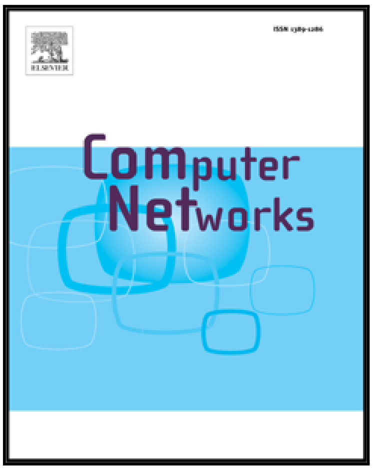
Accepted Manuscript



Systematic Survey of Big Data and Data Mining in Internet of Things

Shabnam Shadroo , Amir Masoud Rahmani

PII: S1389-1286(18)30157-9
DOI: 10.1016/j.comnet.2018.04.001
Reference: COMPNW 6463

To appear in: *Computer Networks*

Received date: 22 October 2017
Revised date: 25 March 2018
Accepted date: 2 April 2018

Please cite this article as: Shabnam Shadroo , Amir Masoud Rahmani , Systematic Survey of Big Data and Data Mining in Internet of Things , *Computer Networks* (2018), doi: 10.1016/j.comnet.2018.04.001

Systematic Survey of Big Data and Data Mining in Internet of Things

Shabnam Shadroo [1], Amir Masoud Rahmani [1,2,*]
[1] *Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.*

[2] *Computer Science, University of Human Development, Sulaimanyah, Iraq.*
Email: shabnam.shadroo@srbiau.ac.ir, rahmani@srbiau.ac.ir

*** Corresponding Author**

Abstract
In recent years, the Internet of Things (IoT) has emerged as a new opportunity. Thus, all devices such as smartphones, transportation facilities, public services, and home appliances are used as data creator devices. All the electronic devices around us help our daily life. Devices such as wrist watches, emergency alarms, and garage doors and home appliances such as refrigerators, microwaves, air conditioning, and water heaters are connected to an IoT network and controlled remotely. Methods such as big data and data mining can be used to improve the efficiency of IoT and storage challenges of a large data volume and the transmission, analysis, and processing of the data volume on the IoT. The aim of this study is to investigate the research done on IoT using big data as well as data mining methods to identify subjects that must be emphasized more in current and future research paths. This article tries to achieve the goal by following the conference and journal articles published on IoT-big data and also IoT-data mining areas between 2010 and August 2017. In order to examine these articles, the combination of Systematic Mapping and literature review was used to create an intended review article. In this research, 44 articles were studied. These articles are divided into three categories: Architecture & Platform, framework, and application. In this research, a summary of the methods used in the area of IoT-big data and IoT-data mining is presented in three categories to provide a starting point for researchers in the future.

Keyword: internet of things; systematic survey; big data; data mining

Introduction

In recent years, the Internet of Things (IoT) has grown rapidly, as such, it can identify, control, and monitor each object on earth (usually called things) via the internet [1, 2, 3, 4, 5]. Kevin Ashton invented the concept of IoT in 1999 when the wider device-to-device communicational view has occurred. In the reference [5], the IoT is defined as self-organizing systems of unrestricted devices which provide converged systems that improve the efficiency of processes; it also creates a reference that can identify the objects connected to the Internet. Furthermore, it allows the establishment of types of communication and sharing of data by using IT and provides a variety of services through the interconnection of virtual and physical things based on the interoperability of information and communication technologies. Marjani et al. explain the concept of IoT by placing a sensor on anything that has an internet connection and intelligent layers using the computational methods as shown in Figure 1 [6].

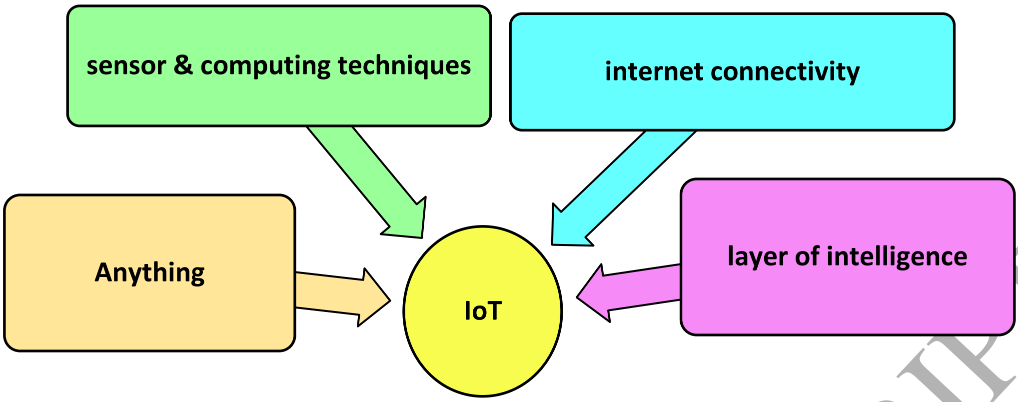


Fig.1. The concept of IoT [6]

A large number of communication devices are inside the sensor devices on the IoT. Data collection systems find the data and transfer them using embedded communication devices. The chain of devices and things are interconnected via various communication solutions such as Bluetooth, WIFI, ZigBee, and GSM. The communication devices transfer the data and receive the commands that allow integrating data from remote control devices. More than 50 billion devices including smartphones, laptops, sensors, and game consoles are forecasted to be connected to the internet via several disparate access networks, such as Radio Frequency Identification (RFID) and Wireless Sensor Networks. If we do not have any solution to this increasing volume of data over the time, gradually, this amount of data will stop us [7]. Therefore, we must think about the processes of transferring, processing, and storing data. The volume progress and data diversity and what we know as the big data nowadays are shown in Figure 2 [8].

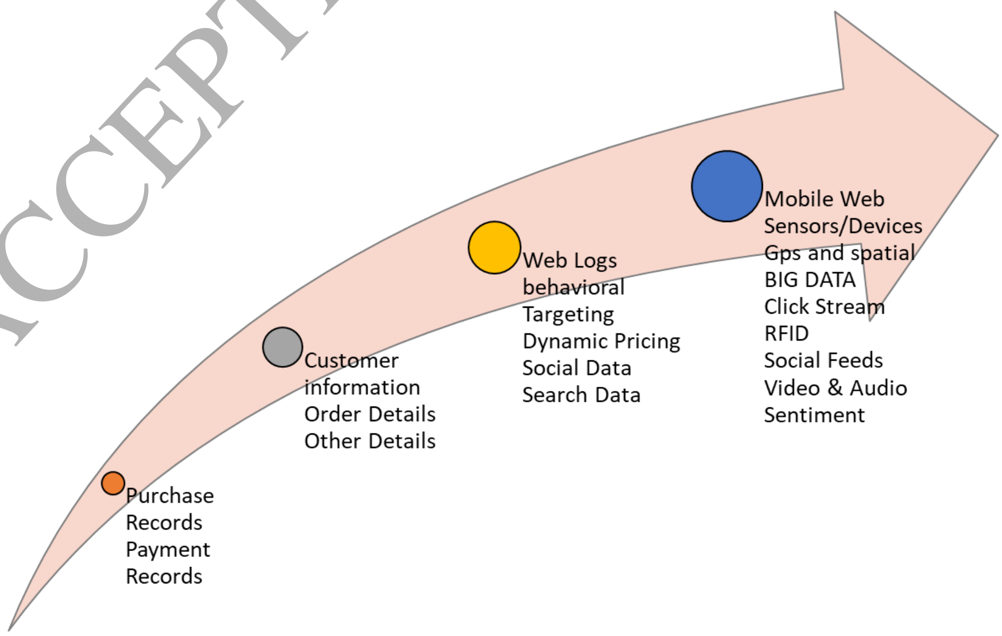


Fig.2. Data evolution in application Software [8].

The traditional database and big data have been compared by Kelian in a chapter of his book. Table 1 compared the two methods [8].

Table 1. Comparing relational database systems to big data solutions [8]

	RDBMS	**Big data Solutions**
Data Size	Gigabytes	Petabyte
Data type	Structured	Semi-structured or unstructured
Access	Interactive and batch	Batch
Update pattern	Read/Write many times	Write once, read many
Structure	Static schema	Dynamic Schema
Integrity	High (ACID)	Low
Scaling	Nonlinear	Linear

This comparison shows that the big data area is used for the large volume of data which has no specific structure. Therefore, we can use big data on IoT, which has numerous and various device connections that lead to the creation of a large volume of data in different variations and high generation velocity. By studying IoT and big data articles, we concluded that we could use big data in storage, transmission, processing, and analysis. Figure 3 shows the taxonomy based on the tools used in this area.

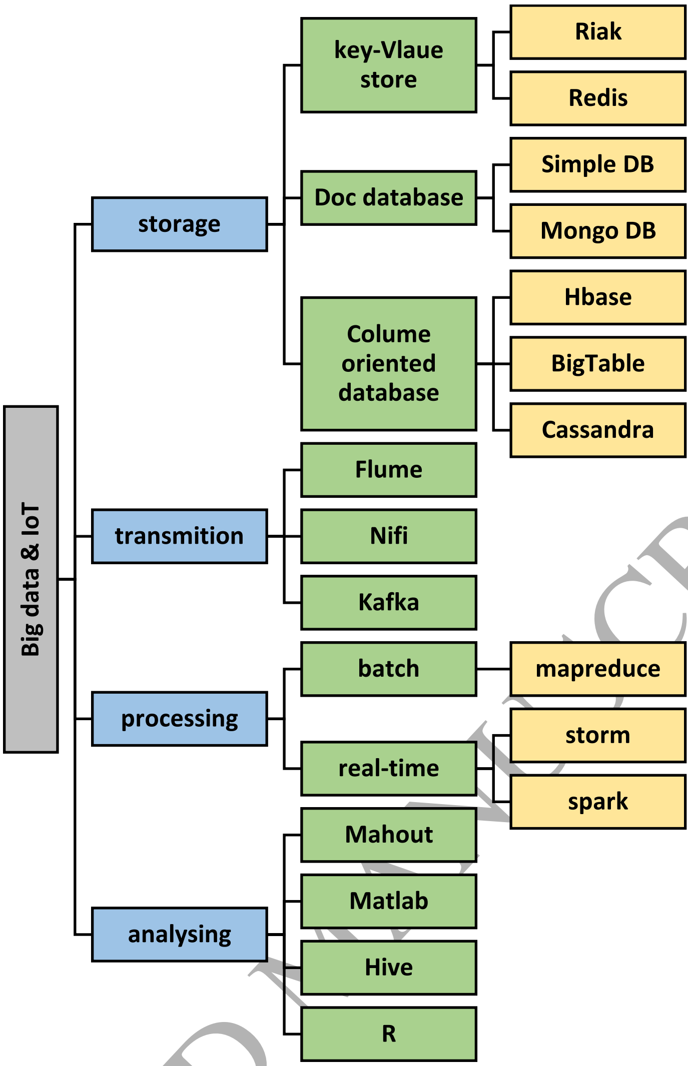


Fig.3. The taxonomy based on the tools used in IoT & Big Data

In Figure 3, the topics of big data and IoT fields are divided into four sections. The major focus of the storage section is for storing data, and the applications of IoT have tried to use databases that are proportional to the size and variety of data. A flume is a useful tool in transmission field of IoT articles that can transfer data into the storage bases. The processing sector deals with data processing in both batch and real time. These two types of processing are commonly found in the IoT articles. They are selected and used in the analysis section according to the application of analytical tools.

Furthermore, we can use some analysis methods to reduce the amount of data storage. In the big data area, there are tools for analyzing these data, which usually use data mining methods. In the data mining field, if we do not analyze or extract any knowledge of these data, we can conclude that the generated data is worthless. In addition, the devices themselves can manage the data by using intelligent tools. Data mining techniques are used to analyze the data collected [9], [10] and enhance IoT intelligence [7], [10]. Figure 4 shows a taxonomy of used data mining in IoT

In the presented taxonomy, data mining [11] is divided into three sections such that one or more sections are used proportional to the considered application in IoT.

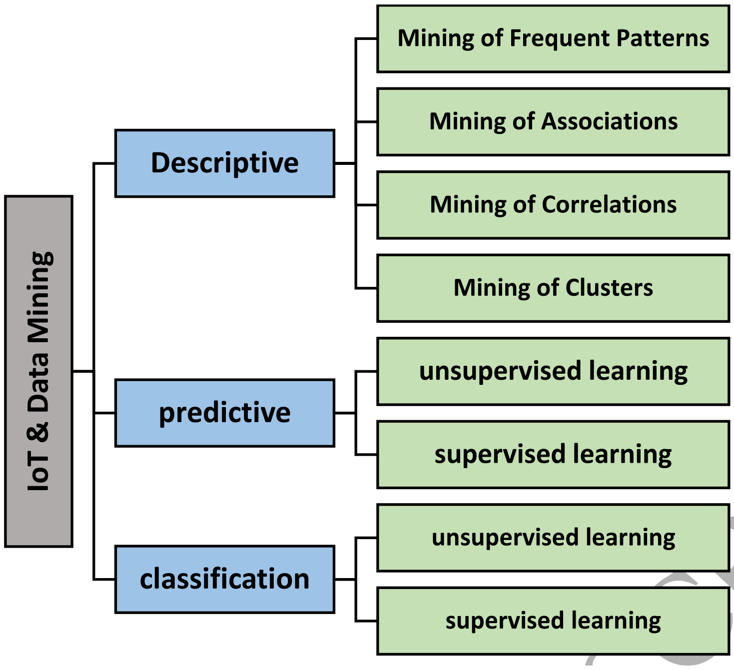


Fig.4. The taxonomy of Data Mining usage in IoT

We will systematically investigate the articles in the IoT's field and its combination with big data and data mining because the IoT has not been examined systematically so far. Therefore, this research can be useful to researchers. At the beginning of the article, we will ask some questions and categorize the available articles according to the methods presented in them into three categories of architecture and platform, the framework, and expression of application, and compare them in a table. Finally, we will answer the questions raised at the beginning of the research with respect to the obtained analysis. In order to answer the questions raised, we investigate the completing and ongoing projects, as well as active countries and universities in this field based on the papers presented in this area, and present new challenges and ideas.

This article is organized into six sections. Section 2 states the suggested method of research and research questions. Section 3 deals with the classification and comparison of the main studies in this research, and the answer to the questions raised in the research is in Section 4 and finally, conclusions are given in Section 5.

2. Research Methodology

This research examines the combined techniques with the IoT systematically. The steps used in this study are in accordance with the steps explained in the Cruz-Benito paper [12] and other research papers [13, 14, 15, 16, 17, 18, 19, 20]. The method of providing SMS (Systematic Mapping Study) and SLR (Systematic Literature Review) articles is mentioned in this article. We will describe the steps used in this research.

2-1. RQs (Research Questions)

The systematic study is designed in a way to provide a general overview of the topic by categorizing and reviewing the categories. This method involves investigating existing studies about the subject and reporting the coverage rate of the issue in the publications. By reviewing articles published, we can find available gaps on the subject, and these articles can help researchers a lot in presenting their new ideas.

These questions are as follows:

RQ1: How many types of articles are published on SMS and SLR and surveyed by data mining and big data methods from 2010 to August 2017 on IoT's area?

The answer to this question helps researchers to know about the history of combining the IoT, big data, and data mining.

RQ2: How many articles are there in IoT's area and what percentage of the articles is related to data mining- IoT and big data -IoT from 2010 to August 2017?

The answer to this question introduces the other areas that are combined with the IoT to the researchers and shows the percentage of its presentation.

RQ3: What are the most popular tools and simulators in the IoT, big data and data mining areas?

The answer to this question helps researchers know about the tools of the areas.

RQ4: What are the most applied applications on the IoT with data mining and big data methods?

The answer to this question helps researchers to recognize the applications that have been used in these areas and provide new applications.

RQ5: Which of the universities are valid in the IoT, data mining, and big data areas? The answer to this question introduces credible universities in this area to researchers.

RQ6: Why are the methods of IoT combined with the methods of big data and data mining?

With the answer to this question, the researcher would ensure that these two methods are suitable for the considered application on IoT.

RQ7: What are the main challenges of IoT with big data and data mining?

The answer to this question helps new researchers to recognise the challenges of this area.

RQ8: What are the open issues of IoT with big data and data mining?

The answer to this question helps new researchers to determine the future path of big data-IoT and data mining-IoT.

2-2. Data Base

This research focuses mainly on searching in scientific databases, and it is assumed that the books have been used in these articles. Therefore, we have not searched for books that explained the subject.

The electronic databases used in this research are:

– ACM Digital Library (http://portal.acm.org).

– IEEE Xplore (http://www.ieee.org/web/publications/xplore/).

– ScienceDirect – Elsevier (http://www.elsevier.com).

– SpringerLink (http://www.springerlink.com).

– Wiley InterScience (http://www3.interscience.wiley.com).

2-3. Search Terms

In order to search the article, we used the search term that was mentioned in Table 2. This term was the same in all databases. Since innovations are usually presented in the title of the article, we searched on the title of articles in all databases. To find survey articles, we searched some words such as Survey, State of Art, Challenges, Review, Literature, and Mapping in the title of articles. After studying the abstract of the articles, they were labeled as survey articles. We ignored articles that did not use words such as Survey, State of Art, Challenges, Review and Literature, SMS and SLR in their titles and merely introduced and reviewed the methods.

Table 2. Searches Terms

S1	((Internet of Things and big data) OR (IoT and big data))
S2	((Internet of Things and big-data) OR (IoT and big-data))
S3	((Internet of Things and data mining) OR (IoT and data mining))

2-4. ICS (Inclusion Criteria) and ECS (Exclusion Criteria)

The purpose of creating the Inclusion Criteria and the Exclusion Criteria is to examine more carefully the studies related to the research. We investigated all English articles from journals and conferences published between 2010 and early August 2017 and ignored studies related to the IoT only and have not used the combined methods. Table 3 describes the criteria for Inclusion and Exclusion articles.

Table 3. Inclusion Criteria and Exclusion Criteria

Inclusion Criteria
English peer-reviewed studies that provide answers to the research questions. Studies that focus on IoT or internet of things and data mining Studies that focus on IoT or internet of things and big data Studies are published between 2010 to August 2017
Exclusion Criteria
Studies, which are in the form of books and technical reports Studies whose full text is not available Studies are not in English Studies that are not related to the research questions Studies in which claims are not-justified or ad hoc statements instead of based on evidence Duplicated studies

2-5. Review Phases

Selection of articles was done via the following steps:

1- First, the articles are searched in the various databases by the search terms.

2- Some of the articles are deleted by the exclusion criteria.

3- Irrelevant articles are deleted after reading the title and abstract of the article.

4- Finally, the main articles are provided by reading the article completely.

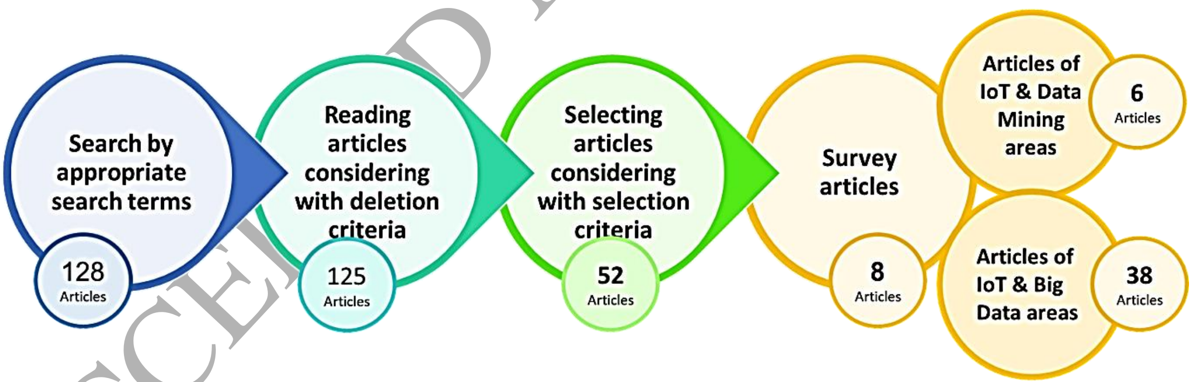


Fig.5. The adopted processes of article selection in the study

In the discussion of IoT and big data, there were 125 articles in the mentioned databases. In order to examine the trend, we used all articles in these two areas. Figure 6 shows the number of articles by year. As the trend shows in the diagram, the trend of use of big data in IoT is increasing, and today, we use a combination of IoT and big data and also data mining and IoT.

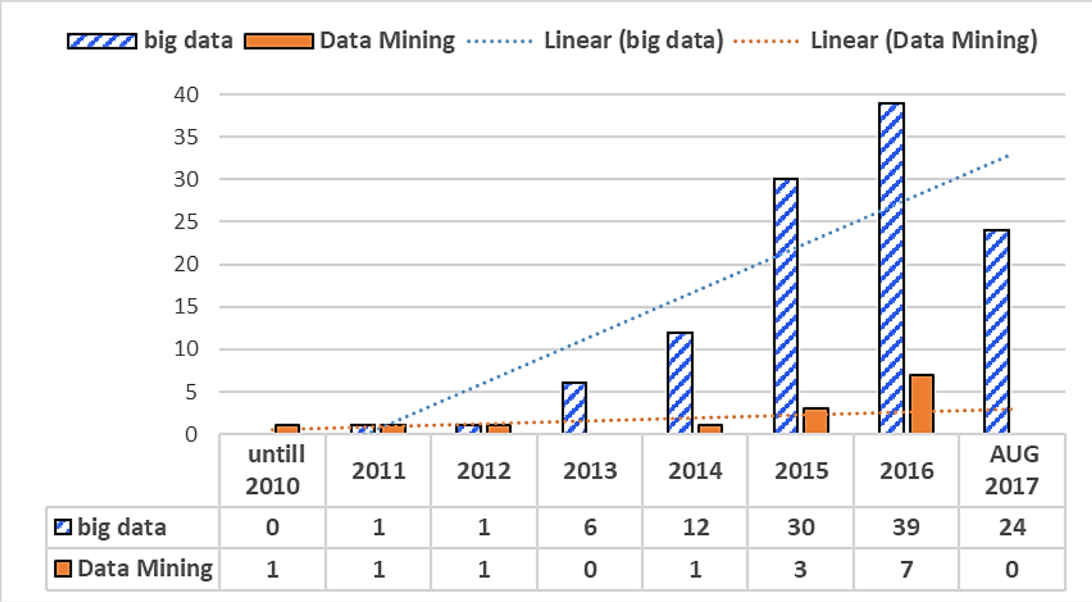


Fig.6. The number of articles by year

At first, we study the dispersion rate from 2010 to August 2017, and compare the number of articles in terms of conference, journalism and publisher type, and then examine each category of the articles.

As shown in Figure 6, IoT and big data have not been proposed together before 2010. However, the ascending trend can be seen for combining IoT and big data after 2011. To ensure this, the keywords have been searched on the abstract, titles, and the keywords of the articles in 2010 and 2011, but there was no change in the results. Thus IoT and big data were not proposed together before 2010, and this is a subject that has been raised since 2011. After reviewing 44 articles, the number of articles per year was determined as shown in Figure 7.

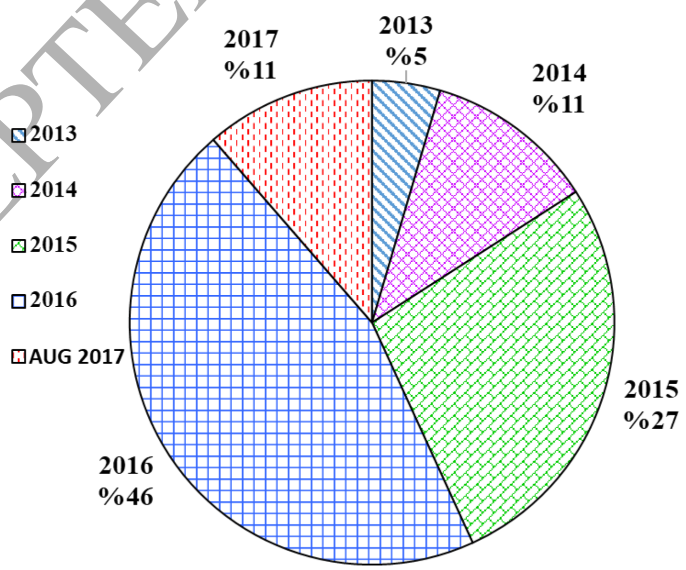


Fig .7. Dispersion rate of publications by year, from 2010 to August 2017

The number of articles and the article type presented by publishers are shown in Figure 8.

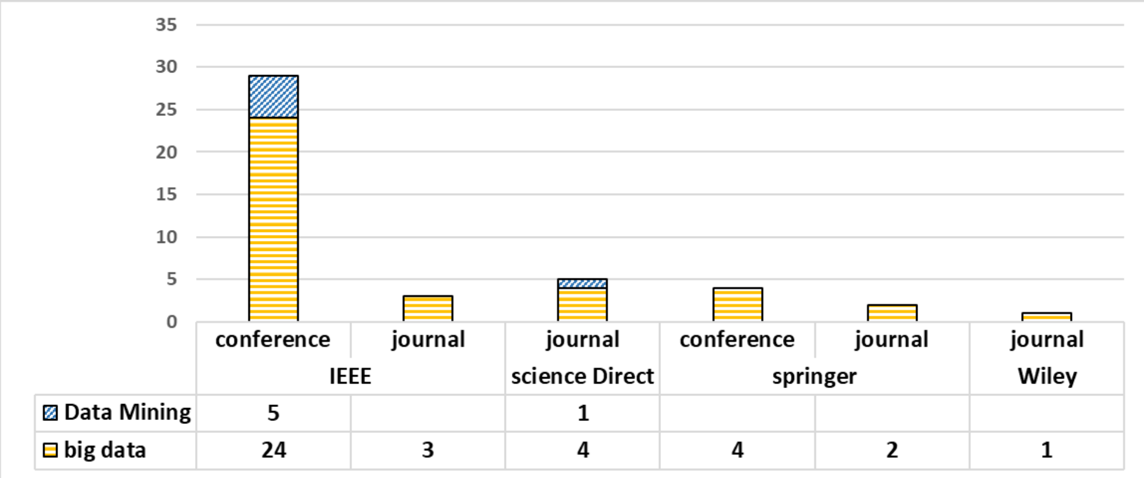


Fig .8. The number of the articles by publishers and the article type

3. Primary Studies Based on Classification Schemes

In this section, we selected the articles and then examined them, so that 44 articles were thoroughly studied. We divided them into three categories. The first category is the articles related to architecture and platform or which used the word "architecture" and "platform" in their titles, or their proposed method is in the form of architecture and platform. The second category has been explained in the framework articles, which introduced framework. In the articles, the “framework” is either in the title or their proposed method is in the framework. The third category is articles related to applications, that its idea is different from previous ones. In most of the articles, the author's main emphasis is on the application or scenario discussed in the articles.The number of articles that are studied in each category is shown in Figure 9.

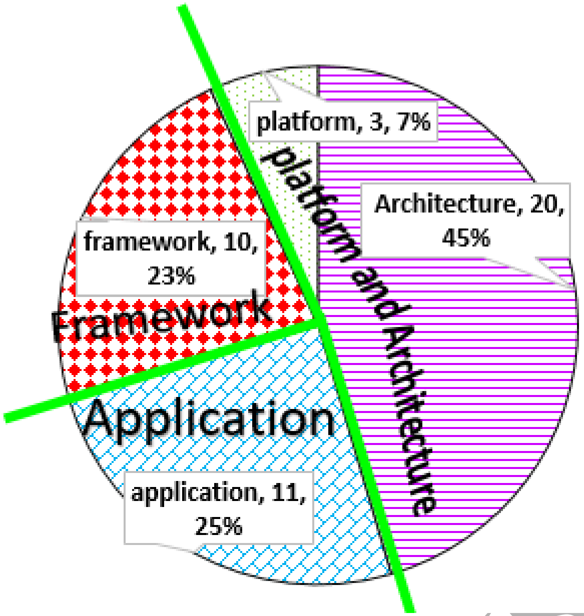


Fig.9. The number of articles is studied in each category

3-1. Primary Studies of Architecture and Platform

The IoT architecture should be able to communicate internally with millions and perhaps billions of non-homogeneous objects via the internet. Therefore, flexible layers are required in this architecture. Various projects have been defined in the field of IoT architecture, including IoT-A (The most remarkable research is the three-year European Lighthouse Integrated Project called IoT-A (Internet-of-Things)).

The IoT-A project [21], [22] presents an architectural reference model that explains the principles and standards for producing IoT structures and promoting collaboration with IoT solutions. The IoT-A-compliant architecture ensures that the knowledge generated is modular and reusable within the scope using certain use cases. However, the IoT-A addresses the design problem of architecture and does not focus on whether existing big data platforms can provide tools and services to support the implementation of IoT-compliant IoT systems.

The Architecture Reference Model is derived from the European research in IoT architecture field (IoT-A). The ARM is a useful tool that creates a common language among potential stakeholder of the M2M or IoT system. It can also provide a starting point for creating a real system architecture when certain conditions are created such as the needs of the stakeholders, design conditions, and design principles.

In the studied research, IoT-A ARM includes four parts [21], [22]:

1- Vision

   The insight summarises the principle of providing a reference model of architecture for the IoT. Simultaneously, basic concepts such as incentives are discussed. It also explains how to use the architecture reference model, the methodology used to model architecture, and business and stakeholder scenarios.

2- Business & stakeholder scenarios

Stakeholders and business scenarios are the architecture work authority. With the businesses tendencies knowledge, a comprehensive vision of the IoT architecture can be concluded.

3- The IoT Reference Model:
This section provides the highest level of abstract architecture's definition of the IoT Reference Model.

4- The IoT Reference Architecture

The reference architecture of the IoT is the reference to construct compatible IoT architecture. In order to complete the research in the field of architecture and IoT platform, some of the platforms proposed in the IoT are presented in Table 4 [23] [24]:

Table 4. some of the platforms proposed in the IOT

Name of platform	URL
AirVantage	https://airvantage.net/
Arkessa	http://www.arkessa.com
ARMmbed	https://www.mbed.com/en/platform/
Brillo Carriots	https://carriots.com
Etherios	http://www.etherios.com
Exosite	https://exosite.com
IoTivity	https://www.iotivity.org
IBM Watson	https://www.ibm.com/internet-of-things
LinkSmart	https://www.linksmart.eu/
OpenRemote	http://www.openremote.com
OpenMTC	https://www.openmtc.org/
OpenIoT	https://www.openiot.eu/
Thethings.io	https://thethings.io
ThingWorx	https://thingworx.com
ThingSpeak,	https://thingspeak.com
Pentaho	https://www.pentaho.com

These platforms mainly support the management of data and device. However, most do not support the management and analysis of big data. Platforms such as AirVantage, Pentaho, and IBM Watson support the management of big data, but do not support learning tools and do not achieve a specific pattern in analyzing actual data. A platform like IBM Watson includes learning tools for data analysis, but there are not many machine learning methods available [23].

In the following, we will investigate and analyze primary studies that are focused on Architecture and Platform in the field of IoT. As mentioned earlier, 45% of primary articles under investigation are focused on the Architecture and 3% are focused on the Platform.

In this article, GOVA [25] is introduced as a data analysis architecture for the IoT. This architecture includes the three sections of the virtual entity, the microservices, and the graph database. The proposed architecture used the "noninvariate data" statement instead of "information", "knowledge", or "insight" statements, and introduced a list of noninvariate data for the smart grid applications. There is a section called multi-temperature data management on this architecture that puts the data in high-speed and low-speed memory, with consideration of the number of data usage. The graph database is also used for storage. This architecture has a storage compatible mechanism. It can be scalable in the areas of computing power, data storage, and system planning. It is designed to analyze simultaneously the history of data and new data concepts in real time.

Rathore et al. [26] implemented the IoT-based intelligent systems to analyze the smart city's data in real time using the architecture that was presented. They installed different sensors in different locations to collect data from smart homes, smart parking, weather and water systems, traffic, environmental populations, and surveillance systems. The architecture of this system has a preprocessing section, which includes the data collection, data filtering, and data categorization, and a data processing section where offline data is processed by MapReduce in the HDFS, and online data is processed by the Spark. Then the analysis section analyzes the data using machine learning and other analysis tools. The results indicate that the system provides effective outcomes even on Big data sets. By increasing the data volume, the throughput system increases. The security issue has not been addressed in this architecture, so Rathore et al presented a different architecture and considered the security subject. They proposed a system based on the four-tier architecture for urban planning and building smart cities on the IoT and the big data analysis in the article [26]. The system has various sensors that include the intelligent home sensor, vehicle network sensor, weather sensor, and parking sensors in the bottom layer, whose responsibility is the resources and the data collection. The other layer is responsible for all types of communication among sensors, stations, and the Internet, and the third layer is responsible for managing data and processing it using the Hadoop framework. The highest layer analyzes the data. This system was created by Hadoop with Spark, VoltDB and Storm for real-time processing of the smart city. The previous offline information is analyzed in Hadoop using MapReduce for urban planning or future development of the city. Ultimately, this system was tested based on effective performance, with consideration of the processing time and efficiency. The system provides effective results even on big data sets. The throughput system increases with increasing data volumes. The proposed system could secure the city via the urban monitoring and the real-time decisions. It was more scalable and efficient than existing systems.

Behera et al. [27] used the cloud on the IoT for storing, as such could be effective in storing and analyzing data generated on IoT. In this architecture, all data collected from the various sensors of IoT subsystem are stored for more processing in the cloud sensor. Useful information is extracted from data stored in the cloud. After the data have been analyzed, the extracted information is stored in the other cloud that will be available to end user or consumer of the sensor information. They used the architecture in the application of smart power system management and used the HDFS in Hadoop and also pig code in data analysis.

Din et al. [28] proposed a big data-based dynamic architecture. This architecture is independent of node hardware, as such, it can perform a wide range of functional activities. The integration unit of this architecture locates the data generated from the same resources in one node and makes a data block after final preprocessing. This aggregation is done based on the identification number of resources. Since the collected information is not in a proper format, it needs to be processed to convert it to a proper format. Then a sequence of data blocks is created. Data blocks are logged into storing system that acts as a decision model, the system filters the real-time data and prepares them for processing. If the data is to be offline, it sends them to the storing server. The storing server is able to share a large amount of data and distribute them equally across the Hadoop processing servers. The final step of this architecture consists of two parts: combining the results of the storage device and the decision server. Although the data segmentation is not required for the small size data, this architecture has a similar function for all data sizes. However, according to this article, each level of the architecture can be changed in various applications depending on the user's needs.

Sun et al. [29] presented an architecture called Tresight for improving the smart tourism in the city of Trento, Italy. Tresight is a content-based recommender system on the basis of FI-WARE technology. Similar to the expressed architecture, it has four layers. The article's authors emphasized on intelligent IoT by analyzing big data, so that they can analyze and make decisions like humans. The authors introduced IoT architecture that has four layers of sensing, responding, interconnecting and data for the smart and connected community (SCC). This paper considered each section of the architecture as equivalent to a part of the human body. For example, the human brain equivalent is the data layer and the human nervous system equivalent is an interconnecting layer. Tresight architecture is similar to this architecture. The sensing/responding layer will gain access to three data sections. These data are static data, including Trento City Attractions and restaurants, and dynamic data that could be received through CrowdSensing using tourists' handcuffs. After processing received data, the data are stored in the MongoDB and analyzed by Hadoop. In this architecture, the analysis section consists of three parts: insight analysis (understanding the depth of data using statistical computations), outsight analysis (understanding of social and external aspects such as climate impact), and foresight analysis (forecasting and protection). In the architecture, users access to their profiles are protected via the site or smartphone by encryption. The service layer presents the offers and services with regard to the content and analysis that have been done.

Sowe et al. [30] proposed the architecture for IoT on big data platform. The purpose of this architecture is the management of services and communications of heterogeneous and complex sensors. The architecture consists of the SCN middleware in the cloud substructure. The SCN middleware organizes the elements of the cloud service, consequently, environmental scientists can discover, manage, and share their knowledge of air pollution or solid particles (PM2.5). The big data platform helps to identify and manage data easily from different sensors and share their knowledge and experience related to the effects of air pollution.

Paul et al. [31] proposed an architecture for Smart Buddy that will be able to receive the useful information from the smart city and smart homes, etc. This architecture has three areas: object, SIoT, and application. On object layer, the various data are collected from heterogeneous devices and sent to SIoT layer server for storing and processing. In SIoT layer, data are preprocessed and distributed among servers, and offline and online data analysis takes place. This architecture used Hadoop for data analysis and MapReduce and Spark for offline and real-time processing. After analyzing, the information is stored in a location. The application layer is divided into four parts: security, cloud server, results in storage devices, and a data server that one can use depending on the user's need.

Rathore et al. [32] designed the 5-layer architecture for the real-time drug response system. The layers of this architecture are collection & filtering layer, communication layer, processing layer, management layer and the service layer. The system's smart core is in processing layer, where data are divided into small chunks. Each chunk is individually processed and their results are stored in a space. This architecture utilized MapReduce and HDFS for data parallel processing. Management layer manages all types of smart output. This layer has a pharmaceutical expert system that analyzes the results stored in processing layer and offers the necessary activity. This architecture used spark for data real-time analysis. Rathore et al. displayed the system's efficiency when faced with offline or real-time data as input and real-time response of the system.

The proposed architecture for monitoring chronic patients and taking care of people, outdoors and indoors, and also related services, is presented by Páez and his colleagues. The most important parts of this architecture are smartphones, message platform, website platform, and a module introduced as Virtual Cloud Carer (VCC) project which is in the field of health information. In this architecture, received information from a person is stored on the smartphone device, and after reaching a specific volume, it is stored in private cloud service, so that the data are processed on time close to real time. The architecture used Riak database which is a distributed and scalable database. This architecture has many and various sensors that can access the latest information from any sensor in the shortest possible time by the query. The authors of this article focused on data storing and did not go into analysis topic [33].

Other researches proposed the architecture for cloud-based health application that protects the two topics of IoT and big data. Perception layer, transfer layer, and the big health cloud service layer are the layers of this architecture. The big health cloud service layer is divided into two sublayers which are the cloud service and the cloud service application. The main task of the cloud service layer is to collect physical data from the physical world and the human data that are compressed, formatted, stored and analyzed. This layer is the main protector of the service and the health application software on IoT. The layer of cloud service application is an interface between public health network and users that directly show the economic and social estimates generated by IoT. This layer is responsible for controlling all sensor nodes and visual representation of data and business flow [34].

Souza et al. [35] used big data and IoT to identify outliers. They created a layer in the LinkSmart middleware, which performs categorization, identification, and estimation using pattern recognition functions. They identified outlier using HDFS and MapReduce along with the K-means algorithm in the Mahout library.

Suciu et al. [36] used the M2M remote telemetry architecture with big data platform for the viticulture. This architecture is used for Radio communications, convenient installation, remote control, and long-life. It is a web-based and low-power system. The system consists of three parts: M2M, CloudIoT, and big data processing. The CloudIoT section uses Slapos and the distributed cloud computing system is based on master/slave architecture. The Exalead cloud view is used to collect information from all sources and formats and to process big data. After installing this system on the grape farm and analyzing, it could successfully predict the disease of the farm, as well as improve the product status. Some of the authors of this article examined the existing components and the methods of securing big data processing by Remote Telemetry-based M2M cloud systems in another paper. They presented their architecture, which they had suggested in the electronic health application. This method used the Exalead cloud view as a search platform that provides access to existing information on the infrastructure level for applications based on online search and on an enterprise level [37], [38].

Cecchinel et al. [39] presented a complete software architecture that protects all data generated by sensors on IoT. In this architecture, the data are collected from sensors and directed to storage centers. The architecture has tried to consider data storing challenges, sensors heterogeneity, and avoid bottlenecks processing and high throughput. Cecchinel and his colleagues implemented the proposed architecture for smart campus projects whose purpose was equipping the university campus using information on university environment applications. This paper discusses two subjects of the parking space and the setting up of the heating and cooling systems. In this architecture, the sensor board collects the data from sensors connected to it and sends them to the bridge connected to it. The bridge collects the data streams of various boards. Then the bridge is connected to the Internet and distributes incoming streams to reach receiver API. Receiver middleware has three different APIs (API for received data from Bridge, Configuration API to support reading rate, and data APIs), and its task is to support the received data and distribute the configuration mode for sensors related to bridge. This middleware includes the database and the global sensor configuration.

Kholod et al. [7] presented the two different architecture of IoT for data mining system. This architecture includes the perception layer, gateway layer, network layer, middleware layer, and the application layer. In the centralized mode, the two layers of data gathering and data processing are located in the middleware layer, and the data mining layer is in the application layer. In this architecture, the cloud is used to store data. The distributed architecture uses the cloud and the Fog node for closing a part of computations to the data source. In this architecture, the data layer is located in the network layer, the data processing layer is located in the middleware layer, which performs storage by cloud, and the data mining layer is in the application layer. In this paper, the data mining algorithm is written in functional blocks on the actor model. Actor model was designed to provide parallel and distributed computing. The parallel data mining algorithm is implemented on a set of actors with the ability to exchange messages with the main actor. Actors can be distributed and run in the cloud and fog nodes. The results which were presented in this paper show that the runtime algorithm in IoT with a distributed architecture based on the actor model when faced with big data set was faster than the centralized architecture type because it does not take time to load data in centralized storage space. Moreover, the closeness of computations to the source increase system performance.

Nigam et al. [40] presented an architecture based on IoT and data mining. This architecture can be used for the smart advertising system in stores and in different places. This system works based on data mining and Sales Trend and a data list that is presented by the POS device in real time. The purpose of this system is to give the offers and deals better with the customers. This architecture includes sales details, database, inventory, the network between devices, board IoT, and billboard. First, the purchase information is collected via the RFID Reading Module and is saved as billing information on the database. The list of products with their ID, the price of the specific customer ID, and other information are stored in the database. Then, using the association technique and the Apirior algorithm in sales data, they found the patterns in the purchase of certain products, and appropriate products are suggested by the system for advertising and displaying on a billboard.

To gain a unified and comprehensive information system, Tracey et al. [41] presented a set of requirements for SNs and associated services. They proposed an architecture that is called holistic because it examines the data stream from sensors to services. This architecture presents a set of the abstractions for various types of services and sensors. It was designed for implementing a resource-limited node and can be expanded for server environments. In this article, a ‘C’ implementation of the core architecture is proposed which includes services on Linux and Contiki (using the Constrained Application Protocol (CoAP)) and a Linux service to unify with the Hadoop HBase data repository.

Kim et al. [42] presented a platform called SWOAP based on the SWO architecture. The SWO's architecture consists of three layers of SWOGW that are responsible for communicating with devices, SWOMP which is a protocol collects information via the Web, and SWOAP is a home analysis service that combines device information, user data, flow data, and physical space information. The Kim’s provided platform consists of two layers. The resource core layer collects smart devices and common devices information via the management platform. Each domain resource is created by core resource layer and the communications between resources are built. The Knowledge Framework layer performs home data analysis based on the concept of sources. This section used descriptive models by PMML for analyzing and data mining. The extracted data is stored in the knowledge map, which has good information about home and residents. This platform has been tested on the application of discovering food habituation by analyzing with the cooking device.

The COT platform was presented by Lee et al. [43]. The COT gate is installed on devices that process information in specific fields. Before sending them to a server in order to be analyzed, preprocessing which consists of message filtering and collecting is done. In addition, the edge is responsible for collecting the necessary criteria from the COT gate. This platform has COT server, analysis server, COT scheduler, Task manager, and Edge analysis sections. It has been performed on video image analysis application. Lee et al. showed that less bandwidth will be used for this application.

Lee et al. [44] proposed an intelligent content platform for the big data analysis on industrial applications. The platform has five layers of IoT, infrastructure, data, analytical, and presentation. This is used to examine the information related to the industry in the field of location and unstructured data and sensor signals for data mining of big data. This paper provides using HDFS for industrial applications in the data layer and fuzzy for analyzing and extracting data.

We thoroughly investigated and analyzed the primary studies in the field of big data & IoT and data mining & IoT that focused on the architecture and platform. Our observations are summarised in Tables 5 and 6. The analysis in Table 5 consists of the Name section, Main Idea, Advantages, and Disadvantages.

Table 5. An overview of existing primary studies focusing on the architecture and Platform

#	Architecture Name	Section Name	Main Idea	Advantages	Disadvantages
1	GOVA [45]	Usr_API Repositories Message substructure Service layer Data stream layer	- Introducing the statement on invariant - Big data analytics architecture for smart grid	- Horizontal scalability - Scalable in storage and system planning - Use of the graph database	- Negative impact on the flexibility of the graph database - Data Compatibility should be maintained to have a significant impact on performance
2	Rathore Architecture et al. [25]	Data preprocessing Data processing Data analysis	Proposed and developed a smart city system based on IoT using big data analytics	-Scalable (throughput and processing time) - Efficient	No testing the accuracy of the system, considering security issues
3	IV-Tier Architecture [26]	- Responsible for resources and downloads -Responsible for communications -Data management and data analysis processing	Propose a combined IoT-based system for smart city development and urban planning using big data analytics	-Security is achieved by the proposed system by continuous monitoring the video of the whole city - Scalable (throughput and processing time) - Efficient	Comparison with non-Hadoop systems has not done.
4	Behera Architecture [27]	- Received the cloud data -Analysis the cloud data	Generalized architecture which uses Cloud Computing and big data for effective storage and analysis of data generated	• Built-in cloud computing • Efficient data management • Reduction of data acquisition cost • Innovations • Applications Real-time data for decision making and policy making	In this architecture, the data security issue is not considered.
5	Din Architecture [28]	Obtained preprocessing and aggregation of decision processing storage	- Provides a combination of data architectures for big data	-Ability to collect data -Use of parallel processing of the Hadoop	The use of the Hadoop is suitable for high-throughput applications. This issue is not

					considered in this architecture.
6	Tre Sight Architecture [29]	Sensing/responding layer Interconnecting layer Data layer Service layer	Recommender Content-based system on the basis of FI-WARE technology	Cost optimization Intelligent IoT	Not mentioned in the relevant article
7	Sowe Architecture [30]	Service overlay *Declarative Service Networking* *Network Control* *Protocol Stack*	An integrated IoT architecture that combines the functionalities Of service-controlled networking (SCN) with cloud computing	The management of services and communications of heterogeneous sensors	Does not cover mobile sensing.
8	Smart buddy Architecture [31]	Object domain SIoT domain Application domain	This article has envisioned the fundamental Role of integrating IoT with social networking in achieving human dynamics based on big Data analytics.	This work analyzes the interaction of the human with the electronic appliances in the IoT environment and social networks from the perspective of human dynamics	there is no parallel processing of the data, which results in minimum throughout.
9	Rathore Architecture [32]	Collection & filtration layer Communication layer Processing layer management layer Service layer	Propose a Real-time Medical Emergency Response System That involves IoT-based medical sensors deployed on the human Body	Perform real-time action in necessary cases	Using the Hadoop is not suitable for low data volume.
10	VCC Architecture [33]	-Smart Mobile Devices - Message platform - Web Site platform -A module for health information	Propose one potential solution for creating those new e-Health services, based on big data processing and IoT concepts.	Use of Riak database	In this article, there is no discussion about analytical manner.
11	Health System Architecture [34]	Perception layer Transmission layer Cloud service layer	The big health application system based on the health Internet of Things and big data	This paper presents the cloud to end fusion big health application system architecture which is supported by health IoT and Big data	Not mentioned in the relevant article
12	Souza Architecture [35]	Network layer Service layer Pattern layer	Use of big data and IoT FOR outlier identification	Scalability contextualization Flexibility	This architecture can be applied to a larger dataset.

		Semantic layer Security layer			
13	Architecture for viticulture of Suciu [36]	The architecture of M2M Cloud IoT Big data processing	Use of the CloudIoT for big data preprocessing and Warning on grapes growing	-Use of M2M Advantages -Early diagnosis of disease in grape farm	The big data analysis is not intelligent and is based on IFs.
14	Architecture for E-health of Suciu [37] [38]	The architecture of M2M Cloud IoT Big data processing	Propose a converged E-Health architecture built on Exalead CloudView, a search based application	-Energy efficiency -Scalability -Providing nonfunctional requirements	Not mentioned in the relevant article
15	Cecchinel Architecture [39]	Sensor Sensor board Middleware	Software architecture supporting the collection of sensor-based data in the context of the IoT	The key point is to develop Software application on top of these datasets to support the base scenarios, and open the datasets to the users to let them create their own services	Not mentioned in the relevant article
16	Kholod Architecture [7]	-Centralized architecture -Distributed Architecture -5 main layers Perception layer ، access gateway layer ،network layer ، middleware layer ، Application layer	Running of distributed data mining algorithm	-Use cloud and fog -Increasing the efficiency of the IoT system -Increase processing speed in distributed mode	Not mentioned in the relevant article
17	The intelligent billboard Architecture [40]	Sales detail Database Inventory Network IoT board Billboard	The design of Intelligent advertising system	This system is not targeting individual user the feeling of privacy invasion among people will not be present	Not mentioned in the relevant article
18	Tracey et al. [41]	- Service Abstractions and Data Model Service Layer - The Object Space Layer - Local Instrumentation Layer - The Holistic P2P Protocol (HPP) and	Proposes a set of requirements for achieving a pervasive, the integrated information system of WSNs and associated services	Integration with Contiki and HBase to store sensor data, requiring only simple message reformats without requiring semantic changes or application proxies in an infrastructure of nodes and services.	Not mentioned in the relevant article

		Hpp Channel			
	Platform Name	Number & Section Name	Main Idea	Advantages	Disadvantages
19	SWOAP Platform [42]	Two layers Resource core layer Knowledge framework layer	Smart home Web of Objects Analytic Platform (SWOAP) based on SWO (Smart home Web of Objects) architecture	Data expendability Data relation Multi-user	From the knowledge map, we can extract useful information and use it to improve.
20	COT [46]	Five sections COT server Analysis Server COT Scheduler Task Manager Edge Analyze	Cooperative big data processing engine in the Internet of Things environment	parallelization of Processing Use the edge	We can find suitable algorithms for collective intelligence between the edges.
21	Industrial platform [44]	Five layers IoT layer Infrastructure layer Data layer Analytical layer Presentation layer	Provide a comprehensive solution for industry through research and Development of an Internet of Things (IoT) based Cyber Physical System for Industrial Informatics Analytics.	Use of fuzzy rule set for analyzing industrial information	Not mentioned in the relevant article

Table 6 compares the architecture and platform expressed in terms of the year of publication, their tool, and method, experimental type to find whether the intended plan is numerical analysis or implementation or programming or simulation or design, the application that is presented in the article and taxonomy that is shown in Figures 3 and 4.

Table 6. An overview of existing primary studies focusing on the architecture and platform

#	Name	Year	Article Type	Tool & Method	Experimental Type	Application	big data
							data mining
1	GOVA [45]	2017	Conference	OOP Spark GraphX	Design	Smart grid	St
							--
2	Rathore Architectur e et al. [25]	2016	Conference	HDFS MapReduce Spark	Implementati on	Smart City	St-pr-An
							Pre-Cl
3	IV-Tier Architectur e [26]	2016	Journal	Spark HDFS MapReduce	Implementati on	Urban planning and building smart cities	Pr-An
							Pre-Cl
4	Behera Architectur	2015	Conference	HDFS MapReduce	Design	Smart Power Management	St-Tr-Pr-An

	e [27]			Hive Pig			Pre
5	Din Architecture [28]	2015	Conference	Hadoop MapReduce	Implementation	Health Care	St-An --
6	Tre Sight Architecture [29]	2015	Journal	FI-WARE Technology Hadoop MongoDB	Implementation	Smart Tourism	An Pre
7	Sowe Architecture [30]	2014	Conference	SpatIoTemporal Information Clustering and Knowledge ExtRaction	Implementation	Air Pollution	Pr-An --
8	Smart buddy Architecture [31]	2016	Journal	Hadoop MapReduce Spark	Implementation	Smart Buddy	Pr-An Pre-Cl
9	Rathore Architecture [32]	2016	Journal	MapReduce HDFS Spark	Implementation	Real-time Medical Emergency Response System	St-Pr-An Cl
10	VCC Architecture [33]	2014	Conference	Riak	Design	Health Care	St-Pr --
11	Health System Architecture [34]	2017	Journal	cloud	Design	Health Care	St-Pr-An --
12	Souza Architecture [35]	2015	Journal	Mapreduce HDFS Mahout K-means	Implementation	Outlier detection	St-Pr-An Cl
13	Suciu Architecture [36]	2015	Conference	Slapos Exaleal cloud view	Implementation	Viticulture	St-Pr --
14	Suciu Architecture [37]	2015	Journal	Slapos Exaleal cloud view	Implementation	E-Health	St-Pr --
15	Cecchinel Architecture	2014	Conference	Raspberry Pi Amazon cloud Mongo DB Json	Implementation	Smart Campus	St-Pr --

	[39]						
16	Kholod Architecture [7]	2016	Conference	DXelopes K-means	Implementation	On data from Time Series, Telescope, Breast Cancer Info, Weather, Breast, and Cancer Features	St-Pr-An Cl
17	Intelligent billboard Architecture [40]	2015	Conference	Raspberry piB+ v1. 2 association technique Apirior algorithm	Implementation	Intelligent billboard	An Des
18	Tracey et al. [41]	2013	Conference	'C' Contiki Hbase	Implementation	Not mentioned a specific application in the article	St-Pr-An --
	Name	Publication Year	Article Type	Method	Experimental Type	Application	
19	SWOAP Platform [42]	2015	Conference	PMML Clustering Association rule	Implementation	Eating habit	An Pre
20	COT [46]	2014	Conference	Raspberry Pi OpenCV	Implementation	Face detection	An Pre
21	Industrial Platform [44]	2015	Conference	Fuzzy FCM	Implementation	Industry	An Cl

St: storage, Tr: transmission, Pr: processing, An: Analyzing, Des: descriptive, Pre: predictive, Cl: Classification

We can also classify the studied papers related to architecture according to the taxonomy presented in Figures 3 and 4. This classification is done in two areas of big data and data mining separately. The major focus of big data's articles was on transferring, storing, processing, and analyzing data and sometimes covered all cases. Data mining also appears in the papers of big data when the author intends to analyze the data. Therefore, we categorize such articles with data mining's articles. The categorization of the big data's taxonomy is presented in Figure 10.

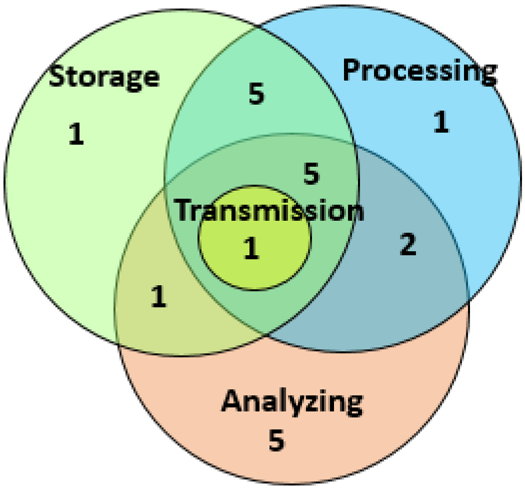


Fig.10. The categorization of papers according to the taxonomy of Figure 3

By reviewing the articles related to architecture, we found that 12 papers used data mining method for analysis. The number of these articles is in accordance with Figure 11 and proportional to the taxonomy presented in Figure 4.

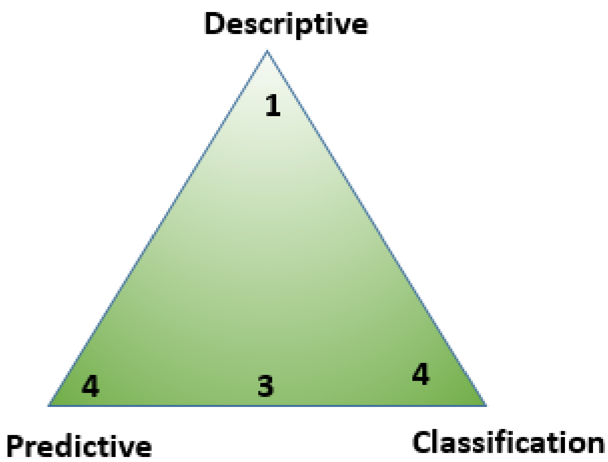


Fig.11. The categorization of papers according to the taxonomy of Figure 4

3-2. Primary Studies of Framework

In this subsection, we will investigate and analyze the primary studies that focus on the framework in the field of IoT. As mentioned earlier, we found that 23% of primary articles under investigation focused on the framework.

Ahab is a framework [47] that processes database on the cloud as scalable and error tolerance. It is a cloud-based framework for the processing of data streams, and suitable for online and offline analysis. This framework using Usr API is able to direct the different data streams from various sources. Data streams, management policies, and actions have the repository of the component, repository of politics, a repository of action, and data stream repository to store the components. Ahab uses the distributed message structure to direct various data streams. The method minimizes the unnecessary traffic of the network and allows the components of the system to select an appropriate communication point freely. It uses the pub/sub mechanism to send and receive data so that it spreads the generated components of data in message substructure to reach the specific routing key. This key includes the component name and type of the published data stream. Ahab architecture is divided into two layers. The data stream which is in the bottom layer directs processed data stream and is implemented as Lambda architecture. The service layer is in the upper layer and is a basis for data stream layer and care for the components of analysis and management. Lambda architecture uses batch processing to provide the comprehensive and accurate views of batch data, whereas synchronizing and running data stream gives a real-time view. This layer includes the sublayers: batch layer, speed layer, and service layer that were proposed by Lambda architecture. First, all data that entered Ahab via the message system are distributed between two layers of the batch and the speed. The batch layer stores the data on HDFS and the data generated from one source are grouped. Then, the batch layer calculates views using all available data on HDFS. The service layer indexes the generated views by the batch layer in order to provide case queries. To compensate for high latency of the service layer, a speed layer can be used. This layer only operates on recent data. The speed layer provides online processing of data streams so that Ahab generates results almost immediately after receiving. Of course, the results may not be accurate or complete as the results of batch layer. The results can be improved by combining batch and real-time views on the service layer. Vögler et al. used this framework in the smart city application. This framework is allowed for online analysis and pattern learning from stored data.

The IBDA framework [48] was created to store and analyze the real-time data generated by the IoT sensors. The Python language and Cloudera platform developed the first version of IBDA. This framework allows the real-time control of the environment by using real-time analysis of data. The scope of IBDA is limited to data generation, data extraction, data storage on HDFS, data visualization, data analysis, and real-time control of intelligent buildings. The framework has three components of the IoT, big data management, and big data analysis. In this study, instead of using different Physical sensors for intelligent building, fifteen virtual sensors are created in different situations via Python code. These sensors produce a large amount of data sent by the Apache Flume to the TCP port and stored. Pyspark is used to analyze data. It analyzes data stored in the HDFS, which includes three sensors. If the numerical range of sensors is within the predefined range, then no action is taken.

Berlian et al. [49] proposed a framework to combine the IoT and big data for the intelligent underwater environment. They used remote devices, portable sensors, and underwater cameras to extract important water parameters to prevent coral bleaching. The system has five subsystems of reception, portable monitoring, coral surveillance, Wireless Mess Network Access Point, and big data analysis. The section of big data analysis includes the database layers, the server application, and user applications. The Hadoop and Hive, metastore, and PostgreSQL are used for the database layer. The server application includes the Oauth2 server, the Web interface, and the Restful API. The client sends data from sensors to this subsystem. For implementing such a system, the sensors are simulated by the Python program. The program gains sensor's data as a sample per second, and stores the data in a CSV file. The files are stored in HDFS and can be directly queried by Hive. In this architecture, data are not directly added to Hive. However, because the Hive performs batch processing, the data is stored in the JSON format in the Redis database. In this study, a comparison between the machine learning statistics was also performed and the time of executing the Bayesian and decision tree with different inputs was compared.

The IoT-StatisticDB framework [50]on a large scale of statistical analysis IoT systems can be a viintol method of transforming data to knowledge and obtaining general information about the physical environment. Most available statistical analyses for sensor data are designed out of the database core and are not suitable for the IoT environment, which include the data types and various statistical queries. In this research, a clustering mechanism for big data is presented. The mechanism performs statistical functions via the statistical operators, alongside the database core. Therefore, statistical queries can be applied in the SQL standard format, and statistical analysis can be implemented in parallel and distributed on multiple servers and improve performance. To design such a system, Ding et al. categorized those that are more common among the complex statistical analysis operators and implemented the same operators for them. They divided the statistical operators into two categories of location group and parametric group. A location group in a dataset can cluster the geographical values extracted from the sensor data into a larger group, which means that, it extracts knowledge in the range of location distribution. The parametric group also allows obtaining of information about the average, minimum and maximum, and counting the sampled values of sensors. In addition, both groups of parameter and location are also divided into two groups of Euclid-based and network-based. The Euclid-based group does not consider the network traffic, whereas the network-based group considers it. This framework has three layers of monitoring and sensor device, IoT raw data storage layer, and statistical layer. Din et al. tested the proposed framework on two applications of collecting data from GPS, 20,000 taxis in Beijing, and sampled sequential data of 200,000 static sensors generated via the simulator. Finally, they compared the response time of the query with statistical methods that work outside of the database core. Using this framework, a query can be performed in parallel and it can use SQL commands to query and perform four types of statistical analysis. This framework can be improved by considering data mining and event detection.

Mishra et al. [46] introduced a cognitive framework based on big data for industrial applications. This framework receives raw data from different sources, and after clearing the data by their quality management, the data are devoted to several clusters based on the behavior and data properties by big data classifier. This big data can be efficiently scalable. The discussion of big data analysis is an important aspect of this research and is done through cognitive and computational tools. For this research to be scalable and real-time, the various sections must be synchronized with each other.

Guo et al. [51] introduced a framework that uses semantic information for media data. It uses the shared knowledge of social users by CSF. CSF introduces crowdsourcing computing for semantic merge. This framework consists of three parts of meaning extraction, a combination of multi-model meanings, and the distribution and storage of meaning. It is an efficient combination plan and a distributed method to fully ensure sharing of semantic information when distributed social media data have been distributed. Guo et al. by examining this framework, showed that the design could have high precision and performance, moreover, it reduces the semantic retrieval problems related to social media information. HBase is used to store semantic information and ensure synchronization. In this method, the data are converted into a key-value format and uses spark to process MapReduce and parallel memory calculations. In this method, the CPU reads data from memory and performs computation and analysis in real time so as to avoid the exchange of information between application software, servers, and hardware and reduces the network conflicts with the input-output device.

The purpose of Dundar et al. [52] was to develop a framework for big data processing based on the necessary fields in IoT. This framework for monitoring and processing of received real-time events includes a processing platform of the distributed real-time data stream. Hence, the proposed framework includes the Provenance service, the Pub/Sub message system, the distributed real-time computing system, and the CEP engine. It also interacts with the self-healing service and predictive maintenance service. The processing of great real events with expressed capabilities does not cost much, but in this paper, the performance of data processing under the increasing load of the system has not been evaluated, and the most emphasis is on self-healing.

Sezer et al. [23] used the combination of big data, the IoT, and semantic web technology to build a framework. Most frameworks and semantic platforms use the existing ontologies, such as SSN and GOname, to solve interoperability problems between sensors and operators. This paper used the neural network algorithm in machine learning to obtain better results and hidden values. In this framework, IoT data are converted into semantic data. The semantic web describes IoT area using standard protocols and vocabularies. This framework has five layers of data acquisition, extraction-transfer-loading (ETL), law-meaning argumentation, learning, and bottom-up practical layer. The task of the first layer is to transmit information to the ETL layer and not to lose it, and for this to be guaranteed, it uses the multi-threaded queue. The ETL layer receives the data and converts them to correct format, and then the data is converted to semantic format via The RDF form. The ontology format of this article specifies the type and size of the data unit. The law - meaning layer checks the rules according to existing semantic information and determines the actions. The learning layer preprocesses the data, extracts the features, and applies learning methods. This research has used NoSQL database to increase scalability and response speed.

Mocanu et al. [53] presented the new IoT framework for performing simultaneously, identification and real-time prediction of flexibility, using deep learning methods of Factored Fourway Conditional Restricted Boltzmann Machines (FW-CRBMsFFW-CRBMs) and Disjunctive FFW-CRBM. The proposed framework was evaluated on energy data sets of ten buildings equipped with smart meters. In this assessment, five building data were used to make the model. The data of the rest buildings were considered as test data. After making the model, it can be used for data buildings, which do not have a smart meter to identify and predict their energy flexibility. This information is used to make a decision on the smart grid or can be a real-time feedback for the building. This framework had a proper performance in the energy breakdown, identification, and prediction of flexibility simultaneously (i.e., Estimation of energy usage real time and time usage of flexible tools).

We thoroughly investigated and analyzed the primary studies in the field of the big data & IoT and data mining & IoT that focused on the framework. Our observations are summarised in Tables 7 and 8. The analysis in Table 7 consists of the Number and Name Layer, and Main idea, Advantages, and Disadvantages sections.

Table 7. An overview of existing primary studies focusing on framework

#	Framework Name	Number & Name Layer	Main Idea	Advantages	Disadvantages
1	Ahab [54]	Four layers Usr_API Repositories Message substructure Service layer Data stream layer	Present Ahab, a generic, scalable, and fault-tolerant data processing framework based on the cloud	-Online and offline analysis -Scalable -Error tolerated	Not yet fully developed and not responsive to machine learning techniques.
2	IBDA [48]	Four layers Ingest Store Analyze Actuate	Proposed IBDA framework for the storage and analysis of real-time data generated from IoT sensors deployed inside the smart building	-Continuous monitoring of stored data	Not mentioned in the relevant article
3	ISES [49]	-Five layers -Data Receiving -Portable Monitoring System -Monitoring system Marjan Wireless Mesh Network -Access Point -big data Analysis	Discuss an integrated smart environment system framework based on IoT and big data that consist of open a platform that processes data from Remotely Operated Vehicle (ROV) and portable sensor with water parameters sensors	-Use Redis instead of storing in the Hive table -More Speed Response Queries Compared to RDBMS	Batch processing of data due to the use of Hive
4	IoT-Static DB [50]	Three layers Layer of monitoring device and sensor	Propose a General Statistical Database Cluster	-Using statistical commands in SOL -Improved performance and reduced response	Limited to implemented commands.

		IoT Raw Data Storage Layer IoT storage layer and statistics	Mechanism for big data Analysis in the Internet of Things)IoT-StatisticDB)	time to the query	
5	COIB Framework [46]	Five layers Application layer Knowledge processing basis Data Layer Transfer layer Receive layer	Propose a Cognitive Oriented IoT Big-data Framework	-Computational and cognitive data analysis -Scalable In real time Storing all access paths to the node	This framework is for industrial applications.
6	CSF Framework [51]	Three layers Extracting Meaning Combining Meaning Distribute and save meaning	Proposes a solution called CSF that makes full use of the collective wisdom of social users and introduces crowdsourcing computing to semantic fusion.	-Efficiency and high precision -A Unified Method for Multi-Modal Meaning Social semantic distribution Decrease semantic retrieval	The average accuracy provided in the article is low and also decreases with increasing number of sample instances. The input of this framework is based on the results of social users and has not used machine learning methods.
7	CEP&Self healing Framework [52]	Four layers Provenance service Message system Real-time calculation system CEP engine	Big data processing The framework that provides self-healing capability on the Internet of Things domain	Self-healing ability Event processing service	The performance of the system during a particular load investigated but performance did not evaluate for increasing loading.
8	SBDA Framework [23]	Five layers Data acquisition Extraction-Transfer -Loading argument-law-meaning Learning Practical layer down to up	Propose a combined framework that brings big data, IoT, and semantic web together to build an augmented framework	Use Semantic Web Use the learning tool Data stream processing	Not implemented on an application.
9	Mocanu et al. [53]	FFW-CRBM framework	Propose a hybrid	Identify and anticipate	Not mentioned in the relevant article

		DFFW-CRBM framework decision Flexibility	approach, which combines sparse smart meters with machine learning methods	simultaneously	

Table 8 compares the framework expressed in terms of the year of publication, their tool and method, experimental type to find whether the intended plan is numerical analysis or implementation or programming or simulation or design, the application that is presented in the article, and taxonomy that is shown in Figures 3 and 4.

Table 8. An overview of existing primary studies focusing on the framework

#	Framework Name	Publication Year	Article type	Tool & Method	Experimental Type	Application	big data	data mining
1	Ahab [54]	2017	Journal	Lambda HDFS Spark Architecture	Implementation	Traffic and transfer system of smart city	St-Pr-An	--
2	IBDA [48]	2016	Conference	Cloudera's platform Flume HDFS pySpark	Implementation	Smart building	St-Tr-Pr-An	--
3	ISES [49]	2016	Conference	PostgreSQL Redis HDFS Hive MapReduce	Implementation	Intelligent underwater environment	St-Pr	--
4	IoT-Static DB [50]	2013	Conference	PostgreSQL Clustering Casandra or MongoDB	Implementation	Beijing taxi Routing Simulation environment	An	Des
5	COIB Framework [46]	2014	Conference	HBase Clustering	Implementation	Industrial application	St-An	--
6	CSF Framework [51]	2017	Journal	HBase MapReduce Spark	Implementation	Social media	St-Pr	--
7	CEP&Self healing Framework [52]	2016	Conference	Kafka Storm PROV-O Contiki Coojaney	Implementation	Cyber Physical System	Tr-Pr	--
8	SBDA Framework [23]	2016	Conference	SPARK MLIB در SPARK	Implementation	Not mentioned a specific application in the article	St-Pr-An	Des
9	Mocanu et al. [53]	2016	Conference	Matlab	Implementation	Smart building Smart grid	An	Pre-Cl

St: storage, Tr: Transmission, Pr: processing, An: Analyzing, Des: descriptive, Pre: predictive, Cl: Classification

We can also classify the studied papers related to framework according to the taxonomy presented in Figures 3 and 4. This classification is done in two areas of big data and data mining separately. The major focus of big data's articles was on transferring, storing, processing, and analyzing data and sometimes covered all cases. Data mining also appears in the papers of big data, when the author intends to analyze the data. Therefore, we categorize such articles with data mining's articles. The categorization of the big data taxonomy is shown in Figure 12.

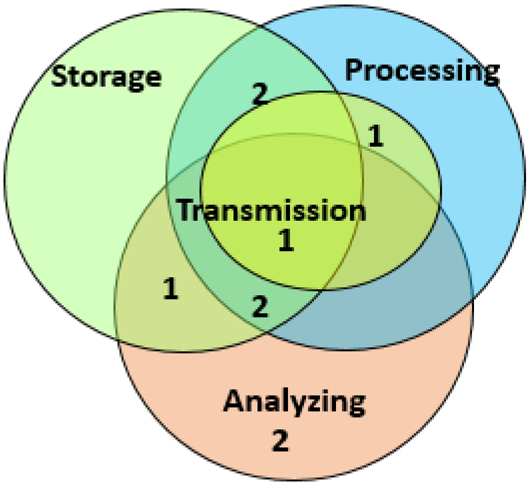


Fig.12. The categorization of papers according to the taxonomy of Figure 3

By reviewing the articles related to the framework, we found that thirteen papers used data mining method. The number of these articles is in accordance with Figure 13 and proportional to the taxonomy presented in Figure 4.

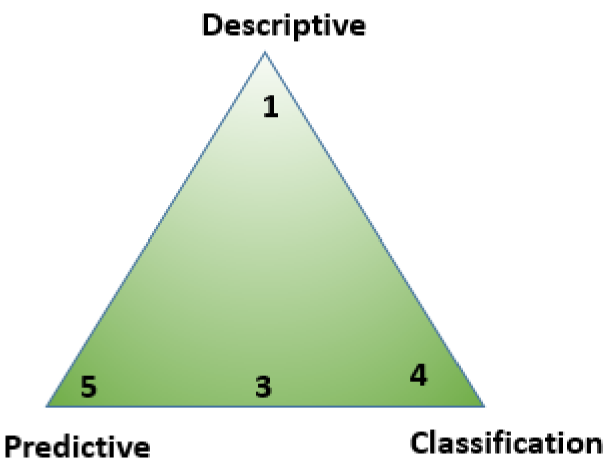


Fig .13. The categorization of papers according to the taxonomy of Figure 4

3-3. Primary Studies of Applications

In this subsection, we will investigate and analyze primary studies that focus on applications in the field of IoT. As mentioned earlier, we found that 25% of primary articles under investigation focused on the applications.

Rizwan et al. [55] used the big data and IoT for intelligent traffic, which is a feature of the smart city. This research presented intelligent real-time and low-cost traffic management and used the IoT for the acquisition of traffic data via sensors installed on the roads and sending traffic data with software installed on the smartphones of passengers. The real-time data stream is sent to the big data analysis section, and in that section, traffic is checked in different locations through the sent information and the created mathematical model. A smartphone application is developed as a user interface for discovering traffic congestion in different locations and provides a suggested method for traffic management.

Khorshed et al. first examined the security of three-layer models and expressed challenges, security issues, and technologies. Then, they tried to attack the three-layer architecture, including data collection layer, receiving layer, and analysis layer. In this research, we tried to use object layer, cloud layer, and the big data layer in the data collection layer. Big data layer tried to reside the Hadoop ecosystem in the cloud. Then, with a variety of experiments on different hardware, we collected performance data from all the three layers, combined them, and finally used the machine learning algorithms to detect 18 different activities and cyber-attacks. The research shows that the Random Forest algorithm can identify 93.9% of attacks and activities in a complex environment. The article states that nobody has ever done anything about cyber-attacks detection. [6]

This system extracts information in a specific time frame. Data trackers or interfaces trace information directly from big data's sources. Data tracker transmits data to the data controllers. Controllers process each data cluster and make them smaller by omitting redundancy. Data controllers process the big data clusters and send them via secure and hidden channels with confidentiality preservation. The proposed framework of this method includes the sections of dynamic query manufacturer, data retrieval via multiple interfaces, two-level data analysis, and data transfer via a secure channel [56].

Embedded sensor networks can produce a significant amount of data. The data must be transferred or sent to the remote cloud server after collection. Since the sensors have limited capacity, they must discard old data or stop producing new products so that old data can be depleted. Mozumdar et al. stated that the data could reach 1GB in a week. A limited capacity orchestrator can handle the space required for a small sensor network for a short period of time. Therefore, it requires data and storage capacity management for large-scale applications. By using the Rame-Douglar-Pevcker compression algorithm on sensor data, we achieved as high as 99.86% of sensor data compression [57].

The repository design based on MongoDB using the embedded plan proposed in [58] for the RFID/Sensor big data is generated from a simulated environment of automotive-parts of the supply chain. The authors developed the event types in the Electronic Product Code Information Standard (EPCIS) and presented compound shard key with the goal of maximizing query speed and the distribution of the same data on the data servers. In this paper, the compound shard key is used for horizontal partitioning and steady distribution of data between data servers. In this method, if a large amount of data are created, the writing will be disturbed. This is a defect of the embedded model. The proposed repository design protects just four event types of EPCIS standard and the query requirements did not satisfy except for any simple query. It can be enhanced to support shard extension.

Niyato et al. [59] began a primary study in the economic analysis of the big data and IoT systems and presented a plan to optimise the pricing on the big data Market model. The big data market model consists of three sections: data source, service provider, and service customers. The data source layer is responsible for collecting data. Data processing, machine learning, and data mining algorithms are performed in the service provider layer. The service offer is performed in the next layer. Ultimately, the consumers share the service proposed by the provider and pay the subscription fee that is specified by the service provider. The service provider purchases the raw data from the data sources. The provider attempts to adjust the price and amount of requested information in order to maximize profits.

Rathore et al. [60] proposed a graph-based intelligent transfer using the combination of big data and IoT. In this application, road sensors are used to obtain all road information and the vehicle network is used to get the location and speed information on vehicles. For this purpose, Giraph and Spark have been used to increase efficiency.

Koo et al. [61] used big data and IoT for water supply. The data are collected by the sensors installed on each device, and transmitted via the internet network system. The IoT application software includes intelligent sensors that collect water data and retrieve any useful information to enhance stability through the virtual data platform and control system.

Dineshkumar et al. [62] stored the information received from sensors in the cloud and used this data that includes all patient information to control the health of patients. In this research, the analysis is performed inside the cloud by HDFS and MapReduce. The proposed system has different sensors (finger pressure, heartbeat rate) than other systems and performs real-time analysis.

Saenko et al. [63] introduced a parallel data processing system for security monitoring of IoT. Hence, a large amount of information about the security events that are received from devices is recorded on a file. The sources of security event create a large amount of heterogeneous traffic that can be called big data. The system has five functional components of data collection component, data storage component, normalization component, analysis component, and data visualization component. The system used HDFS to store data. The advantage of the system is proper scalability and data processing in real time for security events. But in this paper, HDFS is installed on the same hards, so the speed of Hadoop is low to access node at the same location on hard.

Tang et al. [64] designed a complete system for reliability management that performs databases management, online test requests, and test data analysis. This system is based on the browser/server framework which does not require any operating system or remote hardware. It is necessary just for communication. The users work via the web browser. All data are stored on a server, while the algorithm runs on the server. This framework has three layers of the user interface layer, business logic layer, and data access layer, which work with the MYSQL database and law-based algorithms. The rules are discovered from studying of test data. First, the data are collected from the test and streams tools including test requests, process data, control data, and raw data, and after analyzing all of these data records using these rules, RMIS can decide on the priority of each item and recommend the appropriate tools to the technical specialist.

Alam et al. [65] examined and compared the use of well-known data mining algorithms on the IoT. They simulated and compared the eight data mining algorithms consisting of Support Vector Machine (SVM), K-Neighbor (KNN), Linear Discriminant Analysis (LDA), Naïve Bayes (NB), C4.5, C5.0, Artificial Neural Network (ANN), and Deep Learning ANN (DLANN) using three sensor data sets in the UCI site in the R software. The results reveal that C4.5 and C5.0 have good accuracy and efficient memory and higher processing speed, and ANN and DLANN also have high accuracy but have high computational complexity too.

Moreno-Cano et al. [66] in an article investigated the analysis of big data benefits and discovering the knowledge of sensed data in IoT area. They proposed two applications of the smart city. The first application was the intelligent energy management in Murcia University campus and the second application was tramway services and the acquisition of a displacement pattern using the big data techniques. The results of applying complex event processing (CEP) and fuzzy clustering techniques show that these techniques can be very useful in providing more effective services.

We thoroughly investigated and analyzed the primary studies in the field of big data & IoT and data mining & IoT that focused on the application. Our observations are summarised in Tables 9 and 10. The analysis in Table 9 consists of the Number and Name section, and Main idea, Advantages, and Disadvantages sections.

Table 9. An overview of existing primary studies focusing on application

#	Article	Number & Name section	Main Idea	Advantages	Disadvantages
1	Rizwan et al. [55]	Three sections Internet of Things big data User interface	Low-cost Real-Time Smart Traffic Management System	Real-time and low-cost traffic management	The model for calculating the car's motion does not work smartly. In this research, the type of cars and characteristics of cars is not considered.
2	Khorshed [6]	Three layers Data Layer Receive layer Analysis Layer	Check security in the three-layer model and identify cyber attacks	Not mentioned in the relevant article	The obtained accuracy is not reached to 100%
3	IoT-based BD model [56]	Four layers Dynamic query manufacturer Data retrieval via multiple interfaces Two-level data analysis Data transfer via secure channel	Big data processing for Internet of Things (IoT) based data	This system extracts information in a specific time frame. Divide the query into understandable queries and perform them in parallel Reduce time and cost Omit redundancy	Not mentioned in the relevant article
4	Mazumdar et al. [57]	----	Present BigCO: a Big data correlation orchestrator for internet of things how multifaceted data could be interrelated and analyzed using 3D modeling and present a streaming compression algorithm	Provides benefits in big data management for IoT	Not mentioned in the relevant article
5	Kang et al. [58]	Comparison of data distribution and query	Sensor-integrated RFID data repository-implementation model using	Horizontal scalability	-Supports only 4 event types of EPCIS standard -No support for

		efficiency - Compare Mongo-DB to MySQL - Investigate the impact of efficiency with increasing	MongoDB		Shard Extension and complex query
6	Niyato et al. [59]	Data source ، service provider، service customer	-New approach in utilizing large data sets to optimize complex system operations - Optimal a pricing scheme that allows the service provider to determine the amount of data to be acquired to provide services to the users	The numerical examples demonstrate that big data and IoT service provider can achieve the maximum profit through the proposed market model.	Not mentioned in the relevant article
7	Rathore et al. [60]	Data source layer, Communication layer, Graph building, processing Layer, result layer, interpretation layer, application Layer	Proposed a graph-oriented mechanism to achieve the smart the transportation system in the city	Process graphs with near real-time More scalable and efficient	Not mentioned in the relevant article
8	Koo et al. [61]		Conceptual applications utilizing the proposed big data plan for the water industry	1) Automated system condition monitoring, including leak detection; 2) optimizing water supply, production, and energy consumption; and 3) optimizing water consumption	The technology seems not available for water supply system at this moment
9	Dineshkumar et al. [62]	-Sensor -Healthcare -Proxy (IoT agent – Intel Galileo Gen2) -big data Health monitors -Mobile phones	Big data analytics on IoT based Healthcare system is proposed	The response time of the proposed system is less; it is suitable for real-time	Not mentioned in the relevant article

		with GPRS/GSM connections		alerting.	
10	Saenko et al. [63]	Data collection Data storage Data normalization and analysis Data visualization	Implement the parallel system for securing data processing within IoT on the fly basing on complex event processing technology.	Scalability Real-time processing	Not mentioned in the relevant article
11	Tang et al. [64]	Four sections User interface Business logic layer Data access layer MYSQL database	Design of reliability assessment system for product quality assurance	Not mentioned in the relevant article	For the big data discussion, no idea of storage considered.
12	Alam Analysis [65]	-----	Examining the methods of data mining that are suitable for the IoT.	Evaluating the Effectiveness and Efficiency of data mining Algorithms	The dataset used was not very large
13	Moreno-Cano et al. [66]	------	Analyzes the benefits of big data for smart cities and the potential of the knowledge discovery from sensed data	Provide efficiently services like the management of the energy consumption and comfort in buildings, and the transport congestion in the context of smart cities	Not mentioned in the relevant article

Table 10 compares the application expressed in terms of the year of publication, their tool and method, experimental type to find whether the intended plan is numerical analysis or implementation or programming or simulation or design, the application that is presented in the article, and taxonomy that is shown in Figures 3 and 4.

Table 10. An overview of existing primary studies focusing on application

#	Article	Publication Year	Article Type	Tool & Method	Experimental Type	Application	big data / data mining
1	Rizwan et al. [55]	2016	Conference	Make Model Based on Mathematical Formula	Implementation	Intelligent and Real-time Traffic Management	An / Pre
2	Khorshed [6]	2015	Conference	Weka Random Forest C45/J48 LMT Heoffding REPTree	Implementation Programming	Identify 18 types of cyber-attacks on the system	An / Cl
3	IoT-based BD model [56]	2017	Conference	HDFS	Implementation	Not mentioned a specific application in the article	Pr / --
4	Mazumdar et al. [57]	2014	Conference	Raspberry-Pi MySQL database R-DPH algorithm	Implementation	Not mentioned a specific application in the article	St / Des
5	Kang et al. [58]	2015	Journal	-Compound Shard key of MongoDB	Simulation	Supply chain	St-Pr / --
6	Niyato et al. [59]	2016	Conference	Logistic regression and support vector machines (SVMs) with a radial basis function kernel	Numerical analysis	Market model	An / Cl
7	Rathore et al. [60]	2015	Conference	HDFS Giraph Spark	Implementation	Smart Transportation	St-Pr-An / Cl
8	Koo et al. [61]	2015		---	Implementation	Water Supply	St-Pr / --
9	Dineshkumar et al. [62]	2016	Conference	Intel Galileo Arduino programming HDFS MapReduce	Implementation	Healthcare	St-Pr / --

10	Saenko et al. [63]	2017	Conference	Hadoop HDFS Yarn MapReduce	Implementation	Secure monitoring	St-Pr --
11	Tang et al. [64]	2016	Conference	Rule Bayesian	Design	Reliability Management and Index System	An Pre
12	Alam Analysis [65]	2016	Journal	R SVM DLANN ANN C4.5 C5.0 NB LDA KNN	Simulation	Tree real Sensor data set in UCI data repository	An Cl
13	Moreno-Cano et al. [66]	2015	Conference	Regression techniques like Radial Functions Networks (RBs) and Particle Filters (PFs) Fuzzy Clustering	Two examples of services of smart city are Described	-Smart Campus of the University of Murcia - Public Tram service of Murcia City	An Pre

St: storage, Tr: Transmission, Pr: processing, An: Analyzing, Des: descriptive, Pre: predictive, Cl: Classification

We can also classify the studied papers related to the application of IoT based on the taxonomy presented in Figures 3 and 4. This classification is carried out in two areas of big data and data mining separately. The major focus of big data's articles was on transferring, storing, processing, and analyzing data and sometimes cover all cases. Data mining also appears in the papers of big data, when the author intends to analyze the data. Therefore, we categorize such articles with data mining's articles. The categorization of the big data taxonomy is shown in Figure 14. These articles have not regarded transmission subject.

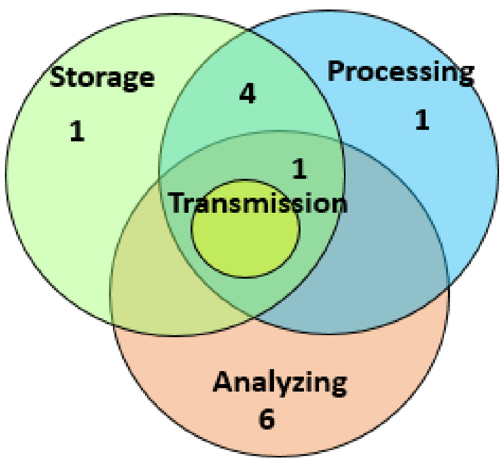


Fig.14. The categorization of papers according to the taxonomy of Figure 3

By reviewing the articles related to the application on IoT, we found that five papers used data mining method for analysis. As shown in Figure 4, two papers are in the prediction section, and three articles are in the classification section.

4. Results

After analyzing and reviewing the articles, this section answers the questions posed in this article.

Answer to Question RQ1: How many types of articles are published on SMS, SLR and survey by data mining and big data methods from 2010 to August 2017 on IoT's area?

To find survey articles, we searched words including Survey, Review, Literature, and Mapping in the titles of articles, and labeled the intended article after studying their abstract. Among the 125 existing articles, we did not find any article titled with SLR and SMS. The number of survey articles expressed in titles is shown in Figure 15.

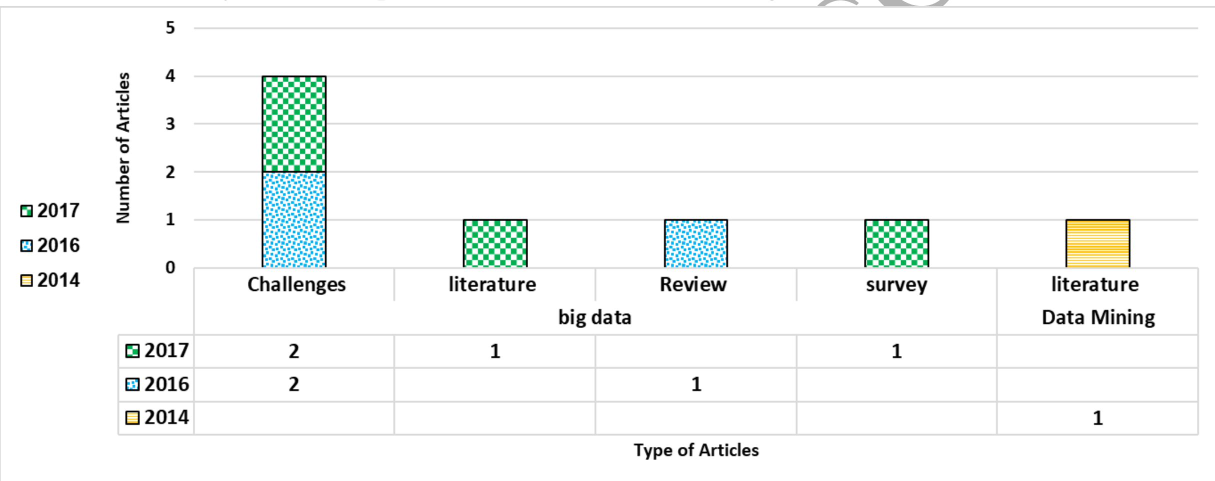


Fig.15. The number of articles by searched words

The list of survey articles is presented in Table 11.

Table 11. List of survey articles

#	Authors	Title	Publisher	Year	Citation
1	U. Ahsan; A. Bais	A Review of big data Analysis and Internet of Things [67]	IEEE conference	2016	0
2	C. Stergiou and K. E. Psannis	Recent advances delivered by Mobile Cloud Computing and Internet of Things for big data applications: a survey [68]	Wiley Journal	2017	9
3	Gustavo Cattelan, NobreElaine Tavares	Scientific literature analysis on big data and internet of things applications on circular economy: a bibliometric study [69]	Springer Journal	2017	2
4	H. Cai; B. Xu; L. Jiang; A. V. Vasilakos	IoT-Based big data Storage Systems in Cloud Computing: Perspectives and Challenges [70]	IEEE journal	2017	5
5	M. Marjani; F. Nasaruddin; A. Gani; A. Karim; I. Hashem; A. Siddiqa; I. Yaqoob	Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges [71]	IEEE journal	2017	0
6	Manan BawaDagmar CaganovaIvan SzilvaDaniela Spirkova	Importance of Internet of Things and big data in Building Smart City and What Would Be Its Challenges [72]	Springer Conference	2016	2
7	R. Gore; S. P. Valsan	big data challenges in smart Grid IoT (WAMS) deployment [73]	IEEE conference	2016	2
8	C. W. Tsai; C. F. Lai; M. C. Chiang; L. T. Yang	data mining for the Internet of Things: A Survey [4]	IEEE journal	2014	170

Answer to Question RQ2: How many articles are there on IoT's area and what percentage of the articles is related to data mining- IoT and big data -IoT from 2010 to August 2017?

The number of articles with IoT or Internet of Things titles based on their titles in the expressed databases was 7023. Figure 16 shows their number of five intended databases.

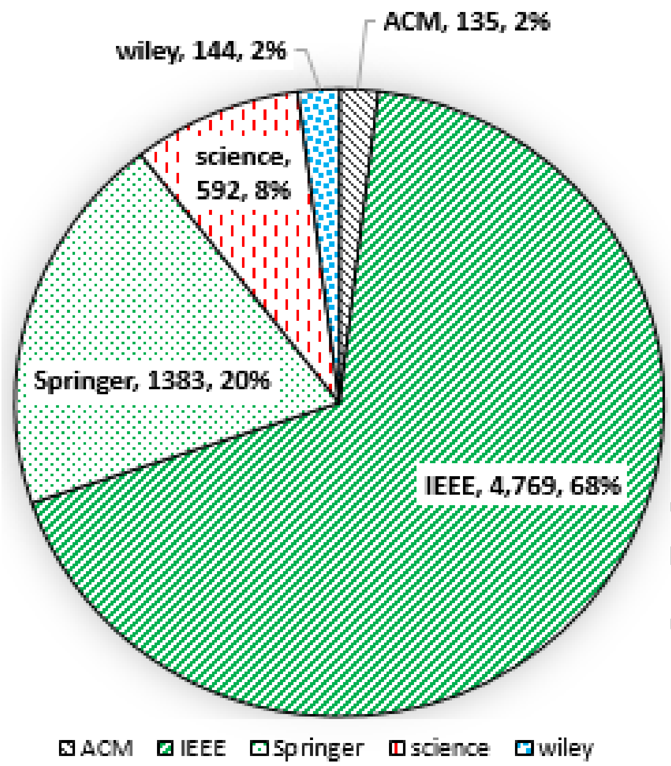


Fig. 16. The percentage of articles by five intended databases.

By dividing the articles into two big data and data mining fields, according to their titles, the number of articles was determined in these areas. There are 125 articles among the 7023 articles in total in these two areas. Figure 17 shows the number in percentage terms.

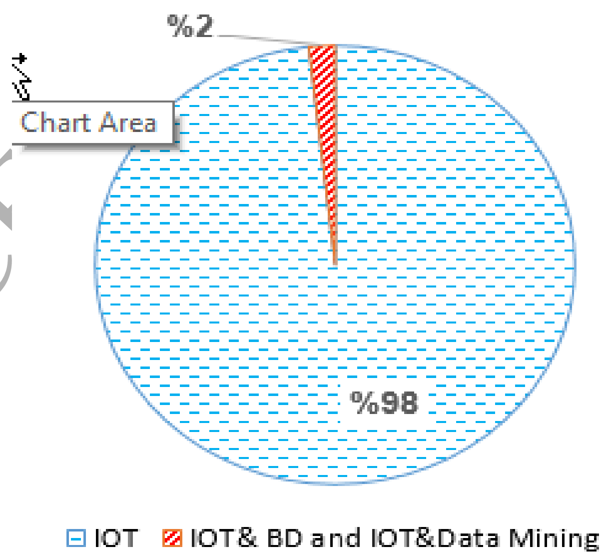


Fig. 17. The percentage of available articles in two big data & data mining fields.

Answer to Question RQ3: What are the most popular tools and simulators in the IoT, big data, and data mining areas?

In the studied articles on IoT and big data, the Hadoop ecosystem was commonly used. This ecosystem continuously includes tools and frameworks for different types of storing, processing, data aggregation, resource management, security, analysis, and search, and data discovery [74] in the main core of Hadoop is comprised of an HDFS storage section and the MapReduce processing section. The Hadoop breaks data into blocks and distributes them between the nodes of one cluster. This process does not require uniform nodes because each data segment can be processed in a separate node [27], [75]. The MapReduce section sends a piece of code for the nodes to be processed in parallel [76], [74]. The IoT must perform an analysis of live data. By MapReduce, the time it takes to process data at the GB level is in seconds [28].

Spark is a very fast clustering computing technology and is designed for high-speed computing. It is based on Hadoop MapReduce and develops the MapReduce model so that it can be effectively used for most types of computing, and it includes the interactive queries and stream processing. The main feature of Spark is to store the computing in memory so that it increases the processing speed of the application [77], [78].

Storm: It processes large volumes of data by horizontal scalability method with error tolerance and analysis in real time. Basically, the framework of Hadoop and storm is used to analyze voluminous data. They complement each other but differ in some aspects. The storm executes all the operations but is not stable. Hadoop is good with MapReduce in all fields but has a delay in real time calculations [79].

Hive: An information storage that runs on Hadoop. The hive manages data stored in the Hadoop and provides an information request interface based on SQL. It is designed for analyzing big data set and responds well to complex queries. The hive is a fast and accessible way to query data stored in Hadoop. It is not designed to be an online transactional platform, hence functions much more slowly than MySQL [79].

Flume is a tool/service for collecting and transferring large volumes of data streams such as event and log file. The flume is a distributed, reliable, available, and configurable service. With the use of flume, you can collect data from several real-time servers.

Kafka: In Apache Storm, the most common way is to get data through Apache Kafka, which is a messaging system.

Kafka is a distributed queue system that has high throughput and can be a good successor to messaging systems due to its good functioning. Kafka in comparison to the older and more traditional systems has better performance, unchangeable and unrepeatable classification, and high failure tolerance. Furthermore, it is suitable for message processing on a wider scale [74]. **Redis**: The database is in the memory and uses a very fast data structure for storing [49].

Mongo DB: A database on disk, rich in the query language and popular in terms of speed, efficiency, and scalability among NoSQLs [58].

The tools that have been used in the studied articles on IoT and big data areas are shown in Table 12.

Table 12. Usable tools in big data & IoT

Distributed file system	Distributed programming			SQL on Hadoop	Data Ingestion	
HDFS	Mapreduce	Spark	Storm	Hive	Flume	Kafka
5	3	4	1	1	1	1
NO SQL Database					New SQL Database	
Column Data Model		Document Data Model	Key-Value Data Model	Graph Data Model	Relational DBMS	
HBase	Cassandra	Mongo DB	Redis	Neo4j	VoltDB	
2	2	3	1	1	1	

In the field of data mining, the Mlib libraries in spark or Mahout in the Hadoop and WeKa and R software are commonly used.

Answer to Question RQ4: What are the most applications applied on the IoT with data mining and big data methods?

From the 44 existing articles about the combination of IoT- big data, and the IoT - data mining presented in this study, 35 articles examined their proposed method of the application. These articles presented their performance on 19 applications. Figure 18 shows the subject of the applications in this field, along with their number of repetitions.

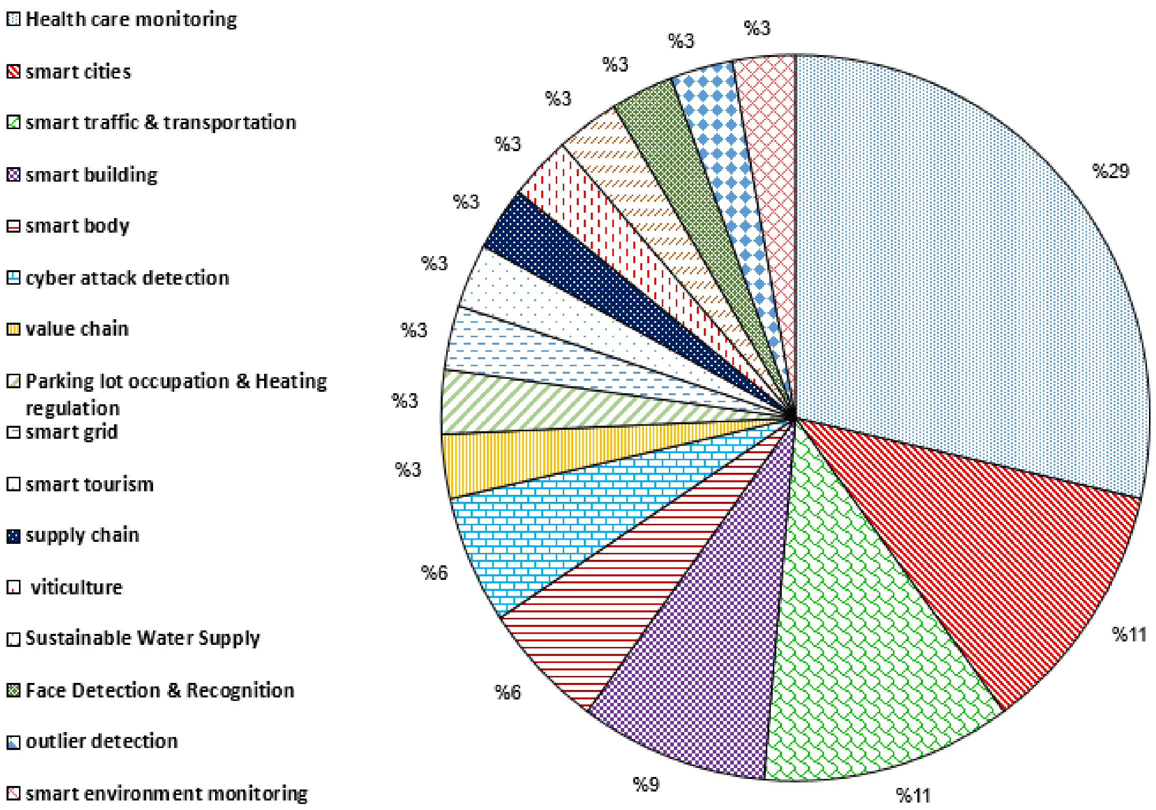


Fig.18. The percentage of used subjects and applications in articles

Answer to Question RQ5: Which of the universities are valid in the IoT, data mining, and big data areas?

In order to answer this question, we first investigated the entire articles based on author affiliation from 2010 to August 2017, and then the universities that provided the 44 papers used in this study. Table 13 introduces some of the universities that have the most publication of the article from 2010 to August 2017 in the IoT field.

Table 13. The universities, which published more than seven articles respectively.

University Name	Country
School of Electrical and Computer Engineering, Georgia Institute of Technology	USA
VTT Technical Research	Finland
Department of Information Technology, University of Turku	Finland
Dept. of Innovation Engineering, University of Salento	Italy
NEC Laboratories Europe	Germany
Center for TeleInFrastruktur (CTIF)	Denmark
Distrib. Syst. Group, Tech. Univ. Wien	Austria
State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications	China
School of Computer and Communication Engineering, University of Science and Technology Beijing	China
Department of Electrical Engineering, Indian Institute of Technology Hyderabad	India
Department of CICE, Hankuk University of Foreign Studies, Yongin-si	Korea

A total of 70 universities contributed to the publishing of about 44 articles, that Kyungpook National University has the largest share. Table 14 shows the name of a number of universities which have published more than one article.

Table 14. The universities, which published more than one article respectively.

University Name	Country	Article Number
Kyungpook National University	Korea	6
Deakin University	Australia	2
University POLITEHNICA of Bucharest	Romania	2
VIT Technical Research University	Finland	2

To give a more comprehensive answer to this question, we examined the countries of the universities that published the articles. Seventy universities mentioned in the articles were from 24 countries and Korea had the largest share. Figure 19 shows the contribution of each country to the published articles.

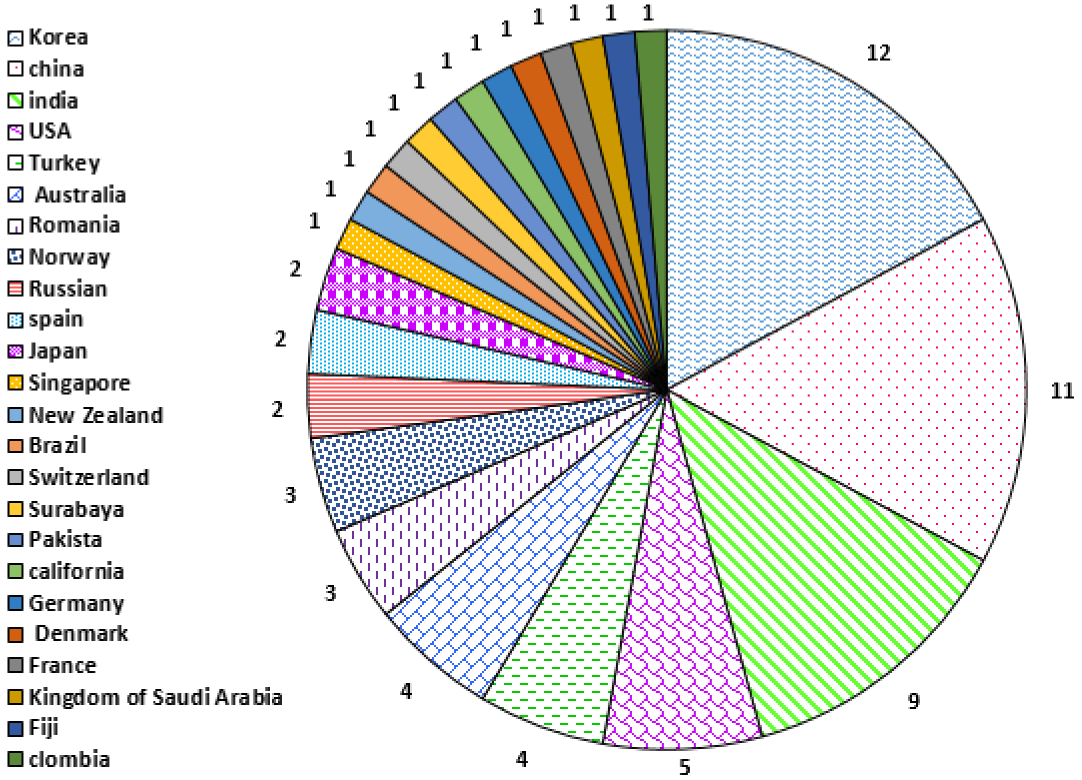


Fig.19. The number of articles by country

We studied the articles that have been investigated in terms of citation and showed ten of the articles with the highest citation in Table 15.

Table 15. The Highest citation articles

Year	Title	Author	Publication	Type of publisher	Citation
2016	Urban planning and building smart cities based on the Internet of Things using big data analytics [26]	M. M. Rathore, A. Ahmad, A. Paul and S. Rho	Science Direct	Journal	57
2016	Internet of Things and big data Analytics for Smart and connected communities [29]	Y. Sun; H. Song; A. J. Jara; R. Bie	IEEE	Journal	50
2014	An Architecture to Support the Collection of big data in the Internet [39] of Things	C. Cecchinel; M. Jimenez; S. Mosser; M. Riveill	IEEE	Conference	41
2015	big data, Internet of Things and Cloud Convergence An Architecture for Secure E-Health [38] applications	George SuciuVictor SuciuAlexandru MartianRazvan CraciunescuAlexandru VulpeIoana MarcuSimona HalungaOctavian Fratu	Springer	Journal	24
2016	MongoDB-Based Repository Design for IoT-Generated RFID/Sensor big data [58]	Y. S. Kang; I. H. Park; J. Rhee; Y. H. Lee	IEEE	Journal	20
2013	IoT-StatisticDB: A General Statistical Database Cluster Mechanism for big data Analysis in the Internet of Things [50]	Z. Ding; X. Gao; J. Xu; H. Wu	IEEE	Conference	11
2016	Smart buddy: defining human behaviors using big data analytics in social internet of things [31]	A. Paul; A. Ahmad; M. M. Rathore; S. Jabbar	IEEE	Journal	10
2015	Towards sustainable water supply: Schematic development of big data collection using Internet of Things (IoT) [48]	D. Koo, K. Piratla, and C. J. Matthews	Science Direct	Journal	10

This section also introduces some of the conducted and ongoing IoT projects. The list of these projects is presented in Tables 16 and 17. Table16 contains some of the presented projects in the sites that were mentioned in reference [80] and [81].

Table 16. Some of the ongoing IoT projects

Current IoT projects			
Name of project	**Description**	**Aim**	**URL**
I-BiDaaS (h2020)	Big data based on industrial as a self-service solution	To easily use big data and interact with it by designing, constructing, and illustrating a unified solution that increases data analysis speed while coping with the data asset growth rate, and assist to cross-domain data-flow for thriving EU economy.	http://itml.gr/i-bidaas
WAZIHUB (h2020)	WAZIHUB is a new project for Africa that provides an IoT OpenHUB and cutting-edge of big data and African solutions, which designed by Africans that the solutions can be adapted to local service requirements.	To repeat and extract value from spinning-off value-added IoT new services based on the technologies developed in WAZIUP and FIWARE projects.	https://wazihub.com.cutestat.com/
symbIoTe (h2020)	It is built on the meaning of virtual IoT environments provisioned on several IoT platforms based on the cloud.	To improve a simplified application of IoT and service improvement process on interworking platforms of the IoT	https://www.symbiote-h2020.eu/
CityPulse (EU FP7)	Processing the Real-Time IoT stream and large-scale data analysis for applications of a smart city.	To design, extend and experiment a distributed framework for semantic finding, processing, and explanation of very real-time IoT and related social data streams for knowledge exploitation in a city environment. The data of smart city is big data	http://ict-citypulse.eu/page/
Frontiercities 2 (H 2020)	It is one of sixteen accelerator projects that designed to advance the future internet research platform of EU, FIWARE is the consortium responsibility to persuade SMEs and startups with new ideas to create a sustainable business.	Investing the ideas involved smart mobility applications and concepts.	http://www.fi-frontiercities.eu/
HYPER/CAT	It is an open, lightweight based on JSON hypermedia catalog format for	To display information about assets of IoT on the web and	http://www.hypercat.io/

	representing URIs set. It is easy to work and authorizes developers to publish linked-data resources descriptions. A HyperCat catalog may display any number of URIs and with any number of statements of RDF-like triple about it.	allow a server to prepare a resources set with a semantic annotations set to a client	
Ikaas (h2020)	It is a project of Research and Innovation Action (RIA) has invested by the EU and a common study platform between the Japan and EU.	To creation an intelligent, privacy protection and safety Smart City platform, which is based on resources of 'Big Data' and an analysis engine constructed on heterogeneous cloud platforms with data that were collected from a variety of sensors of IoT environments.	http://ikaas.com/
SOCIOTAL	Creation a socially conscious citizen-based IoT	Creation participatory and open ecosystems in IoT devices and information flow can be freely shared in a secure way.	http://sociotal.eu/

Table 17 contains some of the completed projects in the sites that were mentioned in reference [80] and [81].

Table 17. Some of the completed IoT projects

Completed IoT projects					
Name of project	**Start - year**	**End - year**	**Description**	**Aim**	**URL**
IOTLab (FP7-ICT)	2013	2016	It is a European project examines the potential of crowdsourcing as a tool to extend current IoT testbed infrastructures for multidisciplinary experiments resulted in more end-user interplay.	It gains by unifying heterogeneous testbeds from the FIRE community and progressing their interplay through using Cloud technologies and virtualization. Testbed scale is developed by containing hundreds of mobile device users that connected via their tablets smartphones, etc.	www.iotlab.eu/
EYEHUB	2013	2014	It is one of eight Ecosystem Demonstrators of IoT that co-funded by Technology Strategy Board of the UK.	To create a 'proving ground' for an IoT ecosystem on the Surrey's campus University and to develop the concepts in the adjoining town (Guildford) in a way that represents how key challenges of IoT can be resolved.	https://connect.innovateuk.org/web/internet-of-things-ecosystem-demonstrator
PROBE-IT	2011	2013	It is a Coordination and Support Action (CSA) protected by the Framework Programme 7 of EU.	To analyze current roadmaps, which are controlling existing IoT research and identifying fragmentation sources that could modify research workings and try to alight the fragmentation.	http://www.probe-it.eu/
FI-WARE	2011	2014	Providing a royalty-free, truly open, and public architecture, and a set of open specifications, which allows service providers, developers, enterprises and other organizations to extend products, which satisfy their requirements while still is innovate and open.	To offer a set of open APIs, which allows developers to avoid connecting with specific vendors, so protecting application investment of developer.	https://www.fiware.org/
OUTSMART (FP7-ICT)	2011	2013	Help Future Internet (FI) by developing innovative ecosystems.	Help Future Internet (FI) by the development of five tolerable business eco-systems with the technology of FI at	http://out-smart.eu/

				their core.	
iCORE (FP7-ICT)	2011	2014	IoT empowering by Cognitive Technologies.	To extend a cognitive framework for using in IoT environment, While automaticity, cognition, and knowledge creation are totally concepts that spread in the world of common fixed and mobile telecommunications, it cannot be laid in the IoT world.	http://www.iot-icore.eu/
IoT.est (FP7-ICT)	2011	2014	It develops a test-based service creation environment for IOT empower business services.	to describe publish and introduce IoT services that extract data across domain borders and help to run-time monitoring that provide the adaptation of autonomous service to context or environment and network parameter changes.	
IoT-i (FP7-ICT)	2010	2012	A harmonic operation that protects developing the European IoT community.	It presented a number of tools (that setting up an International Forum of the IoT was one of them) and measures that decrease the risks and co-ordinate IoT research operations to empower European investments to exhibit IoT leadership while addressing the social issues of the IoT for the citizen of European.	http://www.iot-i.eu/public
SmartSantander (FP7-ICT)	2010	2013	An individual city-scale empirical study easiness in protecting typical services and applications for a smart city.	To develop a large IoT empirical facility individual in its kind and accessible for testing the experimenter's community, end-users, and service providers.	http://www.smartsantander.eu/
IoT-A (FP7-ICT)	2010	2013	A European Lighthouse Integrated Project that explains the IOT Architecture.	It presents the development of an architectural reference model along with the description of a primary set of key building blocks.	http://www.meet-iot.eu/iot-a-deliverables.html

Answer to Question RQ6: Why are the methods of IoT combined with the methods of big data and data mining?
In the IoT and big data discussions, the big data tools are often used to store large volumes of data and parallel processing and real-time analysis and data mining.
However, the goal of data mining methods is to automate existing applications and more to be discussed in intelligent IoT. Sometimes, data mining methods are used to reduce data storage.

Answer to Question RQ7: What are the main challenges of IoT with big data and data mining?

To answer this question, some of the challenges presented in the articles have been considered. Banafa [82] divided the IoT challenges into the following three sections:

1- **Society**:

Customers' needs and demands are constantly changing, and day by day, newer devices with different technologies come into the market. Therefore, IoT should consider these issues and try to adapt itself to the rapid changes that take place.

2- Business:

IoT could be divided into three classes based on its usage and customers:

2-1 **Consumer IoT** contains devices that are connected such as watches, laptops, phones, smart cars, and amusement systems.

2-2 **Commercial IoT** contains items such as medical devices, inventory control, and tracking devices.

2-3 **Industrial IoT** includes items such as flow gauge, electric meter, sewage systems, robots manufacturing, pipeline monitors, and other types of industrial devices and systems connected.

3- Technology

This section includes all the technologies that are considered to create IoT systems operations smoothly as an independent solution or section of the current systems.

3-1 Security

The IoT has now become a serious security issue that has attracted the attention of prominent technology companies and government agencies around the world. Hacking kid monitors, medicine infusion pump, smart fridges, cameras, thermostats, and aggressive guns represent a safety and security nightmare that will be created by the IoT future.

3-2 Connectivity

To connect many devices is one of the most significant challenges for the future of the IoT, and it challenges the structure of the current following technologies and communication models. Presently, we rely on the server/client model for authenticating, authorizing, connecting various nodes in a network, and concentrating.

3-3 Compatibility

IoT is growing in many aspects by competing with various technologies for becoming the standard. This causes problems and requires the development of additional software and hardware when devices are connected.

3-4 Standards

Technology standards that contain data collection standards, network protocols, and communication protocols are the totality of all operations of management, processing, and to store data that are collected from sensors. This collection enhances the data value by increasing the frequency, domain, and scale of data that are available for analyses.

- Standards for the management of unstructured data
- Technical skills to power newer collection tools

3-5 Intelligent Analysis & Actions

The final step of IoT implementation is the extraction of intelligence data for analyzing that is controlled by accompanying models and intelligent technologies that assist in using the intelligent technologies.

Challenges facing the adoption of intelligent operations within IoT

- Operations of machines in unpredictable conditions
- Privacy and security of information
- Interoperability of machine
- Moderate feedback on human behavior
- Gradual adaption of recent technologies

Some of the challenges are expressed in Tables 18 and 19.

Table 18. Big data Challenges

Ref #	Big Data Challenges
[66]	1- Real-time management of large amounts of data 2- The interoperability between various devices 3- The integration of many proprietary protocols and communication standards 4- Big data analysis 5- Size and heterogeneity of data
[83]	**Technical and Managerial challenges** 1- Data management 2- Data mining 3- Privacy 4- Security 5- Chaos.
[5] - [84]	Privacy
[74]	**Technological challenges** 1- Complexity 2- heterogeneity (hardware, scalability, the volume of data) 3- security 4- privacy
[79]	**Technological challenges** 1- Security 2- Privacy 3- Interpretability 4- Data Quality Issues: **Adoption challenges** 1- Model Reliability, Validation, and Adaptation 2- Integration and Reconciliation with Our Physical Understanding of the World 3-Human-Analytics Interaction: 4- Potential for Systemic Errors and Failures 5- Personalization Versus Limitation of Choice 6- Combining and analyzing heterogeneous data 7- Availability of storage space 8- Compatibility of systems
[41]	1- The existence of real-time data that was irregular and unpredictable
[45]	1- Volume 2-Velocity 3- Real-time analytics
[29]	1-Reasonable data organization model 2-Security and Privacy 3-Technology challenges for protecting featured cultures and heritages 4-Robustness of applications
[85]	1- Human Dynamics 2- Privacy 3- Security 4- Volume 5- Velocity 6- Variety

	7- Veracity 8- Value
[86]	1- Architecture 2- Power Is Critical 3- Security 4- Complexity 5- Sensor's 6- Creating Knowledge and Big data 7- Data is captured, but not used fully 8- Cost 9- Adaptability 10- Self - Learning 11- Deployment 12- Maintenance 13-privacy

Table 19. Data Mining Challenges

Ref #	Data Mining Challenges
[71]	1- data 1-1 Volume 1-2 Accessibility 1-3 Correctness 1-4 Heterogeneity 2- Knowledge Discover 2-1 complexity 3- Processing 3-1 Sequential 3-2 Parallel
[87]	1-energy-efficiency 2- misreading 3- repeated reading 4- fault-tolerance 5- data filtering 6- communications 7- Heterogeneity of data 8- Depositing data at different 9- Data processing problems
[65]	Timely analysis big data
[88]	Transmission of data from sensors in large volumes
[40]	Data management

In total, the challenges of the IoT fields, big data, and data mining can be summarised as follows:

Volume: The volume is how much information and data we have – before measured in Gigabyte size and now measured in Zettabytes (ZB) or even Yottabytes (YB). IoT is making a descriptive growth in data.

Variety: Variety explains one of the greatest challenges of IoT and big data. It can be unstructured and included a variety of kinds of data from XML to video to SMS. Organisation of data in a meaningful manner is not an easy task, especially when data changes quickly.

Velocity: Velocity is the speeding of access to data.

Value: After looking at the volume, variety, visualization, and velocity, which take time, resources, and endeavor, you want to make sure your organization obtains value from the data.

Security, Privacy:

The issue of security and privacy is a major bottleneck for the "Internet of Things." For example, the possibility of an online attack on personal information and hacking it is one of those challenges. This challenge can be focused on the following three areas:

- Vulnerability to Hacking
- Trust Consideration
- Data Protection

Visualization: Visualization is very important in today's world. The use of graphs and charts to visualize great amounts of complex information (data) is more effective in translating the meaning than spreadsheets and informs that it is full of numbers and formulas.

Extracting knowledge:

In a large amount of data, it is necessary to be able to extract knowledge from the data so that the data can be more usefully stored.

Real-time analysis:

Since data in the IoT are increasing day by day, it could be possible to analyze them in real time for more effective analysis. For example, by real-time analysis of data and extracting important features, the amount of data storage can be reduced, or in the event of a security problem, the necessary solutions can be applied. Therefore, the real-time analysis will have a special significance among the challenges posed.

Answer to Question RQ8: What are the open issues of IoT with big data and data mining? Different ideas can be expressed in the Big Data-Internet of Things and Data Mining-Internet of Things. Therefore, after studying related articles in this field, we will discuss some of the issues that can be studied and analyzed further. These topics are categorized as follows:

1- Data collection

1-1. Reducing the data volume

In IoT, there is a discussion on the volume and variety of data. We can use the methods of data dimension reduction for data, and instead of storing the original data, the data should be stored after dimension reduction. Different methods can be employed to reduce the dimension and extract features in data mining before storing the data. Of course, this discussion needs the ability of real-time processing of online data for aggregation data.

1-2 Data categorization

The data are graded according to their importance, and labels can be assigned to show their degree of confidentiality. Then the labels carry out different data mining algorithms. Data can also be stored in accordance with the degree of confidentiality in different locations, and supported by confidential data as well as by providing more protection.

2- Data analysis

In most of the studied papers, the spark has been used for data analysis. However, there was no article that compared the spark, storm, and other analytical methods for analysis of a particular application.

3- New business models

The Internet of Things can help companies make new applications available to their customers and make customer service easier and more intelligent. In this section, we will describe some of these applications that can be created using data mining methods.

- In the mailbox transport section, the IoT discussion can be made. Therefore, the client can track and monitor the process of sending and delivering the packet to the receiver and by using the intelligent algorithm, the packet route is selected to be the shortest and least costly route.
- Lighting of highways can be controlled by using the methods of pattern recognition, and at the times when the highway is off, it reduces brightness. With this clever way, you will save a lot of power in cities.
- By collecting vital system information using data mining methods, you can perform analyses on the system and if you see the influence or prediction of failure, you will send the necessary warnings and, in the event of proper planning, you will automatically solve the problems.
- In the busy discussion, with intelligent traffic control using data mining algorithms, it is possible to always provide green way for buses that are significantly effective in reducing fuel consumption and reducing air pollution as well as citizens' satisfaction.

4- Global visibility

The IoT can also be used for intelligent monitoring of specific applications. For example, in the Smart City discussion, it can be established that all of the city's data is stored in a location and analyzed on it, and it identifies patterns and behaviors of customers, and observes any behavior outside the normal patterns, along with sending the necessary alarms. The appropriate security solutions are also proposed.

5- Security and privacy

The security debate by layers in the IoT is of particular importance. The increase of automation and digitization is causing security concerns. Therefore, security has been addressed in various aspects in the IoT, and hackers' penetration and unauthorized entry and destructive activities should be avoided. By using data mining methods, users can identify behavior patterns, or by encrypting important data, can provide the necessary security system for specific applications.

Figure 20 shows a summary of challenges and opportunities

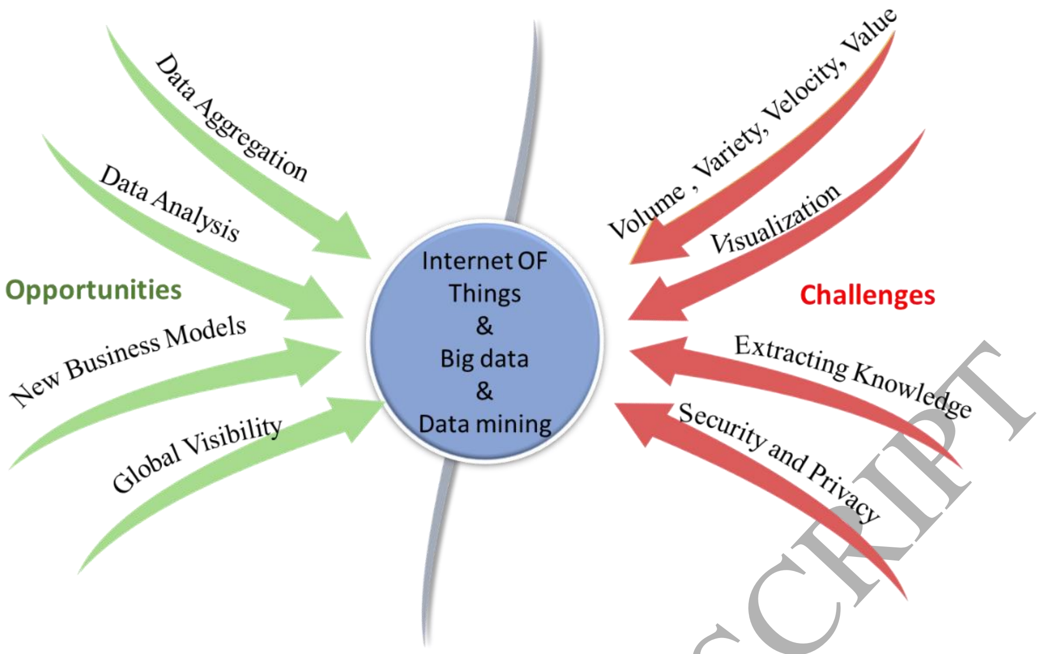


Fig .20. challenges and opportunities

5. Conclusion

In this paper, the results of the SMS study that combined with SLR on IoT with big data and IoT with data mining are presented. The main articles are divided into three categories based on the major focus of the articles. Most of the subjects are related to architecture and platform (52%, 23 articles). This paper examines the publishers and the universities and countries which were effective on the publication of these articles. The most publication in this area belonged to IEEE, and the Kyungpook National University had the most articles. The highest number of articles respectively belonged to Korea, China, India, and the United States.

This study also has some limitations. The performed searches were on the titles of the articles, and other articles can be accessed by developing this search. Also in the field of data mining, only the term "data mining" has been used and to increase the number of articles in this field, the names of data mining methods can be used to search the articles. This study focuses on English articles and has not been searched in non-English articles.

References

[1] L. Atzori, A. Iera, G. Morabito, "The Internet of Things: A survey," *Computer Networks,* vol. 64, no. 15, pp. 2787-2805, 2010.

[2] M. Alaa, A.A. Zaidan, B.B. Zaidan, M. Talal, M.L.M. Kiah, "A review of smart home applications based on Internet of Things," *Journal of Network and Computer Applications,* vol. 97, pp. 48-65, 2017.

[3] B.B. Zarpelão, R.S. Miani, C.T. Kawakani, S.C. de Alvarenga, "A survey of intrusion detection in Internet of Things," *Journal of Network and Computer Applications,* vol. 84, pp. 25-37, 2017.

[4] C.W. Tsai, C. F. Lai, M.C. Chiang and L.T. Yang, "Data Mining for Internet of Things: A Survey," *IEEE Communications Surveys & Tutorials,* vol. 16, no. 1, 2014.

[5] C. D. Gómez Romero, J. K. Díaz Barriga and J. I. Rodríguez Molano, Big Data Meaning in the Architecture of IoT for Smart Cities, Springer International Publishing, 2016.

[6] M.T. Khorshed, N.A. Sharma, K. Kumar, M. Prasad, A. B. M.S. Ali and Y. Xiang, "Integrating Internet-of-Things with the power of Cloud Computing and the intelligence of Big Data analytics : A three layered approach," in *Computer Science and Engineering (APWC on CSE), 2015 2nd Asia-Pacific World Congress on*, Fiji, 2015.

[7] I. Kholod, M. Kuprianov and I. Petukhov, "Distributed data mining based on actors for Internet of Things," in *2016 5th Mediterranean Conference on Embedded Computing (MECO)*, 2016.

[8] S. Klein, "The World of Big Data and IoT," in *IoT Solutions in Microsoft's Azure IoT Suite: Data Acquisition and Analysis in the Real World*, Apress, Berkeley, CA, 2017, pp. 3-13.

[9] A. Siddiqa, I. A.T. Hashem, I. Yaqoob, M. Marjani, Sh. Shamshirband, A. Gani, F. Nasaruddin, "A survey of big data management: Taxonomy and state-of-the-art," *Journal of Network and Computer Applications,* vol. 71, pp. 151-166, 2016.

[10] F. Chen,et al., " Data Mining for the Internet of Things: Literature Review and Challenges," *International Journal of Distributed Sensor Networks,* vol. 2015, p. 14, 2015.

[11] "tutorials point," 2017. [Online]. Available: https://www.tutorialspoint.com/data_mining/dm_classification_prediction.htm.

[12] J. Cruz-Benito, "Systematic Literature Review & Mapping," nov 2016. [Online]. Available: https://doi.org/10.5281/zenodo.165773.

[13] B. Shojaiemehr, A.M. Rahmani, N.N. Qader, "Cloud computing service negotiation: A systematic review," *Computer Standards & Interfaces,* 2017.

[14] E.J. Ghomi, A.M. Rahmani, N.N. Qader, "Load-balancing algorithms in cloud computing: A survey," *EJ Ghomi, AM Rahmani, NN Qader - Journal of Network and Computer Applications,*

2017, vol. 88, pp. 50-71, 2017.

[15] E. Engström, P. Runeson, "Software product line testing – A systematic mapping study," *Information and Software Technology,* vol. 53, no. 1, pp. 2-13, 2011.

[16] K. Petersen, S. Vakkalanka and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Information and Software Technology,* vol. 64, pp. 1-18, 2015.

[17] H.P. Breivold, I. Crnkovic, M. Larsson, "A systematic review of software architecture evolution research," *Information and Software Technology,* vol. 54, no. 1, pp. 16-40, 2012.

[18] A. Patel, M. Taghavi, K. Bakhtiyari, J.C. Júnio, "An intrusion detection and prevention system in cloud computing: A systematic review," *Journal of Network and Computer Applications,* vol. 36, no. 1, pp. 25-41, 2013.

[19] S. Tanwar, N. Kumar, Joel J.P.C. Rodrigues, "A systematic review on heterogeneous routing protocols for wireless sensor network," *Journal of Network and Computer Applications,* vol. 53, pp. 39-56, 2015.

[20] A. Cocchia, "Smart and Digital City: A Systematic Literature Review," in *Cocchia A. (2014) Smart and Digital City: A Systematic Literature Review. In: Dameri R., Rosenthal-Sabroux C. (eds) Smart City.*, Progress in IS. Springer, 2014, pp. 13-43.

[21] Carrez F, ed, "Final architectural reference model for the IoT v3.0. Internet-of-Things Architecture (IoT-A) Project Deliverable D1.5.," 2013. [Online]. Available: http://www.iot-a.eu/public/public-documents/d1.5/view.

[22] Jan Holler, Vlasios Tsiatsis, Catherine Mulligan, Stefan Avesand, Stamatis Karnouskos and David Boyle, From Machine-to-Machine to the Internet of Things, Academic Press, Oxford, 2014.

[23] O. B. Sezer, E. Dogdu, M. Ozbayoglu and A. Onal, "An extended IoT framework with semantics, big data, and analytics," in *IEEE International Conference on Big Data (Big Data)*, 2016.

[24] Partha PratimRay, "A survey of IoT cloud platforms," *Future Computing and Informatics Journal,* vol. 1, no. 1-2, pp. 35-46, 2016.

[25] M. M. Rathore, A. Ahmad and A. Paul, "IoT-based smart city development using big data analytical approach," in *IoT-based smart city development using big data analytical approach M. M. Rathore, A. Ahmad and A. Paul 2016 2016 IEEE International Conference on Automatica (ICA-ACCA)*, 2016.

[26] M. M. Rathore, A. Ahmad, A. Paul and S. Rho, "Urban planning and building smart cities based on the Internet of Things using Big Data analytics," *Computer Networks,* 2016.

[27] R.K. Behera, S. Gupta and A. Gautam, "Big-data empowered cloud centric Internet of Things," in *International Conference on Man and Machine Interfacing (MAMI)*, 2015.

[28] S. Din, H. Ghayvat, A. Paul, A. Ahmad, M. M. Rathore and I. Shafi, "An architecture to analyze big data in the Internet of Things," in *9th International Conference on Sensing Technology (ICST)*, 2015.

[29] Y. Sun, H. Song, A. J. Jara and R. Bie, "Internet of Things and Big Data Analytics for Smart and Connected Communities," *IEEE Access,* 2016.

[30] S. Sowe, T. Kimata, M. Dong, K. Zettsu, "Managing Heterogeneous Sensor Data on a Big Data Platform: IoT Services for Data-intensive Science," in *IEEE 38th Annual International Computers, Software and Applications Conference*, Sweden, 2014.

[31] A. Paul, A. Ahmad, M. M. Rathore and S. Jabbar, "Smartbuddy: defining human behaviors using big data analytics in social internet of things," *IEEE Wireless Communications,* 2016.

[32] M. M. Rathore, A. Ahmad, A. Paul, J. Wan and D. Zhang, "Real-time Medical Emergency Response System: Exploiting IoT and Big Data for Public Health," *Journal of Medical Systems,* 2016.

[33] D.G. Páez , F. Aparicio, M. de Buenaga, J.R. Ascanio , "Big Data and IoT for Chronic Patients Monitoring," in *Hervás R., Lee S., Nugent C., Bravo J. (eds) Ubiquitous Computing and Ambient Intelligence. Personalisation and User Adapted Services*, Springer, Cham , 2014, pp. 416-423 .

[34] Y. Ma, Y. Wang, J. Yang, Y. Miao and W. Li, "Big Health Application System based on Health Internet of Things and Big Data," *IEEE Access,* vol. 5, pp. 7885 - 7897, 2017.

[35] A. M. C. Souza and J. R. A. Amazonas, "An Outlier Detect Algorithm using Big Data Processing and Internet of Things Architecture," *Procedia Computer Science,* vol. 25, pp. 1010-1015, 2015.

[36] G. Suciu, A. Vulpe, O. Fratu and V. Suciu, "M2M remote telemetry and cloud IoT big data processing in viticulture," in *International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2015.

[37] G. Suciu, V. Suciu, S. Halunga and O. Fratu, "Big Data, Internet of Things and Cloud Convergence for E-Health Applications," *New Contributions in Information Systems and Technologies. Advances in Intelligent Systems and Computing,* vol. 353, pp. 151-160 , 2015.

[38] G. Suciu, V. Suciu, A. Martian, R. Craciunescu, A. Vulpe, I. Marcu, S. Halunga and O. Fratu, "Big Data, Internet of Things and Cloud Convergence – An Architecture for Secure E-Health Applications," *Journal of Medical Systems,* pp. 39-141, 2015.

[39] C. Cecchinel, M. Jimenez, S. Mosser, M. Riveill, "An architecture to support the collection of big data in the internet of things," in *IEEE world congress on services (SERVICES)*, 2014.

[40] S. Nigam, S. Asthana,P. Gupta, "IoT based Intelligent Billboard using Data Mining," in *2016 1st International Conference on Innovation and Challenges in Cyber Security (ICICCS 2016)*, 2016.

[41] D. Tracey, C. Sreenan, "A Holistic Architecture for the Internet of Things, Sensing Services and Big Data," in *13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing*, 2013.

[42] J. Y. Kim, H. J. Lee, J. Y. Son and J. H. Park, "Smart home web of objects-based IoT management model and methods for home data mining," in *2015 17th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, 2015.

[43] Y.-J. Lee, H. D. Park and O. Min, "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.

[44] C. K. M. Lee, C. L. Yeung and M. N. Cheng, "Research on IoT based Cyber Physical System for Industrial big data Analytics," in *Industrial Engineering and Engineering Management (IEEM), 2015 IEEE International Conference on*, 2015.

[45] T. H. Dang-Ha, D. Roverso and R. Olsson, "Graph of virtual actors (GOVA): A Big Data analytics architecture for IoT," in *IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2017.

[46] N. Mishra, L. Chung-Chih and C. Hsien-Tsung, "A Cognitive Oriented Framework for IoT Big-data Management Prospective," in *IEEE International Conference on Communiction Problem-solving*, 2014.

[47] M. Strohbach, H. Ziekow, V. Gazis, and N. Akiva, "Towards a Big Data Analytics Framework for IoT and Smart City Applications," in *Xhafa F., Barolli L., Barolli A., Papajorgji P. (eds) Modeling and Processing for Next-Generation Big-Data Technologies. Modeling and Optimization in Science and Technologies*, vol. 4, vol 4, Springer, Cham , 2015.

[48] M. R. Bashir and A. Q. Gill, "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)*, 2016.

[49] M. H. Berlian, T. E. R. Sahputra, B. J. W. Ardi, L. W. Dzatmika, A. R. A. Besari, R. W. Sudibyo, "Design and implementation of smart environment monitoring and analytics in real-time system framework based on internet of underwater things and big data," in *Electronics Symposium (IES), 2016 International*, 2016.

[50] Z. Ding, X. Gao, J. Xu and H. Wu, "IOT-Statistic General Statistical Database Cluster Mechanism for Big Data Analysis in the Internet of Things," in *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things*, 2013.

[51] K. Guo, Y. Tang and P. Zhang, "CSF: Crowdsourcing semantic fusion for heterogeneous media big data in the internet of things," *Information Fusion,* vol. 37, 2017.

[52] B. Dundar, M. Astekin and M. S. Aktas, "A Big Data Processing Framework for Self-Healing Internet of Things Applications," in *12th International Conference on Semantics, Knowledge and Grids (SKG)*, 2016.

[53] D. C. Mocanu, E. Mocanu, P. H. Nguyen, M. Gibescu and A. Liotta, "Big IoT data mining for real-time energy disaggregation in buildings," *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC),* 2016.

[54] M. Vögler, J. M. Schleicher, C. Inzinger and S. Dustdar, "Ahab: A cloud-based distributed big data analytics framework for the Internet of Things," *Software: Practice and Experience,* 2017.

[55] P. Rizwan, K. Suresh and M. R. Babu, "Real-time smart traffic management system for smart cities by using Internet of Things and big data," in *International Conference on Emerging*

Technological Trends (ICETT), 2016.

[56] A. Bera, A. Kundu, N. R. De Sarkar and D. Mou, Experimental Analysis on Big Data in IOT-Based Architecture, Springer Singapore, 2017.

[57] M. Mozumdar, A. Shahbazian and N. Q. Ton, "A big data correlation orchestrator for Internet of Things," in *IEEE World Forum on Internet of Things (WF-IoT)*, 2014.

[58] Y. Kang, I. Park, J. Rhee, Y. Lee, "MongoDB-Based Repository Design for IoT-Generated RFID/Sensor Big Data," *IEEE Sensors Journal,* vol. 16, no. 2, 2016.

[59] D. Niyato, M. A. Alsheikh, P. Wang, D. I. Kim and Z. Han, "Market model and optimal pricing scheme of big data and Internet of Things (IoT)," in *2016 IEEE International Conference on Communications (ICC)*, 2016.

[60] M. M. Rathore, A. Ahmad, A. Paul and G. Jeon, "Efficient Graph-Oriented Smart Transportation Using Internet of Things Generated Big Data," in *11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, 2015.

[61] D. Koo, K. Piratla and C. J. Matthews, "Towards Sustainable Water Supply: Schematic Development of Big Data Collection Using Internet of Things (IoT)," *Procedia Engineering,* vol. 118, pp. 489-497, 2015.

[62] P. Dineshkumar, R. SenthilKumar, K. Sujatha, R. S. Ponmagal and V. N. Rajavarman, "Big data analytics of IoT based Health care monitoring system," in *IEEE Uttar Pradesh Section International Conference on Electrical,Computer and Electronics Engineering*, Varanasi, India , 2016.

[63] I. Saenko, I. Kotenko and A. Kushnerevich, "Parallel Processing of Big Heterogeneous Data for Security Monitoring of IoT Networks," in *25th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, 2017.

[64] T. J. Tang, A. Chung, A. Zhao,J. Yang, J. Zhang, "An loT Inspired Semiconductor Reliability Test System Integrated with Data-Mining Applications," in *2016 2nd International Conference on Cloud Computing and Internet of Things (CCIOT)*, 2016.

[65] F. Alam, R. Mehmood, I. Katib and A. Albeshri, "Analysis of Eight Data Mining Algorithms for Smarter Internet of Things (IoT)," in *Procedia Computer Science*, 2016.

[66] V. Moreno-Cano, F. Terroso-Saenz and A. F. Skarmeta-Gómez, "Big data for IoT services in smart cities," in *IEEE 2nd World Forum on Internet of Things (WF-IoT)*, 2015.

[67] U. Ahsan and A. Bais, "A Review on Big Data Analysis and Internet of Things," in *2016 IEEE 13th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, 2016.

[68] C. Stergiou and K. E. Psannis, "Recent advances delivered by Mobile Cloud Computing and Internet of Things for Big Data applications: a survey," *International Journal of Network Management,* vol. 27, no. 3, 2017.

[69] G. C. Nobre and E. Tavares, "Scientific literature analysis on big data and internet of things applications on circular economy: a bibliometric study," *Scientometrics,* vol. 111, no. 1, p. 463–

492, 2017.

[70] H. Cai, B. Xu, L. Jiang and A. V. Vasilakos, "IoT-Based Big Data Storage Systems in Cloud Computing: Perspectives and Challenges," *IEEE Internet of Things Journal,* vol. 4, no. 1, 2017.

[71] M. Marjani, F. Nasaruddin, A. Gani, A. Karim, I. A. T. Hashem, A. Siddiqa and I. Yaqoob, "Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges," *IEEE Access,* 2017.

[72] M. Bawa, D. Caganova, I. Szilva and D. Spirkovas, "Importance of Internet of Things and Big Data in Building Smart City and What Would Be Its Challenges," in *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, 2016.

[73] R. Gore and S. P.Valsan, "Big Data challenges in smart Grid IoT (WAMS) deployment," in *2016 8th International Conference on Communication Systems and Networks (COMSNETS)*, 2016.

[74] N. Jesse, "Internet of Things and Big Data – The Disruption of the Value Chain and the Rise of New Software Ecosystems," *IFAC-PapersOnLine,* vol. 49, no. 29, pp. 275-282, 2016.

[75] E. Ahmed, I. Yaqoob, I. A.T. Hashem, I. Khan, A.I. A. Ahmed, M. Imran, A. V. Vasilakos, "The role of big data analytics in Internet of Things," *Computer Networks,* 2017.

[76] A.R.M. Forkan, I. Khalil, M. Atiquzzaman, "ViSiBiD: A learning model for early discovery and real-time prediction of severe clinical events using vital signs as big data," *Computer Networks,* vol. 113, pp. 244-257, 2017.

[77] A. Baer, P. Casas, A. D'Alconzo, P. Fiadino, L. Golab, M. Mellia, E. Schikuta, "DBStream: A holistic approach to large-scale network traffic monitoring and analysis," *Computer Networks,* vol. 107, pp. 5-19, 2016.

[78] R. Stackowiak, A. Licht, V. Mantha, L. Nagode, "Big Data Solutions and the Internet of Things," in *Big Data and the Internet of Things: Enterprise Information Architecture for a New Age*, CA, Apress, 2015.

[79] Shah, M., Big Data and the Internet of Things, Springer International Publishing, 2016.

[80] "CORDIS," 15 1 2018. [Online]. Available: http://cordis.europa.eu/programme/rcn/701892_en.html.

[81] "Institute for Communication Systems," 15 1 2018. [Online]. Available: https://www.surrey.ac.uk/institute-communication-systems/research-areas/internet-things.

[82] Ahmed Banafa, "3 Major Challenges Facing the Future of IoT," Datafloq , 2017. [Online]. Available: http://works.bepress.com/ahmed-banafa/59/.

[83] C. Ifrim, A.-M. Pintilie, E. Apostol, C. Dobre and F. Pop;, The Art of Advanced Healthcare Applications in Big Data and IoT Systems, Springer International Publishing, 2017.

[84] Charith Perera, Rajiv Ranjan, Lizhe Wang, Samee U. Khan, Albert Y. Zomaya, "Big Data Privacy in the Internet of things Era," *IEEE Computer Societ y,* 2015.

[85] Sumedha Chauhan Neetima Agarwal Arpan Kumar Kar, "Addressing Big Data Challenges in Smart

Cities: A Systematic Literature Review," *info,* vol. 18, no. 4, 2016.

[86] R. G. Lubhna Kaul, "INTERNET OfTHINGS AND Big DA TA - CHALLENGES," in *Online International Conference on Green Engineering and Technologies (IC-GET)*, 2016.

[87] Shen Bin, Liu Yuan, Wang Xiaoyi, "Research on Data Mining Models for the Internet," in *2010 International Conference on Image Analysis and Signal Processing (IASP)*, 2010.

[88] James Jin Kang, Sasan Adibi, Henry Larkin, Tom Luan, "Predictive Data Mining for Converged Internet of Things: A Mobile Health Perspective," in *International Telecommunication Networks and Applications Conference (ITNAC)*, 2015.

[89] E. Cavalcante, J. Pereira, M. P. Alves, P. Maia, R. Moura, T. Batista, F. C. Delicato and P. F. Pires, "On the interplay of Internet of Things and Cloud Computing: A systematic mapping study," *Computer Communications,* 2016.

Shabnam Shadroo was born in Mashhad, Iran, in 1983. She received the B.S. degree in hardware computer engineering from the University of Sajad, Mashhad, Iran, in 2005, and the M.S degrees in artificial intelligence engineering from the Islamic Azad University, Mashhad, student in software computer engineering in the Science and Iran, in 2007. She is PHD Research Branch of the Islamic Azad University, Tehran, Iran, in 2016. She joined the Department of software computer Engineering, Azad University of Mashhad, as a Lecturer, in 2006 and promoted to a Faculty member in 2009. Her current research interests include machine vision, Big data and natural language processing. She is a Life Member of the Computer Vision Society of Iran.

Kabir University, in Computer Engineering from Amir **Amir Masoud Rahmani** received his BS Sharif University of Technology, MS in Computer Engineering from Tehran, in 1996, the University, Tehran, in Computer Engineering from IAU Tehran, in 1998 and the PhD degree IAU Professor in the Department of Computer Engineering at the Currently, he is a in 2005. publications in technical journals University. He is the author/co-author of more than 150 hoc and research interests are in the areas of distributed systems, ad and conferences. His computing. wireless sensor networks and evolutionary