The Journal of Supercomputing
https://doi.org/10.1007/s11227-022-04594-1

A systematic survey of data mining and big data analysis in internet of things

Yong Zhong[1] · Liang Chen[2] · Changlin Dan[1] · Amin Rezaeipanah[3]

Accepted: 8 May 2022

Abstract

The Internet of Things (IoT) is an emerging paradigm that offers remarkable opportunities for data mining and analysis. IoT envisions a world where all smartphones, vehicles, public services facilities, and home appliances that can be connected to the internet act as data sources. Even today, a significant portion of electronic devices, including watches, emergency alarms, parking doors, and many appliances can be linked to IoT systems and remotely controlled. Big data analysis and data mining methods can be utilized to improve the performance of IoT systems and address their challenges in the area of data storage, processing, and analysis. Extensive studies on IoT with big data can make it possible to accumulate tremendous data and transform it into valuable knowledge using data mining techniques. With this background, this paper provides a systematic survey of the literature on the use of big data analytics and data mining methods in IoT. This review aims to identify the lines of research that should receive more attention in future works. To achieve this goal, the articles published between 2010 and 2021 on the subjects of IoT-based big data and IoT-based data mining (60 articles) have been reviewed. These articles fall into four general categories in terms of focus: architecture/platform, framework, applications, and security. The paper provides a summary of the methods used in IoT-based big data analysis and IoT-based data mining in these four categories to highlight the promising avenues of research for future works.

✉ Liang Chen
tangtongxun_0001@126.com

Yong Zhong
zhongyong23tom@163.com

Changlin Dan
danchanglin2021@126.com

Amin Rezaeipanah
rezaeipanah@pgu.ac.ir

[1] Chengdu Jincheng College, Chengdu, Sichuan 611731, China

[2] Business School of Chengdu University of Technology, Chengdu, Sichuan 610059, China

[3] Department of Computer Engineering, Persian Gulf University, Bushehr, Iran

Published online: 06 June 2022

Keywords Internet of things · IoT · Big data · Data mining · Systematic survey

1 Introduction

The ascendance of the Internet and computers has marked the beginning of a new era, where a progressively increasing number of people engage in information exchange over the Internet using personal computers, laptops, tablets, cell phones, and other data transmission and reception gadgets [1]. In recent years, this trend has stimulated the growth of a technology known as the Internet of Things (IoT) [1, 2], which is based on the idea that any object on earth can be identified, controlled, and monitored via the Internet [3]. IoT was first introduced by Kevin Ashton in 1999 when he observed the breadth of communication and information exchange between devices. IoT-based systems can detect objects that are connected to the Internet and provide a platform for a variety of communications and data sharing using information technology. Figure 1 illustrates the concept of IoT, which is envisioned as having a sensor on anything that has an Internet connection.

For successful implementation, the IoT technology requires standardization to ensure interoperability, compatibility, reliability, and effectiveness on a global scale [4, 5]. It should also be remembered that the growth of cloud computing and IoT architecture is rapidly increasing the volume of data that needs to be stored or processed. Every day, a huge number of sensors continuously collect and transmit environmental, geographic, astronomical, logistical, and other types of data for storage and processing in the cloud. For IoT, the main sources of information are mobile phones, public services, and home appliances. It is expected that eventually, the size of data will exceed the capacity of the existing IT architectures and infrastructures. This will also have significant implications in terms of computing capacity because many applications of this technology require real-time data processing and analysis [6, 7]. The question of how to manage this increasingly large amount of data can also become a major social challenge. Figure 2 shows the forecasted change in the size of data over the coming years [5].

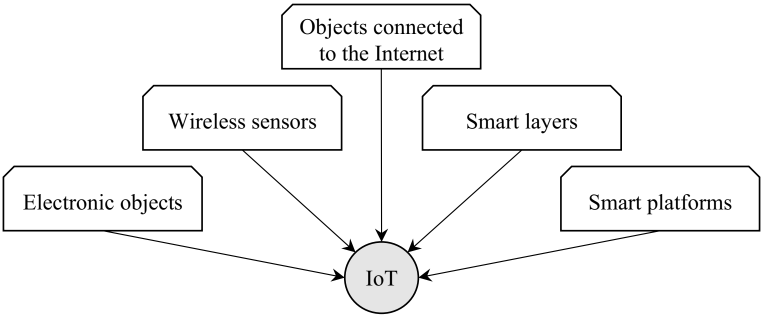


Fig. 1 Concept of the Internet of Things (IoT)

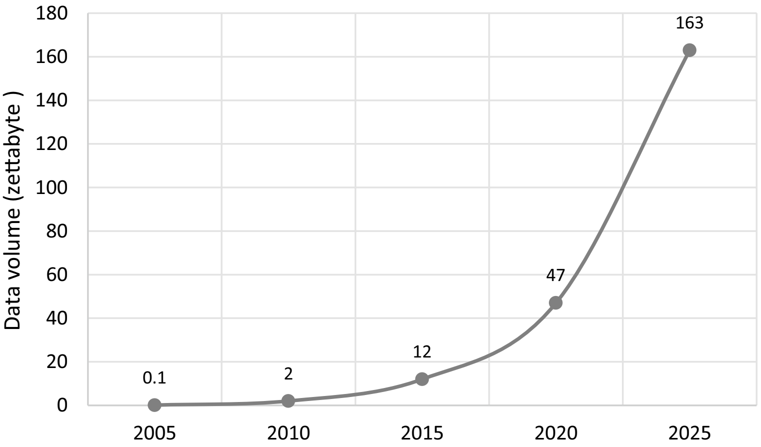


Fig. 2 Forecasted increase in the size of data over the coming years (one zettabyte equals one trillion gigabytes)

Considering the features of the data generated from IoT sources, this data fall in the category of "Big Data". However, big data generated by IoT sources have unique characteristics which can be attributed to the variety of these sources. The most common characteristics of these data are heterogeneity, diversity, unstructured form, noise, and high redundancy [6, 8]. The number of active sensors worldwide is projected to reach one trillion by 2030, and a significant portion of these sensors will be producing big data [7]. In the absence of a solution for handling this increasing volume of data, the existing systems and infrastructures will gradually run into a problem and may even stop working in some areas. The evolution and diversity of what is called big data are shown in Fig. 3 [9].

Big data can be analyzed by a variety of methods including classification, clustering, association rule mining, and regression [6], which are illustrated in Fig. 4. There are many algorithms for data modeling and analysis through each of these data mining techniques. For example, the Bayesian network, support vector machine (SVM), and k-nearest neighbor (KNN) take the classification approach [10, 11]. In any case, various artificial intelligence and data mining techniques have been applied to wireless sensors for performance improvement or to achieve specific goals. These techniques serve as a means of introducing an intelligent learning model to the IoT technology or in other words Wireless Sensor Networks (WSNs) [12].

The big data generated by many IoT devices is sensitive and confidential. In general, the IoT environment for big data publishing is heterogeneous and complex [13, 14]. Therefore, data analysis classical algorithms are not valid for security at all levels. For this reason, providing security and trust for them is inevitable. The cloud and the IoT can store large amounts of data temporarily to reduce the complexity of the analysis process. However, some security issues seriously affect data privacy. There are several proposed challenges and solutions in the literature [15–17]. Approaches such as instantaneous IoT data analysis, homogeneous encryption, and differential privacy are commonly used to address such challenging issues. Therefore, solving

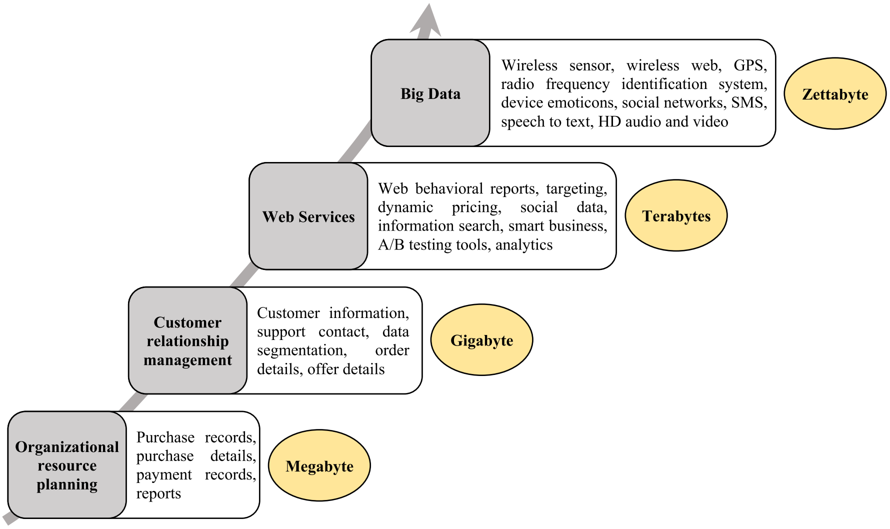

Big Data
Wireless sensor, wireless web, GPS, radio frequency identification system, device emoticons, social networks, SMS, speech to text, HD audio and video
Zettabyte
Web Services
Web behavioral reports, targeting, dynamic pricing, social data, information search, smart business, A/B testing tools, analytics
Terabytes
Customer relationship management
Customer information, support contact, data segmentation, order details, offer details
Gigabyte
Organizational resource planning
Purchase records, purchase details, payment records, reports
Megabyte

Fig. 3 Evolution and diversity of big data

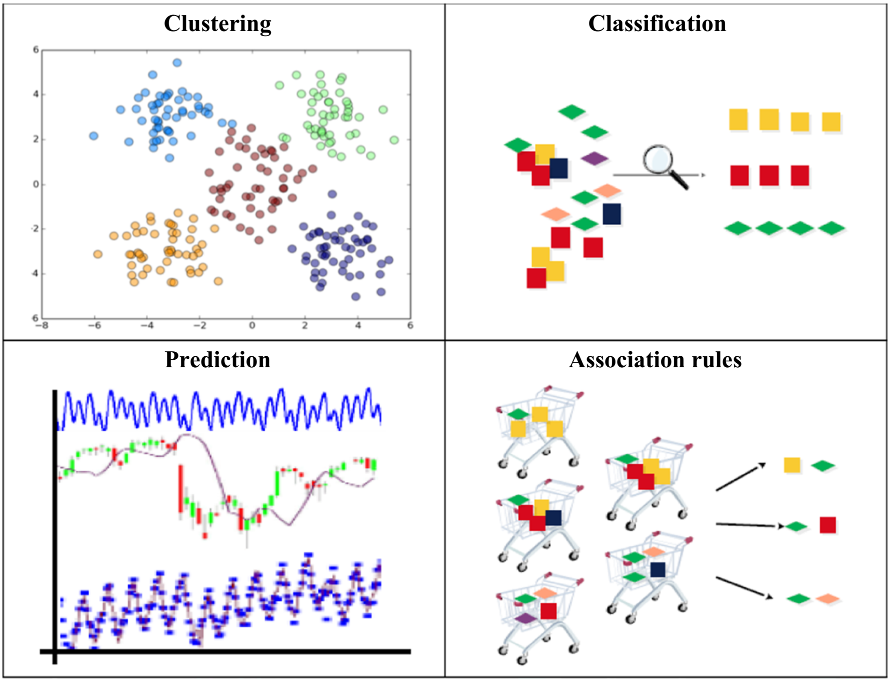


Fig. 4 Schematic diagram of big data analysis methods

IoT security issues with the advent of new technologies or the integration of existing technologies is essential. Machine Learning (ML) has shown promising results in previous studies to detect security breaches [15, 18]. In addition, IoT devices generate a large amount and variety of data. Therefore, with the use of big data technologies, higher performance and better data management can be achieved.

In this paper, the articles in the field of IoT and those in the field of data mining and big data that concern this technology are systematically reviewed. Initially, articles are divided into four categories based on their focus: architecture/platform, framework, applications, and security. Then, the articles of each category are separately reviewed and analyzed. Finally, the paper summarizes the challenges of the field and potential avenues for progress according to the reviewed articles. In this survey, the major contributions of our work can be highlighted as follows:

Systematic and detailed review of data mining and big data analysis used in IoT applications.

Providing an overview of IoT big data in order to understand the importance of IoT data mining.

Analysis of IoT-based big data and IoT-based data mining in four categories: architecture/platform, framework, applications, and security.

The rest of this paper is organized as follows. Section 2 describes the concept of Data Mining and Big Data in IoT, Sect. 3 explains the process of searching for relevant articles, Sect. 4 provides and discusses the categories of the reviewed articles, Sect. 5 presents the results of the review, and Sect. 6 concludes the paper.

2 Data mining and big data in IoT

Data mining is the discovery of a "model" of data [19]. With recent advances in communication technology, people and objects are becoming increasingly interconnected. The availability of the Internet allows the connection of different devices that can communicate with each other and share data. IoT is a new concept that allows users to connect various sensors and smart devices to collect real-time data from the environment. Big Data is the vast amount of data that is collected from the Internet of Things and is used for information that is not processed or analyzed using traditional tools [20, 21]. Every organization faces more challenges in accessing a wealth of information and how to obtain value from a wide variety of data. Traditional software extraction algorithms are only applicable to small-scale IoT data. Also, many IoT applications deal with analyzing data from different devices and correlating them to predict possible machine failures at production sites or emergencies in smart buildings in a home security application. Data mining techniques must manage the heterogeneity of IoT data, the large amount of data, and the speed at which they are available. Next year there is an estimated 20–25 billion IoT devices coming online [22]. All these devices will produce raw data which we aren't able to process into intelligence. This data can be turned into knowledge by tools such as big data analysis and data mining, as shown in Fig. 5.

The relationship between data mining, big data, and IoT is a synergistic interdependence that allows companies to access insights, analytics, and performance

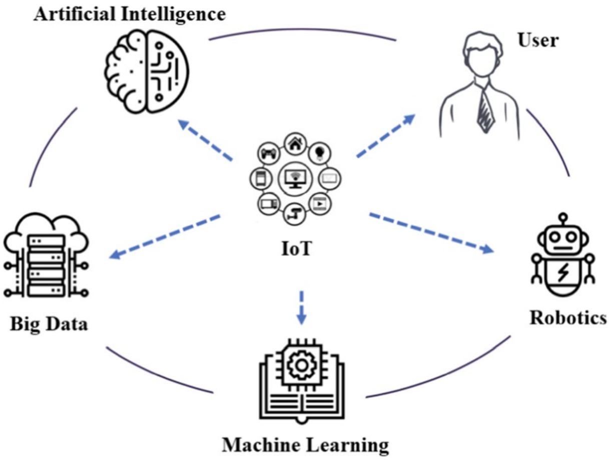


Fig. 5 IoT data analysis by big data and data mining

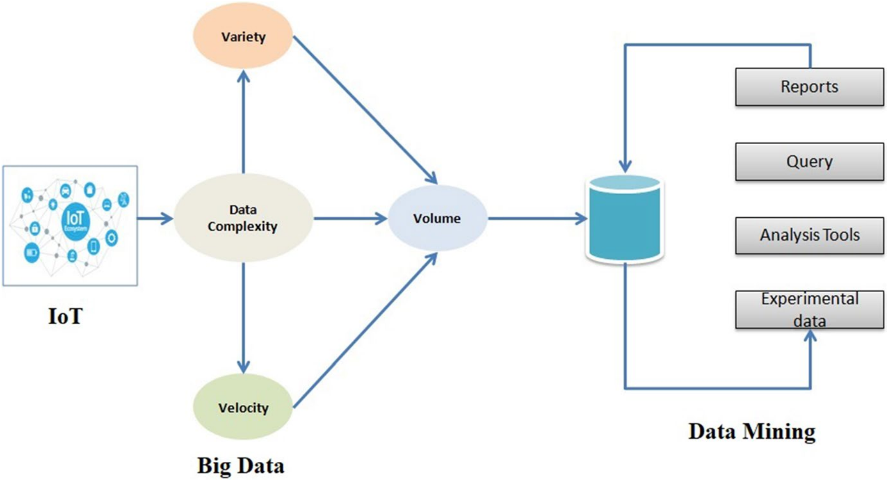


Fig. 6 Relationship between data mining, big data and IoT

reports. The main advantage of IoT with data mining and big data is that it is a scalable, reliable and agile solution for businesses. The integration of data mining, IoT, and big data analytics is becoming a technology hub that supports a variety of applications, including better customer experience, accurate forecasting, and better supply chain management [23]. Figure 6 provides an overview of the relationship between data mining, big data and IoT.

Data mining is important from the IoT and big data, and technology trends that affect large companies around the world. IoT focuses on developers who develop software platforms and applications that enable organizations to manage their IoT devices and the data they generate. Big Data analysis captures unstructured data, traffic patterns, home performance, and information collected by IoT devices and organizes them into digestible datasets [24, 25].

3 Research methodology

In this paper, previous articles on IoT and its combination with data mining and big data are systematically reviewed. This review can help researchers come up with new ideas for expanding the field of IoT. The search for articles was performed in Scholar, ACM, and Scopus databases. The search was configured to find the combinations of the terms “Internet of Things” or “IoT” with the terms “big data” and “data mining” (e.g., IoT + big data) in the title. The search was limited to the articles written in English and published in journals and conference proceedings between 2010 and 2021. Review articles were also included in this literature review. After reviewing the titles of the found articles, only those who had used the terms review, challenges, state of art, survey, Systematic Literature Review (SLR), Systematic Mapping Study (SMS) and literature were tagged as review articles.

The methodology used in this paper is a systematic survey based on [25, 26]. This means that we select the article based on the framework defined by them. This includes the process of searching, selecting articles, and ignoring invalid articles. Figure 7 shows the process of this methodology. This methodology can find articles related to "IoT and big data" or "IoT and data mining" for review. The proposed systematic survey steps are described below.

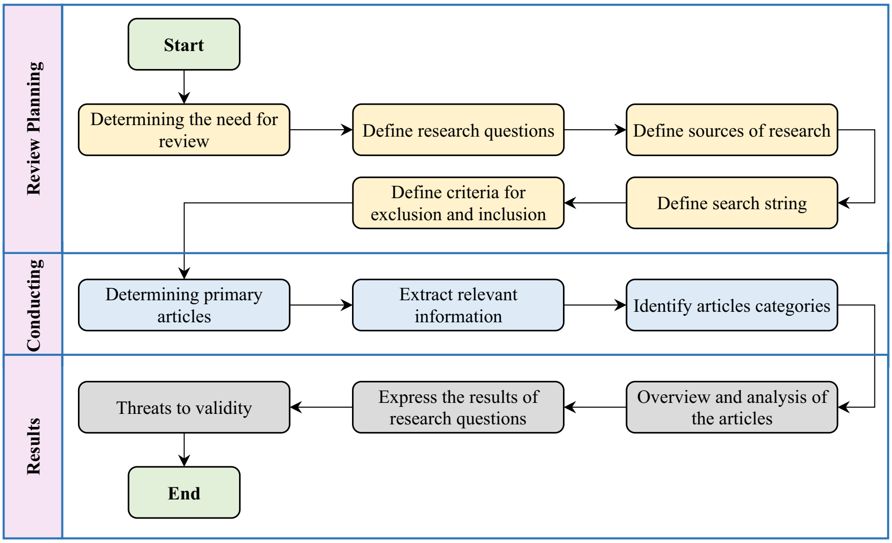

Review Planning
Start
Determining the need for review
Define research questions
Define sources of research
Define criteria for exclusion and inclusion
Define search string
Conducting
Determining primary articles
Extract relevant information
Identify articles categories
Results
Threats to validity
Express the results of research questions
Overview and analysis of the articles
End

Fig. 7 Systematic survey methodology

3.1 Motivation

As discussed earlier, the use of big data approaches in IoT as well as data mining techniques in IoT promises to improve security and reduce processing costs while increasing flexibility [27]. In general, the rapid growth of IoT can improve the level of communication and interaction with various devices, however, IoT-based devices suffer in terms of security and trust [28]. Therefore, solving IoT security issues based on new technologies is essential. Data mining techniques show promising results in previous studies to improve IoT security [29–31]. In addition, the rapid growth of IoT devices has led to large volumes and diversity of data. Nonetheless, big data technologies can achieve more efficient data management and performance. Approaches such as big data in IoT as well as data mining techniques in IoT are often used to address such challenging issues. In addition, reviewing articles related to these topics can greatly help researchers come up with new ideas in the field of IoT.

3.2 Define research questions

The main questions of this research are as follows [25]:

RQ1: How many articles are in IoT field from 2010 to 2021 and what is the rate of these articles related to IoT-data mining and IoT-big data?

RQ2: How many types of IoT review articles have been presented through data mining and big data approaches between 2010 and 2021?

RQ3: What are the main challenges of IoT when using data mining and big data approaches?

RQ4: What is the reason that encourages IoT to combine data mining and big data approaches?

RQ5: Which universities are more active in IoT-data mining and IoT-big data fields?

RQ6: What are the most IoT-related applications when used with data mining and big data approaches?

RQ7: What are free IoT domains through the use of data mining and big data approaches?

RQ8: What are the most popular IoT tools and simulators for data mining and big data?

3.3 Define sources of research

In this systematic survey, three digital science databases have been used to find articles related to IoT-data mining and IoT-big data. These databases include Scholar,[1] ACM[2] and Scopus.[3] The search order is applied as Scopus, Scholar and

[1] https://scholar.google.com/.

[2] https://dl.acm.org/.

[3] https://www.scopus.com/.

ACM, so that articles found in one database will not be considered if viewed in the next database.

3.4 Define search string

In order to search for articles in the three defined databases, the terms "big data" and "data mining" are used together with the term "internet of things" or "IoT" (i.e., IoT and big data, IoT and data mining, internet of things and big data and internet of things and data mining). In addition, different order of these terms was used for search (i.e., big data and IoT, data mining and IoT, big data and internet of things, and data mining and internet of things). All of these terms are used in the same way in all three databases to search for articles. Meanwhile, finding relevant articles is based only on the presence of these terms in the title, because the title of the article usually indicates the innovation and the main contribution of that article.

This systematic survey of articles highlights the type of review. In this regard, articles that use terms such as review, challenges, state of art, survey, SLR, SMS and literature in the title or abstract are defined as this type of articles.

3.5 Define criteria for exclusion and inclusion

In order to limit the scope of research, only articles published in journals and conferences in the last 12 years (from 2010 to 2021) are considered. In addition, only articles written in English are considered. All selected papers should focus on IoT in data mining and big data scenarios. We emphasize here that articles related to IoT applications are not considered in various fields of ML and artificial intelligence such as self-driven cars, smart city and smart homes. In addition, articles that are only relevant to the IoT domain and all books or technical reports are ignored.

3.6 Determining primary articles

In the review process, we remove articles that were found to be irrelevant after reading the title and abstract. At the end of the search, 429 articles were found, as shown in Fig. 8. The search results showed that the terms big data and IoT have not been used together before 2010 (they first appear together in 2011). After 2011, there has been an upward trend in the number of articles with both terms in the title.

According to the search policy defined in this article, out of 486 articles found between 2010 and 2021. After removing duplicate and short articles, all their abstracts were reviewed based on defined exclusion and inclusion criteria. This was done manually and finally 429 articles related to IoT-big data or IoT-data mining were selected, where 98, 117 and 214 articles are from Scholar, ACM and Scopus, respectively. Due to the large number of articles, we selected only 60 articles for a systematic and analytical review. The selection of these articles is based on experience, where the selection criteria are the quality of the journal/conference publisher, the number of pages and the quality of the articles (e.g., number of datasets used, accuracy of results, algorithms used, number of evaluation criteria and so on).

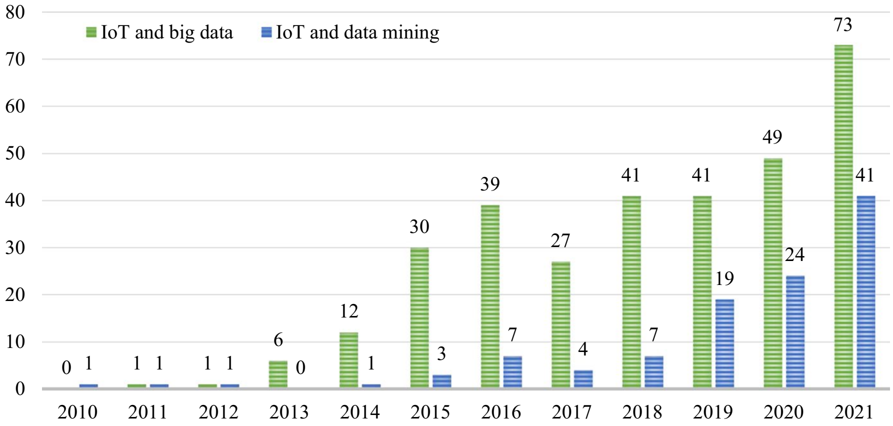
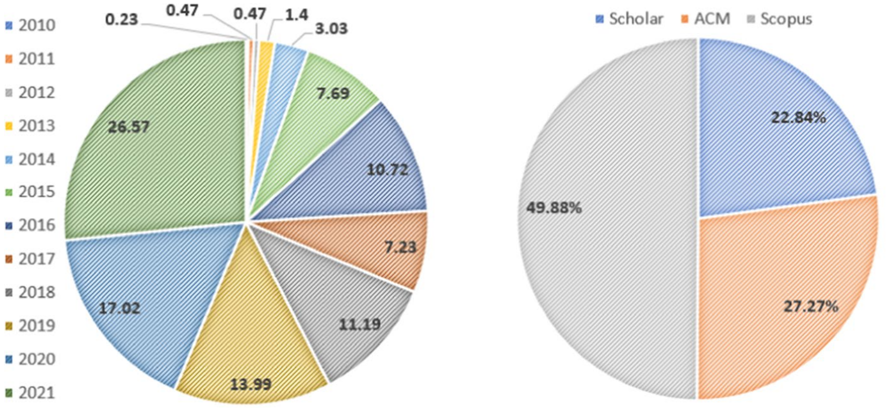


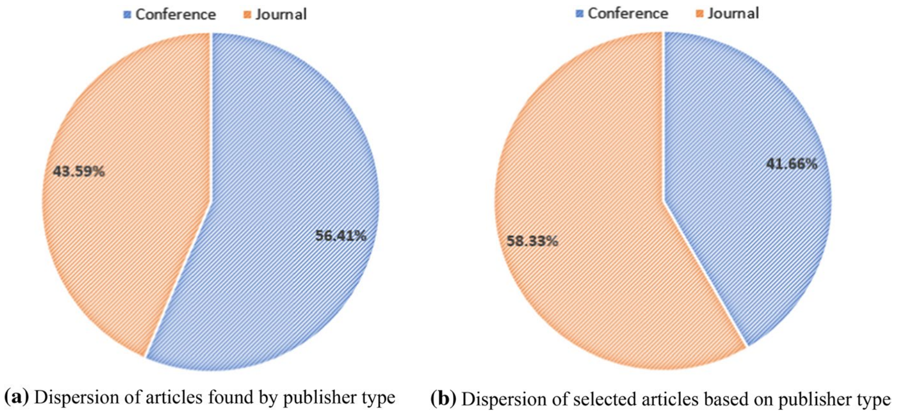
Fig. 8 Number of articles found by year of publication



(a) Dispersion of articles based on year **(b)** Dispersion of articles based on database

Fig. 9 Dispersion of articles found by year and database

The dispersion of articles report can show the upward trend of research on these fields. Also, this may indicate the importance of these fields for researchers in recent years. In this regard, details of the dispersion of articles found by year and database are given in Fig. 9. The search process for the number of articles by publisher type (i.e., conference or journal) shows that 242 conference type articles and 187 remaining articles have been published in journals (429 articles in total). Also, out of 60 articles selected for review, 25 are conference articles and 35 are journal articles. Details of the dispersion of articles found by publisher type are given in Fig. 10.



(a) Dispersion of articles found by publisher type **(b)** Dispersion of selected articles based on publisher type

Fig. 10 Dispersion of articles found and selected for review by publisher type

3.7 Extract relevant information

After reviewing the 429 articles found and also reading the 60 selected articles, we extracted relevant information that may be appropriate to answer the research questions. This activity is done manually and without the use of any tools.

3.8 Identify articles categories

We divide the selected articles into four categories: architecture/platform, framework, applications, and security. The first category consists of the articles focused on architecture/platform as indicated by containing the terms “architecture” or “platform” in the titles. The second consists of the articles focused on the framework as indicated by having the term “framework” in the titles. The third category consists of the articles focused on applications, which include the articles that have proposed different applications and ideas than those in previous works. The fourth and final category includes articles focusing on data mining techniques to big data security generated by the IoT. Out of the 60 articles, 13 are focused on architecture/platform, 7 are focused on the framework, 10 are focused on applications, and 30 are focused on security, as shown in Fig. 11.

3.9 Overview and analysis of the articles

All 60 articles selected for review were read in full. The results of the systematic survey and analysis of these articles are presented in Sect. 4. This work is based on the defined categories of articles and aims to answer research questions.

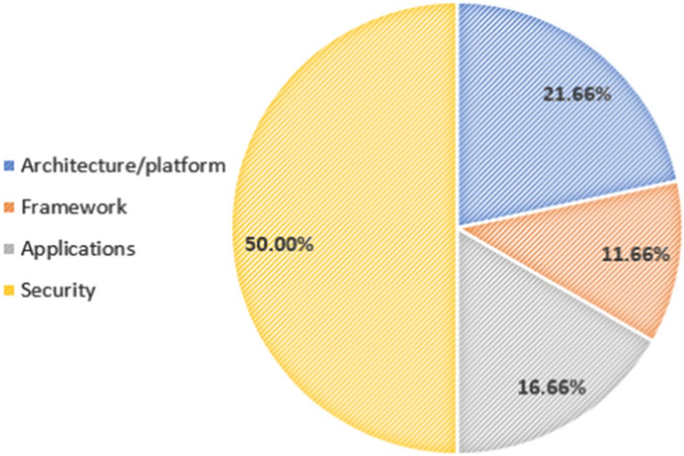


Fig. 11 Dispersion of articles selected for review by articles categories

3.10 Express the results of research questions

Based on a review of selected articles, we use our findings to answer research questions. The results of this activity are given in Sect. 5. To answer the questions, ongoing projects as well as countries and universities active in this field are studied and new challenges and ideas are presented.

3.11 Threats to validity

In this section, the limitations of this systematic survey are presented. We have identified the following threats to the validity of our article:

We formulated our search string using IoT, big data, and data mining terms. Although these are widely used terms in the field, some articles may not have been included in our systematic survey.

Only three databases, Scholar, ACM and Scopus, were used to search for articles. However, there are other databases such as IEEE, ScienceDirect, Wiley, and SpringerLink that may contain more articles related to IoT-big data or IoT-data mining.

In this paper, only articles that use the term IoT (or internet of things) along with data mining or big data in the title are considered for review. In fact, the aim was to review articles that consider the features associated with these fields. However, other articles that were not included in the search process may have approaches related to the scope of this paper.

Given the large number of articles related to the field under review, we only reviewed a small number of articles. The selection criteria for these articles were experimental. However, some important and efficient articles may not be considered in this selection.

4 Article categories and analysis

In this section, we review 60 out of the 429 articles found from the search. In the following, these articles are reviewed in four categories: architecture/platform, framework, applications, and security.

4.1 Architecture/platform-focused articles

The IoT architecture must be able to communicate with millions and possibly billions of heterogeneous objects via the Internet. Over the years, there have been many projects in the field of IoT architecture, one of which is the European Lighthouse Project (IoT-A ARM) [32]. This project is a reference architecture model presenting standards for building IoT structures and improving the compatibility of IoT solutions. The design principles of IoT-A ARM can be a starting point for building a true IoT architecture. The project consists of four components: Vision, Business, IoT Reference Model, and IoT Reference Architecture. In the vision component, basic concepts such as motivations are discussed. The business component discusses the validity and viability of the architecture and how aligning knowledge with business can offer a complete vision of the IoT architecture [33]. The IoT reference model provides the highest level of architectural definition, and the IoT reference architecture is a reference for building an IoT-compatible architecture. Some of the platforms presented in IoT include AirVantage, Arkessa, ARMmbed, Pentaho, IBM Watson, and Exosite [34]. Most of these platforms support data management but not big data management and analysis. Some of these platforms such as AirVantage, Pentaho, and IBM Watson support big data management but do not support learning tools, which means they cannot detect specific patterns when analyzing real data [34].

In [35], researchers have discussed the concept of big data analysis for IoT and future architectures, opportunities, and challenges in this area. This article names smart homes and cities as one of the opportunities and challenges of future studies. Also, this article proposed a new architecture for big data analysis for IoT that takes the value of objects into account. This architecture provides a reference model that defines the relationships between different IoT objects in contexts such as smart traffic, smart home, smart transportation, and smart city. Finally, the article states that challenges in the areas of privacy, big data mining, visualization, and integration can be interesting subjects for future works. In [36], researchers have presented IoT-based smart systems for real-time analysis of smart city data with their proposed architecture. In this architecture, a large number of sensors are installed to collect data from smart homes, smart parking lots, and surveillance systems at various locations as well as weather data, traffic data, and population data. The architecture of this system has a preprocessing phase consisting of data collection, data filtering, data classification, and also two other phases for processing and analysis. In the processing phase, offline data are processed by MapReduce in HDFS and online data are processed by Spark. The analysis phase involves analyzing the data with ML and other tools. The results of this article showed that the proposed system is very effective, even in large datasets. Moreover, the operational capacity of the system increases with the data volume.

In [37], an architecture called Tresight for improving smart tourism in Trento, Italy has been introduced. This architecture is a four-tier content suggestion system based on FI-WARE technology. Tresight focuses on analyzing big data on IoT in order to replicate human decision-making. Layers of this architecture include measurement, response, connectivity, and data. After processing the received data, the data is stored in MongoDB and analyzed by Hadoop. The analysis consists of three phases: vision, perspective, and foresight. Vision refers to the process of gaining a deep understanding of data using statistical calculations, perspective refers to understanding social and external aspects such as climate impact, and foresight refers to prediction and conservation. This architecture provides users with suggestions and services based on the content and the analysis. In [38], researchers have studied the concept of big data in IoT architecture for smart cities. This study aimed to clarify the importance of big data as an information technology for organizing and managing information in the complex systems used in smart cities. This article presents an architecture called Smart Cities Architecture (SCA) for intelligent cloud-based applications that encompasses both fields of IoT and big data. SCA is comprised of multiple layers including users, storage devices, and cloud computing. In general, this architecture offers a general framework for creating a smart city and specifies the role of users from the beginning to the end of the cycle and the important elements of each component of the architecture.

In [39], researchers have presented a smart city architecture for real-time data processing and decision making to achieve enhanced big data analysis. This architecture operates at three levels: (1) the data generation and collection level for the generation and collection of heterogeneous data related to city operations, (2) the data management, analysis, and storage level for real-time decision making, and (3) the application level for receiving decisions for validation purposes. The main purpose of this study was to use realistic intelligent architecture to increase the efficiency of data processing for real-time decision-making. Several datasets were tested in the Hadoop ecosystem to measure the performance of the architecture with different thresholds. These assessments showed that the proposed architecture provides useful insights for improving the current architecture of smart cities. In [40], an IoT-based big data software architecture for smart city services has been presented. This service-oriented software architecture addresses two key operational activities in an intelligent tool: (1) the IoT structure for managing resources and data, and (2) the application platform for decision making. Layers of this architecture include communication, data access, analysis, decision making, and resource management. The protocol and software platform used in this architecture leads to a design based on data service. This protocol can integrate big data and cloud platforms. These researchers have used Open Web standards and evolving network protocols, cloud resources, and large data platforms to design this architecture. The architecture has been tested in the field of smart water management in the Indian Institute of Science, Bangalore.

In [41], researchers have proposed a smart architecture for universities as scaled models of small cities. The purpose of this architecture is to determine the best locations for bus stations. The proposed architecture covers the needs of students by determining the time and routes of each bus based on the analysis of available student data. The first layer of this architecture is responsible for collecting data from sources such as wireless sensors, video surveillance, and university management. When the right data is available, the big data architecture allocates the necessary resources to process the information. This leads to data storage followed by analysis with Hadoop. Layers of this architecture in Hadoop include data collection, storage, filtering, transfer, analysis, and presentation. This architecture was designed to measure the distance students travel to board a bus. The results of this study show that the traveled distance as a fixed variable provides sufficient information for determining the location of bus stations. In [42], a smart big data and IoT-based city architecture called SCDC has been proposed. This article describes the concept of smart city architecture based on the latest technologies. It also discusses the application of IoT-based technologies in designing the architecture of smart city infrastructure. The proposed smart city infrastructure model is based on digital infrastructures such as big data, cloud computing, and IoT. This architecture can control people's information automatically. To address the challenges of storing and managing the big data in smart cities, these researchers have integrated IoT networks, cloud-based applications, big data, and existing wireless communications services with the following goals: 1) real-time service customization, 2) creating a friendly environment and 3) Optimal use of available resources. This architecture consists of eight main components for creating a smart city, which include smart city, smart environment, smart energy, smart security, smart buildings, smart management, smart transportation, and smart industries.

In [43], an architecture called the Block-IoT-Intelligence has been introduced. This architecture is based on data mining and intelligent IoT augmented with blockchain. As an emerging technology, blockchain supports decentralized architecture, where the challenges of data mining can be overcome by secure data sharing. The main purpose of this study was to design and develop an IoT architecture with blockchain technology and data mining to provide an effective analysis of big data. The proposed architecture is analyzed from both qualitative and quantitative aspects. In the qualitative evaluation, the process of using data mining and blockchain in IoT is described as "data-mining-based blockchain" and "blockchain-based data mining". The quantitative evaluation involves evaluating the performance of the BlockIoTIntelligence architecture in the cloud, based on parameters such as accuracy, latency, security and privacy, computational complexity, and energy. In [44], researchers have presented an IoT-based big data platform designed to predict building's energy demand through hybrid learning. This model uses a combination of k-means clustering and Artificial Neural Network (ANN) to predict the energy demand. This article states that given the temperature difference between the walls, windows, roof, and interior spaces, IoT sensors need to be installed in different places of the building. Correlation analysis is used to determine the input variables of the data-oriented prediction model. Daily energy demand is predicted by this model by using measurements from IoT sensors and the latest weather forecast. The results of this article showed a mean absolute percentage error of 3% for learning and 8% for testing.

In [45], IoT and big data have been used for real-time monitoring and processing of health data. In this article, researchers have used the data gathered by IoT sensors to design a three-layer platform model. These layers include data collection, data processing, and data visualization. Researchers in [46] have introduced an enhanced data mining method based on a duplicate item extraction algorithm for the data collected from IoT sources. This article has been focused on big data in IoT. It states that with the development and increasing popularity of 5G technology, 5G-based IoT technology is expected to grow, which will result in increased big data usage. The first thing that must be done with IoT data is to extract relevant items and find association rules between these items. In this article, this is done by an improved version of the Apriori algorithm. Regarding the time series analysis, since IoT data are coarse-grained, it is difficult to find frequent items in these data. Therefore, these data are divided into different subsets based on different times, and then an initial threshold is set for each part individually. Finally, the developed Apriori algorithm (AdAA) is used to extract frequent items. In [47], big data tourism resources based on 5G and IoT networks have been developed. Here, the use of big data with neural network analysis is investigated. Because the IoT has so much information, more and more people prefer to travel intelligently through the IoT. The authors use some classification methods in data mining to better select tourism resources. The authors use the Field Programmable Gate Array (FPGA) to solve this problem.

Table 1 provides a summary of the reviewed articles in the fields of IoT, big data in IoT, data mining in IoT that have been focused on platform/architecture. The information presented in this table includes the name of the architecture, its components, its Main idea, and its advantages and disadvantages.

Table 2 compares the articles reviewed above in terms of year of publication, tools and methods, and the general approach, i.e., whether the article has been focused on numerical analysis, implementation, programming, simulation, design, or application.

While the majority of big data articles have been focused on data transmission, storage, processing, or analysis, some of them have tried to cover all of these areas. Those big data articles that concern data analysis have also discussed data mining in this field.

4.2 Framework-focused articles

In this subsection, the articles in the field of IoT that have been focused on frameworks are reviewed and analyzed. In [48], researchers have proposed an efficient framework for the smart city using big data and IoT technologies. This framework focuses primarily on the problems of an envisioned smart city in relation to real-time decision-making. This article also discusses the various principles and requirements of a smart city to improve people’s living standards. The proposed framework is based on the parallel processing of distributed data storage. This framework is

Table 1 An overview of the articles focusing on platform/architecture

Reference	Architecture name	Components	Main idea	Advantages	Disadvantages
[35]	Marjani	Sensor layer, IoT gate, and big data layer	Presenting a method based on a new meta-model for object integration	Having a reference model for defining relationships between different IoT objects as well as appropriate decision-making support for complex tasks	Dependence of the decision-making process in the architecture on the code execution
[36]	Behera	Collection and analysis of cloud data	Using cloud computing and big data for storage and effective analysis of collected data	Efficient data management; reduced costs of using information and real-time applications for decision making	Does not consider the issue of data security
[37]	Tresight	Sensor layer, connection layer, data layer, and service layer	Developing a content-based system using FI-WARE technology	Smart IoT cost optimization	N/A
[38]	SCA	User's layer, storage layer, and cloud computing layer	Interpreting information flow cycle in the overall structure of IoT for smart cities	Using cloud environment to store data and successfully access IoT applications in smart cities	Platforms lack high-speed data processing
[39]	Nathali Silva	Data collection layer, information management layer, and application layer	Utilizing a realistic smart architecture to increase data processing efficiency for real-time decision making	The control system increases independent decision making and reliability through heterogeneous access technology	Incompatibility of the model with multiple smart city architectures
[40]	Simmhan	Communication layer, data access layer, decision making layer, and resource management layer	Service-oriented software architecture for resource management and decision making	Can be extended to other domains of smart tools and enables rapid development through API service and IoT middleware	Poor security mechanism and the problem of web-based authentication for billions of mobile phones

Table 1 (continued)

Reference	Architecture name	Components	Main idea	Advantages	Disadvantages
[41]	Smart Campus	Collection, storage, filtering, transfer, analysis, and presentation	Transformation of the physical environment of the university infrastructure	The distributed, multilevel mechanism provides an efficient starting point for a sustainability-oriented smart environment	The decision-making process depends on the integration of many variables
[42]	SCDC	Collection, processing, analysis, and control	Integration of IoT networks, cloud-based applications, big data and wireless communication services	Considers common frameworks for smart city services	Combining different frameworks increases computational complexity
[43]	BlockIoTIntelligence	Smart cloud, smart fog, smart edge, and smart device	Providing IoT architecture with blockchain and data mining for big data analysis	Enhanced scalability, interoperability, resource management, data stream integration, and architectural scalability with machine intelligence concepts	Ineffective in energy saving and delay prevention
[44]	sub-ANN	Sensor, storage, analysis, and service support	A hybrid data-driven prediction model based on k-means clustering and ANN	low error prediction of building energy demand	Model accuracy depends on the number of sensors installed in the building
[45]	Luo	Collection, processing, and visualization	Real-time monitoring and processing of healthcare data streams	Processability for data-driven controllers	Only tested for hypoglycemic scenarios
[46]	ADAA	IoT layer, infrastructure layer, data layer, analysis layer, and visualization layer	Extracting frequent items from big data	Analyzes data mining challenges in current IoT environments and improves operational efficiency	Sensitivity to the set threshold value
[47]	FPGA	Data, process, storage and information	Big data development of tourism resources	This method greatly ensures access to the data while reducing costs	FPGA Xilinx software-based system does not provide

Table 2 Comparison of the articles focusing on platform/architecture

Reference	Year of publication	Publication type	Tools and methods	Experimental type	Application	Big data/data mining
[35]	2017	Conference	ZigBee	Design	Improving interactivity with the processing engine	Storage, processing, analysis, and API management
[36]	2016	Conference	HDFS, MapReduce, Hive and Pig	Design	Smart power management	Storage, transmission, processing, and analysis
[37]	2016	Journal	FI-WARE technology	Implementation	Smart tourism	analysis
[38]	2016	Conference	Cloud environment	Design	Smart Cities	Storage, processing, analysis, and management of information
[39]	2017	Journal	Hadoop and Kalman filter	Implementation	Developing and planning future smart cities based on available data	Collection, storage, processing, and application
[40]	2018	Journal	Open Web	Implementation	Smart water management	Communication, storage, analysis, and decision making
[41]	2019	Journal	Hadoop	Implementation	Smart Universities	Storage, analysis, and prediction
[42]	2018	Journal	Cloud environment	Design	Automation and smart city planning	Collection, processing, analysis, and control
[43]	2020	Journal	deep chain and G-coin	Design	Healthcare industry	Storage, transmission, processing, and analysis
[44]	2019	Journal	k-means and ANN	Implementation	Predicting heating and cooling demand of the building	Storage, processing, and analysis
[45]	2019	Journal	Kaa, Storm, and S4	Implementation	Real-time healthcare	Storage, processing, analysis, and visualization
[46]	2019	Conference	Apriori	Implementation	5G Technology	Storage, transmission, and analysis
[47]	2021	Journal	FPGA Xilinx	Implementation	IoT and 5G Technology in tourism	Storage, processing, analysis, and visualization

divided into several layers. The first layer is responsible for communication and data production. The second layer is tasked with collecting and storing data in the distributed environment. The third layer includes a distributed and parallel processing model with MapReduce. The fourth layer is the analysis layer, which allows people and devices to interact directly to make real-time decisions. The result of the analysis may be used to generate forecasts, produce reports and make recommendations for the smart city.

In [19], the challenges of real-time or high-speed storage and analysis of large volumes of data of smart buildings are discussed. This paper proposes a framework called IBDA to fill the gap in the area of big data analytics. The data of this framework are collected by IoT sensors placed in smart buildings. The initial version of the IBDA framework was developed using Python and the Cloudera data platform. The applications of this framework include automatic management of oxygen levels and monitoring of hazardous gases in different parts of a smart building. The IBDA domain encompasses data generation, data mining, data storage in HDFS, data visualization, data analysis, and real-time control of smart buildings. IBDA has three components: IoT, big data management, and big data analysis. Instead of using different physical sensors, the framework was tested with fifteen virtual sensors created with Python code, which were instructed to generate large amounts of data and send them to the TCP port by Apache Flume for storage. The main contribution of this paper is the integration of big data analytics and IoT to address the challenge of real-time handling of large amounts of data generated in smart buildings. In [49], researchers have proposed a framework called Ahab for analyzing large cloud-based distributed data for IoT applications. Ahab is a general, scalable, and error-tolerant cloud-based data processing framework that allows operators to analyze collected data online and offline. This framework can help us better understand and optimize the behavior of existing infrastructure in smart cities. Ahab has been designed for easy integration of new data sources. This framework provides an extensible API for performing analysis tasks and a specific language for defining adaptation rules based on analysis results. This framework was simulated for the application of automated intersection management in a smart city. The results showed that Ahab can automatically optimize the application deployment topology by distributing the processing load over the available resources based on an online analysis of the current state of the environment and a model of historical data. Ahab consists of two layers: a streaming layer for controlling and processing the data stream, which is implemented as a Lambda architecture, and a service layer, which controls the streaming layer.

In [50], researchers have introduced a framework for using the semantic information of media data in urban environments. This framework consists of three parts: semantic extraction, semantic fusion of several models, and semantic storage and distribution. This framework is a distributed method to ensure that semantic information is shared when data is distributed on social media. This study showed that the proposed scheme has good accuracy and performance and addresses the issues of semantic retrieval for social media information. Researchers in [51] have proposed a framework for implementing a Smart Traffic System (STS) in smart cities using IoT and Big Data. The purpose of this framework is to provide better and less expensive options for the immediate updating of urban traffic information. In this scheme, low-cost sensors are permanently installed at every 500 m to detect vehicles. IoT is used to quickly access and send public traffic data for data processing. The data stream is transmitted in real-time for big data analysis. This strategy reduces the activity cycle for points with less movement and improves it for points with more movement. In this system, at least five sensors are connected to each other and communicate with a single IoT kit. All kits are connected to the Internet and share network access information. The vehicle monitoring is continuous and updates are sent regularly for big data storage and analysis.

In [52], an integrated IoT-big data framework for energy-aware industrial software has been presented. This framework is an accurate and efficient way to manage resources in the cloud and takes advantage of software-defined data centers (SDDCs) to minimize energy use. Specifically, SDDC refers to the logical processing, network, and storage resources whereby data configuration is done in real-time based on workload demand. The contributions of this article include: (1) designing an SDDC-based model for optimizing the process of deploying virtual machines and allocating network bandwidth to reduce energy consumption and guarantee service quality; (2) developing a multi-objective optimization problem for the inference of optimal resource allocation; and (3) introducing an efficient scheme for presenting the results of the formulated multi-objective optimization problem. In [53], a life cycle framework for IoT-based green agriculture is proposed. In this article, researchers state that creating green IoT systems throughout the agricultural production cycle will have a great impact on farmers' interest in IoT techniques. However, energy concerns in relation to the implementation of IoT systems also need to be addressed. With a life cycle framework, issues such as finance, operations, and emerging management practices can be incorporated into the implementation of green IoT systems in agriculture. These issues are especially important for innovative agricultural production and new types of agricultural trade.

Table 3 provides a summary of the articles in the fields of IoT, big data in IoT, data mining in IoT that have focused on the framework. The information presented in this table includes the name of the frameworks, their layers, their main idea, and their advantages and disadvantages.

Table 4 compares the reviewed framework-focused articles in terms of year of publication, tools and methods, and the approach. The last item refers to whether the articles have been focused on numerical analysis, implementation, programming, simulation, or design.

4.3 Application-focused articles

In this subsection, the articles in IoT field that are focused on applications are reviewed and discussed. In [54], big data and IoT technologies are used for smart traffic management. This article offers a low-cost method for real-time smart traffic management. In this method, IoT is used to access the traffic data collected from the sensors installed on the streets and from the software installed on drivers' smartphones. The collected data are immediately sent to a big data analysis center, where

Table 3 An overview of the articles focusing on the framework

Reference	Framework name	Layers	Main idea	Advantages	Disadvantages
[48]	Mohbey	Service layer, storage layer, processing layer, and analysis layer	Real-time management of generated data for smart city decision-making	Using the concept and features and IoT and big data technologies to strengthen smart cities	N/A
[108]	IBDA	Generation, extraction, storage, visualization, analysis, and control	Real-time control and management of data received from IoT sensors in smart buildings	Integration of big data and IoT analyses to overcome the challenges arising from data volume and speed	Limited to TCP port for data exchange
[49]	Ahab	Streaming layer and service layer	Online and offline analysis of collected data with cloud-based big data processing	General, scalable, and error-tolerant cloud-based data processing	Online and offline analysis of collected data requires sequential synchronization
[50]	CSF	Semantic extraction, semantic fusion, and semantic distribution	Mapping mass computations to semantic fusion	High efficiency and accuracy, multi-semantic social distribution, and improved semantic retrieval	Average accuracy decreases as the number of samples increases; not using ML methods
[51]	STS	Extraction layer, collection layer, storage layer, analysis layer, and prediction layer	Using IoT to immediately access public traffic data and transfer them for processing	Ability to analyze traffic congestion and provide solutions through prediction	Inability to recognize the nature and capacity of the vehicle for more flexibility
[52]	SDDC	Storage, scheduling, analysis, and service	Taking advantage of the benefits of SDDCs to minimize energy consumption	Real-time request configuration based on workload	Poor allocation of resources for critical applications
[53]	Ruan	IoT finance, supply chain, big data finance, node charging and repair, and IoT data management	Designing an IoT-based green agricultural life cycle for addressing management issues	Considering different scenarios of cultivation in controlled and open environments	N/A

Table 4 Comparison of the articles focusing on the framework

Reference	Year of publication	Publication type	Tools and methods	Experimental type	Application	Big data/data mining
[48]	2019	Conference	MapReduce	Implementation	Smart city services	Storage, processing, analysis, and service
[108]	2016	Conference	Python, Cloudera, and Apache Flume	Implementation	Control and monitoring of smart buildings	Storage, processing, transmission, and analysis
[49]	2017	Journal	MapReduce	Simulation	Intersection management in smart cities	Storage, processing, transmission, analysis, and prediction
[50]	2017	Journal	HBase, MapReduce, and Spark	Implementation	Physical cyber environment	Transmission and processing
[51]	2017	Conference	Cloud Computing	Implementation	Smart traffic management system	Storage, processing, analysis, and prediction
[52]	2019	Journal	Kafka, Storm and PROV-O	Design	Reducing energy consumption in cloud centers	Storage, processing, transmission, and analysis
[53]	2019	Journal	MATLAB	Implementation	Energy-efficient smart agriculture	Storage, processing, transmission, analysis, and prediction

they are analyzed based on a mathematical model. The article presents a smartphone application as a user interface for detecting traffic in different places and offering route recommendations for traffic management. In [55], another application of IoT and big data in a smart city has been introduced. This article proposes using a multitude of sensors to collect information for vehicle management, weather data, water data, and object observations and then using the collected data to intelligently control different urban areas with the help of Hadoop with Spark, VoltDB, and Tempest. This information is eventually reviewed and tested for climate analysis, vehicle speed control, and management of smart machines in homes. The basic idea of this article is to get the right data from different places and use the right tools to make the city smarter so that residents can get the assistance they need as quickly as possible.

In [41, 56], researchers have proposed a big data architecture for a university as the model of a smart city. In other words, they have used the university as the socio-economic representation of a small city for testing purposes. While the concept of smart universities is not novel, these researchers made some innovations in the introduction and integration of new technologies in the university environment. The method of this research involved determining all factors that play a role in transforming a traditional university into a smart university. It was assumed that data collection is done primarily through IoT technologies and data management and analysis are done based on big data technologies. The distributed and multilevel analysis mechanisms presented in this article can be a powerful starting point for finding a safe and efficient solution for building a smart environment based on sustainability concepts. In [57], researchers have developed a system for monitoring patients' health. In this system, data received from various sensors placed on patients are stored in the cloud and then analyzed by HDFS and MapReduce. This system uses various sensors to measure blood pressure and heart rate, and analyzes the readings in real-time.

In [58], the applications of secure IoT and big data technologies in smart cities are discussed. This article examines the development of a smart city in Romania and offers several suggestions for improvement. It states the aforementioned technologies offer advantages in the areas of resource utilization, quality of life, transparency, and autonomy of citizens. A smart city can use digital or ICT technologies to improve the quality and effectiveness of municipal services to the benefit of its citizens. In this article, key indicators such as smart economy, smart life, smart citizens, smart management, smart mobility, and smart environment have been used to rate the city. This article also discusses how factors such as interoperability, security, smart jobs, effective public administration, tourism, healthcare, transportation, and smart e-government services contribute to the development of smart cities. In [59], the applications of big data taken from social networks and IoT sources in smart cities are discussed. This paper first reviews algorithmic advances in the area of big data analysis in the context of smart cities and then presents a platform based on the service-oriented architecture for the retrieval and analysis of large datasets from social networks and IoT sources. The data of this platform are collected by smart city applications and socially aware data services. The proposed platform, which is called RADICAL, is tasked with analyzing sentiments associated with combined IoT/social network data stored in a SQL database. RADICAL uses the components and tools of the SocIoS and SmartSantander projects to integrate social media and IoT data so that they can be used to support innovative smart city services. This platform uses algorithmic configurations and enhancements to optimize data processing latency and retrieval performance.

In [60], emerging trends in IoT and big data analytics for medical and healthcare purposes are discussed. This article examines the theoretical, methodological, and practical implications of using IoT and big data in the medical field. First, it discusses the use of IoT and big data solutions in the analysis of medical databases with the help of ML algorithms and then proceeds to discuss the process of producing structural information for telemedicine. The article also reviews the applications of telemedicine with artificial intelligence methods in robotic health care and also briefly discusses the advances in robotic surgery and the importance of the Internet of Robotic Things (IoRT). Finally, the article describes the recent developments in these technologies and their use in modern health care systems and in related biomedical research. In [61], researchers have proposed a cloud-based IoT solution for the simultaneous batch processing of large datasets in health monitoring applications. Health monitoring applications, where medical sensors collect data from patients and send it to the cloud, often encounter two big data-related problems. The first problem concerns the real-time analysis of the collected data, which becomes more difficult with the increases in the rate of data generation, especially from IoT sources. These data should be analyzed in real-time so that appropriate action can be taken as soon as possible. The second problem is that while the medical data collected from patients over time provides a dataset for training ML models, this leads to another issue regarding the batch processing of very large and highly complex datasets. To address these issues, this paper implemented a cloud-based IoT approach on Amazon Web Services (AWS).

In [62], finance risk is highlighted in corporate management. Here, various financial indicators are selected based on data mining in IoT. Then, finance risk indices are extracted based on a set of fuzzy rules. Frequent fuzzy option set was determined the most appropriate rules by parallel mining algorithm, parallel rules and fuzzy cluster method, parallel rules. These methods lead to the identification of fuzzy association rules with minimal fuzzy validity. In [63], the impact of big data technologies and IoT on higher education is analyzed according to the current situation and future prospects. Here, the role of big data analytics in improving the learning process is identified and the challenges associated with data mining, storage, and security are described. Authors help to improvement of education process in higher education by combining big data technologies and IoT.

Table 5 provides a summary of the reviewed application-focused articles in the fields of IoT, big data in IoT, and data mining in IoT. The information presented in this table includes the focus of the paper, main idea, advantages, and disadvantages.

Table 6 compares the reviewed application-focused articles in terms of year of publication, tools and methods, and the approach, which as before, refers to whether the articles have been focused on numerical analysis, implementation, programming, simulation, or design.

4.4 Security-focused articles

The rapid growth of IoT technology enables communication and interaction with various devices [13]. However, the IoT has been shown to be vulnerable to security breaches. Therefore, maintaining the security of existing data is considered as one of the major challenges in IoT-based environments. The ML and data mining techniques provide promising results to improve security. In addition, the use of big data technologies is recognized as one way to improve security in IoT. In general, the use of big data technologies in data mining can achieve higher performance and better management of IoT data. In this section, IoT security issues are discussed in three sections: data mining in IoT security, big data in IoT security, and data mining and big data in IoT security [13, 15].

Table 7 provides a summary of the reviewed security-focused articles in the fields of IoT, big data in IoT, and data mining in IoT. The information presented in this table includes the main content, main idea, advantages, and disadvantages.

Table 8 compares the reviewed security-focused articles in terms of year of publication, tools and methods, and the approach, which as before, refers to whether the articles have been focused on numerical analysis, implementation, programming, simulation, or design.

4.4.1 Data mining in IoT security

The purpose of this section is to briefly analyze data mining techniques to improve the security of IoT systems. In general, many research studies have been conducted using a wide range of data mining techniques to find the best solution to improve IoT security [13, 15]. In the following, we will analyze some of the approaches that have been used repeatedly by researchers and have provided worthy results in improving IoT security.

Support Vector Machine (SVM) is one of the popular classification techniques in data mining that often offers appropriate accuracy [64]. In [65], the C-support SVM (c-SVM) optimization algorithm based on Radial Basis Function (RBF) was introduced to detect IoT malicious traffic on the network. The reason for using the RBF kernel is the superiority of its classification results over other linear functions such as sigmoid. In [66], the c-SVM algorithm for IoT security is parsed, where C-support values can cause data classification errors. Here, higher C values mean that all training data must be properly classified. In general, this algorithm similar to SVM is always looking for a model that creates the highest margin in the hyperplane. In [67], a hierarchical approach to intrusion detection in IoT devices is presented. Here, deep learning and SVM are used to detect and classify the types of intrusions, where the objective is to balance the detection efficiency with the overhead of local resources. In [68], a method for detecting distributed denial-of-service (DDoS) attacks for IoT networks is presented, where feature selection is done by several data mining techniques. Here, decision tree techniques, KNN, random forest and SVM are compared, which show the results of SVM superiority.

Random Forest (RF) is another popular data mining technique that performs ensemble classification based on boosting and bagging techniques [64]. Researchers

Table 5 An overview of the articles focusing on the application

Reference	Focus	Main idea	Advantages	Disadvantages
[54]	IoT, big data, and user interface	Introducing a smart traffic management system	Low-cost real-time traffic management	Failure to consider the type and specifications of the vehicle
[55]	Collecting the right data from different places and using the right tools to make the city smarter	Using sensors to collect data and linking them to smart frameworks	Using smart databases and frameworks to collect and evaluate data from different parts of the city	No specific process for determining the best layer to use for Hadoop with Spark
[41, 56]	Collection, storage, filtering, transmission, analysis, and presentation	Building a smart university as a small-scale city	Recommending the best activities based on data analysis and expanded learning and creating comfortable and environment-friendly ecosystems	The decision process depends on the integration of many variables
[57]	Sensors, IoT agents, big data-based health monitoring using mobile phones with GPRS/GSM connections	Analyzing IoT-based big data and creating a smart health system	Reduced system response time; suitable for real-time alerts	False alerts
[58]	Simulation, analysis, and comparison	Using different factors to rate smart cities	Raises awareness of local events and history to make the city smarter	Lack of ideas for storing big data
[59]	IoT devices, social networking sites, smart city applications	Integrating big social network data and IoT data for use in smart city infrastructure	Reduced storage, retrieval, updating, and processing time	Not using an appropriate algorithm to minimize data processing delays
[60]	Comparison and analysis	Investigating the emerging trends in IoT and big data analyses in the medical field	Using machine intelligence to predict health issues	Failure to check the system performance under heavy loads
[61]	Sensors, cloud server, storage, analysis, and monitoring	Introducing a cloud-based IoT approach on AWS for health monitoring	Good adaption to changes in the speed and volume of data; guaranteeing a certain response time	Not guaranteeing proper processing in long-term scenarios
[62]	Big data, IoT, and financial indicators	Early warning of enterprise finance risk	Analysis of financial risks of companies with different parameters	Not yet fully developed
[63]	Big data, IoT and higher education	Improvement of education process in higher education	A new approach to combination of big data and IoT	Need to analyze synchronization with convergent educational platform

Table 6 Comparison of the articles focusing on the application

Reference	Year of publication	Publication type	Tools and methods	Experimental type	Application	Big data/data mining
[54]	2016	Conference	Mathematical model	Implementation	Real-time smart traffic management	Analysis
[55]	2018	Conference	Hadoop with Spark, VoltDB, and Tempest	Implementation	Building a smart city	Storage, processing, transmission, and analysis
[41, 56]	2016	Journal	Hadoop and cloud computing	Implementation	Building a smart university	Storage, processing, and prediction
[57]	2016	Conference	HDFS and MapReduce	Implementation	Health monitoring	Storage and processing
[58]	2017	Journal	Jobs-Nearby and CityDrop	Simulation	Smart living in a smart city	Storage and processing
[59]	2016	Journal	SocIoS and SmartSantander	Design	Innovative smart city infrastructure	Storage, aggregation, synthesis, and analysis
[60]	2020	Journal	HDFS and Spark	Design	Designing a system for personal and targeted medication	Storage, processing, and analysis
[61]	2019	Journal	AWS and Raspberry	Implementation	Health monitoring	Storage, processing, transmission, and analysis
[62]	2021	Journal	N/A	Implementation	Financial risks of companies	Storage, aggregation, synthesis, and analysis
[63]	2021	Journal	N/A	Design	Responding to the needs of digital transformation	Storage and processing

have used RF in several areas of research, including finding anomalies, malicious network traffic, and improving IoT security. In [69], several data mining techniques are analyzed to detect DDoS attacks on IoT devices. Here, the results of logistic regression, decision tree, SVM and RF are compared. The results show the superiority of the RF algorithm for detecting DDoS attacks with 99.17% accuracy, where it performs better than other compared data mining techniques. In [70], various data mining techniques have been evaluated to identify Mirai botnets in IoT traffic. These techniques include KNN, Naïve Bayes (NB) and RF. The simulation results show that the production of classification models by RF with 99.1% accuracy is slightly better than other data mining techniques. In [71], different classification techniques of data mining are compared to identify DDoS attacks in IoT traffic. These techniques include SVM, KNN, decision tree, RF, and ANN. In this comparison, the results report the best performance by RF with 99.2% accuracy for detecting malicious IoT traffic.

Deep Learning (DL) is a branch of ML and data mining and uses the method used by the human brain to learn specific topics for analysis and decision making [72]. DL performs better in the face of large data than conventional data mining techniques. In [73], they used deep neural networks with a large number of layers to improve the accuracy of DL models. By adding Graphics Processing Units (GPUs) to the layers, this method can provide the use of thousands of processing units and repetitive multiplication of matrices. Here, both traditional DL and ANN models are used to detect large-scale attacks, where DL with Rectified Linear Unit (ReLU) activation function with 99.01% accuracy offers better performance than ANN. In [74], a model for improving IoT security based on Genetic Programming (GP) and Deep Belief Network (DBN) is presented. This method detects attacks by reducing the complexity of the DBN, where GP improves network structure by optimally searching for the number of layers and neurons. In [75], a DL-based method for detecting attacks in IoT networks is proposed. This method uses Bi-directional Long Short-Term Memory Recurrent Neural Network (BLSTM RNN) based on binary classification.

Convolutional Neural Networks (CNNs), like other ANNs, are composed of weighted and biased neuronal layers as well as learning capabilities. CNN consists of one or more convolutional layers and layers for classifying input data, including images [72]. In [76], an extensive study has been conducted to find efficient CNN for the security of IoT devices. Here, a balanced, accurate, low-cost CNN model is provided. This model requires CNN to reduce the complexity of locating an IoT device that requires knowledge and trial-and-error testing. In [77], another method is proposed to reduce the computational overhead when analyzing large IoT data on fog computers. This method provides appropriate efficiency in detecting malicious network traffic, because fog computing resources are close to the end users. Besides, the authors used a combined CNN and LSTM model to classify malicious IoT data traffic. In [78], a CNN-based Intrusion Detection System (IDS) is introduced in IoT environment that selects features as multi-objective optimization. This system is a combination of CNN and LSTM algorithms for classifying different attacks in big data applications. This hybrid technique has been able to DDoS attack detection in IoT with 99.03% accuracy. In [79], an approach to detect infected software and

Table 7 An overview of the articles focusing on the security

Reference	Main content	Main idea	Advantages	Disadvantages
[65]	Diagnosis of abnormalities in IoT	Using real IoT network traffic	Up to 100% detection accuracy for Blackhole and Sinkhole attacks	IDS dependence on high-energy nodes due to computational cost of c-SVM
[66]	Anomaly detection in IDS	Using SVM to security attacks and detecting intrusions	Identify the limitations of anomaly detection approaches with primary capabilities	Uncertainty is not included in the intrusion detection process
[67]	Intrusion detection protocol	Combination of SVM and DL for intrusion detection	Providing a compromise between intrusion detection efficiency and resource overhead for security	Lack of analysis of various anomalies
[68]	Automatically detect IoT attack traffic	Using IoT-specific network behaviors to detect DDoS	Analysis of a wide range of types of ML algorithms	There is no mechanism to counter the detected DDoS attack
[69]	DDoS attack traffic detection	Using different data mining techniques such as SVM and RF	Perform defensive actions in addition to detecting DDoS attack	Low detection accuracy in detecting DDoS attacks
[70]	Diagnosis of Mirai-alike botnet	A novel profile scheme based on domain name system	Analysis of ML classifiers in identifying effective features in botnets detection	Failure to review the computational cost of the techniques used
[71]	DDoS attack detection	Analysis of the latest ML approaches to detect DDoS attacks	Expressing the advantages, disadvantages and performance of different ML approaches in identifying DDoS attacks	The number of approaches examined is limited
[73]	Understanding the GPU utilization of DL workloads	Determining the GPU utilization of DL workloads through a prediction engine	Determining GPU utilization without the need for in-depth or isolated online profiling	Not mentioned in the relevant article
[74]	Maintain the security of the IoT system	An IDS model based on improved DBN	Creating an adaptive DBN structure through multiple iterations of the genetic algorithm	Failure to consider all deep network parameters
[75]	Detecting attacks within IoT network	A novel deep learning technique for detecting attacks using BLSTM RNN	Focusing on binary classification of normal and attack patterns to reduce complexity	Failure to analyze the values of the model parameters

Table 7 (continued)

Reference	Main content	Main idea	Advantages	Disadvantages
[76]	Improved security on resource-limited IoT devices	Reduce large network models or use small network models	An efficient and accurate CNN for IoT devices	Failure to check the fine-tuning of the meta-parameter to reduce memory footprint
[77]	DL analysis in industrial IoT	A mechanism to optimize DL model for reduce computational power	Reduce network traffic overhead for industrial IoT	The big data analysis is not intelligent
[78]	IDS against DDoS attacks in IoT networks	Combining DL and multi-objective optimization methods to detect DDoS attacks	reduction data dimension and using LSTM to improve attack detection	Failure to investigate various attacks based on IoT networks
[79]	Cyber security threats detection in IoT	A hybrid DL approach to detect pirated software and malware-infected files in IoT	Using tokenization and weighting feature methods to filter noisy data	Not yet fully developed
[80]	classifying malicious traffic at the packet level in IoT	Building the malicious classification system with the primary support of word embedding	Using DL to generate IDS without detection delay	The average accuracy decreases with increasing number of samples
[81]	DDoS attack traffic detection in IoT	A novel DDoS detection scheme based on LSTM	Fast detection by checking only a small number of network stream packets	Limited to implemented commands
[82]	Intrusion detection in IoT	Using ANN based threat detection for IoT to solve the authentication issues	This approach has apt robustness, accuracy and precision on large and heterogeneous datasets	The performance did not evaluate for increasing loading
[83]	Privacy in the cloud	Providing an efficient homomorphic encoding scheme	Extracting frequent items used for privacy using cloud databases	Lack of evaluation of the effectiveness of this scheme in safety computing applications
[84]	Encryption scheme for secure data sharing in the cloud	Using random keys and decryption to share data	Solving the trust problem in unlicensed identity-based cryptography	Limitation control of privacy settings and trust in the cloud environment
[85]	IoT privacy protection	Introduction of Hilbert curve cryptography technique	Search out-of-source databases at the same time as validating the data	This scheme does not support query processing and user access control

Table 7 (continued)

Reference	Main content	Main idea	Advantages	Disadvantages
[86]	IoT data processing and big data security	Introducing a new system for big data security in IoT	Flexibility by encrypting separate parts of the database	Analysis of big data concepts and aspects of IoT is not included
[87]	Academic topics of privacy and its organization	A multi-stakeholder approach to convenient privacy in big data	Development privacy regulations, protect users, and promote big data applications	How to control big data and privacy is not clearly stated
[88]	Troubleshoot IoT through cloud integration	Covering some cloud gaps such as limited domain	Troubleshooting limited storage space in IoT through cloud integration	Public key encryption cannot be applied to all IoT layers due to computational power limitations
[89]	Intersection of security and privacy in IoT	Security limitations of big data resulting from different IoT designs	Effective intersection control through the process of iteration and redesign	Failure to consider the contrast between constraints in different contexts
[90]	Distributed IoT security monitoring algorithm	Extracting information of multidimensional time series data on big data scale	Extracting the dependence of flow data on time and place from multidimensional big data	Detection error analysis for DoS attack only
[91]	Parallel big data processing system for IoT security	Increase storage reliability and processing of requests	Considering the computational constraints of the IoT for implementation	Existence of computational resources when running in Hadoop
[92]	IDS development in IoT using big data analytics	Reducing complexity by selecting features and reducing dimensions	Comparing different data mining techniques and using feature selection to analyze big data	Database balance is not included to identify attacks
[93]	Maintain and improve privacy on the WBAN	Using the assured data deletion approach for health data privacy	Using the DL to analyze big data and monitor patient health data	This system does not consider a balance between security, efficiency, flexibility and usability
[94]	Development of hybrid IDS to counter IoT security threats	Combining MLP and FFNN based on big data technologies and clustering platform	Apply DL model based on host and network-based features	This system does not provide comprehensive information on the structure and specifications of the malware
[95]	A framework for monitoring IoT mobile security	Using ensemble classification based on majority voting, weighted voting and soft voting	Proper integration of big data and data mining techniques by the proposed framework	This framework has not been tested in special software environments such as Hadoop and Spark

Table 8 Comparison of the articles focusing on the security

Reference	Year of publication	Publication type	Tools and methods	Experimental type	Application	Big data/data mining
[65]	2019	Conference	N/A	Simulation	IoT-based intrusion detection	Storage, processing, and prediction
[66]	2021	Journal	MATLAB	Implementation	Anomaly detection schemes	Storage, processing, and prediction
[67]	2019	Conference	Castalia	Simulation	Detect attacks against WSN and IoT networks	Analysis and prediction
[68]	2018	Conference	Python	Simulation	Improving IoT-based network traffic security	Storage, processing, and prediction
[69]	2019	Conference	Apache	Simulation	Security of IoT devices	Analysis and prediction
[70]	2019	Conference	N/A	Implementation	Detection of botnets in IoT-based networks	Storage, processing, and prediction
[71]	2019	Conference	MATLAB	Simulation	Detection on DDoS attacks of IoT systems	Storage, processing, transmission, and analysis
[73]	2020	Conference	N/A	Design	Improved security of GPU based systems	Storage, processing, transmission, and analysis
[74]	2019	Conference	Python	Simulation	Maintain the security of the IoT system	Analysis and prediction
[75]	2018	Conference	Spyder	Implementation	Improving the security of IoT networks	Storage, processing, and prediction
[76]	2019	Journal	CIFAR-10	Simulation	deploying the CNN on resource-constrained IoT devices	Analysis and prediction
[77]	2020	Journal	Google cloud	Simulation	Security of industrial IoT systems	Storage, processing, transmission, and analysis
[78]	2020	Conference	TensorFlow/Keras	Simulation	IoT-based intrusion detection	Storage, processing, and prediction
[79]	2019	Journal	C + +, Java, Python	Implementation	Cyber security threats in IoT	Analysis and prediction
[80]	2019	Journal	TensorFlow/Keras	Implementation	Malicious traffic on IoT networks	Storage, processing, transmission, analysis, and prediction
[81]	2019	Conference	N/A	Simulation	DDoS detection in IoT	Storage, processing, and prediction

Table 8 (continued)

Reference	Year of publication	Publication type	Tools and methods	Experimental type	Application	Big data/data mining
[82]	2019	Conference	MATLAB	Implementation	Intrusion detection in IoT	Analysis and prediction
[83]	2016	Journal	N/A	Implementation	Consumption of resources in the cloud	Storage and processing
[85]	2012	Conference	CloudSim	Simulation	Reduce cloud computing costs	Storage, processing, and prediction
[86]	2016	Journal	N/A	Implementation	Reduce the cost of cloud communications	Storage and processing
[87]	2015	Journal	N/A	Implementation	Current technologies for big data security	Storage, processing, transmission, and analysis
[88]	2015	Conference	N/A	Design	Evaluation the benefits of cloud users	Storage and processing
[89]	2018	Journal	CloudSim	Simulation	Public key encryption in IoT layers	Storage, processing, and prediction
[90]	2019	Journal	N/A	Design	Security and privacy of IoT big data	Analysis
[91]	2019	Journal	BeagleBone Black	Implementation	Security of multidimensional streaming data	Storage, processing, and analysis
[92]	2017	Journal	Hadoop and Spark	Implementation	Security monitoring in IoT networks	Storage, processing, transmission, and analysis
[93]	2017	Conference	Apache Spark	Simulation	Intrusion detection on IoT based networks	Storage, processing, and prediction
[94]	2020	Journal	–	Design	Health monitoring	Storage and processing
[95]	2019	Journal	Apache Spark	Implementation	Intrusion detection on IoT based networks	Storage, processing, and prediction
[96]	2018	Journal	MATLAB	Implementation	Mobile IoT security systems	Storage, processing, transmission, and analysis

malware on IoT devices is proposed. This approach uses CNN to analyze the software source code. The results of this approach show 96.2% accuracy in detecting DDoS attacks when malware is transmitted.

Recurrent Neural Networks (RNN) is a type of deep neural network that is commonly used to process and find patterns in time series data [64]. One of the famous RNNs is called Long Short-Term Memory (LSTM) which has the ability to store information for a long time. In [80], LSTM is used to detect malicious IoT network traffic only by examining closed header information. The results of this method show a promising accuracy of 97.22% in experiments. Similarly, in [81], LSTM and packet loss information are used to detect attacks in IoT. The results of this method provide better performance than conventional data mining techniques. In [82], RNN was used to identify attacks in an IoT network and to solve the authentication problem. This method reports appropriate results on the NSL-KDD3 and UNSW-152 datasets.

Table 9 presents the evaluation of reviewed articles using evaluation metrics in the field of data mining in IoT security. These criteria include accuracy, precision, recall, f-measure and runtime.

4.4.2 Big data in IoT security

In this section, we look at some of the big data approaches in IoT security. In [83], a scheme for extracting frequent patterns used for privacy using cloud data servers is proposed. In [84], the use of an unlicensed proxy encryption scheme is proposed, which uses random keys and re-encryption to share data. This method limits the control of privacy settings and trust in the cloud environment. In [85], a Hilbert curve-based cryptographic technique for IoT privacy protection is proposed. This technique provides the ability to improve the search of out-of-source databases at the same time as the validity of the data. In [86], the existing processes in IoT data processing as well as the security aspects of big data are examined. Here, a new system for big data security in IoT is proposed. In [87], a multi-stakeholder approach is proposed to create appropriate privacy in big data. This approach focuses on academic discussions about privacy regulation and how it is organized. In addition, the authors introduce a new framework for regulating privacy, protecting users, and promoting social data applications.

In [88], it is shown that through the integration of IoT and cloud computing, some IoT disadvantages such as limited storage space can be eliminated. The IoT can also cover some cloud computing gaps, such as limited scope. Therefore, when IoT applications move toward cloud computing, concerns about big data analytics are addressed due to lack of trust in the service provider. The authors have shown that public key encryption cannot be applied to all IoT layers due to computational power constraints. In [89], the contradiction in the collection, use and management of big data at the intersection of security and privacy in the field of IoT is investigated. In this study, the security data requirements and limitations of different IoT systems designs are considered. The authors argue that by identifying distinct objectives in

Table 9 Comparison of evaluation criteria used in studies related to data mining in IoT security

Author(s) name	Accuracy	Precision	Recall	F-Measure	Runtime
Ioannou and Vassiliou (2019)	√	×	×	×	√
Hosseinzadeh et al. (2021)	√	×	×	×	×
Yahyaoui et al. (2019)	√	×	×	√	√
Doshi et al. (2018)	√	√	√	√	×
Chaudhary et al. (2019)	√	×	×	×	×
Dwyer et al. (2019)	√	√	√	√	×
Wehbi et al. (2019)	√	×	√	×	×
Yeung et al. (2020)	√	×	×	×	×
Li and Zhang (2019)	√	×	×	×	×
Roy and Cheung (2018)	√	√	√	√	√
Lawrence and Zhang (2019)	√	×	×	×	×
Liang et al. (2020)	√	×	×	×	√
Roopak et al. (2020)	√	√	√	√	√
Ullah et al. (2019)	√	×	×	√	×
Hwang et al. (2019)	√	√	√	√	√
Liang and Znati (2019)	×	√	√	√	×
Hanif et al. (2019)	√	√	×	×	×

big data security and IoT, more effective control over the intersection of these fields is possible through an iterative process of review and redesign.

In [90], an Online Distributed IoT Security monitoring algorithm (ODIS) on large data scales is proposed. ODIS uses an advanced operator to extract important information from multidimensional time series data, where it can extract data from distributed IoT sensors based on spatial and temporal dependence. In addition, the algorithm includes an accurate data structure model for storing IoT system behaviors, where it is able to solve the scalability problem. In [91], a parallel big data processing system is proposed for security monitoring in IoT networks. Given the limited computing capabilities of IoT networks, the authors introduce the proposed architecture based on the Hadoop and Spark platforms. The system architecture consists of five components: data collection, data storage, data aggregation, data normalization and analysis, and data visualization. The data is stored in a distributed system, which increases the storage reliability and processing speed of the data request. Besides, this architecture is capable of timely management of large data streams in IoT networks.

Table 10 presents the evaluation of reviewed articles using evaluation metrics in the field of big data in IoT security. These criteria include energy consumption, efficiency coefficient, prediction error and runtime.

4.4.3 Data mining and big data in IoT security

The purpose of this section is to review data mining techniques and recent advances in big data analytics that can be used to develop advanced security methods for IoT

Table 10 Comparison of evaluation criteria used in studies related to big data in IoT security

Reference	Energy consumption	Efficiency coefficient	Prediction error	Runtime
Li et al. (2016)	√	√	×	√
Xu et al. (2012)	×	√	×	√
Kim et al. (2016)	×	×	×	√
Toshniwal et al. (2015)	√	√	×	×
Will (2015)	×	√	×	×
Stergiou et al. (2018)	√	√	×	√
Sollins (2019)	√	√	×	√
Li et al. (2019)	×	×	√	√
Kotenko et al. (2017)	×	×	×	√
Li et al. (2016)	√	√	×	√

systems. Here, the relationship between the three salient components of this paper, namely data mining, big data, and IoT security, is explained. However, past studies often include data mining and IoT security or data mining and big data technologies [92–95]. Therefore, very few studies have been conducted on all three components of data mining, big data, and IoT security. This clearly represents a special research field for future researchers. However, we found several studies that discuss all three components simultaneously. The following are the details of these studies.

In [92], a large data framework is designed for intrusion detection using classification techniques such as SVM, Deep Belief Network (DNN), NB, decision tree and RF. Here, the criteria of accuracy, recall, specificity, false rate and execution time are used for evaluation. Besides, the implementation is based on Apache Spark in IoT networks using big data analysis. The results show better performance of DNN algorithm. However, the reported accuracy is less than 80% and DNN has more runtime than other algorithms. In [93], a privacy system is presented for the analysis of big data on Wireless Body Area Networks (WBAN). This system performs forecasting and warning work to monitor patients' health by collecting data from sensors and using DL approaches. In this study, the assured data deletion approach is used to protect the privacy of health data, where the data owner can deny some users access to their health data.

In [94], advances in hardware, software and network topology are discussed. These are issues related to IoT and the security threats required to implement modern methods. Hence, an IDS based on DNN algorithm is proposed which uses Multi-Layer Perceptron (MLP) and Feed Forward Neural Network (FFNN). The system is based on big data technologies and the Apache Spark clustering platform. Apache Spark cluster computing is configured via the Apache Hadoop Yet Another Resource negotiator (YARN). Here, accuracy, correctness, recall, f-measure, TPR and FPR are used as evaluation criteria. The simulation results show that the performance of this system is better than the classical data mining techniques in HIDS and NIDS. However, the accuracy of this system for multi-class datasets is less than 90%. In [95], a framework for monitoring mobile IoT security is proposed, which is based on big data processing and data mining. This framework identifies several data mining mechanisms to solve classification tasks. Here, ensemble classification approaches are used, where the amalgamation process is compared on the basis of majority voting, weighted voting and soft voting. In framework simulation, the performance and accuracy of the weighted voting method have been proven.

Table 11 presents the evaluation of the reviewed articles using evaluation criteria in the field of data mining and big data in IoT security. These criteria include accuracy, precision, recall, f-measure, runtime, and so on.

5 Results and discussion

This section discusses and presents an analysis of the reviewed articles. Our goal is to answer the questions posed in this paper, which we will address below.

Answer for RQ1: How many articles are in IoT field from 2010 to 2021 and what is the rate of these articles related to IoT-data mining and IoT-big data?

The searched digital databases had 17,548 articles with “IoT” or “Internet of Things” in the title. After separating the articles related to the two fields of data mining and big data, it was found that only 429 out of these 17,548 articles are related to these two fields. The percentage of articles related to these fields is shown in Fig. 12.

Answer for RQ2: How many types of IoT review articles have been presented through data mining and big data approaches between 2010 and 2021?

In the systematic survey to find articles of review type, the fields of Review, Challenges, State of Art, Survey, SLR, SMS and Literature were used. Among the 429 available articles, there was no article titled SLR and SMS. However, we found 68 related review articles, as shown in Fig. 13.

Answer for RQ3: What are the main challenges of IoT when using data mining and big data approaches?

Here, we discuss the challenges and opportunities mentioned in the reviewed articles. In [96], IoT-related challenges were divided into three parts: society, business, and technology. Given the ever-changing needs and wants of societies, IoT must be able to adapt to these changes. From the business perspective, IoT applications can be divided into three categories: consumer, commercial and industrial. The consumer IoT refers to the use of interconnected devices such as smartwatches, laptops, smartphones and smart machines. The commercial IoT refers to the use of IoT in medical machines, inventory controllers, and tracking devices. Finally, the industrial IoT refers to the use of IoT in pressure gauges, electricity meters, sewer systems, robots, pipeline monitors, and other industrial devices and systems. From the technological perspective, one can discuss the security of IoT devices, how to connect multiple devices, compatibility with software and hardware, large-scale data analysis, and smart IoT operations.

In general, the challenges in the fields of IoT, big data, and data mining can be summarized as follows:

Table 11 Comparison of evaluation criteria used in studies related to data mining and big data in IoT security

Reference	Accuracy	Precision	Recall	F-Measure	Specificity	Sensitivity	TPR	FPR	RSME	Runtime
Vimalkumar et al. (2017)	√	×	√	×	√	×	×	√	×	√
Ge et al. (2020)	×	×	×	×	×	×	×	×	√	×
Vinayakumar et al. (2019)	√	√	√	√	×	×	√	√	×	√
Kotenko et al. (2018)	√	×	×	×	×	×	√	√	×	√

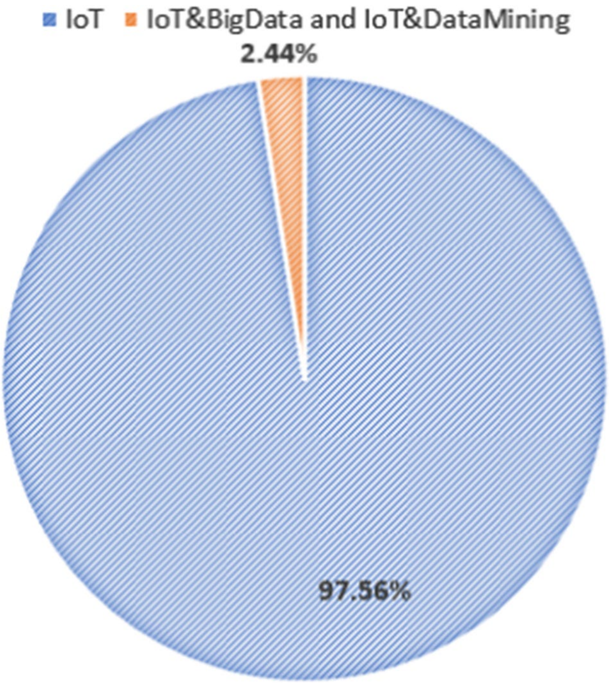


Fig. 12 Percentage of IoT articles related to data mining and big data

Data volume: this refers to the amount of data that is collected or extracted. While this parameter is commonly measured in gigabytes (GB), in future, we will be using zettabytes (ZB) and yottabytes (YB) for this purpose.

Data diversity: this is one of the biggest challenges in the field of IoT, because it is difficult to organize rapidly changing data.

Speed: data access rate has always been an important challenge in IoT.

Data value: This refers to the challenge of ensuring that data organization always leads to value extraction.

Security: Privacy and security are key issues for IoT applications. This challenge is focused on minimizing vulnerability to hacking and high reliability in data protection.

Data visualization: this refers to the challenge of using graphs and charts to visualize large amounts of complex information and make them comprehensible.

Knowledge extraction: Considering the huge amount of data collected in IoT systems, knowledge extraction techniques can be used to separate more useful parts of the data for storage.

Real-time analysis: Many applications require real-time data analysis to minimize the data storage implications of constantly increasing IoT data.

Answer for RQ4: What is the reason that encourages IoT to combine data mining and big data approaches?

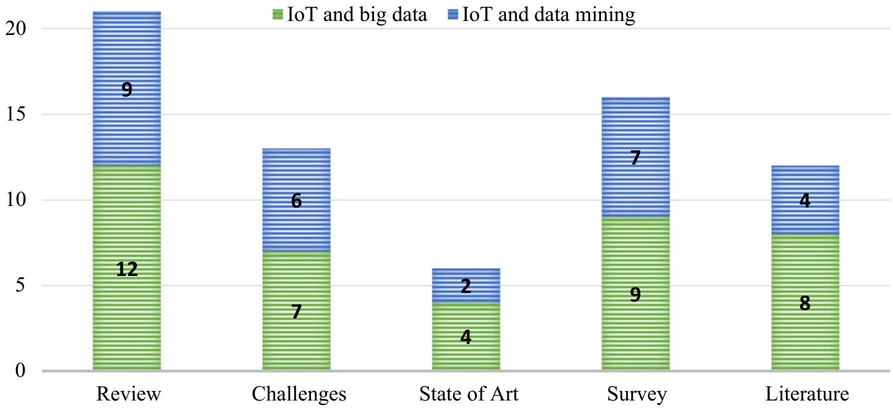


Fig. 13 Number of articles found by type of review

The availability of the Internet allows the connection of different devices that can communicate with each other and share data. IoT allows users to connect smart devices to collect real-time data from the environment. Big data is the vast amount of data that is collected from the IoT devices. However, such data are not useful without analytical power. Big data tools can provide analytical solutions to gain valuable insights into this amount of data generated by the IoT by storing large amounts of data. In addition, in IoT discussion, data mining tools can be used to process and analyze IoT data in real time.

Answer for RQ5: Which universities are more active in IoT-data mining and IoT-big data fields?

Hereon, all articles were classified according to their affiliation with universities and their country of origin. Table 12 shows the lists of the universities with the highest number of IoT articles published between 2010 and 2021. Out of 54 universities that have contributed to the publication of these 60 articles, the greatest contributor has been the Kyungpook National University of South Korea with 8 papers.

Answer for RQ6: What are the most IoT-related applications when used with data mining and big data approaches?

Based on 60 reviewed articles related to IoT-big data or IoT-data mining, 39 articles focused on applications. These articles present their content in 18 different applications. Figure 14 shows the 10 most common in these articles along with their rate of repetitions.

Answer for RQ7: What are free IoT domains through the use of data mining and big data approaches?

In addition to the challenges, there are also various promising avenues for improvement in the field of IoT based on big data handling and processing and data mining methods. In this regard, topics such as data collection, data volume reduction, data classification, data analysis, security and privacy, smart monitoring,

Table 12 Universities with more than seven articles

No	University name	Country
1	School of Electrical and Computer Engineering, Georgia Institute of Technology	United States of America
2	School of Systems Engineering, National University of Defense Technology, Changsha	China
3	VTT Technical Research Centre	Finland
4	School of Technology and Design, Canberra Institute of Technology	Australia
5	Department of Engineering for Innovation, University of Salento	Italy
6	NEC Laboratories Europe	Germany
7	School of Computer and Communication Engineering, Beijing University of Science and Technology	China
8	Hankuk University of Foreign Studies, Yongin-si	South Korea

and development of new commercial models have received more attention in the reviewed articles.

Answer for RQ8: What are the most popular IoT tools and simulators for data mining and big data?

A great number of reviewed IoT and big data articles have used the Hadoop ecosystem. This ecosystem provides a set of reliable tools and frameworks for different varieties of storage, processing, data aggregation, resource management, security, analytics, and search needs. In general, data collection in the core of Hadoop consists of an HDFS storage component and a MapReduce processing component [97]. The use of MapReduce in this system allows it to process gigabytes of data in a few seconds [98]. Spark is a very fast cluster computing technology designed for high-speed computing, which operates based on Hadoop and MapReduce and can be considered as an extension of the MapReduce model. The main feature of Spark is the storage of calculations in memory, which results in higher processing speed [99]. In addition to Hadoop and Spark, Storm is an effective tool for big data analysis. Storm processes big data based on fault tolerance and real-time scalability. In general, Storm can perform all operations, although it is not always stable. However, Hadoop and Spark perform optimally in most fields, although they greatly delay real-time applications [100].

In the field of data mining, it is common to use Mlib libraries in Spark or Mahout in Hadoop using WeKa and R software [101]. In the field of IoT and big data, it is common to use big data tools for storage, parallel processing, and real-time analysis and data mining [102]. However, the goal of these data mining methods is to automate the discussed applications in the context of Smart IoT. Sometimes, data mining methods are used to simply reduce data storage.

In addition to the tools introduced, there are other tools in the field of IoT through data mining and big data. These tools include Hive, Flume, Kafka, Redis and Mongo DB. Hive is a technique for storing and managing information that is implemented in Hadoop [103]. Flume is a reliable, accessible and distributed service for collecting large amounts of real-time data [104]. Kafka is a distributed queuing system for

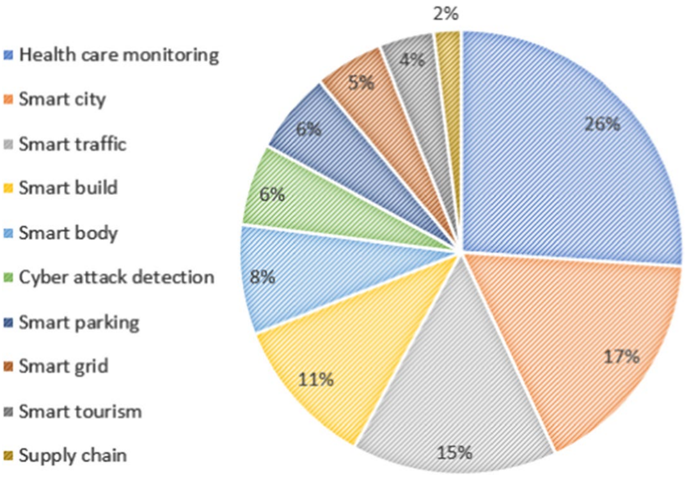


Fig. 14 Details of 10 common applications used in reviewed articles

receiving data in the Apache Storm [105]. Redis is configured based on a database that is in memory and has a very fast and powerful data structure for storage [106]. Mongo DB is a scalable database that is configured on disk and has apt speed and performance in the query language [107].

6 Conclusion

While today's world is called the world of big data, the popularization of IoT technologies is expected to cause an even greater explosion of data in future. The huge network of IoT devices generates a new type of data known as the IoT big data. Only big data-based frameworks can hope to control this huge amount of diverse data. The existing big data frameworks can effectively collect and store sensor data so that they can be analyzed with data mining methods. The biggest challenge in today's data mining world comes with several issues like security, privacy, management, data storage, and processing limitations such as real-time/streaming data. However, data analysis is a challenging task requiring the proficient use of IoT and big data analytics tools to reveal hidden patterns, trends, and correlations, a process that helps us better understand the data and use them for smart decision making. Basically, in IoT analytics, big data is the fuel and data miner are the brain of the operation. In this paper, the literature on big data in IoT and data mining in IoT was reviewed. The selected articles were divided into four categories based on their focus: architecture/platform, framework, applications, and security. Most of the reviewed articles were focused on architecture/platform, and the Kyungpook National University of South Korea was found to have the greatest contribution to the examined literature. The countries with the highest number of articles in this field were found to be South Korea, China, India, and the United States in that order.

One limitation of the study is that the search was limited to the titles of the articles, which means other articles can potentially be found by expanding the search. Also, since we only used the term “data mining” in the search, it may be possible to find a greater number of relevant articles by using the name of data mining methods. Furthermore, this review was focused entirely on English articles, but future reviews may find interesting works among non-English articles. Lastly, the search process can be expanded by including books and reports in the search. Besides, future directions could include an overview of the architecture of data mining and big data in IoT with the assisted of the cloud computing.

Funding No funding to declare.

Availability of data and material All data and materials generated or analyzed during this study were included in this published article.

Code availability There is no code for this study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Atzori L, Iera A, Morabito G (2010) The internet of things: A survey. Comput Netw 54(15):2787–2805
2. Rezaeipanah A, Nazari H, Ahmadi G (2019) A hybrid approach for prolonging lifetime of wireless sensor networks using genetic algorithm and online clustering. J Comput Sci Eng 13(4):163–174
3. Alaa M, Zaidan AA, Zaidan BB, Talal M, Kiah MLM (2017) A review of smart home applications based on Internet of Things. J Netw Comput Appl 97:48–65
4. Rezaeipanah A, Amiri P, Nazari H, Mojarad M, Parvin H (2021) An energy-aware hybrid approach for wireless sensor networks using re-clustering-based multi-hop routing. Wireless Pers Commun 120(4):3293–3314
5. Guo H, Wang L, Chen F, Liang D (2014) Scientific big data and digital earth. Chin Sci Bull 59(35):5066–5073
6. Ahmed M, Choudhury S, Al-Turjman F (2019) Big data analytics for intelligent internet of things. In: Artificial intelligence in IoT (pp. 107–127). Springer, Cham
7. Shahidinejad A, Ghobaei-Arani M, Esmaeili L (2020) An elastic controller using Colored Petri Nets in cloud computing environment. Clust Comput 23(2):1045–1071
8. Shakarami A, Shahidinejad A, Ghobaei-Arani M (2021) An autonomous computation offloading strategy in Mobile Edge Computing: A deep learning-based hybrid approach. J Netw Comput Appl 178:102974
9. Klein S (2017) The world of big data and IoT. In: IoT solutions in Microsoft’s azure IoT suite (pp. 3–13). Apress, Berkeley, CA
10. Ghobaei-Arani M, Shamsi M, Rahmanian AA (2017) An efficient approach for improving virtual machine placement in cloud computing environment. J Exp Theor Artif Intell 29(6):1149–1171
11. Berahmand K, Mohammadi M, Faroughi A, Mohammadiani RP (2022) A novel method of spectral clustering in attributed networks by constructing parameter-free affinity matrix. Clust Comput 25:869–888
12. Ghobaei-Arani M (2021) A workload clustering based resource provisioning mechanism using Biogeography based optimization technique in the cloud based systems. Soft Comput 25(5):3813–3830

13. Zhang J (2021) Distributed network security framework of energy internet based on internet of things. Sustain Energy Technol Assess 44:101051
14. Berahmand K, Nasiri E, Li Y (2021) Spectral clustering on protein-protein interaction networks via constructing affinity matrix using attributed graph embedding. Comput Biol Med 138:104933
15. Forouzandeh S, Berahmand K, Nasiri E, Rostami M (2021) A hotel recommender system for tourists using the Artificial Bee Colony Algorithm and Fuzzy TOPSIS Model: a case study of tripadvisor. Int J Inf Technol Decis Mak 20(01):399–429
16. Ghobaei-Arani M, Shahidinejad A (2021) An efficient resource provisioning approach for analyzing cloud workloads: a metaheuristic-based clustering approach. J Supercomput 77(1):711–750
17. Nasiri E, Berahmand K, Rostami M, Dabiri M (2021) A novel link prediction algorithm for protein-protein interaction networks by attributed graph embedding. Comput Biol Med 137:104772
18. Li Y, Song Y, Rezaeipanah A (2021) Generation a shooting on the walking for soccer simulation 3D league using Q-learning algorithm. J Ambient Intell Hum Comput, 1–11. In press.
19. Mohindru G, Mondal K, Banka H (2020) Internet of Things and data analytics: a current review. Wiley Interdiscipl Rev Data Min Knowl Discov 10(3):e1341
20. Yan C, Gong B, Wei Y, Gao Y (2020) Deep multi-view enhancement hashing for image retrieval. IEEE Trans Pattern Anal Mach Intell 43(4):1445–1451
21. Yan C, Li Z, Zhang Y, Liu Y, Ji X, Zhang Y (2020) Depth image denoising using nuclear norm and learning graph model. ACM Trans Multimedia Comput Commun Appl (TOMM) 16(4):1–17
22. Jesus EF, Chicarino VR, de Albuquerque CV, Rocha AADA (2018) A survey of how to use blockchain to secure internet of things and the stalker attack. Secur Commun Netw 2018:9675050
23. Yan C, Hao Y, Li L, Yin J, Liu A, Mao Z, Gao X (2021) Task-adaptive attention for image captioning. IEEE Trans Circuits Syst Video Technol 32(1):43–51
24. Yan C, Teng T, Liu Y, Zhang Y, Wang H, Ji X (2021) Precise no-reference image quality evaluation based on distortion identification. ACM Trans Multimedia Comput Commun Appl (TOMM) 17(3):1–21
25. Shadroo S, Rahmani AM (2018) Systematic survey of big data and data mining in internet of things. Comput Netw 139:19–47
26. Santos GL, Bezerra DDF, Rocha ÉDS, Ferreira L, Moreira ALC, Gonçalves GE, Endo PT (2022) Service function chain placement in distributed scenarios: a systematic review. J Netw Syst Manage 30(1):1–39
27. Aggarwal PK, Jain P, Mehta J, Garg R, Makar K, Chaudhary P (2021) Machine learning, data mining, and big data analytics for 5G-enabled IoT. In: Blockchain for 5G-Enabled IoT, pp 351–375. Springer, Cham
28. Li C, Niu B (2020) Design of smart agriculture based on big data and Internet of things. Int J Distrib Sens Netw 16(5):1550147720917065
29. Kobusińska A, Leung C, Hsu CH, Raghavendra S, Chang V (2018) Emerging trends, issues and challenges in Internet of Things, Big Data and cloud computing. Futur Gener Comput Syst 87:416–419
30. De Francisci Morales, G., Bifet, A., Khan, L., Gama, J., & Fan, W. (2016, August). Iot big data stream mining. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 2119–2120).
31. Chen F, Deng P, Wan J, Zhang D, Vasilakos AV, Rong X (2015) Data mining for the internet of things: literature review and challenges. Int J Distrib Sens Netw 11(8):431047
32. Saemaldahr R, Thapa B, Maikoo K, Fernandez EB (2020) Reference Architectures for the IoT: a survey. In: International conference of reliable information and communication technology, pp 635–646. Springer, Cham
33. Nauman A, Qadri YA, Amjad M, Zikria YB, Afzal MK, Kim SW (2020) Multimedia Internet of Things: a comprehensive survey. IEEE Access 8:8202–8250
34. Ray PP (2016) A survey of IoT cloud platforms. Future Comput Inf J 1(1–2):35–46
35. Rathore MM, Ahmad A, Paul A (2016) IoT-based smart city development using big data analytical approach. In: 2016 IEEE international conference on automatica (ICA-ACCA), pp 1–8. IEEE
36. Rathore MM, Ahmad A, Paul A, Rho S (2016) Urban planning and building smart cities based on the internet of things using big data analytics. Comput Netw 101:63–80
37. Sun Y, Song H, Jara AJ, Bie R (2016) Internet of things and big data analytics for smart and connected communities. IEEE Access 4:766–773
38. Ma Y, Wang Y, Yang J, Miao Y, Li W (2017) Big health application system based on health internet of things and big data. IEEE Access 5:7885–7897

39. Souza AM, Amazonas JR (2015) An outlier detect algorithm using big data processing and internet of things architecture. Procedia Comput Sci 52:1010–1015
40. Kholod I, Kuprianov M, Petukhov I (2016) Distributed data mining based on actors for Internet of Things. In: 2016 5th mediterranean conference on embedded computing (MECO), pp 480–484. IEEE
41. Nigam S, Asthana S, Gupta P (2016) IoT based intelligent billboard using data mining. In: 2016 international conference on innovation and challenges in cyber security (ICICCS-INBUSH), pp 107–110. IEEE
42. Lee YJ, Park HD, Min O (2016) Cooperative big data processing engine for fast reaction in internet of things environment: greater than the sum of its parts. In: Mobile and wireless technologies 2016, pp 145–149. Springer, Singapore
43. Singh SK, Rathore S, Park JH (2020) Blockiotintelligence: A blockchain-enabled intelligent IoT architecture with artificial intelligence. Futur Gener Comput Syst 110:721–743
44. Luo XJ, Oyedele LO, Ajayi AO, Monyei CG, Akinade OO, Akanbi LA (2019) Development of an IoT-based big data platform for day-ahead prediction of building heating and cooling demands. Adv Eng Inform 41:100926
45. Kharbouch A, Naitmalek Y, Elkhoukhi H, Bakhouya M, De Florio V, El Ouadghiri MD, Blondia C (2019) IoT and big data technologies for monitoring and processing real-time healthcare data. Int J Distrib Syst Technol (IJDST) 10(4):17–30
46. Wang Z, Liang W, Zhang Y, Wang J, Tao J, Chen C, Men T (2019) Data mining in IoT era: a method based on improved frequent items mining algorithm. In: 2019 5th international conference on big data and information analytics (BigDIA) (pp 120–125). IEEE
47. Gao H (2021) Big data development of tourism resources based on 5G network and internet of things system. Microprocess Microsyst 80:103567
48. Strohbach M, Ziekow H, Gazis V, Akiva N (2015) Towards a big data analytics framework for IoT and smart city applications. In: Modeling and processing for next-generation big-data technologies, pp 257–282. Springer, Cham
49. Berlian MH, Sahputra TER, Ardi BJW, Dzatmika LW, Besari ARA, Sudibyo RW, Sukaridhoto S (2016) Design and implementation of smart environment monitoring and analytics in real-time system framework based on internet of underwater things and big data. In: 2016 international electronics symposium (IES), pp 403–408. IEEE
50. Guo K, Tang Y, Zhang P (2017) CSF: Crowdsourcing semantic fusion for heterogeneous media big data in the internet of things. Information Fusion 37:77–85
51. Sezer OB, Dogdu E, Ozbayoglu M, Onal A (2016) An extended IoT framework with semantics, big data, and analytics. In: 2016 IEEE international conference on big data (big data), pp 1849–1856. IEEE
52. Kaur K, Garg S, Kaddoum G, Bou-Harb E, Choo KKR (2019) A big data-enabled consolidated framework for energy efficient software defined data centers in IoT setups. IEEE Trans Industr Inf 16(4):2687–2697
53. Ruan J, Wang Y, Chan FTS, Hu X, Zhao M, Zhu F, Lin F (2019) A life cycle framework of green IoT-based agriculture and its finance, operation, and management issues. IEEE Commun Mag 57(3):90–96
54. Rizwan P, Suresh K, Babu MR (2016) Real-time smart traffic management system for smart cities by using Internet of Things and big data. In: 2016 international conference on emerging technological trends (ICETT), pp 1–7. IEEE
55. Bera A, Kundu A, De Sarkar NR, Mou D (2017) Experimental analysis on big data in iot-based architecture. In: Proceedings of the international conference on data engineering and communication technology, pp 1–9. Springer, Singapore
56. Niyato D, Alsheikh MA, Wang P, Kim DI, Han Z (2016). Market model and optimal pricing scheme of big data and Internet of Things (IoT). In: 2016 IEEE international conference on communications (ICC) (pp 1–6). IEEE
57. Dineshkumar P, SenthilKumar R, Sujatha K, Ponmagal RS, Rajavarman VN (2016) Big data analytics of IoT based Health care monitoring system. In: 2016 IEEE Uttar Pradesh section international conference on electrical, computer and electronics engineering (UPCON), pp 55–60. IEEE
58. Saenko I, Kotenko I, Kushnerevich A (2017) Parallel processing of big heterogeneous data for security monitoring of IoT networks. In: 2017 25th Euromicro international conference on parallel, distributed and network-based processing (PDP), pp 329–336. IEEE

59. Alam F, Mehmood R, Katib I, Albeshri A (2016) Analysis of eight data mining algorithms for smarter Internet of Things (IoT). Procedia Comput Sci 98:437–442
60. Banerjee A, Chakraborty C, Kumar A, Biswas D (2020) Emerging trends in IoT and big data analytics for biomedical and health care technologies. In: Handbook of data science approaches for biomedical engineering, pp 121–152. Academic Press
61. Taher NC, Mallat I, Agoulmine N, El-Mawass N (2019) An IoT-Cloud based solution for real-time and batch processing of big data: application in healthcare. In: 2019 3rd international conference on bio-engineering for smart technologies (BioSMART), pp 1–8. IEEE
62. Shang H, Lu D, Zhou Q (2021) Early warning of enterprise finance risk of big data mining in internet of things based on fuzzy association rules. Neural Comput Appl 33(9):3901–3909
63. Mkrttchian V, Gamidullaeva L, Finogeev A, Chernyshenko S, Chernyshenko V, Amirov D, Potapova I (2021) Big data and internet of things (IoT) technologies' influence on higher education: current state and future prospects. Int JWeb-Based Learn Teach Technol (IJWLTT) 16(5):137–157
64. Shon T, Moon J (2007) A hybrid machine learning approach to network anomaly detection. Inf Sci 177(18):3799–3821
65. Ioannou C, Vassiliou V (2019) Classifying security attacks in IoT networks using supervised learning. In: 2019 15th International conference on distributed computing in sensor systems (DCOSS), pp 652–658. IEEE
66. Hosseinzadeh M, Rahmani AM, Vo B, Bidaki M, Masdari M, Zangakani M (2021) Improving security using SVM-based anomaly detection: issues and challenges. Soft Comput 25(4):3195–3223
67. Yahyaoui A, Abdellatif T, Attia R (2019) Hierarchical anomaly based intrusion detection and localization in IoT. In: 2019 15th international wireless communications and mobile computing conference (IWCMC), pp 108–113. IEEE
68. Doshi R, Apthorpe N, Feamster N (2018) Machine learning ddos detection for consumer internet of things devices. In: 2018 IEEE Security and Privacy Workshops (SPW), pp 29–35. IEEE
69. Chaudhary P, Gupta BB (2019). Ddos detection framework in resource constrained internet of things domain. In: 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), pp 675–678. IEEE
70. Dwyer OP, Marnerides AK, Giotsas V, Mursch T (2019) Profiling IoT-based Botnet Traffic using DNS. In 2019 IEEE global communications conference (GLOBECOM), pp 1–6. IEEE
71. Wehbi K, Hong L, Al-salah T, Bhutta AA (2019) A survey on machine learning based detection on DDoS Attacks for IoT systems. In: 2019 SoutheastCon, pp 1–6. IEEE
72. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
73. Yeung G, Borowiec D, Friday A, Harper R, Garraghan P (2020) Towards {GPU} utilization prediction for cloud deep learning. In: 12th {USENIX} workshop on hot topics in cloud computing (HotCloud 20)
74. Li P, Zhang Y (2019) A novel intrusion detection method for internet of things. In: 2019 Chinese control and decision conference (CCDC), pp 4761–4765. IEEE
75. Roy B, Cheung H (2018) A deep learning approach for intrusion detection in internet of things using bi-directional long short-term memory recurrent neural network. In: 2018 28th international telecommunication networks and applications conference (ITNAC), pp 1–6. IEEE
76. Lawrence T, Zhang L (2019) IoTNet: An efficient and accurate convolutional neural network for IoT devices. Sensors 19(24):5541
77. Liang F, Yu W, Liu X, Griffith D, Golmie N (2020) Toward edge-based deep learning in industrial Internet of Things. IEEE Internet Things J 7(5):4329–4341
78. Roopak M, Tian GY, Chambers J (2020) An intrusion detection system against DDoS attacks in iot networks. In: 2020 10th annual computing and communication workshop and conference (CCWC), pp 0562–0567. IEEE
79. Ullah F, Naeem H, Jabbar S, Khalid S, Latif MA, Al-Turjman F, Mostarda L (2019) Cyber security threats detection in internet of things using deep learning approach. IEEE Access 7:124379–124389
80. Hwang RH, Peng MC, Nguyen VL, Chang YL (2019) An LSTM-based deep learning approach for classifying malicious traffic at the packet level. Appl Sci 9(16):3414
81. Liang X, Znati T (2019) A long short-term memory enabled framework for DDoS detection. In: 2019 IEEE global communications conference (GLOBECOM), pp 1–6. IEEE

82. Hanif S, Ilyas T, Zeeshan M (2019) Intrusion detection in IoT using artificial neural networks on unsw-15 dataset. In: 2019 IEEE 16th international conference on smart cities: improving quality of life using ICT & IoT and AI (HONET-ICT), pp 152–156. IEEE
83. Li L, Lu R, Choo KKR, Datta A, Shao J (2016) Privacy-preserving-outsourced association rule mining on vertically partitioned databases. IEEE Trans Inf Forensics Secur 11(8):1847–1861
84. Xu L, Wu X, Zhang X (2012) CL-PRE: a certificateless proxy re-encryption scheme for secure data sharing with public cloud. In: Proceedings of the 7th ACM symposium on information, computer and communications security, pp 87–88
85. Kim HI, Hong S, Chang JW (2016) Hilbert curve-based cryptographic transformation scheme for spatial query processing on outsourced private data. Data Knowl Eng 104:32–44
86. Toshniwal R, Dastidar KG, Nath A (2015) Big data security issues and challenges. Complexity 2(2):15–20
87. Will MG (2015) Privacy and big data: the need for a multi-stakeholder approach for developing an appropriate privacy regulation in the age of big data. Available at SSRN 2634970
88. Stergiou C, Psannis KE, Gupta BB, Ishibashi Y (2018) Security, privacy and efficiency of sustainable cloud computing for big data & IoT. Sustain Comput Inf Syst 19:174–184
89. Sollins KR (2019) IoT big data security and privacy versus innovation. IEEE Internet Things J 6(2):1628–1635
90. Li F, Xie R, Wang Z, Guo L, Ye J, Ma P, Song W (2019) Online Distributed IoT Security Monitoring With Multidimensional Streaming Big Data. IEEE Internet Things J 7(5):4387–4394
91. Kotenko IV, Saenko I, Kushnerevich A (2017) Parallel big data processing system for security monitoring in Internet of Things networks. J Wireless Mobile Netw Ubiquitous Comput Dependable Appl 8(4):60–74
92. Vimalkumar K, Radhika N (2017) A big data framework for intrusion detection in smart grids using apache spark. In: 2017 International conference on advances in computing, communications and informatics (ICACCI), pp 198–204. IEEE
93. Ge C, Yin C, Liu Z, Fang L, Zhu J, Ling H (2020) A privacy preserve big data analysis system for wearable wireless sensor network. Comput Secur 96:101887
94. Vinayakumar R, Alazab M, Soman KP, Poornachandran P, Al-Nemrat A, Venkatraman S (2019) Deep learning approach for intelligent intrusion detection system. IEEE Access 7:41525–41550
95. Kotenko I, Saenko I, Branitskiy A (2018) Framework for mobile Internet of Things security monitoring based on big data processing and machine learning. IEEE Access 6:72714–72723
96. Singh D, Tripathi G, Jara AJ (2014) A survey of Internet-of-Things: Future vision, architecture, challenges and services. In: 2014 IEEE world forum on Internet of Things (WF-IoT), pp 287–292. IEEE
97. Jesse N (2018) Internet of Things and Big Data: the disruption of the value chain and the rise of new software ecosystems. AI & Soc 33(2):229–239
98. ur Rehman MH, Yaqoob I, Salah K, Imran M, Jayaraman PP, Perera C (2019) The role of big data analytics in industrial Internet of Things. Future Gener Comput Syst 99:247–259
99. Li J, Li X, Peng Y (2019) Application of big data in agricultural internet of things. Rev Fac Agron 36:1521–1529
100. Gore R, Valsan SP (2016) Big Data challenges in smart Grid IoT (WAMS) deployment. In: 2016 8th International conference on communication systems and networks (COMSNETS), pp 1–6. IEEE
101. Dahdouh K, Dakkak A, Oughdir L, Ibriz A (2019) Large-scale e-learning recommender system based on Spark and Hadoop. J Big Data 6(1):1–23
102. Elshawi R, Sakr S, Talia D, Trunfio P (2018) Big data systems meet machine learning challenges: towards big data science as a service. Big Data Res 14:1–11
103. Ahmad M, Kanwal S, Cheema M, Habib MA (2019) Performance analysis of ECG big data using apache hive and apache pig. In: 2019 8th international conference on information and communication technologies (ICICT), pp 2–7. IEEE
104. Birjali M, Beni-Hssane A, Erritali M (2017) Analyzing social media through big data using infosphere biginsights and apache flume. Procedia Comput Sci 113:280–285
105. Le Noac'HP, Costan A, Bougé L (2017) A performance evaluation of Apache Kafka in support of big data streaming applications. In: 2017 IEEE International Conference on Big Data (Big Data), pp 4803–4806. IEEE
106. Hu L, Xia X (2021) 5G-Oriented IoT big data analysis method system. Mob Inf Syst 2021:3186696

107. Seth S, Johari R (2019) Statistical survey of data mining techniques: a walk-through approach using MongoDB. In: International conference on innovative computing and communications, pp 145–158. Springer, Singapore

108. Bashir MR, Gill AQ (2016) Towards an IoT big data analytics framework: smart buildings systems. In: 2016 IEEE 18th international conference on high performance computing and communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), pp 1325–1332. IEEE

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.