. In the realm of traffic and transportation, IoT systems optimize traffic flow, reduce congestion, and enhance public transit operations. Additionally, IoT-driven environmental monitoring helps cities address challenges related to pollution, waste management, and resource sustainability. By exploring these key sectors, this review highlights how IoT is driving the future of urban living, fostering efficiency, sustainability, and improved quality of life.

### 4.1 Smart Home

Smart homes are one of the integral components in smart cities, as they promote living comfort, enhance security, reduce energy consumption, and efficient resource management, such as gas, water, and electricity. IoT enabled devices help to monitor the usage of utilities and optimize the resource utilization. IoT can also be used in detecting leaks or preventing short circuits.

A smart home system which utilizes the IoT architecture, which is named iHAS, was proposed by S. M. Shabber, M. Bansal, P. M. Devi and P. Jain [8] as illustrated in Figure 3. This framework enables users to remote monitor the status of their home appliances and interact with them, such as turning on and off the appliances. This can be done by the users through the Blynk app in their smartphones which the app is connect to the Blynk server. The implementation of the system uses an ESP8266 microcontroller, and a relay board to connect the input of appliances to the microcontroller.

Besides, the aging of the population in the city is also another reason why smart homes should be implemented, as elderlies might hurt themselves due to a fall, when there's nobody at home. Therefore, it's important that mart homecare solutions are being integrated as a part of the smart home system.

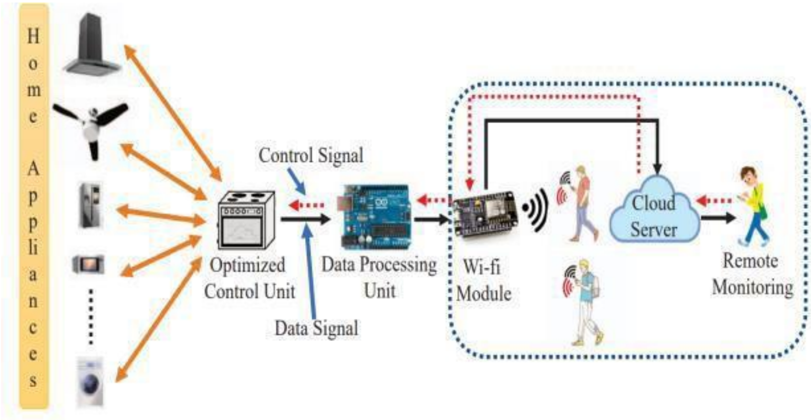


Figure 3. iHas system's proposed flow [8]

To support the comfort and safety of elderly people, I. Zavalyshyn, A. Legay, A. Rath and E. Riviére proposed a smart homecare system, with the smart hub as its central component [9]. The smart hub would collect data from various sensor devices being installed at home or from a wearable device, and then process the data locally in the smart hub without sending it to external parties, which enforces the protection of privacy as shown in Figure 4. If anomalies are detected according to the sensor data or certain criteria met, alerts and notifications will be sent to the respective social circles which are responsible for that item. For example, the family will have full access to the elderly's data, while the doctor can only access aggregated data instead. The internal of the smart hub consists of device drivers, middleware such as Zigbee or BLE, APIs for the application to interact with the sensor data, and applications to show historical or real-time data of the elderly at home.

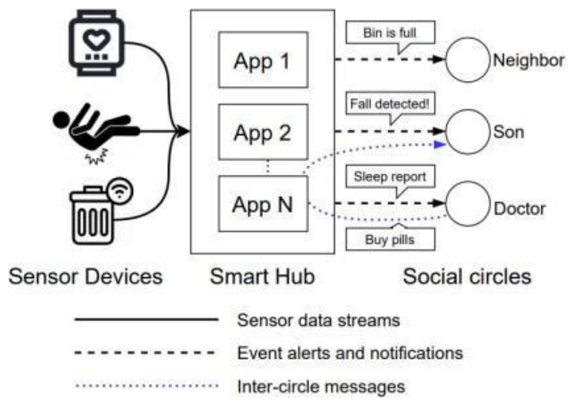


Figure 4. Overall view of the proposed smart homecare system [9]

## 4.2 Smart Traffic and Transportation

IoT are being used in smart traffic and intelligent transportation systems (ITS), which real-time traffic data and physical transportation systems can be obtained and monitored to provide safer and efficient traffic management and transportation systems to the citizens.

In [10], the architecture of smart transportation is being discussed. Communication between smart vehicles usually relies on the vehicle-to-everything network, which the vehicle can communicate with vehicles, pedestrians, road infrastructure and to the network. In Figure 5, the use application scenarios of smart transportation are stated. Global Navigation Satellite System (GNSS) which includes GPS as a subset, can help to localize the vehicle within 1cm-1m accuracy. Sensors such as LiDAR and Radar are used to help in vehicle navigation by detecting the distance between vehicles and obstacles. When there's potential for collisions, real time safety warnings will be sent through V2X communications to avoid collisions and prevent traffic accidents. Autonomous vehicles which require connecting with the V2X network to collect data from various infrastructure about information on the road to make informed decisions during autonomous driving. Infotainment can also make use of V2X communications to deliver high quality in car entertainment system.

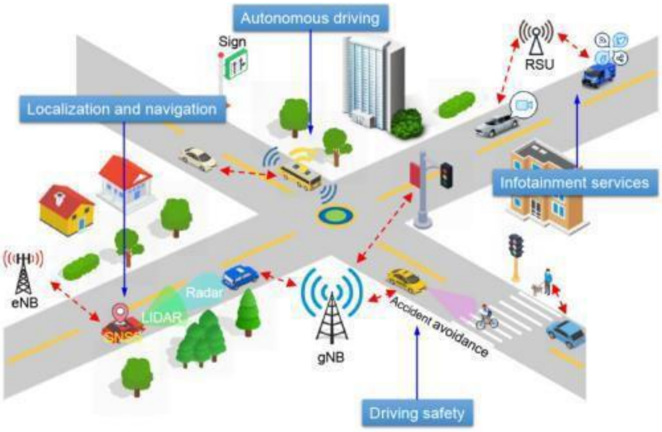


Figure 5. Smart Transportation Use Cases [10]

M. Derawi, Y. Dalveren and F. A. Cheikh had proposed a solution where vehicle-to-infrastructure (V2I) communication is being integrated in smart traffic systems to share data from multiple IQRF sensors into a single system to improve traffic safety [11]. Traffic lights, light poles, and other infrastructures are being connected through radio frequency to the IQRF network, and they will be used to collect data and send to the cloud.

Europe has developed a smart city project - Stardust Project. It implements IoT based smart buses into the smart city project and monitors the data generated by the electric buses [12]. The electric buses communicate with the infrastructure nodes distributed in the city, such as lamp posts or buildings via wireless communications. They communicate through the LoRa/LoRa WAN communication, which provides communication within long ranges while only consume low power, and it's easy to integrate into the cloud enabled infrastructures. They are using a star topology for their network, where the buses will communicate directly to multiple infrastructure nodes in the city. The data collected by the buses, including the location data, the flow of passengers at every stop, temperature and humidity in the bus, bus's battery status and traffic incidents will be recorded and sent to a cloud-based system for monitoring purposes. The bus will then improve their operation by informing the users about the expected arrival time, and to obtain insights of traffic flows for the smart city's traffic optimization. The data are processed using big data tools like Apache Kafka, stored in a MySQL database, and are analysed using Machine Learning models.

## 4.3 Smart Environment

The rapidly increasing urbanization and development in cities, and the increasing of vehicles and expanding industrial areas has caused air pollution issues in cities, which brings harmful impact to both human and environment. The poor quality of air due to the pollution would slowly harm the health conditions of humans and lead to reduction in life span. The environment, such as air, or the agriculture industry would also be affected. Therefore, smart city should address these issues by implementing a smart environment framework to monitor the pollution levels and solve the pollution issues in the city.

Y. P. Reddy, T. Parameswaran and R. Sathiyaraj has then proposed a framework to monitor the air quality level in traffic, industrial and residential areas [13] as illustrated in Figure 6. The framework involves sensors to measure temperature, humidity and multiple pollutants, such as mononitrogen oxides (NOx), carbon monoxide (CO) etc. to compute the Air Quality Index (AQI). Moreover, as the vehicle emission is one of the main contributors to air pollution, therefore this framework also involves intelligent traffic control to avoid traffic congestion and helps to reduce fuel combustion. Sensors including piezo sensors, road tube connectors, microwave detector and proximity sensors are being used to detect the traffic information. The data collected will then be stored and analysed to be visualised in graphical user interface, and to be used as input for the intelligent traffic control system. It helps to re-route the traffic which minimizes the traffic congestion, which leads to reducing the pollutants produced by vehicle emissions.

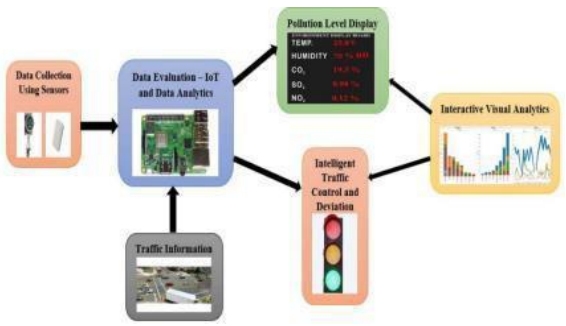


Figure 6. Proposed Framework for Smart Environment Monitoring [13]

Another example of usage of IoT in smart environment monitoring which is related to monitoring of water contamination has been given in [14]. Cloud-connected smart aqua sensors are being placed in water sources from different location, and the data collected are being transmitted to the cloud. The system can then analyse and monitor if the water is clean or being contaminated by using the insights and visualization from the processed data.

## 5. Review: Big Data in Smart City

The increasing complexity of urban environments calls for innovative solutions, and Big Data is emerging as a key enabler in the development of smart cities. This review delves into the diverse applications of Big Data, focusing on four critical areas: its role in smart homes, the Public Energy Living Lab (PELL) Smart City Platform, predicting crowdedness levels in Mass Rapid Transit (MRT) systems, and geospatial Big Data architecture. In smart homes, Big Data enhances energy management, security, and user personalization. The PELL platform leverages data for optimizing public energy consumption and sustainability efforts. Predicting crowdedness levels in MRT systems, facilitated by Big Data frameworks, improves public transit efficiency and user experience. Lastly, geospatial Big Data architecture provides real-time insights for urban planning and resource allocation. This review explores how Big Data is transforming smart city infrastructures, driving informed decision-making and improving urban life.

### 5.1 Big Data in Smart Home

In recent years, it is noticeable that more and more homeowners transform their traditional homes into smart homes, due to the awareness of the convenience and comfort that smart home solution brings to life. However, there are vast amounts of data being produced by the IoT devices in a smart home, which requires processing to provide valuable insights. Therefore, big data solutions are integrated into smart homes to collect data and provide analysis for different aspects, such as energy consumption, behavioural patterns and much more.

In [15], it discusses the overview of integrating big data service into a smart home. The system will first collect data from MQTT brokers in the acquisition layer. Next, Apache Kafka is used to store the streamed raw data temporarily in the collection layer. Data is then passed to the storage layer, where data is stored persistently in Apache Cassandra to ensure that no data loss occurs. The analytic layer will then utilize Apache Spark to aggregate and transform the data to different types of data output according to the use case in the application. Data can be processed in real-time, periodically, and on demand. Lastly, the serving layer provides data warehouse services, where smart home applications can make use of the data, and data API, where it serves as a gateway for third-party applications to communicate and access the data being generated.

### 5.2 Public Energy Living Lab (PELL) Smart City Platform

A Big Data Processing architecture for PELL Smart City Platform was introduced by M. Ali and F. Moretti, to develop a big data driven processing algorithm to collect, process and analyse data, which helps in the understanding of the behaviour of the lighting plant and assesses the performance of public street lightning [16]. As shown in Figure 7, the architecture of the proposed solution is split into 3 layers, firstly the data sources and ingestion, secondly the data integration and processing, and lastly presentation. 3rd party apps which use a general middleware, PELL portal with manually uploaded information which sends data through SQL protocols, PELL Gateway where machine to machine communicate through RESTful APIs, and broker which uses the MQTT protocol to stream data are the data sources of the proposed big data processing architecture.

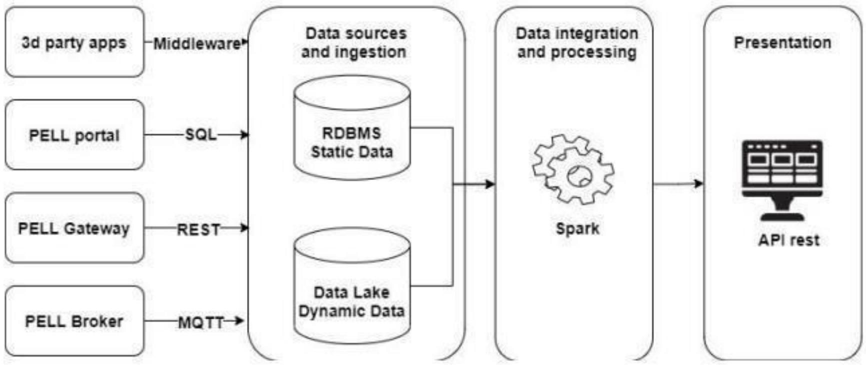


Figure 7. Proposed architecture for Big Data Processing in PELL Smart City Platform [16]

The data collected are then stored into two different database systems, static data are utilizing the relational database, while dynamic data is stored in unstructured data lake repository. The data integration phase will then merge these stored data from both domains, presented through XML format or through MQTT broker in JSON format, to calculate the Key Performance Indicators (KPI) of the public street lightning’s electric energy consumption. The data will then be passed on to perform data integration, where Apache Spark library is used to aggregate data into a new output data and transport to the final presentation layer in Apache Parquet format. Lastly, the presentation layer monitors and does analysis of the KPI results using the data, and JavaScript libraries are being used to visualise the data in the web based front-end.

## 5.3 Predicting Crowdedness Level of MRT Using Big Data Framework

As public transport is key role in the development of smart cities, therefore it is important that the public transportation provided is reliable and cost-effective. Big data comes in place as it provides insights about the crowd density that utilise public transport, which is crucial for urban planning, and to avoid overcrowded of transport. F. Liu, S.P.R. Asaithambi and R. Venkatraman had proposed a big data framework, which is based on Hadoop, as seen in Figure 8 to predict and analyse the level of crowdedness of the MRT station platforms in Singapore [17].

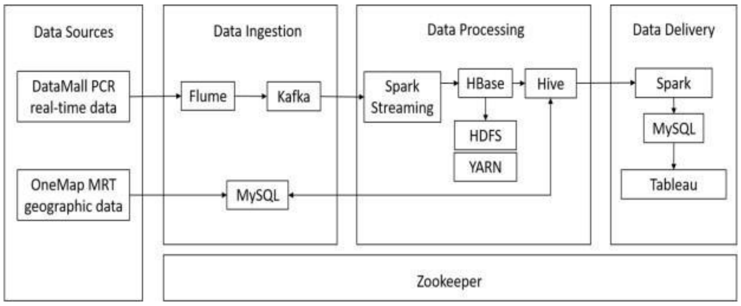


Figure 8. Conceptual Architecture of the proposed Big Data Framework [17]

The data used in this framework are obtained from 2 sources, which is the MRT platform crowd density, which is a dynamic data, and the geographical data, which is a static data. The architecture of the framework includes 4 stages of big data, which includes data ingestion, where Apache Flume is used for real-time crowd density data, and MySQL is used for static geographical data. Spark Streaming is then being used to read the data being streamed, then process it through Kafka. Both the raw and streaming data are then being stored in HBase and HDFS to perform data processing by using Hive and SparkSQL to convert HBase data to MySQL. Lastly, machine learning model is being trained by using Apache Spark ML and the output are stored in MySQL database for data visualization.

The framework uses Recursive Feature Elimination (RFE) to identify the number of data need to be remained for the classification model, which are the stations with at least one occurrence of medium or high occupancy. Besides, an approach of SMOTE with Balanced Random Forest Classifier has been used to solve the imbalanced distribution of data to obtain a more accurate data.

## 5.4 Geospatial Big Data Architecture

The rapid expansion of cities has led to numerous environmental challenges, including increased pollution, the intensification of the greenhouse effect, and disruptions to ecological balance. As urban populations grow, these issues become more pronounced, affecting public health and long-term sustainability. To address these challenges, smart cities have emerged as a critical solution, utilizing data-driven technologies like IoT and Big Data to monitor and manage environmental conditions. By integrating smart infrastructures, cities can reduce pollution, optimize energy use, and promote eco-friendly practices such as renewable energy and efficient waste management.

A key component of smart cities is geospatial big data architecture, which provides valuable insights into the spatial relationships between urban elements and the environment. By analyzing geospatial data, cities can make informed decisions on land use, resource allocation, and environmental protection. This technology enables more accurate environmental models, helping to mitigate risks associated with pollution and climate change. By leveraging such advanced tools, smart cities can optimize transportation systems, enhance green spaces, and ensure a more sustainable and ecologically balanced urban environment for future generations.

A general big data architecture has been proposed by Y. Gacha, M.A.B Rhaiem and T. Abdellatif for a geospatial system for a green smart city [18] as shown in

Figure 9. Two sources of data are being collected, which are streaming data from IoT devices and social medias, and bound data, which are static data from existing geospatial databases. Data collection is then done by using Apache Kafka and Apache Flume to collect the large amount of data. Preprocessing of the data collected, and data labelling is done during this stage. Next, the data can be stored in non-relational database, such as MongoDB and Apache HBase and so on, according to types of data. The spatial property of data stored can then be managed through spatial indexing and spatial querying. Next, the data are being pass on to the data processing stage, where analysis are being carried out through either stream processing for real-time data, or batch processing for existing static data. Lastly, these analysed data will then be provided to different service to represent the data through visual interpretation.

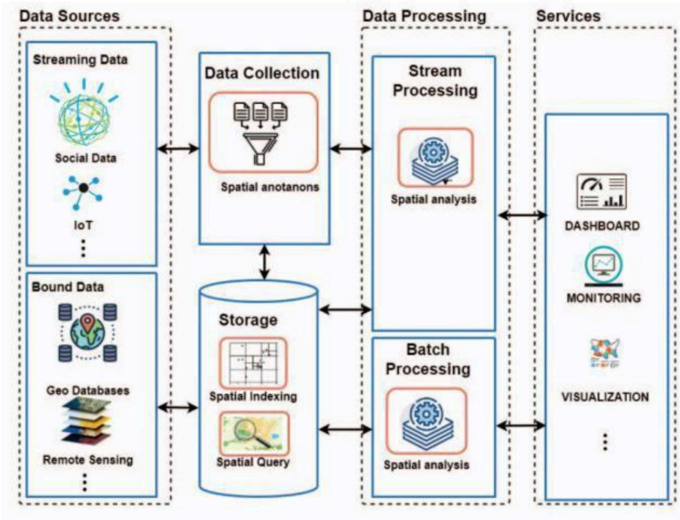


Figure 9. Big data based Geospatial System Architecture [18]

## 6. Solution Approaches and Method

The method being used to identify the role of IoT and big data in smart city is based on Systematic Literature Review approach. Various relevant studies that are published are gathered and analysed in this paper to provide a clearer picture about what are the role of IoT and Big Data in Smart Cities. They are gathered from multiple sources, mainly from IEEE, and some from Science Direct, Springer, and Google Scholar.

First level searching was done through both IEEE Xplore and Google Scholar. Keywords related to the topic, for example, smart city, IoT, Big Data, and their combination, such as IoT in Smart City, Big Data in Smart City and so on as shown in Figure 10. The search results are being filtered to only include the recent 5 years of papers, from 2020 to 2024. Possible candidates are then being chosen according to the title of the topic, which has a higher match with the title of this review paper's topic.

Next, the second layer of filtering is done via screening through the abstract in the possible candidate papers. Then, the abstract that shows high relevancy to this topic are then being selected for review.

In the third layer, a read through of the contents of the candidates such as its results stated, graphs and diagrams are done. Finally, 20 papers are being reviewed and cited in this paper. The final paper selections are reviewed comprehensively and chosen based on these criteria:

i. Research explains about the definition or concept of IoT, big data or smart city in general.
ii. Research incorporates the details of what are the roles of IoT or big data in smart cities.
iii. Research contains framework of IoT or big data in smart cities.

## 7. Results

Smart Home, Smart Traffic and Transportation and Smart Environment are being chose to review the implementation of IoT and big data in smart cities. These applications are being chose as they are more common to the citizens, as they usually involve in all these applications in their daily lives. By reviewing how IoT and big data can bring benefits to these applications, it helps the urban planners to make informed decisions accordingly.

The literature reviewed which states different IoT frameworks implementation in different applications proposed few concepts that can be referred in future implementation. According to the data summarised in Table I, V2X and V2I communication are highly implemented in smart traffic and smart transportation systems, instead of just relying on the sensors of the individual vehicle. By implementing V2X communication, it extends the perception of the vehicle beyond its sensor range, as the vehicle can communicate and exchange data directly with other vehicles, pedestrians, or road infrastructure as desired [19]. Moreover, the road infrastructures, especially those located in high traffic flow areas are suitable locations for sensor placement, to track traffic flows and perform V2I/V2X communications. It can also be concluded that in terms of smart home, wireless communication with low power consumption and easy for deployment, such as Zigbee and BLE is suitable for smart home context, where there’s no need for long range data transmission. Whereas for long range data transmission, LoRa WAN or IQRF can be implemented as they are able to achieve long-range communication, which are suitable for communication between infrastructures. The sensors chosen for the IoT implementations are usually based on the purpose of the implementation, so that correct data that are related to the proposed system are collected.

Table 1. Implementation of IoT in Smart Cities

| Application | Existing Use Cases | Purpose | Sensors Used | Communication / Network Technology Used |
|---|---|---|---|---|
| Smart Home | Home appliances monitoring system [8] | To monitor the status of home appliances remotely | Home appliances | Not mentioned |
| | Smart homecare system [9] | To support the comfort and safety of elderly people at home | Not mentioned, multiple sensors used | Zigbee, BLE |
| Smart Traffic and Transportation | Smart vehicles [10] | Localization and navigation, accident avoidance, autonomous driving, better infotainment services | GNSS, LiDAR, Radar | V2X |
| | Smart traffic systems [11] | Improve safety on the road | IQRF sensors | V2I |
| | Smart buses [12] | Provide better user experience to bus passengers, traffic optimization in smart city | Not mentioned | V2I, LoRa/LoRaWAN |
| | Intelligent traffic control system [13] | Minimize the traffic congestion, which leads to reducing the pollutants produced by vehicle emissions | Piezo sensors, road tube connectors, microwave detectors, proximity sensors | Not mentioned |
| Smart Environment | Smart air pollution detection system [13] | Monitor air pollution level | Nitrogen Oxide Sensor, Sulphur Dioxide Sensor, Carbon Monoxide sensor, Electrochemical Toxic Gas sensor, temperature sensor, humidity sensor | Not mentioned |
| | Smart water pollution detection system [14] | Determine contaminated water source | Smart aqua sensors | Not mentioned, connected to cloud |

Big data is the key instrument for city management purposes, which helps to enhance efficiency, effectiveness, transparency and accountability, and provides more effective communication between government bodies and citizens [20]. In Table 2, it can be concluded that Apache Hadoop gains trust from multiple users, as it is open source, and it provides scalability, reliability, and distributed computation power for massive data sets.

Table 2. Implementation of Big Data in Smart Cities

| Application | Existing Use Cases | Purpose | Big Data Technology Used |
|---|---|---|---|
| Smart Home | Smart Home IoT Network Diagnostics [15] | Process large amount of data generated in smart home | Apache Kafka, Apache Spark, Apache Cassandra |
| Smart Traffic | PELL Smart City Platform [16] | Understanding of the behavior of the lighting plant, analyze the performance of public street lightning in smart city | Apache Spark, Apache Parquet, SQL Protocol broker |
| Smart Transportation | Crowd density prediction in MRT [17] | Obtain MRT platform crowd density for better public transport planning and utilization | Apache Flume, MySQL, Spark Streaming, Apache Kafka, Apache HBase, Apache HDFS, Hive, SparkSQL, Apache Spark ML |
| Smart Environment | Geospatial System [18] | To obtain insights to form a green city | Apache Kafka, Apache Flume, MongoDB, Apache HBase |

## 8. Conclusion

In summary, this review paper has provided a comprehensive introduction to the concepts of the Internet of Things (IoT), Big Data, and their roles in the development of smart cities. As cities continue to face challenges stemming from rapid urbanization, the implementation of smart city initiatives has emerged as a promising solution. These initiatives rely on the seamless integration of IoT and Big Data to address urban issues effectively. This review has explored the application of IoT and Big Data in various smart city domains, including smart homes, smart traffic management, smart transportation systems, and smart environmental monitoring. The discussion covers the different technologies and frameworks used to implement IoT and Big Data across these applications, demonstrating how these tools work together to enhance the functionality and sustainability of urban environments. The vast amount of data generated from these smart city applications offers valuable insights that contribute to the continuous improvement and development of cities. By analyzing this data, urban planners and policymakers can make more informed decisions, optimizing resource allocation, improving energy efficiency, reducing pollution, enhancing public safety, and ensuring a higher quality of life for residents. The integration of IoT and Big Data is not just a technical advancement but a critical driver for the future of sustainable urban living, paving the way for smarter, more resilient cities.

## Acknowledgments

This research was supported by Faculty of Artificial Intelligence, Universiti Teknologi Malaysia, Malaysia.

## References

1. I. H. Khan, Mohd. I. Khan, and S. Khan, “Challenges of IoT Implementation in Smart City Development,” *Lecture Notes in Civil Engineering*, pp. 475–486, 2020, doi: 10.1007/978-981-15-2545-2_40.

2. M. Talebkhah, A. Sali, M. Marjani, M. Gordan, S. J. Hashim and F. Z. Rokhani, "IoT and Big Data Applications in Smart Cities: Recent Advances, Challenges, and Critical Issues," *IEEE Access*, vol. 9, pp. 55465-55484, 2021, doi: 10.1109/ACCESS.2021.3070905.

3. C. Inibhunu and A. Carolyn McGregor, "A Privacy Preserving Framework for Smart Cities utilising IoT, Smart Buildings and Big Data," *2020 IEEE 22nd International Conference on High Performance Computing and Communications; IEEE 18th International Conference on Smart City; IEEE 6th Interna tional Conference on Data Science and Systems (HPCC/SmartCity/DSS)* , Yanuca Island, Cuvu, Fiji, 2020, pp. 1096 -1103, doi: 10.1109/HPCC-SmartCity-DSS50907.2020.00197.

4. M. A. Bakar, S. M. Sam, A. Azizan, Y. M. Yusof, N. Mohamed and N. Shafie, "The Role of Io T and Big Data in Shaping Smart Cities: A Review," *2023 IEEE Symposium on Wireless Technology & Applications (ISWTA)* , Kuala Lumpur, Malaysia, 2023, pp. 78-83, doi: 10.1109/ISWTA58588.2023.10249252.

5. R. Yadav and V. Kumar, "A Literature Review of Challenges in Internet of Things," *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)* , Greater Noida, India, 2021, pp. 729-733, doi: 10.1109/ICAC3N53548.2021.9725470.

6. M. N. Bhuiyan, M. M. Rahman, M. M. Billah and D. Saha, "Internet of Things (IoT): A Review of Its Enabling Technologies in Healthcare Applications, Standards Protocols, Security, and Market Opportunities," in *IEEE Internet of Things Journal*, vol. 8, no. 13, pp. 10474-10498, 1 July1, 2021, doi: 10.1109/JIOT.2021.3062630.

7. H. Yahui, "Research on Intelligent Mode of Public Service Driven by Big Data," *2021 International Conference on Public Management and Intelligent Society (PMIS*) Shanghai, China, 2021, pp. 12-16, doi: 10.1109/PMIS52742.2021.00010.

8. S. M. Shabber, M. Bansal, P. M. Devi and P. Jain, "iHAS: An Intelligent Home Automation Based System for Smart City," 2021 *IEEE International Symposium on Smart Electronic Systems (iSES* ), Jaipur, India, 2021, pp. 48-52, doi: 10.1109/iSES52644.2021.00023.

9. I. Zavalyshyn, A. Legay, A. Rath and E. Riviére, "Smart Home Care: Towards Supporting Elderlies in the Comfort and Safety of their (Smart) Homes," 2021 *10th Latin-American Symposium on Dependable Computing (LADC),* Florianópolis, Brazil, 2021, pp. 1-6, doi: 10.1109/LADC53747.2021.9672565.

10. Y. Wu, H. -N. Dai, H. Wang, Z. Xiong and S. Guo, "A Survey of Intelligent Network Slicing Management for Industrial IoT: Integrated Approaches for Smart Transportation, Smart Energy, and Smart Factory *," IEEE Communications Surveys & Tutorials, vol. 24, no. 2, pp. 1175 -1211, Secondquarter 2022*, doi: 10.1109/COMST.2022.3158270.

11. M. Derawi, Y. Dalveren and F. A. Cheikh, "Internet-of-Things-Based Smart Transportation Systems for Safer Roads," *2020 IEEE 6th World Forum on Internet of Things (WF-IoT),* New Orleans, LA, USA, 2020, pp. 1-4, doi: 10.1109/WF-IoT48130.2020.9221208.

12. J. J. Astrain, F. Falcone, A. J. Lopez-Martin, P. Sanchis, J. Villadangos and I. R. Matias, "Monitoring of Electric Buses Within an Urban Smart City Environment," in *IEEE Sensors Journal,* vol. 22, no. 12, pp. 11364- 11372, 15 June15, 2022, doi: 10.1109/JSEN.2021.3077846.

13. Y. P. Reddy, T. Parameswaran and R. Sathiyaraj, "A Smart Environment Monitoring Framework Using Big Data and IoT," *2021 IEEE Mysore Sub Section International Conference (MysuruCon),* Hassan, India, 2021, pp. 399-404, doi: 10.1109/MysuruCon52639.2021.9641609.

14. S. L. Ullo and G. R. Sinha, “Advances in smart environment monitoring systems using IoT and sensors,” Sensors, vol. 20, *no. 11, p. 3113, May 2020, doi: 10.3390/s20113113.*

15. N. Vajagic and M. Antic, "Smart Home IoT Network Diagnostics using Big Data Services," *2022 30th Telecommunications Forum (TELFOR),* Belgrade, Serbia, 2022, pp. 1-4, doi: 10.1109/ TELFOR56187. 2022.9983792.

16. M. Ali, P. Scandurra, F. Moretti, L. Blaso, M. Leccisi and F. Leccese, "From Big Data to Smart Data -centric Software Architectures for City Analytics: the case of the PELL Smart City Platform," *2021 IEEE International Conference on Smart Data Services (SMDS),* Chicago, IL, USA, 2021, pp. 95-104, doi: 10.1109/SMDS53860.2021.00023.

17. F. Liu, S. P. R. Asaithambi and R. Venkatraman, "Predicting Crowdedness Level of the Mass Rapid Trans it (MRT) Platform Using Big Data Framework: A Case Study in Singapore," *2023 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM),* Singapore, Singapore, 2023, pp. 1007-1011, doi: 10.1109/IEEM58616.2023.10406356.

18. Y. Gacha, M. A. B. Rhaiem and T. Abdellatif, "Geospatial Big Data for a Sustainable and Green Smart City," *2023 International Conference on Cyberworlds (CW),* Sousse, Tunisia, 2023, pp. 217-224, doi: 10.1109/CW58918.2023.00039.

19. N. Ambrosy, J. Rainer, M. Niebisch, R. German and L. Underberg, "Self-Navigating Automotive Production: Using 5G for Automated Driving in Car Assembly," *2022 IEEE 18th International Conference on Factory Communication Systems (WFCS),* Pavia, Italy, 2022, pp. 1-4, doi: 10.1109/WFCS53837.2022.9779168.

20. M. N. I. Sarker, M. N. Khatun, G. M. Alam and M. S. Islam, "Big Data Driven Smart City: Way to Smart City Governance," *2020 International Conference on Computing and Information Technology (ICCIT-1441),* Tabuk, Saudi Arabia, 2020, pp. 1-8, doi: 10.1109/ICCIT-144147971.2020.9213795.