Digital Communications and Networks 11 (2025) 280–298

Contents lists available at ScienceDirect

Digital Communications and Networks

journal homepage: www.keaipublishing.com/dcan

Integration of data science with the intelligent IoT (IIoT): Current challenges and future perspectives

Inam Ullah [a], Deepak Adhikari [b], Xin Su [c], Francesco Palmieri [d], Celimuge Wu [e], Chang Choi [a,*]

[a] Department of Computer Engineering, Gachon University, Seongnam, Sujeong-gu 13120, Republic of Korea
[b] University of Electronic Science and Technology of China, Chengdu 611731, China
[c] College of Information Science and Engineering, Hohai University, 213000, China
[d] Department of Computer Science, University of Salerno, Italy
[e] Graduate School of Informatics and Engineering, The University of Electro-Communications 1-5-1, Chofugaoka, Chofu-shi, Tokyo, 182-8585, Japan

ARTICLE INFO

Keywords:
Data science
Internet of things (IoT)
Big data
Communication systems
Networks
Security
Data science analytics

ABSTRACT

The Intelligent Internet of Things (IIoT) involves real-world things that communicate or interact with each other through networking technologies by collecting data from these “things” and using intelligent approaches, such as Artificial Intelligence (AI) and machine learning, to make accurate decisions. Data science is the science of dealing with data and its relationships through intelligent approaches. Most state-of-the-art research focuses independently on either data science or IIoT, rather than exploring their integration. Therefore, to address the gap, this article provides a comprehensive survey on the advances and integration of data science with the Intelligent IoT (IIoT) system by classifying the existing IoT-based data science techniques and presenting a summary of various characteristics. The paper analyzes the data science or big data security and privacy features, including network architecture, data protection, and continuous monitoring of data, which face challenges in various IoT-based systems. Extensive insights into IoT data security, privacy, and challenges are visualized in the context of data science for IoT. In addition, this study reveals the current opportunities to enhance data science and IoT market development. The current gap and challenges faced in the integration of data science and IoT are comprehensively presented, followed by the future outlook and possible solutions.

1. Introduction

The Internet of Things (IoT) is a cutting-edge technology that is transforming our daily lives and businesses in areas such as mobile phones, transportation systems, food production, housing, healthcare, clothing, and remote monitoring. Various “things” enable customers to change their habits and even make their lives easier [1]. According to the McKinsey Global Institute (MGI), the IoT will be $3.9-11.1 trillion in output by 2025 across nine different environments, including retail, cities, and factories, and the number of IoT devices is expected to reach 754100 million, equivalent to a global addition of 127 IoT devices per second since 2020 [2]. The operation of an IoT system can be summarized into three stages: sensor deployment for data collection, conversion of the collected information into useful information along with its storage and retrieval, and information transformation into domain knowledge that the IoT system controller will use for user feedback or system response [3]. IoT becomes an Intelligent IoT (IIoT) system by implementing three operations using intelligent methods (i.e., machine learning and deep learning), which expand to boost operational efficiency and avoid unintentional interruption [4,5]. Artificial Intelligence (AI) impacts the IoT with a deluge of investment, new technologies, and business placements. The potent combination of AI and IoT technologies is assisting businesses in avoiding unintentional downtime, increasing operational proficiency, enabling creative goods and services, and enhancing risk management. Corporations developing an IoT strategy, evaluating a prospective new IoT project, or attempting to maximize the value of an existing IoT deployment should consider the role of AI. AI can extract insights from IoT data more precisely than traditional business intelligence tools. Using the two technologies together can bring significant benefits to businesses, including preventing costly unexpected suspensions, enabling new and enhanced goods and services, increasing operational efficiency, and improved risk management.

* Corresponding author.
E-mail address: enduranceaura@gmail.com (C. Choi).

https://doi.org/10.1016/j.dcan.2024.02.007
Received 21 June 2023; Received in revised form 7 February 2024; Accepted 29 February 2024
Available online 6 March 2024

Table 1
Abbreviations list.

Abbreviation	Description	Abbreviation	Description
AI	Artificial Intelligence	IoV	Internet of Vehicles
APIs	Application Programming Interfaces	ITS	Intelligent Transportation Systems
AP	Access Point	M2M	Machine-to-Machine
CC	Climate Corporation	MEC	Mobile Edge Computing
CAD	Computer-Aided Design	MGI	McKinsey Global Institute
CNN	Convolutional Neural Networks	MPI	Message Passing Interface
CRM	Customer Relationship Management	OpenMP	Open Multi-Processing
D2D	Device-to-Device	P2P	Peer-to-Peer
DoS	Denial of Service	QoE	Quality of Experience
FPP	Food Production Process	RFID	Radio Frequency Identification
FPGA	Field-Programmable Gate Arrays	RoI	Region of Interest
GPS	Global Positioning System	UPS	United Parcel Service
IIoT	Intelligent IoT	VANETs	Vehicular Ad hoc Networks
ICT	Information and Communication Technology	VR	Virtual Reality
IoT	Internet of Things	WSN	Wireless Sensor Network

On the other hand, data science is an interdisciplinary scheme in identifying, extracting, and presenting vision into data using data collection, storage, access, analysis, and communication technologies. Data science skills include diagnostic, descriptive, and predictive capabilities, that help users and managers incorporate data science to determine what happened and why and how to deal with the foreseeable consequences. Automation of IIoT is only possible through the proper use of data science and techniques to deal with the challenges encountered by IIoT systems [6,7]. The main difficulties for an IIoT are a) data collection for IoT application functions and how to organize sensor placement and linkages through communication networks, b) determining the technique to deploy AI and machine learning to evaluate and interpret the data obtained from IIoT [8], c) efficiently communicating the evaluated results to IIoT user devices.

1.1. Motivations

IoT data requires efficient and effective analytical methods for a variety of applications. While there are methods to effectively analyze general IoT data, they are mostly focused on specific data types (such as trajectory data). Because IoT devices are often considered for specialized applications (such as traffic monitoring), such systems are beneficial in various settings. IoT data analysis is required for different types of devices considered by different companies, as well as for a single type of device considered by a single company. IoT contexts such as smart cities will increase the prevalence of such a wide variety of devices. This demonstrates the growing need for efficient and effective IoT big data analysis that can handle the peculiarities through the novel innovation of the new systems. Existing systems need to be analyzed to determine what functions they provide and how they lack the functions needed for IoT big data analytics to develop new systems, which is the main motivation of this research.

1.2. Methodology

In this section, we describe the process we used to gather cutting-edge research to address data science claims for IoT and related algorithms and methodologies.
Scope of research: This article aims to provide an overview, categorization, and analysis of all relevant data science techniques for IoT and the challenges existing in data science and big data. Consequently, the purpose of this survey article is to provide answers to the following research questions:

- **RQ1:** What are the sources of data science in the IoT domain? How should a large volume of data be handled in an IoT environment?
- **RQ2:** What platforms or applications are used to handle big data in IoT communication systems?
- **RQ3:** What are the key challenges or issues concerning the data science techniques in IoT?
- **RQ4:** What are the possible future perspectives or possible solutions for handling data in an IoT environment?

The above highlighted research concerns are addressed by extracting key evidence from numerous databases, including Research Gate, Google Scholar, AC, IEEE, Springer, Elsevier, etc. The most relevant and high-quality articles were selected from a large body of literature available in various databases dealing with data science for IoT or big data analytics based on keywords such as "IoT, big data, communication systems, data science, networks, and security." Several articles were excluded from the study because their methods or applications were repeated, not extended, or already used in other applications. Finally, 144 articles were selected based on the scope of this paper. The abbreviations used in this survey are listed in Table 1.

1.3. Contribution

This work's key contributions are as follows:

- To expand the scope of our research to data science meets IIoT. We also provide a full description of the data science modeling concept, from business challenges to data products and automation, in order to understand its relevance and deliver intelligent services in real-world situations. We also have a brief conversation about what makes data science useful for the IoT.
- To present the full scope of data science for the IoT, including advanced analytics approaches that can be applied to advance the intelligence and capabilities of IoT applications.
- To address the relevance and usefulness of IoT-based approaches in diverse real-world application areas.
- To discuss and highlight the possible applications of data science meets IIoT in many real-world scenarios ranging from industry to personalized everyday life applications where excellent analytics and machine learning models can deliver predictable results.
- To highlight and outline current challenges, potential future research directions, and potential solutions.

1.4. Organization

The remainder of this paper is organized as follows: Section 2 presents the related work. Section 3 introduces the IIoT, while Section 4 focuses on IoT data. Section 5 presents an overview of data science, while Section 6 presents the importance of integrating data science and IIoT. Section 7 presents the performance indicators, and the applications of data science for IoT are illustrated in Section 8. Section 9 and its subsections present the discussion, key challenges, future prospects,

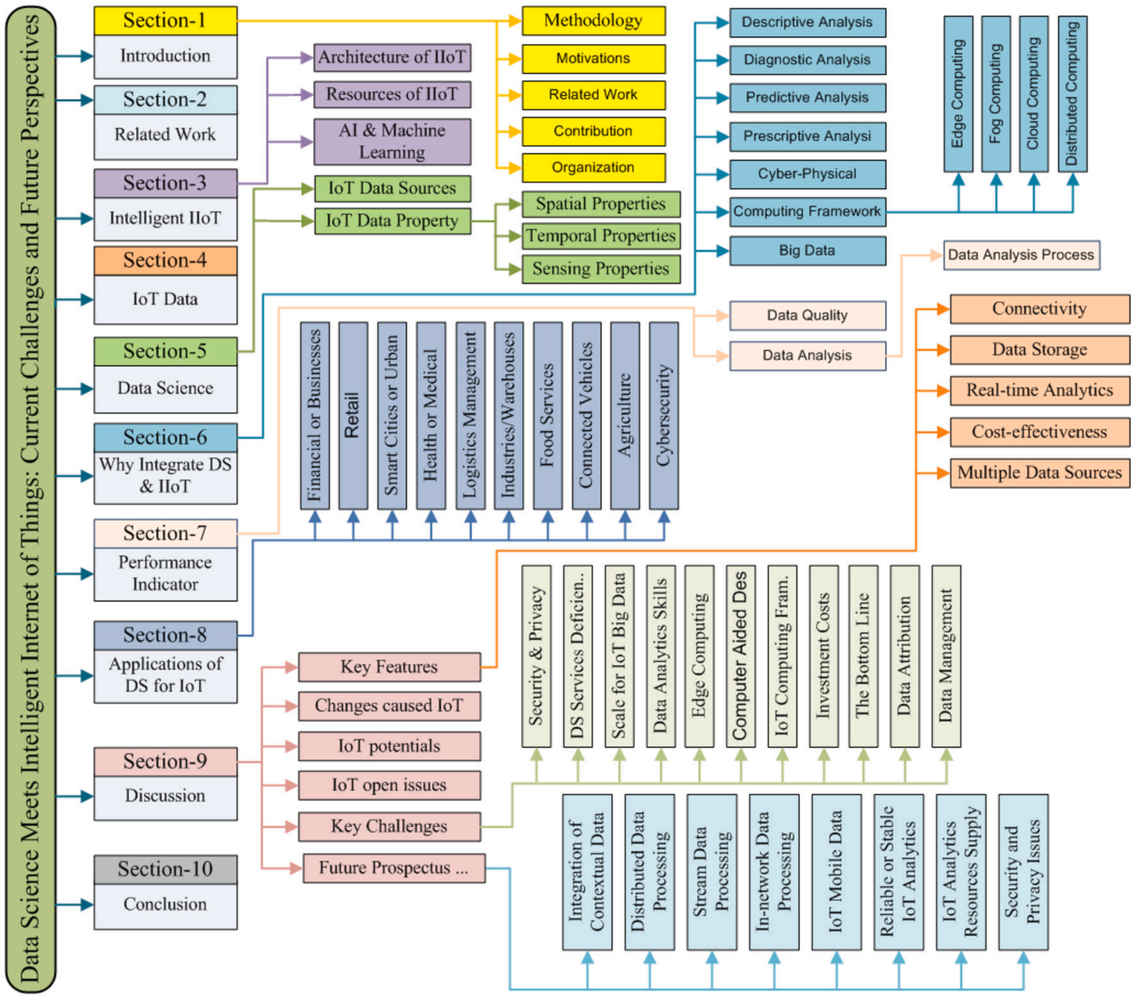

Fig. 1. Taxonomy of data science for IoT.

and possible solutions. Finally, Section 10 concludes the article. A taxonomy of data science for IoT is illustrated in Fig. 1.

2. Related work

To the best of our knowledge, there is no literature that integrates data science with IIoT, along with its applications and challenges. There are several surveys and reviews related to IoT and data science independently. Cao [22] focused on the overview of data science; however, it does not focus on IoT and intelligent approaches. There are a few works that illustrate data analytics in the IoT as well as applications in IoT. Tsai et al. [16,23] reviewed data mining techniques in IoT along with the services and IoT infrastructure. In addition, the main focus was only on offline data mining. Surveys on the IoT architecture, services, and its operating system are addressed in [24,17]; however, data science was excluded. Some [25,26] included IoT data but were limited to certain topics, for instance, spatial data, storage, and computing platforms. System designs, including batch and stream processing systems, were outlined by Marjani et al. [27], but specific existing analytics systems, their features, and their limitations were not covered. One crucial component of IIoT big data was the availability of effective algorithms for processing IoT data [6]. A variety of algorithms had been applied in IoT contexts to maximize the utility of IoT big data, including predictive algorithms for data mining [28], streaming data analytics [13], analytics [10], [11], smart cities [29,9], smart buildings [14], smart healthcare [19], and deep learning [4]. None of the above literature related to IoT focuses on data science. Therefore, this review focuses on filling the gap in the literature by integrating data science with IoT along with its importance, challenges, and applications. One of the critical components of the IoT is the availability of effective algorithms for data processing using intelligent techniques. These algorithms are part of the analytics, which is not part of the survey. A comparative analysis of the existing surveys and reviews is presented in Table 2.

3. Intelligent Internet of Things (IIoT)

IoT is a broad concept, and a variety of definitions are available, some of which are listed in [30–32]. In general, IoT refers to an adaptive and self-configuring network that allows different objects or things, such as sensors, Radio Frequency Identification (RFID) tags, phones, and actuators, to cooperate with each other through an exclusive addressing structure to achieve a common goal [31]. IIoT integrates IoT with intelligent techniques, including deep learning and machine learning, for reliable and secure monitoring and decision-making. In addition, some of the key features of IoT are listed below [33–35]:

- **Sensing capability:** In IoT networks, sensing is the most important capability where the “things” (in the IoT context) can accomplish their sensing responsibilities.
- **Heterogeneity:** An IoT network can be heterogeneous because it can support different networks, for example, wireless, wired, and cellular, and various devices for communication (e.g., Access Point (AP)-based and Peer-to-Peer (P2P)-based styles).

Table 2
Summary of the existing surveys and reviews that are focusing on some of the areas of data science and IoT.

Ref.	Year	IoT	Data Science	Analytics/ Data Mining	Architec- ture	Computing Platform	Applica- tions	Key Challenges	Future Directions	Remarks (focused on)
[4,9]	2018	√		√	√	√	√	√	√	Deep learning and IoT big data analytics
[10]	2018	√		√			√	√		IoT data analytics
[11]	2021	√		√		√	√	√		IoT big data analytics
[12]	2013	√		√				√	√	IoT data mining techniques
[13]	2017	√		√	√		√	√	√	Network methodologies for real-time IoT data
[14]	2020	√		√			√			IoT smart building
[15]	2018	√		√		√	√	√		IoT and machine learning methods
[16]	2018	√		√	√		√	√	√	IoT data analytics
[17]	2015	√			√		√	√		IoT architecture and applications
[18]	2021	√		√			√	√	√	IoT big data analytic for smart things
[19]	2021	√		√			√	√	√	IoT health care
[20]	2018	√		√			√	√	√	IoT agriculture
[21]	2020	√		√	√		√	√	√	Supply chain decision making using IoT big data analytics
Our	2024	√	√	√	√	√	√	√	√	Integration of IoT and data science

- **Addressing modes:** IoT can address different modes, such as unicast/anycast/multicast/broadcast transmissions.
- **High reliability:** An IoT assures connectivity and decent transmissions based on diverse solutions.
- **Self-capabilities:** IoT self-capabilities include (a) high configuration sovereignty, (b) self-adaptation and organization to varying situations, and (c) self-processing of massive data.
- **Secure environment:** IoT also guarantees strength in security issues such as network attacks (e.g., hacking and Denial of Service (DoS)), confidentiality of data transfer, authentication, data integrity, data privacy, and a reliable atmosphere.

3.1. Architecture of IIoT

To support the characteristics discussed above, a variety of IoT architectures have been presented in the literature [31,32,36–38]. Fig. 2 shows a general IoT architecture with the following distinct layers:

- **Devices:** This layer includes lower level devices such as sensors, smart devices, and RFID tags. Since these devices have limited storage and processing resources, their roles are to accomplish only basic operations, i.e., to collect data from the physical things (e.g., environmental conditions or surveillance Region of Interest (RoI) [39]. Further actions of the devices are to connect to Internet gateways for data aggregation/forwarding [40].
- **Communications and networking:** This layer comprises data transmission and networking infrastructures that transport data from physical layer devices to higher-level layers, such as the cloud.
- **Platform and data storage:** This layer comprises hardware and platforms in data centers or cloud services to provide data storage and access facilities.
- **Data management and processing:** This layer works as application software that offers access facilities for IoT users.

3.2. Resources or services of IIoT

Since IoT is often a large-scale heterogeneous system with numerous services and resources, resource management is crucial to efficiently deliver IoT services and resources to clients [41]. Generally, the following IoT resources and services can be managed by using pricing and market models [42–46]:

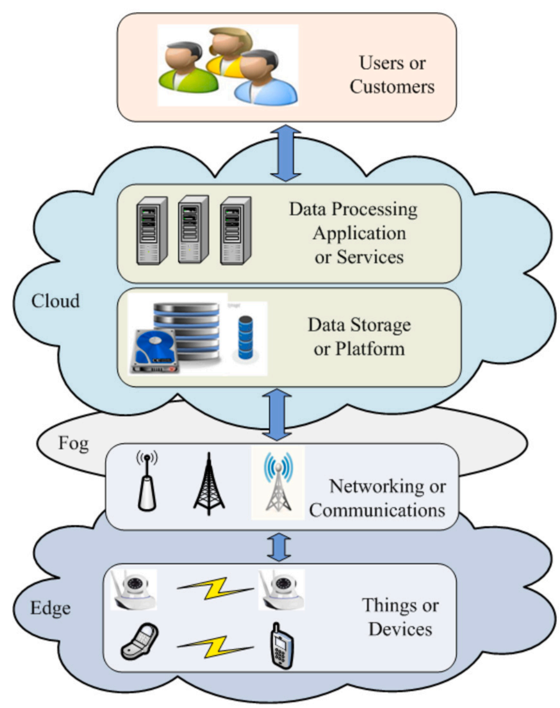

Fig. 2. IoT general architecture along with computing platforms.

- **Sensing data:** Sensing data from IoT devices is acquired at the physical layer. The sensing information generated from the sensing data can be traded or priced to maximize the owner's earnings.
- **Energy:** Sensors, APs, base stations, servers, and other IoT components all need energy to function. Energy providers can increase revenues in monopoly markets by lowering the energy costs of IoT components. Renewable energy harvesting has recently emerged as a potential method to provide "self-powering" of IoT sensors via energy cooperation when used with the smart grid.
- **Cloud services:** This service could be offered to IoT customers as cloud computing and data storage services [47].
- **Network and spectrum bandwidth:** Spectrum and bandwidth are valuable data transmission resources in wireless networks. For example, in cognitive networks, the spectrum is dynamically shared

between licensed and unlicensed users to increase revenue and optimize spectrum usage. In fact, caching can help conserve spectrum and network capacity in a large-scale IoT network by utilizing sensory data and information stored in the cache, as long as the information remains up-to-date and short-term accurate, thereby optimizing resource usage.
- **Data and information services:** These services can be provided and combined to help IoT applications provide services to IoT consumers, such as data mining, searching information, and data security.
- **Location-based services:** The IoT provides location services to consumers, businesses, and the government by using real-time geographical data from user devices, such as tablets or smartphones. These include indoor and outdoor location-based services, such as locating a person or event and identifying nearby venues, such as coffee shops and restaurants.

3.3. AI and machine learning

AI is a branch of computer science in which intelligent machines are trained to behave like humans. Machine learning is a key component of AI that seeks to predict outcomes without being programmed. These algorithms are used in IIoT to predict machine outcomes using big data. The IoT is changing the technology trend with its highly connected devices or things, data collection and analysis. AI and machine learning are two critical technologies that improve the user experience, system responsiveness, and automation of IoT. Machine learning allows machines to learn how to perform tasks depending on new data inputs. Machine learning enables computers to evaluate data more quickly and identify patterns for future predictions. Machine learning-based methods and technologies have emerged in AI in recent years, and the convergence of ML and IoT will complement each other to improve the impact and accessibility of a wide range of services, including agriculture, healthcare, supply chain, finance, transportation, retail, and power.

4. IoT data

In practice, all “things” in the IoT can generate a data overflow, including numerous forms of important information or data. However, in recent years, technical concerns and challenges have emerged regarding the handling of this data and the extraction of usable information. Some of the key IoT data sources, IoT data properties, and temporal and sensing properties are discussed in the following subsections.

4.1. IoT data sources

IoT data can be generated in many forms and collected at different time scales and rates [48]. Data is typically generated from two sources: Things (sensors and Global Positioning System (GPS)) and People (social network services). Things (sensors) provide quantitative observations and assess physical phenomena at different levels of accuracy. In addition, these measurements can be produced in a variety of formats, such as camera images, GPS text, and satellite audio. Social sensors, on the other hand, provide qualitative scenario observation that is concise and fast. An IoT application (real-time flood prediction and warning) must incorporate social sensor data and machine learning to provide balanced and corroborative evidence. This cumulative data can be intelligently annotated to trigger events of interest to the participants. A major challenge in data science research is classifying IoT data sources that are best suited for specific IoT applications. To adequately answer this question, the five problems outlined by Baltrusaitis and colleagues [49] must be addressed:

1. Representation: To enable numerous modalities, arrange and portray the data while taking advantage of the complementarity and severance of different data sources.
2. Translation: To construct data from one medium to another, i.e., provide a translator that permits the modalities to engage with one another to allow data transmission.
3. Alignment: To detect the relationship between modalities, which necessitates the discovery of correlations between different types of data.
4. Fusion: To combine data from several modes.
5. Co-learning: To transport knowledge across modes. This looks at how modality knowledge can benefit a computational model trained in dissimilar ways.

4.2. IoT data property

IoT data can be classified based on its qualities, such as spatial, temporal, and sensory [11]. Each property is explained first, followed by different examples. IoT data can be treated as spatiotemporal Web data, such as e-commerce, social media, weather, logs, Web services, etc. As a result, these spatiotemporal data are incorporated into IoT data.

4.2.1. Spatial properties

These IoT data properties have different types that affect the optimal system selection. In IoT data analysis, initially, it is important to acknowledge whether an IoT device is mobile or fixed. For example, the street lamp’s environmental sensors are fixed, while cars have mobile spatial information. Secondly, the shape of IoT data needs to be considered. The spatial data comprise points (i.e., latitude and longitude), line strings, areas (i.e., polygons and rectangles), and various points and areas that are not linked. For example, weather information/data and road networks are correspondingly characterized by strings of lines and polygons. Third, the spatial data information is 1) 2D or 3D and 2) relative or absolute locations. For example, drones have 3D absolute locations, and cars in workshops have 2D relative locations. Finally, spatial information has limitations (i.e., cars can only drive on the roads). These limitations are often exploited by preprocessing methods to avoid incorrect analysis results.

4.2.2. Temporal properties

IoT data contains temporal information, and data updates should be addressed for temporal data aspects. First, whether the data is updated regularly depends on whether there are useful approaches for storing IoT data and load balancing it for data stream processing [50]. The IoT data can be conveniently partitioned without temporal variation if the data is updated regularly. Otherwise, special approaches to partitioning IoT data may be required to avoid temporal variation. Second, the update frequency might take on a variety of values. Data compaction can work successfully when data is updated regularly and frequently since the same or comparable data will probably be attained. To maintain minimal latency, IoT devices with high-frequency data gathering should be handled favorably in stream processing.

4.2.3. Sensing property

IoT data typically has sensing features that vary widely in type, such as text, photos, numerical values, and labels. In addition, because IoT devices are distributed differently for each society, the names of the characteristics (e.g., temp or temperature) and units (e.g., Fahrenheit or Celsius) are often different, even when the IoT devices are measuring the same data. The ability to sense values is also a necessary characteristic. Some devices or sensors are quite precise, whereas others are not. Normally, sensing values contain mistakes, noise, and erroneous data. To correctly investigate IoT data, IoT systems must be clever enough to manage these inaccurate numbers.

Examples of IoT data include connected vehicles or traffic management (e.g., AI-enabled cameras, motion sensors, and onboard computers), smart grids, SmartSantander projects for IoT experimentation, data aggregation schemes (i.e., cryptographic accumulator), environmental/weather monitoring (e.g., spatial, temporal, sensing), smart homes

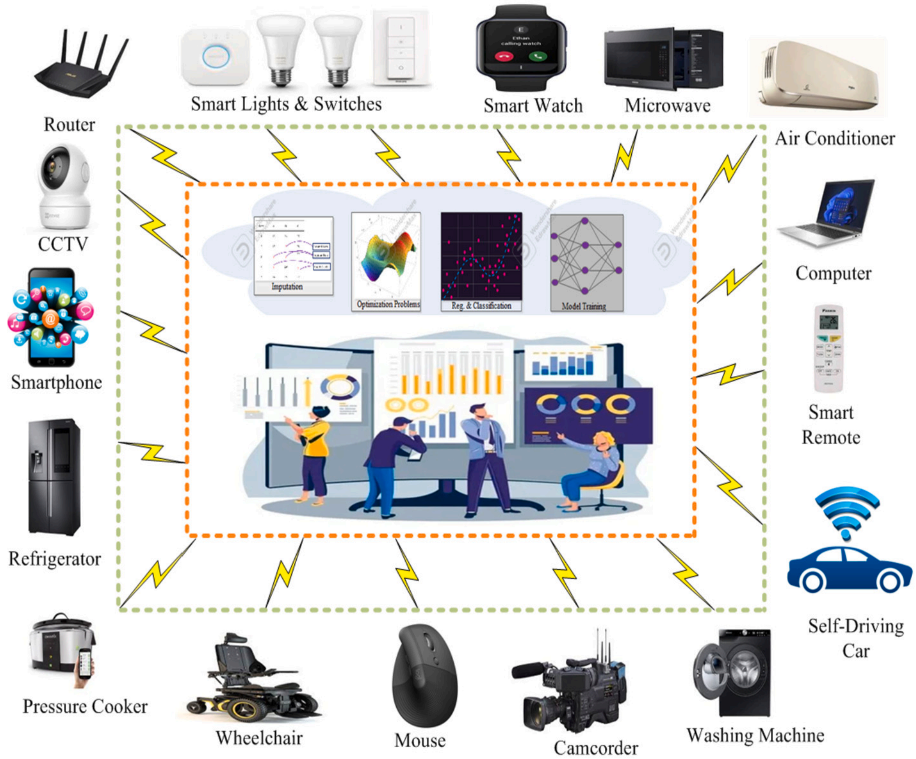

Fig. 3. The architecture incorporating IoT and data science. The outer layer is the IoT data collection layer, the middle layer shows the network layer, and the inner layer shows the computational layer.

and buildings, smart cities, industrial, agricultural management, supply chain management, healthcare, construction, telecoms, retail, and so on [51,52].

5. Data science

Data science is the process of extracting actionable perspectives from raw corporate data. These perspectives enable firms to increase revenue, cut costs, find new possibilities, and improve customer experiences. Data science is thriving in IoT initiatives, providing the tools and strategies to transform raw data into valuable knowledge with the potential to enhance business processes, optimize operations, and generate new revenue streams. A general architecture incorporating IoT and data science is shown in Fig. 3. The outer layer is the data collected through IoT, the middle layer is the network layer and the inner layer is the computational layer, where the data is processed and visualized after intelligent processing. Data science can drive business outcomes in a number of ways, some of which are as follows:

- **Modernize operations:** IoT data helps monitor equipment, facilities, and processes. Data scientists can create models that detect patterns and trends in order to identify potential difficulties, forecast future performance, and keep things running smoothly.
- **Increase security:** IoT devices can be vulnerable to cyber-criminal attacks. Data scientists use data analysis techniques to detect anomalies and potential security issues.
- **Improve customer experiences:** IoT data allows us to better understand our customers' interests and behaviors. Data scientists use this data to personalize experiences, improve products, and discover new revenue streams.
- **Overcome data processing issues at scale:** IoT initiatives generate large data that must be examined and analyzed quickly. Data scientists use tools such as distributed computing and cloud computing to ensure that an IoT project scales up smoothly.
- **Identify new business opportunities:** IoT data can uncover untapped business goldmines and aid in developing innovative products and services. Consider data scientists to be treasure hunters who use data to uncover fascinating new prospects.

The phrase "data science" was most likely first used in the preface of Naur's book in 1974 [53]. That defined data science as "the science of dealing with data, once they have been established, while the relationship of the data to what they represent is assigned to other fields and sciences." Another term, "datalogy" was first used in 1968 [54] to describe "the science of data and processes". These two definitions are more particular than the ones we have been discussing. They have, however, encouraged today's substantial shift toward complete scientific content and progress.

The phrase "data analytics" refers to the ideas, techniques, and tools that are capable of in-depth knowledge and finding actionable insight into data. Data analytics is divided into three categories: descriptive, predictive, and prescriptive. On the other hand, "data mining" is the method of extracting meaningful data, patterns, and trends from raw data. Data analysis is a method for investigating, analyzing, and demonstrating data in order to discover relevant information. The data mining output reveals the data pattern. The key variances between data analytics and data mining are listed in Fig. 4.

6. Why integrate data science and IIoT?

IoT is primarily concerned with machines and computers using networks to "exchange" with one another, a process entirely based on transmitting relevant data. As a result, if data is fuel to power IoT, data science IoT algorithms convert that fuel into something useful. As a result, we feel that data science plays an essential value-generating role in IoT systems. Data science takes and uses data collected by IoT systems and technology to transform it into something that can create value for an organization or business through analysis and visualization. The data

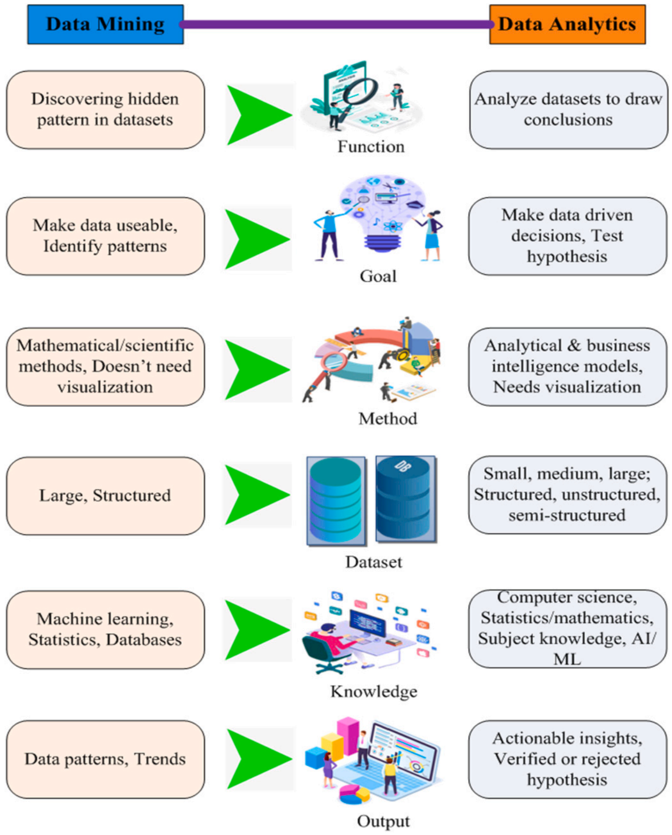

Fig. 4. Key variances between data mining and data analytics.

science component enables value to be derived and understood from using and deploying IoT technology.

Machine learning is used by virtual assistants such as "Amazon Alexa" to fuel their speech recognition functions. The advanced hardware and networking configurations that Alexa devices rely on would be useless if data scientists were not present to allow their fundamental voice-command capabilities. Data science supports several IoT-centric capabilities, including speech recognition. IoT devices use a wide range of data science skills to influence the basic operations required.

The purpose of AI, IoT, and data science in digital communication networks is to stimulate corporate and academic collaboration in changing networks and systems by using the capabilities of the three technologies. Advances in IoT, mobility, and fixed communication protocols have paved the way for a data-driven society. Increased M2M connectivity has generated new workloads, requiring well-organized and stable infrastructures. Because of the variety of applications and traffic, agile networks with self-improvement capabilities will be required to ensure high reliability and extremely quick response times. Network devices, agents, sensors, meters, intelligent vehicles and systems generate large amounts of data, putting security and reliability measures to the test. Such challenges need cutting-edge technologies for successful services and analytics management tools to predict network activity before problems develop. Furthermore, using data mining to detect Quality of Experience (QoE) signals might boost customer satisfaction.

While there are many fascinating data science applications in an IoT world, they also present various additional challenges compared to standard data science applications. Utilizing data science, IoT can be controlled profusely and enormously for various real-time applications that have been connected with real-time data and computing breakthroughs (Fig. 5), as explained below [55–57]:

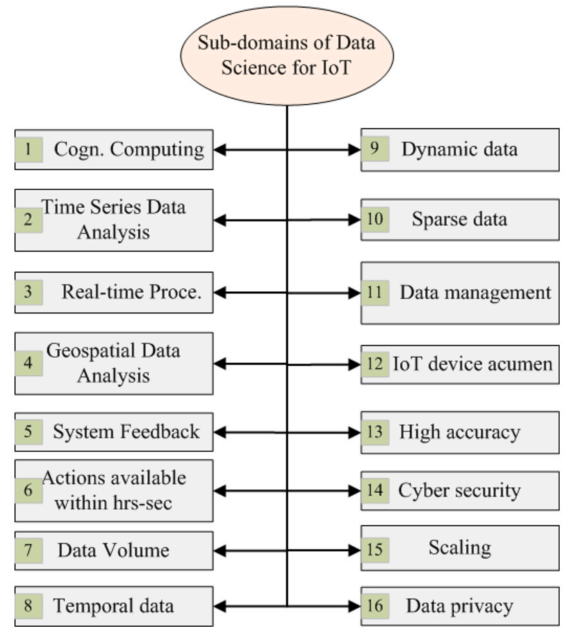

Fig. 5. Sub-domains of data science for IoT.

Cognitive Computing: Cognitive computing extends this research approach to address areas that were previously impossible, such as business comprehension and estimation.

Time Series Data Analysis: It is a data collection organized chronologically. The most widely used way to depict time series data is an essential outline, where the level hub plots the changes in time and the vertical hub depicts the estimated variable.

Real-time Processing: It involves a continuous information process and data output. Real-time data processing and analytics are critical for a company to act rapidly when acting within min or sec.

Geospatial Data Analysis: Geospatial data science addresses several critical emerging innovation and financial progress concerns.

Systems feedback: Data insights alone are insufficient because IoT systems necessitate machine input in response to predefined activities. Insights must be interpretable and intelligible by machines and/or people to turn them into actions instantly and unambiguously, and processes must be built so that they can act directly on those insights.

Actions must be available within hrs/min/sec: Automation is essential at every stage of the data science process. Feedback and actions must be returned to or generated by the IoT device regularly and continuously. Manual therapies are not appropriate currently. The acts are mostly part of Machine-to-Machine (M2M) interactions.

Data volume: Not many fields of data science have to deal with such a large amount of data. IoT gadgets produce more data than social media. The entire data science process, from data collection to pre-processing, algorithmic analysis, and the generation of insights and recommendations, must process large amounts of data in real time.

Temporal data: Every piece of information or data is time-dependent. Models and algorithms must take this into account. As a result, approaches from time series analysis and signal processing are frequently combined.

Dynamic data: In real-time, data is collected and evaluated. This means that data is always changing in real time. Self-acting modifications must be reflected in algorithms. The models must incorporate real-time continuous learning from the data.

Sparse data: Although there is a lot of data, most of it is noise. Anomalies that forecast the breakdown of a production machine many weeks before the incident or a slight divergence from the required production quality are examples of signals of interest. Finding these insights and converting them into precise actions necessitates novel approaches to analyzing sparse data.

Data management: There are many challenges to managing IoT data. It begins with a vast volume of data. It progresses through the heterogeneity of the data from various sensors and devices, efficient management between cloud and edge, instant processing through all data science phases, and feedback to the devices [58]. A data management system must serve all of these.

IoT device acumen: Every sensor and device has its own characteristics, such as data accuracy and quality and transmission technology (Wi-Fi, Bluetooth, radio frequency, etc.). Due to advances in technology, items produced three months later behave slightly differently than the same type of items produced three months earlier. Accurate insights necessitate an extensive understanding of the properties of specific IoT devices.

High accuracy: The insights must be very accurate, with low false-positive and false-negative rates. People and robots are provided with real-time feedback. The wrong input can endanger a worker or cause equipment or processes to fail.

Cyber security: Cyber-physical IoT systems require adequate protection against cyber-attacks. Nobody wants a complete production plant shutdown or sensitive health data on the darknet. This must be built in from the start, and close engagement with cybersecurity professionals is necessary.

Scaling: When a sensor is replaced or added, the data format, accuracy, kind, and volume will be changed. These scaling issues must be understood and addressed before constructing an IoT ecosystem.

Data privacy: Many countries have privacy rules, particularly for health data. Even without regulation, there are inherent benefits to preventing the misuse of customer data. The resulting reputational damage and client loss can impair the business's profitability, lead to lawsuits, and result in fines.

Data can be utilized in various situations to answer questions and assist decisions. It can help to be familiar with the four forms of data analysis often employed to determine the best technique to examine your data [10]. Data science is used in various applications, the four main ways to analyze data are listed below:

6.1. Descriptive analysis

It examines data to learn about what occurred or is occurring in the data context and differentiates through data representations, such as line graphs, bar charts, pie charts, and tables that produce narratives. For instance, a flight booking service can keep track of the ticket number purchased daily. For this service, the descriptive analysis will reveal booking slumps, booking spikes, and high-performing periods.

6.2. Diagnostic analysis

Diagnostic analytics thoroughly evaluates data to determine why something is happening. Methods such as drill-down, data mining, data discovery, and correlations are used to categorize it [59]. To identify new patterns in these methods, a given dataset can be subjected to many data manipulations and transformations. For example, the airline may focus on a predominantly high-performing day to better understand the booking surge. This could lead to the discovery that several consumers travel to a specific country every month to attend a sporting or other event.

6.3. Predictive analysis

This analysis uses historical data to make accurate predictions about data patterns that may occur in the future. Methods such as machine learning, pattern matching, forecasting, and predictive modeling are used to classify them. In each approach, computers are trained to reverse-engineer causal relationships in the data. For instance, at the start of every year, the airline team might use data science to forecast flight booking patterns for the future year. The computer software or algorithm can use historical data to forecast April booking increases for specific terminuses. After estimating its customers' future travel demands, the corporation might begin targeted marketing for those cities in January.

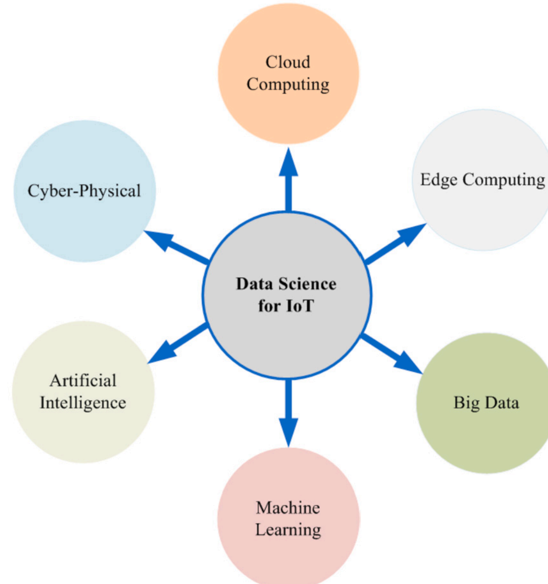

Fig. 6. Data science for IoT enabling technologies.

6.4. Prescriptive analysis

The analysis elevates predictive data to new heights. It not only forecasts what is likely to happen, but it also offers the best way to respond to that outcome. It can assess the possible consequences of several options and recommend the optimal course of action. It uses graph analysis, complex event processing, modeling, neural networks, and machine learning recommendation engines. In the case of flight bookings, the prescriptive analysis can look at previous marketing initiatives to increase the value of the upcoming booking surge. A data scientist could predict booking results for different levels of marketing investment across many marketing channels. These data projections would increase the flight booking company's trust in their marketing policies.

To enable such advanced analysis, a range of technologies are utilized. These technologies aim to monitor, evaluate, and automatize the streamlining process to improve it and reduce production costs. Cloud computing, cyber-physical systems, AI and machine learning, edge computing, and big data are among the key technologies that support the IIoT; see Fig. 6. Data is generated daily in several areas, particularly in the IoT and industrial situations [60,61]. As a result, businesses and professionals are demanding better tools and technologies to discover, collect and assess every insight and anomaly in every process. Data science and machine learning technologies can help us make the most of the data collected by industrial equipment.

6.5. Cyber-physical

A cyber-physical system is a computer system in which a computer-based algorithm controls or monitors the mechanism. Software and hardware are closely intertwined in the system, permitting them to connect and exchange data on the cloud. Data science is a fantastic tool for finding abnormalities, performance concerns, and cost-saving detection in algorithms and machines. For example, when monitoring a cyber-physical process, the algorithm will register data from the hardware

Table 3
Comparison of differences between edge and cloud computing.

Edge Computing	Cloud Computing
Basic data analytics, short-term data historian features	Big data mining
Basic data visualization	Complex analytics
D2D communications	Long time data storage
Data pre-processing, filtering, optimization, and cleaning	Rules of machine learning
Aggregation of some data	Advanced visualization
Data caching, streaming, buffering	Business logic sources

and adapt its behavior based on the task at hand. Occasionally, however, the software or hardware will deviate from its intended function, causing problems throughout the supply chain. The data scientist can identify any outliers, changes in events or even drift in the data set by performing an anomaly detection analysis. In the case of short-term anomalies, identification can save anything from a few seconds to hours of work, which might result in a significant loss of opportunities.

6.6. Computing framework

6.6.1. Edge computing

It is a form of distributed computing in which computer data storage is moved closer to the point of consumption [62]. Unlike cloud computing, edge computing is decentralized data processing at the network's edge [63]. The key benefits of the edge are real-time processing and machine control. It supports basic data visualization/analytics, continuous data flow, pre-processing, and Device-to-Device (D2D) connectivity. Access to an edge computing device permits one to analyze machinery and swiftly determine its functioning and efficiency [64]. In a windmill field, for example, cloud-based solution can provide the critical data needed to analyze the turbine without relying on spotty cellular coverage [65]. A comparison table of edge and cloud computing is shown in Table 3.

6.6.2. Fog computing

Fog computing, also known as fog networking, is moving computer applications, data, and associated services away from the central cloud and into the networked edge's logical stream [66]. Instead of being predominantly governed by the network gateways embedded in the LTE network, the fog networking system aims to build configuration and management via the Internet infrastructure. The fog computing architecture can be defined as a mainly virtualized computing infrastructure that provides classified computing services via edge server nodes. These nodes organize various applications and services to store or process data near end users. “Edge computing” is sometimes used interchangeably with “fog computing.” However, there is a subtle difference between the two theories. Both fog and edge computing require moving processing and intelligence closer to the source of information. The fundamental distinction between the two networks is the location of computers and the amount of intellectual power available. Both topologies send data from the same sources, such as sensors, pumps, motors, etc. All equipment in this setting performs a physical function, such as water pumping, electrical circuits, switching, or sensing the work at hand [67,68].

6.6.3. Cloud computing

The on-demand accessibility of computer system resources for storage and computational power without the need to own the hardware is known as cloud computing. IoT uses cloud computing for complicated analytics, massive data mining, improved visualization, and long-term storage [69]. Data centralization is also one of the key benefits. For example, it would be easier to collect all the data in a windmill field than to travel to each turbine and download it. Typically, data is pushed to a server and retrieved by a client; cloud computing is unsuitable for real-time data in supply chains where speed is critical. On the other hand, the cloud computing system is suited for more intricate research requiring significant computer capacity. Predictive maintenance, for example, is a service that uses cloud computing to detect when a machine demands to be serviced [70].

6.6.4. Distributed computing

Computing has changed dramatically in recent years in terms of platform and environment. With the advent of the Internet, centralized computing became obsolete, and a new computer paradigm known as distributed computing emerged. Distributed computing is a subfield of computer science and engineering that deals with solving computational problems in distributed systems. Distributed computing divides a task into multiple jobs, each of which is solved by one or more computer entities. This system contains many independent computational units with local memory, message-passing communication between computational groups, fault tolerance in each computing group, and each computing element has a limited view of the complete system. When data is generated in one physical location and required in another, distributed computing is critical. Distributed systems have several advantages over single-system computing, including scalability, availability, consistency, transparency, and efficiency.

6.7. Big data

While cloud computing is about computing power and storage, big data is about analyzing a large amount of data that requires a large amount of computing power [71]. Big data computing is critical in an industrial environment because of all the devices attached to machines (visual sensors, heat sensors, communication protocols, Wi-Fi, Bluetooth, etc.). Once the captor had studied the data problem, it became easier to predict the development of the equipment (machine failure, maintenance analysis) and its efficiency (productivity, predicted productivity).

7. Performance indicators

Performance indicators are essential to ensure better quality results. In the case of data science and IoT, different performances are used in terms of achieving data quality, decision-making, and others, which are discussed below:

7.1. Data quality

Data quality is essentially a measure of how well a dataset is suited to perform its unique function [72]. Data quality metrics are based on data quality characteristics such as accuracy, thoroughness, reliability, validity, distinctiveness and timeliness.

Data quality refers to the establishment and implementation of actions that use quality management methods to ensure data is appropriate to satisfy an organization's specific needs in a given scenario [73]. High-quality data has been determined to be fit for its intended use. Data quality issues include duplicated data, inconsistent data, incomplete data, incorrect data, badly ordered data, badly described data, and inadequate data security. The evaluations are carried out by data quality analysts, who look at and analyze each data quality measure, generate an overall data quality score and provide companies with a percentage that reflects the accuracy of their data. A low score on the data quality scoreboard implies poor data quality, which could lead to more effective and correct decisions.

Data quality standards are an integral part of data governance, which is the process of creating and implementing a specific, agreed set of rules and standards for managing all data across an organization. A successful data governance program ought to integrate data from several sources, establish and monitor data usage policies, and eradicate errors and inaccuracies that jeopardize the reliability and integrity of data analytics.

7.2. Data analysis

The process of interacting with data to gain useful information that can then be used to make better decisions is known as data analysis. We can make better decisions when we can derive meaning from data. And we live in an era where there is a lot of data available. Organizations are becoming more aware of the advantages of using data. Data analysis can assist a bank in customizing customer service, a healthcare system to forecast future health needs or a media corporation in developing the next big streaming hit [74,18].

The need for an effective and productive system for capturing the value of data grows in parallel with the number and complexity of data available to enterprises. Typically, the data analysis process consists of several repeated rounds. Let us investigate each of them more closely.

- *Identify* the business question you want to react to. What problem is the company trying to solve? What must be assessed, and how will this be accomplished?
- *Gather* or collect the raw datasets you'll need to solve the desired inquiry. Data can be collected from both internal sources, such as a company's Customer Relationship Management (CRM) software, and other sources, such as government records or social media Application Programming Interfaces (APIs).
- *Clean* the data before analyzing it. Common duties include purifying aberrant and replicated data, reconciling incompatibility, standardizing the structure of data and layout, and tackling white spaces and other syntactic issues.
- *Examine* the existing data. Changing the data with various data analysis methodologies and tools can reveal trends, correlations, outliers, and variations that convey a story. During this step, you can use data mining to discover trends in systems or software for data visualization to assist in transforming data into an understandable graphical format.
- *Evaluate* the findings of your analysis to determine how well the data addressed your initial query. What inferences can you make based on the data? What are the limitations of your conclusions?

8. Applications of data science for IoT

This section will look at the use of data science for IoT, also known as IoT analytics. IoT analytics can help develop new services and business models as we approach the digital age of 50 billion connected devices. If IoT is the backbone of the infrastructure, IoT analytics is the key to extracting valuable insights from the daily amounts of data generated by sensors and devices. Furthermore, the potential of IoT devices and big data analytics is becoming more affordable and widely available, opening up new opportunities for the applications of IoT analytics [10,75] that drive invention and business decisions across industry verticals. While some businesses prefer to regulate their own IoT analytics, there are also IoT data and analytics services businesses specializing in effectively transforming IoT data into business insights. Here are a few examples of IoT analytics and data science use cases:

8.1. Financial or businesses

Generally, financial or business data science is "the study of business or e-commerce data to gain vision about a business that can lead to wise decision-making" [76]. Based on firm historical data, data scientists can develop methods or data-driven models that estimate customer/user actions and find patterns or trends. This allows businesses to cut costs, generate endorsements, and improve service delivery for enhanced decision-making. Finally, corporate intelligence, automation, and efficiency can be attained by data science, which includes numerous sophisticated analytics methodologies and machine learning models based on acquired data. Considering predictive modeling based on machine learning approaches, several online vendors, such as Amazon, can advance inventory management, eliminate out-of-stock conditions, and advance transport and storage [77]. Financial organizations use historical data to make high-stakes business choices, such as fraud prevention, risk management, credit allocation, trading algorithms, tailored services, customer analytics, etc. Overall, data science and IoT technologies have the potential to play a significant role in the next generation of economic sectors, particularly in business intelligence, automation, and intelligent decision-making systems.

8.2. Retail industry

It's no surprise that IoT is transforming the retail industry, with 70% of global retailers planning to invest in IoT to improve their business strategies. Retailers can plan orders and maximize demand coverage by controlling stock, estimating product demand, ensuring customer satisfaction, and increasing customer experiences.

IoT and Big Data are two industry-leading technologies that are having a significant impact on the retail industry. They offer businesses new ways to get to know their customers and provide them with tailored customer journeys with product suggestions and tailored experiences based on their previous choices. According to a 2019 McKinsey report [78], integrating personalized recommendations and triggering messages with these technologies has resulted in actual gains in consumer engagement and spending for retail firms. Some have seen a 5%-15% rise in sales and a 10%-30% boost in marketing spending efficiency. When it comes to assisting business decision-making procedures, IoT analytics have enormous potential [79]. For example, combining IoT with big data in retail delivers more precise customer information: who they are, where they are, their preferences, and much more. Various IoT retail apps leverage this technology to create a comprehensive consumer profile (for example, an application that records client preferences). Sensors in IoT devices can track their habits and behavior, which big data can then collect and review, analyzing the volume, speed, and regularity of various behavioral patterns. The flow of big data in IoT is illustrated in Fig. 7.

8.3. Smart cities or urban

IoT analytics improve urban planning, transportation, crime detection, and sustainability, to name just a few processes and services. Ordinary cities are being transformed into safer, greener, and more efficient communities through the use of sensors, cameras, and data analytics. Nowadays, cities and metropolitan areas house over half of the total population of the globe [80], and they are seen as engines or foci of economic growth, well-being, wealth generation, and social engagement [81]. "Urban area" can also consider adjacent places such as conurbations, suburbs, towns, and cities. As a result, data is collected that records citizens' daily functions, thoughts, perceptions and emotions, which can be broadly classified as personal data, e.g., household, healthcare, employment, education, etc., government data, e.g., statistics about crime in a city, or institutions of government, etc., proprietary data, e.g., banking, online platforms data Open and public data, such as data.gov and government surveys, as well as natural and crowd-sourced data, (i.e., social media and user-generated internet data).

8.4. Health or medical

IoT wearables like smartwatches and fitness trackers are just a few instances of how IoT is assisting in improving our health [82]. Our devices allow us to analyze our behaviors by measuring our sleep, steps, kilometers run, or minutes without moving. Machine learning modeling in healthcare can reduce medical expenses, prevent preventable diseases, anticipate infectious outbreaks, and generally improve life quality [83]. Life expectancy is increasing around the world, creating significant challenges for today's healthcare systems. As a result, health data science modeling can examine existing and historical information

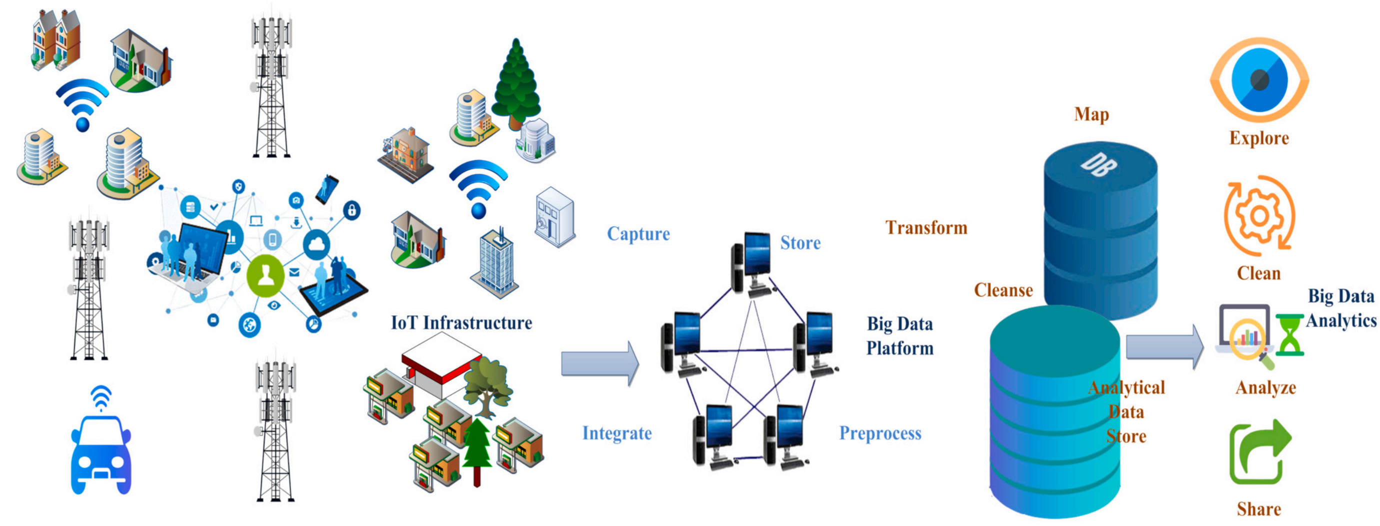

Fig. 7. The big data flow in IoT.

to predict trends, improve services and even track the spread of disease more effectively. It can lead to revolutionary approaches to patient care, diagnosis, clinical knowledge and healthcare management.

8.5. Logistics management

The United Parcel Service (UPS), based in Atlanta, Georgia, has launched a project to improve its IoT delivery performance. Each vehicle in the fleet is equipped with a device that collects sensor data from the environment. Sensors are essential for generating data in the IoT. Sensors of various types can be set to collect data on humidity, temperature, pressure, acceleration, speed, orientation of physical objects in space, smoke, gas, infrared, chemicals, and various other things. Sensors are utilized on UPS vehicles to gather data on about 200 various environmental and operational parameters. UPS has around 110,000 cars in its delivery fleet. Each one contains 200 other pieces of information. It will be combined with metadata (time, date, vehicle number, name, driver, and so on) to make the data more relevant and helpful. Even if the trucks only communicate sensor data from their RFID chips to a centralized RFID reader once per day (which is somewhat helpful but could be more useful), that is at least 22,000,000 different data points every day. This sensor data was used by UPS to gain real-time visibility into its operations. As a result, they chose to send data from the trucks every six seconds, resulting in 14,400 daily uploads and over 300 billion unique data points.

As a result, big data processing and analytics are essential for creating a sense of all this data and obtaining relevant and practical knowledge that can drive business outcomes [84]. UPS boasts on its website that IoT data analytics have helped it save more than $400 million per year. This is largely true for transportation, but there are even more ways to use IoT in logistics and warehousing. Sensors and robotics can be used in warehouses to improve the layout, track catalogs and orders along the supply chain, minimize labor costs, and automate catalog management to minimize errors.

8.6. Industries and warehouses

Industries are moving towards optimization due to IoT analytics in areas such as asset tracking and predictive maintenance. Companies can predict machine failures, reduce maintenance costs, track critical assets to reduce the risk of theft, and optimize maintenance routes. The key categories of industrial data are massive amounts of data devices, corporate operation data, production value chain resources, and cooperation data from external sources [85].

The data must be analyzed, processed, and safeguarded to augment the system's success, scalability, and safety [86]. Thus, data science modeling can be applied in the manufacturing industry to increase output, reduce costs, and increase profits. A data expert is unlikely to work in a large industrial context. Workplaces using beanbags can appear to be a far cry from a factory, mine, or other location where heavy machinery is employed to do duties. Some experts' perspectives will change as a result of the Internet of Things, which describes the projected billions of devices that will collect and transmit data via sensors. With IoT developments, data collected by thousands of sensors on heavy machinery can be captured and processed to provide economic value. Several factors are contributing to the increasing connectivity of heavy machinery and, more broadly, any device equipped with sensors. The availability of Wi-Fi and cellular networks and robust cloud infrastructures enable simple sensor data collection, transmission, storage, and analysis.

8.7. Food services

Restaurants and bars use IoT to monitor inventories and identify more effective ways to handle their everyday operations. I-TAPR2 technologies employ a wireless smart tap that regulates beer flow and assists food service management in determining which items are selling the best, when to buy fresh stock, which beverages to promote, and which to discontinue offerings. As a result, the food sector is looking for automated technologies to help them accomplish more with less and speed up the process with more variety without sacrificing quality to reach a wider consumer base. Using traceability systems provided by Information and Communication Technology (ICT) can advance food safety [87]. The importance of ICT in user-friendliness, simplicity of access, cost-effectiveness, and security is critical in the Food Production Process (FPP). AI, IoT, and cloud computing are only a few technologies created by developing ICT applications with AI. These technologies increase nutrition analysis, quality control, packaging, supply chain management, and FPP, to name a few areas. IoT-enabled food distribution systems may aid in automating the food supply chain. Meanwhile, a mobile-based IoT application can monitor food freshness and quality in real time. IoT has significantly impacted the food industry, with important applications including industrial equipment management, smart refrigerators, energy consumption reduction, stock management, security, and complexity.

8.8. Connected vehicles

The use of IoT analytics in automobiles enables manufacturers to improve a wide range of business processes. For example, predictive maintenance allows manufacturers to identify vehicle maintenance issues before they occur. Because of these sophisticated analytical methods, companies in many industries can use IoT analytics to optimize their processes, reduce operating costs and strengthen brand partnerships. Multiple wireless communication possibilities allow vehicles to communicate with their internal and exterior environments. This type of interconnected car solution is intended to be the next frontier in the automobile transition and a vital source of development toward next-generation Intelligent Transportation Systems (ITSs) [88]. Furthermore, networked vehicles are the foundation for the coming Internet of Vehicles (IoV) [89,90]. Vehicular Ad hoc Networks (VANETs) are another distributed way of linking cars on the move. VANETs have sparked interest due to their prospective roles in IoT applications such as ITS, autonomous driving, etc. [91]. When the data transfer range in VANETs is insufficient to cover all intended receivers, a multi-hop communication protocol is required. Inter-vehicle communication is necessary for various applications, such as collision alert systems and other value-added services. However, because of varying vehicle numbers, vehicle mobility, and the restricted bandwidth of wireless communications, providing a high-packet acceptance ratio and low end-to-end delay is problematic.

8.9. Agriculture

Many farmers have been leveraging data from IoT devices to increase crop production, agricultural planning, and maintenance. The Climate Corporation (CC) employs IoT with sensors to evaluate soil quality and moisture, assisting farmers in determining how to rotate crops and when to rinse them. Farmers also use IoT devices to collect data from farming vehicles and aerial image analytics with IoT drones. IoT nodes are critical in precision agriculture for collecting real-time data [92]. These nodes can make the system appear more accurate by collecting real-time data across crop fields to improve the accuracy of the agricultural system. By applying data analytics and machine learning, the farming system becomes more practical. All these technologies have a wide range of applications in different industries.

Precision agriculture is developing a range of applications for farmers to keep them up to date with the status of their crops. The design of profitable agriculture normally consists of three major stages. The first stage includes a variety of IoT nodes that monitor ambient, plant, and soil conditions; for example, the soil moisture sensor records soil moisture readings, and the soil nutrient sensor checks soil fertility. Second, we must gather this precise data. Depending on the necessity, the data can be kept domestically at the nearest fog node or transferred to the cloud for higher processing and remote monitoring. Analytics methodologies are employed at the architecture's third level to determine crop fields' status. This information is then communicated to the end user, who can then determine whether the reading is above or below the threshold. As a result, they start communicating with the engine, which turns the irrigation system on and off to spray water on the land, or the farmer may need to apply nitrogen, potassium, and phosphate fertilizers to maintain the soil fertility. When a catastrophic circumstance is detected, a reaction mechanism is launched using analytics and actuators. There are several applications of IoT and Wireless Sensor Networks (WSNs) in agriculture: soil selection and planning, crop field irrigation, fertilizers, crop disease, pest management, harvest monitoring, etc. [93].

8.10. Cybersecurity

One of the most important topics of Industry 4.0 is cybersecurity, which protects networks, hardware, systems, and data against digital threats [94]. Data science methods, especially machine learning, have evolved into a critical cybersecurity technology that continuously learns to detect malware in encrypted traffic, detect insider threats, predict where bad neighbors are online, keep people safe while browsing, or protect cloud data by revealing suspicious user activity. For example, deep learning and machine learning-based security models can detect a wide range of anomalies or intrusions [95]. Association rule learning can be very useful in designing rule-based security rules. When employed on a large scale, deep learning-based security models can outperform traditional ones [96]. Thus, by removing illicit insights from security information, data science modeling might enable cybersecurity personnel to avoid dangers and respond in real-time to active attacks proactively.

9. Discussion

This section discusses various incorporation of data science and IoT as an emerging research interest in recent technologies. Similarly, key challenges and future prospects are also discussed.

9.1. Key features

Some of the key features are described as follows:

9.1.1. Connectivity

Connectivity is crucial in data science to permit the collection of large amounts of data from many sources and the subsequent delivery of information to associated managements [97]. One of the key benefits of big data analytics and IoT-based environments for emergency management is the availability of multiple communication technologies. Connectivity between interconnected data nodes and management systems serves as a foundation for ensuring effective moves. Because numerous communication methods are accessible [98], the overall architecture of the environment must adapt to various local and remote communication protocols. Furthermore, with the advancement of post-disaster communication networks, unified connectivity is offered even when other standard communication networks are disrupted in post-disaster settings.

9.1.2. Data storage

It can be more challenging to store diverse data in real time in a traditional management system (e.g., a disaster management system). However, with advancements in big data analytics, large datasets (structured & unstructured) can be efficiently stored on low-cost service platforms [25]. Real-time settings for IoT devices and other data sources with stream storage can improve overall data processing performance and benefit the applications that use them [99]. In addition, big data analytics technology can provide effective data analytics processing with minimal latency while preserving the storage of vast unstructured datasets.

9.1.3. Real-time analytics

Real-time analytics involves using data and related resources for analysis as soon as they arrive in the system. The term "real-time" refers to a level of computer responsiveness that a user perceives as instantaneous or near-instantaneous. Streaming data systems and real-time operational decisions that can be made autonomously via automated robotic processes and policy enforcement are typically associated with the term. Real-time analytics includes CRM, credit scoring, fraud detection, and other applications. Individual retail shoppers are targeted with discounts and incentives while in the store and near the item [13].

Real-time analytics is typically performed at the edge of the network to ensure that data analysis occurs as close to the source of the data as possible. In addition to edge computing, other technologies that enable real-time analytics include the following:

- **In-memory processing** is a chip design that integrates the computer microprocessor into a memory chip to reduce latency.
- **In-memory analytics** is a method of querying data stored in random access memory instead of data stored on physical disks.

- **In-database analytics** is a method of analyzing data within the database itself by applying analytic concepts.
- **Data warehouse equipment** is a collection of software and hardware elements that are built explicitly for analytic computation. An appliance that enables the buyer to instantly build a high-performance data warehouse.
- **Massive parallelism programming** is the synchronized execution of a program by multiple processors working on different parts of the program, each with its own operating system and memory.

9.1.4. Cost effectiveness

Big data analytics tools in data science are often open source, offering a significant cost savings opportunity compared to purchasing proprietary data processing software solutions for some monitoring processes [100]. Cost effectiveness is a critical consideration for concerned management authorities in underdeveloped nations where disaster management is not applied due to a lack of money. Map-reduce is a perfect approach for low-cost storage and reducing overall system processing costs. In addition, with the decreasing cost of IoT software and hardware tools, cutting-edge technologies can be deployed at a limited cost.

9.1.5. Multiple data sources

By integrating IoT environments with numerous data sources such as sensor devices, cameras, mobile phones, etc., multiple data sources can be integrated to gain new and useful insights and information [101,102]. Big data analytics tools assist in data processing [103]. Multiple data sources provide alternative approaches to the problems, which require multidimensional data representations to identify simple patterns for a solution and cannot be obtained from a single data source. Big data analytics and IoT-based management settings can outperform traditional data sources due to the accessibility of multiple data sources.

Most IoT applications contain active, passive, and dynamic data. So the IoT is now “all of the above” and a bag of chocolates. It is time to consider the emerging IoT patterns that provide better ways to define problems and solve them. These patterns focus on the use of data, which is at the heart of the IoT.

- **Passive data** is the ability to receive data from passively communicating sensors. These sensors must be engaged before they can communicate data, and they only provide data when requested (for example, a sensor that measures groundwater saturation only provides existing data when the API is called).
- **Active data** refers to data that is typically streamed from the sensor (e.g., data from a jet engine). These sensors emit data continuously, rather than passive data that we have to inquire.
- **Dynamic data** is the most sophisticated and valuable. These sensor-equipped objects engage in dynamic, two-way communication with IoT software (e.g., smart thermostats). These sensors interact with IoT applications like a conversation.

9.2. Changes caused by IoT

The potential changes caused by IoT are classified into three perspectives: things-oriented, Internet-oriented, and semantic-oriented [31], with the following details:

1. *Things-oriented* is a simple method to distinguish a device if it can inevitably become part of an IoT network, i.e., if it can inevitably be connected to the Internet, both directly and indirectly, despite its capability, via other devices [104]. Another trend is that devices are shrinking in size, which means that battery life can be extended and, as a result, more functionality can be integrated into a single device (e.g., integrating a wireless sensor and RFID into a single device). Such variants allow everything to submit the data it acquired to the server. Then all the linked things or devices on the sphere can come true in the predictable future [12]. As a result of these improvements, data mining technologies can enhance the performance of these devices, giving them basic sovereign decision-making authority. An air conditioner, for example, has the autonomous capability to regulate the temperature and interact with other things automatically.
2. *Internet-oriented*, several IP addresses are required, and a massive volume of data will be generated, potentially overflowing the internet. The good news is that IPv6 can provide a viable solution to the IP address problem, but the massive volume of data will be a major concern. Several experts are focusing on big data issues, as it is clear that IoT data will flood the internet. This data includes data from the server, device, and app interactions as well as data collected from all devices. Another issue is the rapid rate at which devices join and leave the internet environment, as IoT devices may not always be connected to the internet for various reasons. Depending on whether it has data to communicate, a device can switch itself on or off to conserve battery power. Another example is that when automotive electronics are switched on to collect data, they may or may not link to the internet.
3. *Semantics-oriented* is ultimately the so-called smart object. The ability of smart objects to categorize themselves, sense data, make decisions, connect to other devices and services, and communicate information is at their core [105]. In contrast to the standpoint of intelligent things, the author of [31] stated that storing, searching, interconnecting, and organizing data is still a complex problem. In general, the changes in the Internet and things are the result of developments in computer technology; however, the reason for making the IoT smarter is to improve the level of service for this new environment. In conclusion, the changes for the IIoT are determined by what is necessary and what can be done about it, which is no longer limited to incorporating current technology into the system.

9.3. IoT potentials

According to the literature, a variety of IoT applications have been introduced or will be introduced in the near future [106–108]. As the IoT is likely to generate significant benefits in various fields, from advanced research technology to everyday life, research institutions in many countries have turned their attention to IoT revenues. [31] categorized the possible applications into the innovative environment, health care, transportation, and social field. On the other hand, [109] classified the application domain into the industry, community, and environment based on the IoT function on the type of event, such as where it will be applied and who will utilize it. Additionally, the work in [110] presented around 50 applications, which are classified into different smart categories (13 categories) (e.g., smart environment, cities, metering, water, agriculture, farming, security, logistics, retail, industrial management, e-health, and home automation). Currently, most researchers consider simple data mining algorithms to create directions to split the data, but these results are insufficient. Therefore, on the basis of observations, the potential of employing data mining in the IoT is summarized below.

1) To people:
Because the IoT is an intelligent, semi-autonomous system, it can make accurate suggestions to people. Unlike a fully autonomous system, this type of system may require high accuracy (e.g. medical analysis). For systems that require high accuracy, data mining technologies support the user’s decision making rather than participating in the final choice.

2) To themselves:
Thinking by themselves for “things” is one of the encouraging research aspects of the smart object. In practice, data mining has the capacity to filter out unnecessary data and select what type of information must be uploaded that will be relevant for applications on a large scale with finite resources (for example, natural resource monitoring and precision

agriculture). Furthermore, assessing the current situation to provide a possible resolution has additional data mining potential (e.g., animatedly adjusting the air-conditioning for a relaxed environment).

3) To other things:
This is another critical challenge in IoT or M2M connection with other devices or things. However, the emphasis is on something other than passing data to others and performing tasks together. Suppose the objects are analogous to each other and therefore have analogous requirements. In that case, clustering techniques can be used to arrange them into a similar group so that the objects can rapidly determine which devices require the data they carry and which do not.

9.4. IoT open issues

Although the focus of IoT mining challenges is different from traditional mining challenges, they still face several unresolved issues of traditional mining algorithms and challenges in developing IoT mining approaches (e.g., scalability of large datasets). As a result, given today's technologies, some of the open issues are as follows:

1) Infrastructure perspective:
The two features of IoT (heterogeneity and decentralization) significantly affect the development of data mining algorithms [109]. However, most IoT professionals emphasize that IoT data storage and computing must be distributed; most data mining methods are structured in such a way that they require a complete view of the data to find relevant information. As a result, most current research using data mining tools is still based on centralized computing. Therefore, it is critical to investigate how to appropriately decentralize mining technology for IoT systems.

2) Data Perspective:
To address the issue of massive amounts of data entering the system quickly from multiple sources, a range of technologies (e.g., information extraction, data preparation, and information retrieval) are typically examined [111]. However, deciding what kind of data should be stored on sensors and what type of data should be avoided is a difficult challenge for sensors, because the limited memory of the sensor is not the only concern. There are other reasons to filter out unnecessary data to avoid its impact on system effectiveness. Several recent research have attempted to give a practical answer for dealing with the huge data issue from things (i.e., data reduction, contraction, and sampling). For the IoT large data issue, it has been discovered that analysis, deposition, acquisition, and integration contain various outstanding challenges that can significantly affect the IoT system's functionality [112]. Another difficulty is that sensors are utilized for data collection, devices, and RFID; as a result, data is dispersed to various systems and applications.

3) Algorithm perspective:
The data features that must be analyzed are not always the same because sensors/devices can leave/join the system at any instant, and the network architecture can change. Occasionally, the streaming data will not arrive in the system simultaneously [113]. Animated adjustment of mining results and re-running the mining method are two obvious solutions to the problem of a changing environment, but both strategies increase computation time. As a result, we needed a flexible and accurate mining strategy to deal with this problem. Another example is that IoT classification seeks to either dynamically add classifiers to the system or statically update classifiers in response to system conditions. Many systems do not strive to be completely dynamic. The smart house example forces us to evaluate the situation. Is it necessary to dynamically add classifiers when the people entering and leaving are family members? The example also shows that a static system can have a less precise rate since it cannot handle unpredictability, such as a stranger intrusion.

Since improved systems for smart environments or IoT applications stress this type of incorporation (i.e., clustering integration) and classification [114,115], the open problem of how to connect mining methods with other mining technologies remains. One approach is to employ routine pattern mining technology as the front or back end of the other mining modules. As a result, input and output will be more critical than ever. In the worst-case scenario, the algorithm needs to be modified to interact with the other procedures in the system.

Table 4
A comparison between IoT and traditional data.

Subject	IoT Data	Traditional Data
Content	By machine	By human
Consume content	Pushing data	On request
Combine data	By operators	By links
Content value	Data and action	Question answers
What was done?	Data creation	HTML and search engines

9.5. Key challenges

Integrating data science with IoT is challenging since organizations need to consider performance limitations, data kind, volume changes, the expense of qualified data researchers, and dynamic modeling. Administrations working with IoT should carefully consider how to analyze the data collected from the impact of IoT-connected sensors or devices. There are several possibilities for cost savings and improvements in various commercial goals, including product creation and services. Obtaining insights from IoT datasets, however, is not without difficulties. The following are some of the most critical challenges:

9.5.1. Security and privacy

With a rising number of sensors and other associated devices, security and privacy are fundamental concerns in the IoT, as IoT devices expose individuals' limited privacy in the digital world [116,117]. Administrations must take precautions to avoid data loss that could attract negative attention from cybercriminals. Before implementing a successful data science strategy, reduce the likelihood of data storage in insecure areas and provide access to other sources.

IoT creates massive volumes of data, increasing the possibility of hacking or leaking sensitive information [118]. For example, hackers can obtain vital health details if they hijack the connection between the doctor's office and your fitness tracker app. Privacy concerns are a major issue in IoT data science. For example, many organizations have come under fire for disclosing sensitive information about their customers without their knowledge or permission.

9.5.2. Data science services deficiencies

There can not be sufficient qualified data scientists graduating from academia to assist administrators in making better decisions based on IoT data. Data science for IoT is paving the way for larger categories of robots by collecting enormous amounts of data on user behaviors and behavior at the micro-level. As entrepreneurs rush to ship novel new goods, such as Virtual Reality (VR) headsets, cameras, aerial vehicles and self-driving cars full of sensors that collect data and share it with analytics applications, organizations must develop more effective ways to achieve the flow of expanding big data for data science and analytics to derive valuable business insights.

Data science is a heterogeneous subject of today's technology that contains valuable insights and extracting knowledge from organized or unstructured data; nevertheless, the question here is if data science for IoT is different or the same as the traditional data science procedure. The mathematics principles for data science include a variety of notions such as Bayesian statistics, optimization approaches, matrix algebra, and the utilization of supervised or unsupervised models, all of which apply to IoT datasets. Programming with IoT datasets entails applying different tools such as Python, MATLAB, R, etc., as with other datasets but traditionally for time series applications. A comparative analysis between IoT data and traditional data is shown in Table 4.

9.5.3. Scaling for IoT big data

Scaling is another important challenge for big data in the IoT. IoT sensors and devices generate vast volumes of data every second, and organizations have to establish a suitable method for cataloging this massive data. If the data is not managed correctly, the scale of the data may overwhelm even the most robust big data platforms available today.

IoT data science is an important tool, but users may find it difficult to scale it up to suit their needs. If a company wants to add new sensors or integrate an IoT system with other software solutions, significant concerns and challenges will likely arise. This is why upfront planning for the scaling project is critical. In order to properly scale data science activities, everything from the software to the workers needs to be in place first.

The IoT's big data signal quality brings a data researcher's mathematical abilities and talent to the test, as they can be forced to step outside their convenience sphere to give significant observation via advanced digital signal processing approaches. The chronological nature of data from the IoT sensors makes a data researcher's job position working with the IoT realm more difficult, as they must transform a related high-level inquiry into something perceptible, compelling them to have a suitable statistical vocabulary. Correspondingly, time series analysis and anomaly detection are critical components of data science and IoT instead of functionalization, which is sufficient to address other sorts of data science breakthroughs [119]. Data researchers must automate their quantization policies to align with the data speed supplied by IoT devices and market demands. Data researchers interested in working in the IoT sector should have the ability to understand and comply with what works and what does not due to the data science regulations for IoT, which are a significant challenge that must be addressed as soon as possible.

9.5.4. Data analytics skills

Data science for the IoT can be tremendously valuable, but are there enough individuals with suitable analytics skills? Traditional data science contractors dominate the market because IoT analytics have not yet gained widespread adoption [120]. However, as more businesses adopt IoT technology, this could change very quickly. IoT data scientists will need to learn new skills and understand the nuances of the deployment process. In order to accomplish this, they will need to learn the followings:

- **Edge Computing:** It is the technique of processing data as close to the source as feasible in order to improve speed and reduce network congestion.
- **Computer-Aided Design (CAD):** It is critical to understand the reasoning underlying a smart device's physical design for data analytics purposes.
- **IoT Computing Frameworks:** To understand IoT hardware, data scientists must also leverage open-source learning tools considering data analytics.

There is no common strategy for addressing data science difficulties with IoT big data. The key distinction between data science and IoT is that it focuses on time series data analysis, cognitive computing, real-time processing, geo-geographic data analysis, edge computing, deep learning, etc. Data science for the IoT necessitates data researchers who are familiar with numerous ways of integrating hardware and processing complicated events. Working on IoT-critical problems will likely fascinate data researchers because it is an innovatory specialization with a lot of irony, energy, and excitement. For example, IoT data can be gathered using sovereign digital sensors, and data researchers can approach an IoT job with an efficient supposition for gathering high levels of data to guarantee information nature. The IoT sensors gather various data points about user engagement, device operation, and location. Most of these sensor data are logically coded, but in rare circumstances, they are specifically designed with data science in consideration so that a data researcher has a good opening point to progress with data analytics.

Similarly, the other type of IIoT challenge is the lack of advanced machine learning or AI algorithms capable of performing analysis of data with input data damaged by unstable connection lines. But enabling machine learning or AI prospects on IIoT devices is difficult. IIoT devices have modest memory sizes, minimal power consumption, and are distributed.

9.5.5. Investment/operating costs

Another issue with data science for the IoT is the high expense of launching a whole new technology. This is especially true for businesses that want to use it more widely. We anticipate that many enterprises will experience severe budget constraints while using IoT data science technology. Although the cost of implementing IoT in trade development may not be a huge concern, it carries risks for some of them. Admittedly, adopting the Internet of Things takes into account revenue growth, but it should not be ignored that due to the complexity and originality of data science solutions being industrialized, it may increase investment costs.

9.5.6. Data management

IoT data are significant resources, and as several IoT devices, frameworks, and processes grow, new ways of handling big data, such as data lakes, have evolved [121,122]. The data lake contains both structured and unstructured data, without considering how to use this data in the future. This technology does not require the use of language queries and can save any form of data without restriction. However, data stores pose a few issues. Considering that any data can be injected, data transactions may occur in the future. We should monitor data quality, require meta-data insertion, and ensure data derivation to avoid such issues. Secondly, considering data lakes can result in agility loss, especially for big organizations that want to use a huge data pool for rapid investigation and decision-making, but cannot do so professionally due to the numerous steps required before removing something important from the data. Instead, these businesses must distinguish between data that can be applied for near-real-time decision-making and data that can be applied to develop business policies. Because these data will not be used directly, the latter kind is more applicable for data lake storage.

9.5.7. The bottom line

Data science for the IoT is a significant advancement over traditional data analytics. To make data science more robust, powerful, and accurate, additional efforts must be made. The data generation capability of the Internet of Things makes it possible. The web of interconnected devices communicates constantly to offer businesses and organizations massive amounts of user-related data. Data scientists have enough information to derive significant inferences from their databases. Deploying data science for the Internet of Things is difficult, but the returns are too great to ignore. In this situation, we expect the data science of the Internet of Things to become mainstream in the next decade.

Companies with solid data science teams that detect additional data by interpretation are likely to be able to withstand the IoT expansion; otherwise, they should brace themselves to be left behind as the IoT unleashes widespread catastrophe. There are numerous opportunities for data science specialists as big data created by IoT rises at an annual compound rate of expansion of 66% [123]. As businesses scale up their data science approaches to gain access to new data streams generated by IoT devices, more employees will be required to connect with and make sense of data from wearables, smart TVs, smartwatches, and driverless vehicles.

9.5.8. Data attribution

Data attribution is linked to data validity and traceability to determine data holders and modernizers at each stage. However, as big data provides important perceptions and analytics that may lead to some sovereign actions in real-world scenarios, it is necessary to ensure that the data used to create such actions is properly sourced. A number of large programs, such as smart health and cities, intend to make use of big data analytics, exacerbating the problem. Though most existing IoT studies have concentrated on data management, just a few have addressed data attribution [124]. When different parties share their data, having the capability to imply IoT data ownership can be helpful for monetization.

9.6. Future prospectus and possible solutions

Even though many research problems have been overcome, many research concerns remain to be addressed. As a result of our study and investigation of current trends in data science and IoT, we present the following future trends:

9.6.1. Integration of contextual data

The state of the environment cannot be captured by IoT sensor data alone [125]. Therefore, IoT data needs to be combined with other data sources, such as contextual information that enhances awareness of the environment. Due to the limited search space of the reasoning engine, this combination can support fast data analysis and fast reasoning. For example, a smarter camera with face pose recognition capacity can be employed in various scenarios, such as entryways in smart houses or government offices or assistance while driving smart cars. In all these cases, complementing contextual information assists the system in reasoning about the optimal action to take based on the perceived human stance.

9.6.2. Distributed data processing

Amongst the currently distributed data processing algorithms, MapReduce and its substitutes have several benefits (e.g., scalable, fault-tolerant, and simple), [126]. However, the downside of this paradigm is its inefficiency compared to other parallel computing models. The performance of traditional parallel computing models is superior to that of MapReduce-like models. These models include Message Passing Interface (MPI), Open Multi-Processing (OpenMP), GPU-based parallel processing, and Field-Programmable Gate Arrays (FPGA)-oriented programming. Performance may be increased by integrating models like MapReduce with parallel computing models. Furthermore, the existing data analysis algorithm can be improved by utilizing specialized processing platforms. For example, advances in FPGA-oriented Convolutional Neural Networks (CNN) and FPGA deep-learning algorithms have demonstrated superior performance than that of conventional microcomputers.

9.6.3. Stream data processing

Due to the massive amount of real-time data generated by WSNs, it is impractical to maintain and process it all in a single storage. As a result, several typical data analysis algorithms that require access to entire data sets fail in this scenario. As a result, various data stream preparation methods can be employed in communication networks and WSNs [127]. However, there has been very little literature on data analysis of data streams to date [128].

9.6.4. In-network data processing

In large-scale WSNs, numerous wireless nodes are deployed, and data processing requires the inclusion of data generated by remote nodes. However, data fusion over distant networks invariably incurs significant transmission costs. To address this issue, we must do in-network processing across the entire network, with data processed at each node rather than at centralized servers [129]. To further reduce computing costs, nodes are frequently organized into clusters. However, deciding which cluster to use to suit the big data requirements has become a new key challenge.

9.6.5. IoT mobile data

Mobile devices contribute a significant amount of IoT data. One approach to developing better IoT services, particularly in smart city environments, is to examine efficient ways of using big mobile data in conjunction with deep learning algorithms. In [130], the abilities of deep learning models in mobile big data processing were investigated using a distributed learning framework that performs an iterative MapReduce task on many similar Spark workers.

9.6.6. Reliable or stable IoT analytics

As we depend more on Cyber-Physical Systems (CPS) and IoT on a broad scale, the requirement for measures to assure system safety against malicious attacks and malfunctions becomes more critical [131]. Deep learning approaches could be used in these directions by examining vast volumes of log traces from CPS and IoT frameworks to classify and forecast weak points in the system where cyberattacks could occur or operations are broken. This will assist the system in preventing or recovering from problems, increasing the dependability of CPS and IoT schemes.

9.6.7. IoT analytics resources supply

Fast deep learning-based data analytics would require the online provisioning of fog or cloud sources to host the data stream. Due to the nature of IoT data streaming, it is impossible to predict the volume of the data stream. In this sense, a novel algorithms class would be required to work on existing data streams without requiring data stream knowledge. The work of [132,133] offers a deep learning technique and an online auctioning algorithm to facilitate the online supply of cloud and fog resources for IoT applications.

9.6.8. Security and privacy issues

IoT device security and end-to-end data security will continue to pique researchers' curiosity. Further research is required to create IoT devices capable of supporting new security schemes, such as sophisticated encryption algorithms [134,135]. Signcryption combines a digital signature with data encryption to protect sensitive data from eavesdropping and unwanted alterations. Physical attackers and intruders from IoT devices need to be prevented through security approaches and new schemes need to be developed. Security and privacy are critical concerns in WSN big data analytics. Although security and privacy are closely related, they are not the same.

- **Data analytics privacy:** In this phase, one of the important issues is the balance between data privacy and analysis efficiency. To safeguard private user documents, for example, the papers are often encoded and stored on a cloud server [136]. However, actions on encrypted documents are time-consuming, resulting in inefficiency in data analysis [137]. Much work remains to be done in the areas of data publishing, distributed privacy and data mining performance [138,139].
- **Data storage security and privacy:** If a data storage system attack succeeds, additional personal and secret data can be revealed.It is more important to protect the stored data at this stage. Fortunately, it is easier to use encryption techniques to protect data storage than data transmission. However, using privacy-protected methods in data storage is still encouraged, particularly if the data service for storage is offered by a third party [140,141]. Mobile Edge Computing (MEC) can fundamentally provide a solution to data storage privacy by offloading data from an untrustworthy or third-party server to a trustworthy MEC server placed near the user [142–144].

10. Conclusion

This survey comprehensively studies integrating data science and IIoT and reviews multiple characteristics and related key challenges. It also presents the IIoT architecture that is deployed in most IoT applications. Similarly, the various data sources in the IoT and the ways to handle them through multiple computing platforms are discussed. The new paradigms for implementing data science on IIoT devices are examined, and different methods are discussed. Moreover, numerous IoT applications where integrated IIoT data science poses multiple challenges are highlighted. Then, the new perspectives and potential directions for future studies in integrating data science and IoT applications are discussed in detail. The success of IIoT and data science in multiple research and business domains has drawn the attention of scientists and researchers. The integration of IIoT and data science consists of data producers, analyzers, and consumers, in which IIoT generates a tremendous amount of data, from which data science derives essential insights for multiple IoT applications and businesses through intelligent approaches. Based on processing, scalability, security, privacy, and decentralization, future research to overcome the limits of integrating IIoT and data science approaches to enhance their decision-making in real-world situations was recommended.

CRediT authorship contribution statement

Inam Ullah: Writing – original draft, Validation, Software, Methodology, Data curation, Conceptualization. **Deepak Adhikari:** Writing – review & editing, Investigation, Data curation. **Xin Su:** Validation, Formal analysis. **Francesco Palmieri:** Formal analysis, Validation, Resources. **Celimuge Wu:** Visualization, Validation, Resources. **Chang Choi:** Visualization, Validation, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under Grant 62371181, and in part by the Changzhou Science and Technology International Cooperation Program under Grant CZ20230029.

This work was also supported by a National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2021R1A2B5B02087169). This work was also supported under the framework of international cooperation program managed by the National Research Foundation of Korea (2022K2A9A1A01098051).

References

[1] M.N. Bhuiyan, M.M. Rahman, M.M. Billah, D. Saha, Internet of Things (IoT): a review of its enabling technologies in healthcare applications, standards protocols, security, and market opportunities, IEEE Int. Things J. 8 (13) (2021) 10474–10498.
[2] K. Rose, S. Eldridge, L. Chapin, The internet of things: an overview, The Internet Society (ISOC) 80 (15) (2015) 1–53, https://www.mckinsey.com/mgi/overview/in-the-news/by-2025-internet-of-things-applications-could-have-11-trillion-impact.
[3] L. Guo, M. Dong, K. Ota, Q. Li, T. Ye, J. Wu, J. Li, A secure mechanism for big data collection in large scale Internet of vehicle, IEEE Int. Things J. 4 (2) (2017) 601–610.
[4] M. Mohammadi, A. Al-Fuqaha, S. Sorour, M. Guizani, Deep learning for IoT big data and streaming analytics: a survey, IEEE Commun. Surv. Tutor. 20 (4) (2018) 2923–2960.
[5] L. Zhang, P. Wang, Y. Zhang, Z. Chi, N. Tong, L. Wang, F. Li, An adaptive and robust secret key extraction scheme from high noise wireless channel in IIoT, Digit. Commun. Netw. 9 (4) (2023) 809–816.
[6] D. Adhikari, W. Jiang, J. Zhan, Z. He, D.B. Rawat, U. Aickelin, H.A. Khorshidi, A comprehensive survey on imputation of missing data in Internet of Things, ACM Comput. Surv. 55 (7) (2022) 1–38.
[7] D. Adhikari, W. Jiang, J. Zhan, M. Assefa, H.A. Khorshidi, U. Aickelin, D.B. Rawat, A lightweight window portion-based multiple imputation for extreme missing gaps in IoT systems, IEEE Int. Things J. 11 (3) (2024) 3676–3689.
[8] I.H. Sarker, Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective, SN Comput. Sci. 2 (5) (2021) 377.
[9] S.B. Atitallah, M. Driss, W. Boulila, H.B. Ghézala, Leveraging deep learning and IoT big data analytics to support the smart cities development: review and future directions, Comput. Sci. Rev. 38 (2020) 100303.
[10] E. Siow, T. Tiropanis, W. Hall, Analytics for the Internet of Things: a survey, ACM Comput. Surv. 51 (4) (2018) 1–36.
[11] Y. Sasaki, A survey on IoT big data analytic systems: current and future, IEEE Int. Things J. 9 (2) (2021) 1024–1036.
[12] C.-W. Tsai, C.-F. Lai, M.-C. Chiang, L.T. Yang, Data mining for Internet of Things: a survey, IEEE Commun. Surv. Tutor. 16 (1) (2013) 77–97.
[13] S. Verma, Y. Kawamoto, Z.M. Fadlullah, H. Nishiyama, N. Kato, A survey on network methodologies for real-time analytics of massive IoT data and open research issues, IEEE Commun. Surv. Tutor. 19 (3) (2017) 1457–1477.
[14] A. Daissaoui, A. Boulmakoul, L. Karim, A. Lbath, IoT and big data analytics for smart buildings: a survey, Proc. Comput. Sci. 170 (2020) 161–168.
[15] M.S. Mahdavinejad, M. Rezvan, M. Barekatain, P. Adibi, P. Barnaghi, A.P. Sheth, Machine learning for Internet of Things data analysis: a survey, Digit. Commun. Netw. 4 (3) (2018) 161–175.
[16] C.-W. Tsai, P.-W. Tsai, M.-C. Chiang, C.-S. Yang, Data analytics for Internet of Things: a review, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 8 (5) (2018) e1261.
[17] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, M. Ayyash, Internet of Things: a survey on enabling technologies, protocols, and applications, IEEE Commun. Surv. Tutor. 17 (4) (2015) 2347–2376.
[18] Y. Hajjaji, W. Boulila, I.R. Farah, I. Romdhani, A. Hussain, Big data and IoT-based applications in smart environments: a systematic review, Comput. Sci. Rev. 39 (2021) 100318.
[19] W. Li, Y. Chai, F. Khan, S.R.U. Jan, S. Verma, V.G. Menon, X. Li, A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system, Mob. Netw. Appl. 26 (2021) 234–252.
[20] O. Elijah, T.A. Rahman, I. Orikumhi, C.Y. Leow, M.N. Hindia, An overview of Internet of Things (IoT) and data analytics in agriculture: benefits and challenges, IEEE Int. Things J. 5 (5) (2018) 3758–3773.
[21] M. Koot, M.R. Mes, M.E. Iacob, A systematic literature review of supply chain decision making supported by the Internet of Things and big data analytics, Comput. Ind. Eng. 154 (2021) 107076.
[22] L. Cao, Data science: a comprehensive overview, ACM Comput. Surv. 50 (3) (2017) 1–42.
[23] S. Shadroo, A.M. Rahmani, Systematic survey of big data and data mining in Internet of things, Comput. Netw. 139 (2018) 19–47.
[24] O. Hahm, E. Baccelli, H. Petersen, N. Tsiftes, Operating systems for low-end devices in the Internet of Things: a survey, IEEE Int. Things J. 3 (5) (2015) 720–734.
[25] H. Cai, B. Xu, L. Jiang, A.V. Vasilakos, IoT-based big data storage systems in cloud computing: perspectives and challenges, IEEE Int. Things J. 4 (1) (2016) 75–87.
[26] V. Pandey, A. Kipf, T. Neumann, A. Kemper, How good are modern spatial analytics systems?, Proc. VLDB Endow. 11 (11) (2018) 1661–1673.
[27] M. Marjani, F. Nasaruddin, A. Gani, A. Karim, I.A.T. Hashem, A. Siddiqa, I. Yaqoob, Big IoT data analytics: architecture, opportunities, and open research challenges, IEEE Access 5 (2017) 5247–5261.
[28] G. Atluri, A. Karpatne, V. Kumar, Spatio-temporal data mining: a survey of problems and methods, ACM Comput. Surv. 51 (4) (2018) 1–41.
[29] I.H. Sarker, Smart city data science: towards data-driven smart cities with open research issues, Int. Things 19 (2022) 100528.
[30] J.A. Stankovic, Research directions for the Internet of Things, IEEE Int. Things J. 1 (1) (2014) 3–9.
[31] L. Atzori, A. Iera, G. Morabito, The Internet of Things: a survey, Comput. Netw. 54 (15) (2010) 2787–2805.
[32] J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami, Internet of Things (IoT): a vision, architectural elements, and future directions, Future Gener. Comput. Syst. 29 (7) (2013) 1645–1660.
[33] E. Borgia, The Internet of Things vision: key features, applications and open issues, Comput. Commun. 54 (2014) 1–31.
[34] Y. Liu, Y. Kuang, Y. Xiao, G. Xu, Sdn-based data transfer security for Internet of Things, IEEE Int. Things J. 5 (1) (2017) 257–268.
[35] H. Jiang, F. Shen, S. Chen, K.-C. Li, Y.-S. Jeong, A secure and scalable storage system for aggregate data in IoT, Future Gener. Comput. Syst. 49 (2015) 133–141.
[36] N.C. Luong, D.T. Hoang, P. Wang, D. Niyato, D.I. Kim, Z. Han, Data collection and wireless communication in Internet of Things (IoT) using economic analysis and pricing models: a survey, IEEE Commun. Surv. Tutor. 18 (4) (2016) 2546–2590.
[37] A. Gluhak, S. Krco, M. Nati, D. Pfisterer, N. Mitton, T. Razafindralambo, A survey on facilities for experimental Internet of Things research, IEEE Commun. Mag. 49 (11) (2011) 58–67.

[38] M. Luckner, M. Grzenda, R. Kunicki, J. Legierski, IoT architecture for urban data-centric services and applications, ACM Trans. Internet Technol. 20 (3) (2020) 1–30.
[39] A. Shah, B. Ali, F. Wahab, I. Ullah, K.T. Amesho, M. Shafiq, Entropy-based grid approach for handling outliers: a case study to environmental monitoring data, Environ. Sci. Pollut. Res. 30 (60) (2023) 125138–125157.
[40] A.L.R. Madureira, F.R.C. Araújo, L.N. Sampaio, On supporting IoT data aggregation through programmable data planes, Comput. Netw. 177 (2020) 107330.
[41] J. Xu, Y. Andrepoulos, Y. Xiao, M. van Der Schaar, Non-stationary resource allocation policies for delay-constrained video streaming: application to video over Internet-of-Things-enabled networks, IEEE J. Sel. Areas Commun. 32 (4) (2014) 782–794.
[42] D. Ilic, P.G. Da Silva, S. Karnouskos, M. Griesemer, An energy market for trading electricity in smart grid neighbourhoods, in: 2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST), IEEE, 2012, pp. 1–6.
[43] R. Buyya, C.S. Yeo, S. Venugopal, Market-oriented cloud computing: vision, hype, and reality for delivering it services as computing utilities, in: 2008 10th IEEE International Conference on High Performance Computing and Communications, IEEE, 2008, pp. 5–13.
[44] Y. Feng, B. Li, B. Li, Price competition in an oligopoly market with multiple IaaS cloud providers, IEEE Trans. Comput. 63 (1) (2013) 59–73.
[45] P.J. Nesse, S. Svaet, D. Strasunskas, A.A. Gaivoronski, Assessment and optimisation of business opportunities for telecom operators in the cloud value network, Trans. Emerg. Telecommun. Technol. 24 (5) (2013) 503–516.
[46] Z. Chen, F. Xia, T. Huang, F. Bu, H. Wang, A localization method for the Internet of Things, J. Supercomput. 63 (2013) 657–674.
[47] L. Jiang, L. Da Xu, H. Cai, Z. Jiang, F. Bu, B. Xu, An IoT-oriented data storage framework in cloud computing platform, IEEE Trans. Ind. Inform. 10 (2) (2014) 1443–1451.
[48] R. Ranjan, O. Rana, S. Nepal, M. Yousif, P. James, Z. Wen, S. Barr, P. Watson, P.P. Jayaraman, D. Georgakopoulos, et al., The next grand challenges: integrating the Internet of Things and data science, IEEE Cloud Comput. 5 (3) (2018) 12–26.
[49] T. Baltrušaitis, C. Ahuja, L.-P. Morency, Multimodal machine learning: a survey and taxonomy, IEEE Trans. Pattern Anal. Mach. Intell. 41 (2) (2018) 423–443.
[50] S. Sawalha, G. Al-Naymat, Towards an efficient big data management schema for IoT, J. King Saud Univ, Comput. Inf. Sci. 34 (9) (2022) 7803–7818.
[51] L. Sanchez, L. Muñoz, J.A. Galache, P. Sotres, J.R. Santana, V. Gutierrez, R. Ramdhany, A. Gluhak, S. Krco, E. Theodoridis, et al., Smartsantander: IoT experimentation over a smart city testbed, Comput. Netw. 61 (2014) 217–238.
[52] F. Rezaeibagha, Y. Mu, K. Huang, L. Chen, Secure and efficient data aggregation for IoT monitoring systems, IEEE Int. Things J. 8 (10) (2020) 8056–8063.
[53] P. Naur, Concise Survey of Computer Methods, Petrocelli Books, 1974.
[54] P. Naur, 'Datalogy', the science of data and data processes, in: IFIP Congress (2), 1968, pp. 1383–1387.
[55] B. Cheng, A. Papageorgiou, F. Cirillo, E. Kovacs, Geelytics: geo-distributed edge analytics for large scale IoT systems based on dynamic topology, in: 2015 IEEE 2nd World Forum on Internet of Things (WF-IoT), IEEE, 2015, pp. 565–570.
[56] W. Ding, X. Jing, Z. Yan, L.T. Yang, A survey on data fusion in Internet of Things: towards secure and privacy-preserving fusion, Inf. Fusion 51 (2019) 129–144.
[57] H. Fang, Managing data lakes in big data era: what's a data lake and why has it became popular in data management ecosystem, in: 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), IEEE, 2015, pp. 820–824.
[58] T. Wang, Y. Liang, Y. Zhang, X. Zheng, M. Arif, J. Wang, Q. Jin, An intelligent dynamic offloading from cloud to edge for smart IoT systems with big data, IEEE Trans. Netw. Sci. Eng. 7 (4) (2020) 2598–2607.
[59] F. Terroso-Saenz, A. González-Vidal, A.P. Ramallo-González, A.F. Skarmeta, An open IoT platform for the management and analysis of energy data, Future Gener. Comput. Syst. 92 (2019) 1066–1079.
[60] L. Xu, H. Yin, H. Jia, W. Lin, X. Zhou, Y. Fu, X. Yu, Data secure transmission intelligent prediction algorithm for mobile industrial IoT networks, Digit. Commun. Netw. 9 (2) (2023) 400–410.
[61] M. Devi, R. Dhaya, R. Kanthavel, F. Algarni, P. Dixikha, Data science for Internet of Things (IoT), in: Second International Conference on Computer Networks and Communication Technologies: ICCNCT 2019, Springer, 2020, pp. 60–70.
[62] T. Wang, L. Qiu, A.K. Sangaiah, A. Liu, M.Z.A. Bhuiyan, Y. Ma, Edge-computing-based trustworthy data collection model in the Internet of Things, IEEE Int. Things J. 7 (5) (2020) 4218–4227.
[63] I. Ullah, S. Qian, Z. Deng, J.-H. Lee, Extended Kalman filter-based localization algorithm by edge computing in wireless sensor networks, Digit. Commun. Netw. 7 (2) (2021) 187–195.
[64] D. Sun, J. Wu, J. Yang, H. Wu, Intelligent data collaboration in heterogeneous-device IoT platforms, ACM Trans. Sens. Netw. 17 (3) (2021) 1–17.
[65] F. Jameel, W.U. Khan, N. Kumar, R. Jäntti, Efficient power-splitting and resource allocation for cellular V2X communications, IEEE Trans. Intell. Transp. Syst. 22 (6) (2020) 3547–3556.
[66] M.R. Anawar, S. Wang, M. Azam Zia, A.K. Jadoon, U. Akram, S. Raza, et al., Fog computing: an overview of big IoT data analytics, Wirel. Commun. Mob. Comput. 2018 (1) (2018) 1–22.
[67] Y. Ai, M. Peng, K. Zhang, Edge computing technologies for Internet of Things: a primer, Digit. Commun. Netw. 4 (2) (2018) 77–86.
[68] K. Wang, Y. Shao, L. Xie, J. Wu, S. Guo, Adaptive and fault-tolerant data processing in healthcare IoT based on fog computing, IEEE Trans. Netw. Sci. Eng. 7 (1) (2018) 263–273.
[69] R. Goyat, G. Kumar, M. Alazab, M. Conti, M.K. Rai, R. Thomas, R. Saha, T.-H. Kim, Blockchain-based data storage with privacy and authentication in Internet of Things, IEEE Int. Things J. 9 (16) (2020) 14203–14215.
[70] F. Rezaeibagha, Y. Mu, K. Huang, L. Chen, L. Zhang, Toward secure data computation and outsource for multi-user cloud-based IoT, IEEE Trans. Cloud Comput. 11 (1) (2023) 217–228.
[71] A.-T. Fadi, B.D. Deebak, Seamless authentication: for IoT-big data technologies in smart industrial application systems, IEEE Trans. Ind. Inform. 17 (4) (2020) 2919–2927.
[72] Y. Fathy, P. Barnaghi, Quality-based and energy-efficient data communication for the Internet of Things networks, IEEE Int. Things J. 6 (6) (2019) 10318–10331.
[73] T. Luo, J. Huang, S.S. Kanhere, J. Zhang, S.K. Das, Improving IoT data quality in mobile crowd sensing: a cross validation approach, IEEE Int. Things J. 6 (3) (2019) 5651–5664.
[74] L.B. Furstenau, P. Leivas, M.K. Sott, M.S. Dohan, J.R. López-Robles, M.J. Cobo, N.L. Bragazzi, K.-K.R. Choo, Big data in healthcare: conceptual network structure, key challenges and opportunities, Digit. Commun. Netw. 9 (4) (2023) 856–868.
[75] Y. Xu, Z. Wang, H. Gao, Z. Jiang, Y. Yin, R. Li, Towards machine-learning-driven effective mashup recommendations from big data in mobile networks and the Internet-of-Things, Digit. Commun. Netw. 9 (1) (2023) 138–145.
[76] F. Provost, T. Fawcett, Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking, O'Reilly Media, Inc., 2013.
[77] A. Marchand, P. Marx, Automated product recommendations with preference-based explanations, J. Retail. 96 (3) (2020) 328–343.
[78] J. Boudet, B. Gregg, K. Rathje, E. Stein, K. Vollhardt, The future of personalization—and how to get ready for it, Recuperado el. 12 (2019) 1–7, https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-future-of-personalization-and-how-to-get-ready-for-it.
[79] J. Grabis, K. Jegorova, K. Pinka, IoT data analytics in retail: framework and implementation, in: IN4PL, 2020, pp. 93–100.
[80] United Nations, World Urbanization Prospects: the 2014 Revision, United Nations, New York, 2018.
[81] B. Resch, M. Szell, Human-centric data science for urban studies, ISPRS Int. J. Geo-Inf. 8 (12) (2019) 1–9, https://doi.org/10.3390/ijgi8120584.
[82] H. Yoo, R.C. Park, K. Chung, IoT-based health big-data process technologies: a survey, KSII Trans. Int. Inf. Syst. (TIIS) 15 (3) (2021) 974–992, https://doi.org/10.3837/tiis.2021.03.001.
[83] M. Nilashi, O. bin Ibrahim, H. Ahmadi, L. Shahmoradi, An analytical method for diseases prediction using machine learning techniques, Comput. Chem. Eng. 106 (2017) 212–223.
[84] G. Ayoade, A. El-Ghamry, V. Karande, L. Khan, M. Alrahmawy, M.Z. Rashad, Secure data processing for IoT middleware systems, J. Supercomput. 75 (2019) 4684–4709.
[85] A. Al-Abassi, H. Karimipour, H. HaddadPajouh, A. Dehghantanha, R.M. Parizi, Industrial big data analytics: challenges and opportunities, in: Handbook of Big Data Privacy, 2020, pp. 37–61.
[86] X. Yuan, X. Yuan, B. Li, C. Wang, Toward secure and scalable computation in Internet of Things data applications, IEEE Int. Things J. 6 (2) (2019) 3753–3763.
[87] M. Mahesh, S.R. Kawale, M.D. PraveenKumar, S.J. Veeresh, D.N. Sahu, M. Barote, A. Rangampet, Applications of Internet of Things in food and beverage industries, Ann. For. Res. 65 (1) (2022) 3957–3970.
[88] N. Lu, N. Cheng, N. Zhang, X. Shen, J.W. Mark, Connected vehicles: solutions and challenges, IEEE Int. Things J. 1 (4) (2014) 289–299.
[89] B. Ji, M. Zhang, L. Xing, X. Li, C. Li, C. Han, H. Wen, Research on optimal intelligent routing algorithm for IoV with machine learning and smart contract, Digit. Commun. Netw. 9 (1) (2023) 47–55.
[90] W.U. Khan, M.A. Javed, T.N. Nguyen, S. Khan, B.M. Elhalawany, Energy-efficient resource allocation for 6G backscatter-enabled NOMA IoV networks, IEEE Trans. Intell. Transp. Syst. 23 (7) (2021) 9775–9785.
[91] M. Dibaei, X. Zheng, K. Jiang, R. Abbas, S. Liu, Y. Zhang, Y. Xiang, S. Yu, Attacks and defences on intelligent connected vehicles: a survey, Digit. Commun. Netw. 6 (4) (2020) 399–421.
[92] J.M.K. Sri, V. Narendra, V. Pai, Implementing and testing of IoT technology in agriculture, Int. J. Innov. Technol. Explor. Eng. 8 (2S) (2018) 190–194.
[93] R. Akhter, S.A. Sofi, Precision agriculture using IoT data analytics and machine learning, J. King Saud Univ, Comput. Inf. Sci. 34 (8) (2022) 5602–5618.
[94] I.H. Sarker, A. Kayes, S. Badsha, H. Alqahtani, P. Watters, A. Ng, Cybersecurity data science: an overview from machine learning perspective, J. Big Data 7 (2020) 1–29.
[95] I.H. Sarker, Cyberlearning: effectiveness analysis of machine learning security modeling to detect cyber-anomalies and multi-attacks, Int. Things 14 (2021) 100393.
[96] Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, C. Wang, Machine learning and deep learning methods for cybersecurity, IEEE Access 6 (2018) 35365–35381.
[97] S.A. Shah, D.Z. Seker, S. Hameed, D. Draheim, The rising role of big data analytics and IoT in disaster management: recent advances, taxonomy and prospects, IEEE Access 7 (2019) 54595–54614.

[98] J. Lloret, J. Tomas, A. Canovas, L. Parra, An integrated IoT architecture for smart metering, IEEE Commun. Mag. 54 (12) (2016) 50–57.
[99] B. Xu, L. Da Xu, H. Cai, C. Xie, J. Hu, F. Bu, Ubiquitous data accessing method in IoT-based information system for emergency medical services, IEEE Trans. Ind. Inform. 10 (2) (2014) 1578–1586.
[100] P. Basanta-Val, N.C. Audsley, A.J. Wellings, I. Gray, N. Fernández-García, Architecting time-critical big-data systems, IEEE Trans. Big Data 2 (4) (2016) 310–324.
[101] I. Ullah, H.Y. Youn, Intelligent data fusion for smart IoT environment: a survey, Wirel. Pers. Commun. 114 (2020) 409–430.
[102] W. He, G. Yan, L. Da Xu, Developing vehicular data cloud services in the IoT environment, IEEE Trans. Ind. Inform. 10 (2) (2014) 1587–1595.
[103] J. Li, H. Liu, Challenges of feature selection for big data analytics, IEEE Intell. Syst. 32 (2) (2017) 9–15.
[104] W.U. Khan, F. Jameel, A. Ihsan, O. Waqar, M. Ahmed, Joint optimization for secure ambient backscatter communication in NOMA-enabled IoT networks, Digit. Commun. Netw. 9 (1) (2023) 264–269.
[105] T.S. López, D.C. Ranasinghe, B. Patkai, D. McFarlane, Taxonomy, technology and applications of smart objects, Inf. Syst. Front. 13 (2011) 281–300.
[106] C. Xu, W. Zhao, J. Zhao, Z. Guan, X. Song, J. Li, Uncertainty-aware multiview deep learning for Internet of Things applications, IEEE Trans. Ind. Inform. 19 (2) (2022) 1456–1466.
[107] C. Bayılmış, M.A. Ebleme, Ü. Çavuşoğlu, K. Küçük, A. Sevin, A survey on communication protocols and performance evaluations for Internet of Things, Digit. Commun. Netw. 8 (6) (2022) 1094–1104.
[108] J.R. Bhat, S.A. AlQahtani, M. Nekovee, Fintech enablers, use cases, and role of future Internet of Things, J. King Saud Univ, Comput. Inf. Sci. 35 (1) (2023) 87–101.
[109] A. de Saint-Exupery, Internet of things, strategic research roadmap, in: Internet of Things - Global Technological and Societal Trends, 2009, pp. 9–52, https://www.semanticscholar.org/paper/Internet-of-Things-Strategic-Research-Roadmap-De-Saint-Exup%C3%A9ry/b5336e565b38f99fd3527fe68ad002d9e309ecfc, 2009.
[110] A. Asin, D. Gascon, 50 Sensor Applications for a Smarter World, Tech. Rep., Libelium Comunicaciones Distribuidas, 2012, pp. 589–594.
[111] R. Baeza-Yates, B. Ribeiro-Neto, et al., Modern Information Retrieval, vol. 463, ACM Press, New York, 1999.
[112] O.J. Reichman, M.B. Jones, M.P. Schildhauer, Challenges and opportunities of open data in ecology, Science 331 (6018) (2011) 703–705.
[113] J. Wang, R. Zhu, S. Liu, A differentially private unscented Kalman filter for streaming data in IoT, IEEE Access 6 (2018) 6487–6495.
[114] E. Nazerfard, P. Rashidi, D.J. Cook, Using association rule mining to discover temporal relations of daily activities, in: Toward Useful Services for Elderly and People with Disabilities: 9th International Conference on Smart Homes and Health Telematics, ICOST 2011, Montreal, Canada, June 20-22, 2011. Proceedings 9, Springer, 2011, pp. 49–56.
[115] E. Nazerfard, P. Rashidi, D.J. Cook, Discovering temporal features and relations of activity patterns, in: 2010 IEEE International Conference on Data Mining Workshops, IEEE, 2010, pp. 1069–1075.
[116] K.R. Sollins, IoT big data security and privacy versus innovation, IEEE Int. Things J. 6 (2) (2019) 1628–1635.
[117] X. Yao, F. Farha, R. Li, I. Psychoula, L. Chen, H. Ning, Security and privacy issues of physical objects in the IoT: challenges and opportunities, Digit. Commun. Netw. 7 (3) (2021) 373–384.
[118] T. Mazhar, D.B. Talpur, T.A. Shloul, Y.Y. Ghadi, I. Haq, I. Ullah, K. Ouahada, H. Hamam, Analysis of IoT security challenges and its solutions using artificial intelligence, Brain Sci. 13 (4) (2023) 683.
[119] J. Jiang, F. Liu, Y. Liu, Q. Tang, B. Wang, G. Zhong, W. Wang, A dynamic ensemble algorithm for anomaly detection in IoT imbalanced data streams, Comput. Commun. 194 (2022) 250–257.
[120] H. Bi, J. Liu, N. Kato, Deep learning-based privacy preservation and data analytics for IoT enabled healthcare, IEEE Trans. Ind. Inform. 18 (7) (2021) 4798–4807.
[121] R. Hai, S. Geisler, C. Quix, Constance: an intelligent data lake system, in: Proceedings of the 2016 International Conference on Management of Data, 2016, pp. 2097–2100.
[122] 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, Comput. Netw. 129 (2017) 459–471.
[123] ProjectPro, Internet of Things (IoT)'s Impact on Big Data and Data Science, https://www.projectpro.io/article/internet-of-things-iot-s-impact-on-big-data-and-data-science/249, 2022. (Accessed 11 April 2023).
[124] B. Glavic, Big data provenance: challenges and implications for benchmarking, in: Specifying Big Data Benchmarks: First Workshop, WBDB 2012, San Jose, CA, USA, May 8-9, 2012, and Second Workshop, WBDB 2012, Pune, India, December 17-18, 2012, in: Revised Selected Papers, Springer, 2014, pp. 72–80.
[125] C. Perera, A. Zaslavsky, P. Christen, D. Georgakopoulos, Context aware computing for the Internet of Things: a survey, IEEE Commun. Surv. Tutor. 16 (1) (2013) 414–454.
[126] J. Dean, S. Ghemawat, Mapreduce: simplified data processing on large clusters, Commun. ACM 51 (1) (2008) 107–113.
[127] C.C. Aggarwal, Data streams: models and algorithms, in: Advances in Database Systems, vol. 31, Springer Science & Business Media, 2007, pp. 18–354.
[128] G. Krempl, I. Žliobaite, D. Brzeziński, E. Hüllermeier, M. Last, V. Lemaire, T. Noack, A. Shaker, S. Sievi, M. Spiliopoulou, et al., Open challenges for data stream mining research, ACM SIGKDD Explor. Newsl. 16 (1) (2014) 1–10.
[129] O. Diallo, J.J. Rodrigues, M. Sene, J. Lloret, Distributed database management techniques for wireless sensor networks, IEEE Trans. Parallel Distrib. Syst. 26 (2) (2013) 604–620.
[130] M.A. Alsheikh, D. Niyato, S. Lin, H.-P. Tan, Z. Han, Mobile big data analytics using deep learning and apache spark, IEEE Netw. 30 (3) (2016) 22–29.
[131] G. Xie, H. Peng, Z. Li, J. Song, Y. Xie, R. Li, K. Li, Reliability enhancement toward functional safety goal assurance in energy-aware automotive cyber-physical systems, IEEE Trans. Ind. Inform. 14 (12) (2018) 5447–5462.
[132] M. Borkowski, S. Schulte, C. Hochreiner, Predicting cloud resource utilization, in: Proceedings of the 9th International Conference on Utility and Cloud Computing, 2016, pp. 37–42.
[133] A. Gharaibeh, A. Khreishah, M. Mohammadi, A. Al-Fuqaha, I. Khalil, A. Rayes, Online auction of cloud resources in support of the Internet of Things, IEEE Int. Things J. 4 (5) (2017) 1583–1596.
[134] H.U. Khan, M. Sohail, F. Ali, S. Nazir, Y.Y. Ghadi, I. Ullah, Prioritizing the multi-criterial features based on comparative approaches for enhancing security of IoT devices, Phys. Commun. 59 (2023) 102084.
[135] Y. Shi, J. Han, X. Wang, J. Gao, H. Fan, An obfuscatable aggregatable signcryption scheme for unattended devices in IoT systems, IEEE Int. Things J. 4 (4) (2017) 1067–1081.
[136] R. Lu, H. Zhu, X. Liu, J.K. Liu, J. Shao, Toward efficient and privacy-preserving computing in big data era, IEEE Netw. 28 (4) (2014) 46–50.
[137] M.H. Au, K. Liang, J.K. Liu, R. Lu, J. Ning, Privacy-preserving personal data operation on mobile cloud—chances and challenges over advanced persistent threat, Future Gener. Comput. Syst. 79 (2018) 337–349.
[138] J. Zhang, G. Cormode, C.M. Procopiuc, D. Srivastava, X. Xiao, Privbayes: private data release via Bayesian networks, ACM Trans. Database Syst. 42 (4) (2017) 1–41.
[139] R. Mendes, J.P. Vilela, Privacy-preserving data mining: methods, metrics, and applications, IEEE Access 5 (2017) 10562–10582.
[140] B. Wang, B. Li, H. Li, Panda: public auditing for shared data with efficient user revocation in the cloud, IEEE Trans. Serv. Comput. 8 (1) (2013) 92–106.
[141] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, W. Zhao, A survey on Internet of Things: architecture, enabling technologies, security and privacy, and applications, IEEE Int. Things J. 4 (5) (2017) 1125–1142.
[142] J. Liu, M. Ahmed, M.A. Mirza, W.U. Khan, D. Xu, J. Li, A. Aziz, Z. Han, RL/DRL meets vehicular task offloading using edge and vehicular cloudlet: a survey, IEEE Int. Things J. 9 (11) (2022) 8315–8338.
[143] S. Chen, B. Tang, K. Wang, Twin delayed deep deterministic policy gradient-based intelligent computation offloading for IoT, Digit. Commun. Netw. 9 (4) (2023) 836–845.
[144] P. Mach, Z. Becvar, Mobile edge computing: a survey on architecture and computation offloading, IEEE Commun. Surv. Tutor. 19 (3) (2017) 1628–1656.