

## Review Article

# Big Data and Its V's with IoT to Develop Sustainability

Xiao Chu,<sup>1</sup> Shah Nazir ,<sup>2</sup> Kunhao Wang ,<sup>3</sup> Zeqi Leng,<sup>3</sup> and Wajecha Khalil<sup>4</sup>

<sup>1</sup>School of Information Engineering, Changchun University of Finance and Economics, Changchun 130122, China

<sup>2</sup>Department of Computer Science, University of Swabi, Swabi, Pakistan

<sup>3</sup>College of Computer Science, Jilin Normal University, Siping 136000, China

<sup>4</sup>Department of Computer Science and Information Technology, University of Engineering and Technology, Peshawar, Pakistan

Correspondence should be addressed to Shah Nazir; [snsahnzr@gmail.com](mailto:snsahnzr@gmail.com) and Kunhao Wang; [wkhyoyo@jlnu.edu.cn](mailto:wkhyoyo@jlnu.edu.cn)

Received 28 May 2021; Accepted 23 June 2021; Published 3 July 2021

Academic Editor: Zhu Xiao

Copyright © 2021 Xiao Chu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Internet of Things (IoT) has been considered as one of the emerging network and information technologies that can comprehend automatic monitoring, identification, and management through a network of smart IoT devices. The effective use of IoT in different areas has improved efficiency and reduced errors. The rapid growth of smart devices such as actuators, sensors, and wearable devices has made the IoT enable for smart and sustainable developments in the area. Physical objects are interlinked with these smart devices for the progression to analyse, process, and manage the surroundings data. Such data can then be further utilised for smarter decisions and postanalysis for different purposes. However, with the limited IoT resources, the management of data is difficult due to the restrictions of transmission power place and energy consumption, and the processing can put pressure on these smart devices. The network of IoT is connected with big data through Internet for manipulating and storing huge bulk of data on cloud storage. The secure framework based on big data through IoT is the awful need of modern society which can be energy efficient in a sustainable environment. Due to the intrinsic characteristics of sensors nodes in the IoT, like data redundancy, constrained energy, computing capabilities, and limited communication range, the issues of data loss are becoming among the main issues which mostly depend on the completeness of data. Various approaches are in practice for the recovery problem of data, such as spatiotemporal correlation and interpolation. These are used for data correlation and characteristics of sensory data. Extracting correlation data became difficult specifically as the coupling degree between diverse perceptual attributes is low. The current study has presented a comprehensive overview on big data and its V's with Internet of Things to describe the research into the area with in-depth review of existing literature.

## 1. Introduction

IoT paradigm is a mixture of three visions: “things,” “Internet,” and “semantic-oriented” visions. The IoT is a set of disparate networks that are connected and addressed via a common communications protocol, either from a things-oriented or Internet-oriented perspective [1]. It is difficult to provide reliable and realistic insights in industrial Internet applications with sensors that transmit complex data at high speeds. One major issue is the processing of large amounts of data, since the machine's underlying dynamic patterns change over time due to a variety of factors, including degradation. As a consequence, the actionable model has become obsolete, and it is important to update it. The article

proposes the Gaussian-dependent dynamic probabilistic clustering (GDPC), a new deep learning algorithm centred on Gaussian mixed models that is based on the integration and optimization of three well-known methods for use in complex environments. The expectation-maximization (EM) algorithm was used to estimate parameter values, and the Page-Hinkley test with the Chernoff bound had been used to detect concept drifts. In contrast to other unregulated models, GDPC's model assigns membership probabilities to cluster. This can detect the robustness and evolution of the instance assignment each time a concept drift can be detected through a Brier score analysis. In addition, the algorithm operates with very small amounts of data and greatly reduces calculation power to determine

whether to modify the model. The algorithm can be evaluated on synthetic data and data stream from a test bed where various operating conditions are detected automatically, with good results in terms of classification precision, sensitivity, and characteristics [2]. Based on trajectory data, regular behaviour of private cars are extracted [3–5].

In digital technology such as IoT and big data, video-based surveillance plays an important role in the identification of pedestrians in the smart industries. Due to inconsistent sorting factors such as backgrounds, lighting, clothes, occlusion, and object collisions in pedestrians, the detection is a daunting issue. To tackle these challenges, an improved feature extraction is required. Multiple attributes can be driven from varying pedestrians. For feature detection of pedestrian histogram of gradients, Haar that characterizes the ordering of boundary level classifications and scale-invariant feature transform is used. Occlusion extraction feature supports the recognition of pedestrian detection regions. In addition, pedestrians are classified by classifiers such as support vectoring machines and random forests. All of these extraction and foot detection techniques are now streamlined due to deep learning techniques like convolutional neural networks (CNNs). To achieve accurate results, a CNN approach was used to train the model by supplying favourable and unfavourable sets of images as well as larger sets of data. Extensible mark-up language cascade is often used to detect and recognize face of the identified pedestrian [6]. In recent years, the intelligent factory has become a subject of research for both academia and industry in the form of Industry 4.0 and the Internet of Things (IIoT). The requirement for data exchange with different time flows between various smart devices is increasing in IIoT. There is, however, limited research on this topic. The integration of a global unified software-defined network (SDN) and edge computing (EC) in IIoT with EC has been taken into account to resolve the limitations of conventional methods and solve the problem. SDN and EC for IIoT suggested adaptive transmission architecture. The specifications can be divided into two groups of standard and emerging sources, based on data streams with different latency constraints. A coarse-grained algorithm in the short time situation provides all the way to the hierarchical IoT that meets time constraints. Following that, using the path difference degree (PDD), an optimum planning route is selected, considering the time frame, traffic load ratios, and power consumption add-on. If the net grained strategy is far beyond the scenario, a finely crafted procedure is used to set the effective transmission route using an optimum power strategy for low latency in a high-deadline scenario. Finally, simulation evaluates the success of the proposed plan. The results show that the proposed system is above the average time, goods, output, PDD, and download time in relation to the relevant methods. The proposed method offers a better way of handling IIoT data [7]. Various approaches and studies exist with the applications of big data in different fields [8–10].

The proposed study has presented a wide-ranging overview on big data and its V's with Internet of Things to describe the research into the area with in-depth review of existing literature and research work done in the area.

The organization of the paper is described as follows. Section 2 presents the literature of the proposed study. Section 3 elaborates the approaches used in IoT domain. The big data and IoT paradigm is given in detail in Section 4. Section 5 briefly shows the determinations of the IIoT and big data. Existing approaches of IoT and big data in different research areas are given in Section 6. Section 7 shows the analysis of literature about the area of research.

## 2. Literature Review

Industry 4.0 is considered as the compiler of the new technology with conventional production method. Industry 4.0 adds abundant value to corporations' production processes for addressing existing and future competition and challenges. The digital architectures help enterprises to build the digital ordering network between vendors and producers. Companies are looking for an IIoT analytical mapping technique for calculating the success of own partners in consideration of IIoT architectures. On the other hand, companies also emphasize the evaluation of the important IIoT architectures to determine the universal single layer IIoT models that could apply those techniques. In the first step, the authors perform several important literature surveys in IIoT to list only significant IIoT architectures to create the universal model. Furthermore, many literature surveys help authors assess empirical IIoT mapping technology for own partner/suppliers to benchmark. The authors quantify the research with completion of the architectures of the IIoT, i.e., cyberspace, networking, virtual reality, data storage, and security. In order to promote Industry 4.0 undertakings with a proposed universal model cum analytical methodology, a fuzzy-grey relational analysis computational technique for IIoT model was employed to allow companies to map their own partner's achievements under the proposed model. The hypothesis in SA automation technology company's empirical case study justified the actual application of scientific analysis [11]. Industrial IoT produces big data which is useful to gain information from the data analysis, but it is a strain to store all data. The industrial data was compressed by regressing the neural network into a representative vector with loss compression. The dividing and conquer method was applied for compression efficiency. Experiments confirmed that industrial data can be seen by a function and predicted with high precision [12].

There is a surge in curiosity in the Industrial Revolution 4.0, which is focused on information technology. Interconnection, robotics, intelligent systems, optimization, and IoT are being viewed as innovations capable of creating certain goals in addition to the Industrial Revolution 4.0. The IoT was among the most remarkable innovations, and it has the potential to bring in the Industrial Revolution 4.0. However, since various sensors, modules, and drives are spread across the IoT, the use of different protocols constantly leads to problems with heterogeneity. There is now a need for a framework to intelligently and meaningfully use the big data produced by the rapid adoption of IoT devices. For virtualizing different IoT devices, a method known as a

Cloud of Things based on linked data has been suggested. To address heterogeneity issues and organize virtualized objects in linked data types to form interconnected device meshing devices, virtualized devices in the cloud using real-world device metadata are used. This enables self-contained knowledge, such as connection through a linked data-based computer mesh and results via big data linkage. Scenarios demonstrated how this method can be used for the cloud of items based on relevant data [13]. In recent years, optical field image technology had got a lot of attention in the academic community because of its new imaging properties like first and later shooting, variable field depth, and variable perspective. However, existing optical field acquisition equipment can only obtain a small number of discrete angle signals, resulting in alias in optical field images due to the sampling of optical field edge signals. It investigates the technique of optical field imaging and the method of depth estimation based on big data in the IoT collected by the camera array around the angular sampling features of the optical field data collection using the camera array system as a medium. A depth estimation approach is suggested, combining parallax and focusing procedure, in order to analyse the characteristics of different depth indices in the optical field data collection. To commence, the study examines the disparity and focus clues found in the camera array's multiview dataset and light field focused image set, respectively, and highlights the differences and relationships between the two methods of extracting depth clues across the augmented reality frequency field sampling area. The weighted linear image gradient-based fusion approach then fuses the two measurement effects, improving precise and robust depth assessment. Finally, the results of the profoundly assessment trials on various scenes show that the method in this paper is more precise in the measurement of depth in discontinuous scene areas and related texture areas than the method based on one single deep cue [14].

### 3. Approaches Used for IoT Domains

The following sections briefly describe the approaches used of the IoT domains.

*3.1. Industrial Applications and IoT.* Current approaches to location privacy security depend primarily on traditional asymmetric encryption, fuzzy, and cryptography techniques, with limited success in the world of big data, where sensors, for example, pose a serious threat that must be adequately secured. Current technologies, such as "Industry 4.0" and the IoT, collect, store, and exchange massive amounts of security-critical and autonomy data, making them an easy target for hackers. However, in the past, the topic of data protection was overlooked, resulting in privacy violations. To protect the privacy of local data and to improve the usability of data and algorithms in the industrial IoT, a data security strategy was proposed to satisfy differential privacy restrictions. Because of the high value and low data density, a multilevel tree model of location data was introduced to balance utilities and privacy. Furthermore, the differential

privacy index function is used to select data depending on how much the tree node accesses it. Finally, Laplace is used to add noises to the frequency of access to the data selected. This strategy can lead to substantial improvements in safety, privacy, and applicability, as demonstrated in theoretical research and experimental results [15]. A high-order clustering algorithm with quick search and density peak detection has recently emerged to discover latent data structures in big data and will produce great application values in the fields of industrial data management and analysis. The popularity of cloud computing makes Outsourcing Calculations convenient for users and also puts them at risk of confidentiality. Focused at the above issue, the present study proposes a stable high-order cluster algorithm by rapid search and finding of density peaks in the hybrid cloud due to the characteristics of the secure cloud service system. The client will first create the encrypted object tensors, using homomorphic encryption, with user data, and then upload these to the cloud, in order to enforce all of the protocols proposed. Finally, random numbers will return clustering results to the client to eliminate the disturbance. In terms of cluster precision, reliability, and speed-up ratios, the performance of the proposed method is assessed on an intelligent grid data collection. Experimental findings show that the method can cluster data without exposing user privacy with precision and effectiveness, while ensuring that the customer is very flexible. The proposed system with high levels of protection and scalability is therefore ideal for the clustering of IIoT big data [16].

*3.2. Industrial Internet of Things.* The Internet of Vehicles (IoV) is the key to the intelligent transport industry in association with IoT, which enables omnipotent exchange of information and sharing of content among vehicles with little or no human intervention. The research analysis examined the combination of physical and social layer knowledge for rapid content diffusion in IoV-based device-by-device vehicle-to-vehicle (D2D-V2V) networks. The progressive distance of vehicles is modelled as a Wiener process in the physical layer, with the application of the Kolmogorov equation estimating the connection probability of D2D-V2V connections. In social terms, Bayesian non-parametric learning on the basis of social data from reality, which is collected from Sina Weibo's biggest Chinese microblogging service and Youku's largest Chinese video sharing website, represents the tightness of the social relationship that is similar to selecting content. In addition, a price-based iterative matching algorithm under different quality-of-service standards resolves the formulated joint pair discovery, power monitoring, and channel selection problems. In conclusion, numerical results show that, in view of the weighted sum and matched satisfaction benefits, the proposed algorithm is efficient and superior [17].

The study aims to demonstrate the life extension programme of the FSO2 through big data and the Industrial Internet of Things (IIoT). The aim of the study is to explain how big data and IIoT technologies are to be implemented and help in producing advanced technology and predictive

lifespan maintenance of the FSO2 programme. The FSO2 life extension software began to be implemented in 2014. The FSO2 was certified without dry docking by the ABS class for a design life of 15 years. The aim of this project was to develop a programme and solution for the ABS class for a lifetime extension of 10 years without dry docking. The analysis shows the current condition and the preceding steps [18]. The IoT will gather different types of sensor data. There are spatial properties in each sensor node, and a large amount of measuring data that develops over time can also be linked to it. Fundamentally high-dimensional sensor data is present. It is an extremely difficult task to detect outliers in large IoT sensor data. Most of the methods of detection of anomalies are based on vectors. However, large IoT sensor data possesses features which increase the efficiency of tensor methods for information extraction. The methods based on vectors can eliminate original structural information and correlations in broad sensor data, leading to the problem of “dimensionality curses.” In this research, a one-class support Tucker machine (OCSTuM) and an OCSTuM based on tensor factorization and a genetic algorithm known as GA-OCSTuM were proposed. The techniques expand one-class vector machines to space tensoring. The anomaly detection methods for large sensor information are unchecked OCSTuM and GA-OCSTuM. Data structural information has been maintained, and anomaly detection accuracy and reliability increased. Experimental analyses of real datasets have shown that the approach proposed increases anomaly detection accuracy and efficiency while maintaining the structure of large sensor data [19].

A major function in designing the industrial data platform is the larger data acquisition and storage system (ASS). Many big data systems have been compressed and encoded. Such methods cannot satisfy the needs for time-consuming and mass storage of industrial data management. An efficient industrial big data platform based on existing big data systems is designed, to eliminate data processing times while consuming less storage space. The study examines the effects of various compression and encoding approaches on the performance of a big data platform, with the aim of determining the best compression and encoding approach for an industrial data platform. The test results revealed that the platform’s data compression time was reduced by 73.9 percent compared to Hadoop and Spark approaches, with less than 96 percent compressed data, and data serialisation time was reduced by 80.8 percent. Compare it with benchmarking approaches with the growing amount of data [20]. The privacy related to cybersecurity is grouped within the framework of the industrial Internet of Things, and it has been assessed using two specific methods, specifically built to be relevant to industrial environments, and assessment techniques were carried out for protection of Internet IoT devices that will be used in an industrial infrastructure environment. Case studies proved the problems and issues of cybersecurity caused by such specific technologies and revealed how the rules, requirements, and technological mechanisms were resolved. The purpose of such case studies is to demonstrate that regulatory and technological efforts in industrial contexts protect against

cybersafety concerns. The research study also helped in explaining the problems of security and the implementation of standards and tools within industrial environments to solve cybersecurity issues [21].

Earlier blockchain data transmission techniques have low security, high trading centre management costs, and great monitoring difficulties in the IIoT. A safe fabric blockchain-based technology for industrial IoT was proposed for addressing these problems. This technique utilizes the dynamic secret sharing system that is based on blockchain. The power blockchain sharing model, which can also exchange power trading books, produces a stable trading centre. The process of the power data consensus and the dynamic connected storage were built to ensure efficient matching of the power data transfer. Experiments show that high protection and reliability are provided by fabric optimized data storage and transmission. The approach suggested will increase the rate of transmission and receipt of packets by 12 and 13 per cent, respectively. In addition, the technology suggested has strong management and decentralisation supremacy [22]. In an intelligent plant, thousands of IT devices and sensors are mounted on production machines in order to gather big data on machine conditions and to transfer them to a cyber-physical infrastructure in the factory’s cloud centre. The device then uses a number of CBM methods to estimate the time when machines continue to run irregularly and to retain or upgrade components in order to prevent the production of large detecting objects. CBM suffers from idea drifts complications (i.e., fault distribution can alter over time) and data imbalance (i.e., the data with faults accounts for a minority of all data). A high-performance approach to these issues is to incorporate learning that combines the diversity of multiple classifiers. Many firms do not have the capacity to create a sound infrastructure for the development of Internet classifiers in real time but may have offline classifiers on their current networks. However, much of the previous supervised work is concentrated only on online classification promotion. Consequently, a learning algorithm ensemble has been suggested, supporting offline grading to meet the 3-stage CBM with definition drifts and inequality data, which employ the improved Dynamic AdaBoost. The NC grading and the MOTE method to solve the imbalance data are used at stage 1 (training an ensemble classifier); the improved method LFR (linear quarter rates) is used at stage 2 and at stage 3 (creation of new ensemble), a new ensemble is created for detecting concept drifts from imbalance data. The experimental results on datasets with different degrees of imbalance showed that it is possible for the proposed system, with a high accuracy rate of over 94 percent, to detect all concept drifts and to recognize minority class data [23].

Spatial data from satellites, drones, and big data is an essential factor for all operations in disaster management, such as prevention, preparedness, response, and mitigation (mobile CDRs, trajectory data, GPS, wireless sensor network, and IoT). The invention of a global navigation system and wireless communications has revolutionized the way we operate and collect geospatial information. For example, a significant amount of geospatial data streams can be

collected and transferred from the IoT cloud server or central geodatabase as the base map of the database repository. The collection, dissemination, and display of all collected geospatial data are important to efficiently plan and minimize hazard adequate information to property, and life loss recovery teams must be obtained within a minimum time period. The article helps in creating a city geospatial dashboard to capture, exchange, and monitor data from satellites, IoT devices, and other large data gathered by geospatial data. In order to improve spatial analysis and planning processes in disaster management, a set of space analytical tools known as geovisualization, such as a virtually real-time rainfall profiling system, approximate population, and flow direction of mobile CDR, were used to analyse large-scale data for performance evaluation [24].

The existing concept and strategies for building a cross-industrial IoT service network at a group level have been used. The use of fixed geodistributed wireless IoT gates and mobile facilities allows fast deployment of local IoT services to offer social and economic benefit in a city. The idea proposed is consistent with the strategy of “local data output for local data use,” as local yet socially useful data may provide different types of IoT service without the need for a mobile network or for a cloud/data centre for big data. This idea was prototyped by the use of soft drink facilities, i.e., sales machinery, vans, and taxi services facilities, i.e., taxis in Tokyo, Japan, to show a prototype platform. The platform’s role as a data transfer network for socially beneficial data dissemination can be seen, and the result is evaluated [25].

The production of large-scale data gives the plant a huge opportunity to turn the existing production model into smart production. Multisource data modelling and integration issues are nevertheless the current differences between the collected big data and the intelligent applications powered by data. The extensive Internet deployment of products on the shop floor means that the manufacturing big data driven has to be manageable and organized with adequate data modelling and integration methods. The spatiotemporal modelling of the data in temporal, spatial, and attributive dimensions is initially presented in this study. Furthermore, the ontological approach for integration of big data is proposed to handle the manufacturing data of multisource providers and to ensure that the data for various subsequent applications can be easily indexed and easily reused. Finally, through the existing large data analysis and decision-making framework, the proposed data modelling and integration methods are applied and tested [26]. The IIoT is a manufacturing trend and a necessary component of the smart factory. In the industrial IoT, data transmission protection is extremely important. The key contribution of this paper is the implementation of a new chaotic secure communication scheme to deal with the security issue of data transmission. The framework is suggested and analysed by synchronising fractional order chaotic systems with various structures with various orders. In order to validate synchronisation between the fractional order drive and the reaction system, Lyapunov’s stability principle is used. In order to encrypt and decrypt the key data signals, the  $n$ -shift encryption principle is used. The main area of the scheme is

calculated and analysed. The efficiency of theoretical approach is shown through numerical simulations [27].

#### 4. Big Data and IoT Paradigm

The abrupt changing economic standards and revolutionary breakthroughs in the IoT would soon produce a large data explosion. This in turn would include data collection and cloud platform processing in real time. Wide and geographical data centres serve a significant part of distributed data centres (DCs). However, these DCs impose significant costs in terms of exponentially increasing energy usage and damage the environment in turn. In this respect, effective use of resources is often seen as a possible candidate for improving energy efficiencies and reducing the burden on the electricity sector. In most public clouds, however, the resources are usually idle (i.e., underused), since the server load is unpredictable, which leads to a considerable increase in energy use and resource waste. For this reason, a precise and effective resource management methodology is extremely important. The benefits of SDDCs (software-defined data centres) have been used to minimize the use of resources. In particular, SDDC refers to the method of abstracting the logical computing, networking, and storage resources in a programmatic way to potentially develop consolidated models based on SDDC to optimise the processes of VM deployment and network bandwidth assignment, particularly for heterogeneous computer infrastructures in order to achieve an optimization of resource and in addition to formulate a multiobjective optimization problem. The presented work is not optimal based on First Fit Decreasing (FFD). Moreover, the proposed framework shows that the proposed framework achieves reductions in energy usage of about 27.9% compared with existing schemes with negligible QoS breaches (approximately 0.33) [28]. The IoT used in various smart factories is similar in a vertical industry alliance. For instance, most automakers have similar assembly lines and IoT monitoring systems. Industrial information is commonly observed using IoT data-dependent deep learning and data mining. Some information, however, cannot be easily extracted from just one factory’s data, as there are still few samples. If several factories can collect their data together through an alliance, further information can be extracted. However, data protection is the main concern of these factories. Existing matrix-based approaches can ensure data protection within a plant but do not share data between factories due to a lack of correlation and, therefore, their mining efficiency is weak. To address this problem, the research proposed a new federated tensor mining (FTM) system to combine multisource mining data in order to provide security for tensor-based mining. FTM’s main contribution is to share its ciphertext data for security purposes only, and because of its homomorphic attribution, this ciphertext is suitable for tensor-based information mining. Evidence-driven simulation results show that FTM not only exploits the same information as plaintext mining, but it is also permitted by distributed eavesdroppers and unified hackers to protect attacks. FTM improves the mining accuracy by up to 24% in

our typical experiment compared with the matrix-based privacy-preserving compressive sensing (PPCS) [29].

The amount of accessible geospatial data is growing every minute as a result of big data and IoT, which is increasing the acquisition and retrieval of geospatial data. This necessitates the use of a cutting-edge data processing system. On the basis of network virtual reality geographic data, a “building information model (BIM)” with hybrid storage architecture and big data-storage-management approach has been proposed (WebVRGIS). At various stages of urban development, BIM is linked to the integration of spatial and semantic knowledge. A data storage and management model for BIM geospatial big data management was proposed based on the spatial distribution features of BIM geospatial big data. Not Only the Structured Query Language (NoSQL) but also the database and decentralized peer-to-peer processing are critical components of the architecture. The proposed storage model is applied using the same software framework that was used in the previous WebVR study. The experimental results show that the proposed hybrid storage model is less time-efficient than the conventional connection database in geo-big data searches for this study. The incorporation and fusion of BIM large data in WebVRGIS transforms city knowledge management in a revolutionary manner across the entire life cycle. The system is also very promising for storing other geospatial data including traffic information [30]. High-performance analysis tools and algorithms are necessary with advanced wireless communication, the IoT and big data. Data clustering, promising analytical method, is commonly employed as it does not really need labelled datasets to solve IoT and big data-related problems. Metaheuristic algorithms have recently been used in a number of different clustering problems efficiently. However, these algorithms do not react within a desired time to handle large datasets from IoT devices because of the high cost of calculation. The research presented a novel metaheuristic clustering approach to solve big data problems through the use of MapReduce intensity. The methods proposed utilise the military dog group’s quest potential to find the perfect centroids and MapReduce architecture to manage the large datasets. The optimization effectiveness of the proposed approach is validated by 17 benchmarking functions, compared with 5 other recent algorithms, namely, artificial bee colony, bat, multiverse optimization, particle swarm optimization, and whale optimization algorithm; in order to cluster large datasets generated from industrial IoT, the parallel version of the suggested technique will also be implemented using MapReduce (MR-MDBO). In addition, MR-MDBO performance is investigated using 2 UCI benchmark datasets and 3 actual industry-related IoT datasets. The MR-MDBO is compared to 5 other advanced methods with F measurement and computing time. The experimental results indicate that the clustering based on MR-MDBO is superior to the other considered algorithms in terms of cluster precision and calculation time [31].

The IIoT evolved quickly with the emergence of 5G. There has also been widespread interest in the industrial sensor-cloud system (SCS). Mostly in future, several integrated sensors will be added to industrial SCS, which

simultaneously collect multifunctional data. Because of the harsh sensor world, however, the collected large data are not reliable. If the data obtained on the bottom network is downloaded directly to the cloud for processing, the results of the query and data mining would be unreliable, affecting the cloud judgement and feedback seriously. The conventional approach for data cleaning based on sensor nodes is inadequate to handle big data while edge computing offers a good solution. A new data cleaning method based on the mobile edge node during data collection is proposed. At the border node, an angle-based outlier detection method is used to obtain the cleaning model training data that is then defined by means of a support vector machine. In addition, online learning is used to optimize the model. Experimental findings show that multidimensional data cleaning, which takes the form of mobile edge nodes, increases efficiency of the data cleaning while preserving data reliability and integrity [32].

The industrial Internet of Things (IIoT) is rapidly gaining traction as smart sensors, instruments, computers, and applications are increasingly deployed and connected over wired and wireless networks. Industrial practices will be greatly improved, and industrial information will be developed more efficiently, thanks to this integrated hardware software strategy. To recognize and use secret information that is valuable and useful in the manufacturing process, significant developments in IIoT big data processing and analysis are needed. Large-scale, streaming, multiattribute IIoT output data, on the other hand, is unreliable and redundant. As a result, an appropriate data processing technique, such as a tensor train, is required to process these IIoT data. Current tensor-train decomposition methods, on the other hand, are inefficient and unfit for large-scale IIoT big data processing. An incremental computational framework was provided with an advanced (strengthened and super-effective) distributed tensor-train (ADTT) decomposition method for the study of IIoT big data. Finally, tests are run on regular IIoT data that is publicly available. Validation and evaluation of the proposed ADTT system’s performance test data were carried out [32]. Many major difficulties exist in cloud processing, including cloud-based analytics of massive data and decision-making processes that cannot meet the criteria for multiple latency-sensitive applications on the shop floor, in addition to a lack of reconfiguration, transparency, openness, and evaluability of current manufacturing systems to deal with shop floor disorders and changes on the market. The Internet and big data from store floors were not used effectively to automate and upgrade production processes. An open evolutionary architecture of collaborative edge and cloud processing of the intelligent cloud development framework has been suggested. Hierarchical gateways connecting and managing “edge” shops are provided to support latency-sensitive applications to react in real time. Large, cloud and gateway data are processed to help constantly enhance and develop edge-cloud systems to improve performance. As software is dominating manufactory control and decision making like “brain,” it also proposes an AI-enabled Manufacturing Operations (AI-Mfg-Ops) mode with software-defined framing that can help

rapid operation and upgrade of cloud manufacturing systems in an in-line loop with intelligent monitoring, analysis, planning, and implementation. Research may lead to the rapid response of cloud production systems and their efficient functioning [33].

The study of data mining has aroused the interest of both academia and industry. The IoT is defined by the fact that sensor data artificially replaces assembled data. The ability to extract useful information and patterns from a large volume of sensor data is now a worthwhile research subject. For sensor data processing, a dynamic data mining approach was suggested. A sensor model for data mining that can be used for dynamic change has been developed. Different physical structures are seen in various sensor network settings in that model. By collecting historical changes in sensor data, the physical system and its parameters are educated, and the links between different sensor network contexts are discovered by exploiting the links between physical system parameters. In a limited experimental environment, physical quantities such as transmission distance, transmission delay, sensor data, and data changes were considered. The model has been tested on the experimental platform, and the results show that it can mine unpredictable data and find stable patterns. After analysing the experimental results, it was discovered that the model had a reference value for dynamic sensor data mining, and new methods for evaluating industrial big data were expected to be developed [34]. Advanced sensing, data collection, and technology for communication have contributed in recent years to a huge growth of the IIoT, which increases the revolution in electronic asset condition monitoring and maintenance. An open ecosystem was suggested for the future IIoT and the open ecosystem architecture. An open development environment is needed so that users can freely communicate with power devices and servers on user terminals via web or mobile applications, thus enhancing IIoT scalability and flexibility. The core open ecosystem technology for the future IIoT will include robust sensing techniques, wide area communication methods, a large data services platform, algorithms for data processing, and smart maintenance schemes. The potential IIoT ecosystem is then addressed in the management of wind farms. It is shown to increase wind farm maintenance quality and efficiency by supporting an open ecosystem of future IIoT offering a ground-breaking perspective on controlling and maintaining electrical assets with great reliability [35].

## 5. Determinations of the IIoT and Big Data

The virtualization of real time is usually recognized as one of the central promoters of fog computing and industrial Internet of Things (IIoT). Any hypervisor who qualifies as a virtualization solution to be deterrent to the IIoT must meet specifications. An example of the compromise between versatility and deterministic execution is current works in the area of virtualization in real time. There was a shortage of hypervisors that fulfilled all the deterministic virtualization criteria. Preliminary experimental findings comparing ACRN, KVM, and Xen RTDS device latency support

statements for further investigation of deterministic virtualization requirements [36]. To avoid ransomware attacks on IIoT systems, host machine computer operations need a powerful detection model that can reliably detect ransomware behaviour and trigger an alert before the infection spreads to critical control systems. However, detecting models with high-dimensional data, as well as a few qualified observations combined with ransomware dynamics for host machines, is difficult. To address these issues, an effective detection model is essential. To reveal the framework hidden for system operations and ransomware behaviour, the Variational Autoencoder (VAE) model was proposed with a fully connected neural network. To boost the widespread detection model capabilities, a VAE-based data increase method for generating new data was created; it can be used in a fully connected network training. The findings showed that the proposed model is very effective in detecting ransomware [37]. In our personal lives, digital goods and services are a popular spot in which software and its algorithms provide aid. Interactive systems, however, need to match consumer products within an industrial environment, particularly in terms of interactive quality and user experience. The position of human work has been questioned, and the value of collaboration has been highlighted by the increase in automation and data sharing at large scales. New concepts in intelligent factories, where machinery and software perform working tasks, dramatically change work nature from manual labour to increasingly complex tasks in industrial settings. HCI and CSCW in particular have ideas, techniques, and strategies to tackle this disruptive transition to an IIoT. For instance, networked assistance systems may meet the various needs of heterogeneous individual staff. In order to explore the design space of IIoT applications, their impact upon cooperation work and formulation of new research opportunities and new perspectives on HCI and CSCW in industrial contexts were explored in the context of the emerging IIoT [38].

The industrial automation industry undergoes a huge transition to increasingly integrated and globally distributed automation systems with the introduction of Internet of Things (IoT) and cyber-physical system (CPS) concepts. As a result, the industry faces challenges in terms of interoperability between devices and systems that have evolved in recent years as a result of business and technology fragmentation. Due to tighter reliability and real-time constraints, proven IoT integration techniques cannot be completely adapted in IIoT environments. As design models provide a realistic tool for understanding the specific problem more deeply, models are used to create a software architecture appropriate for implementation in future IIoT environments. IoT world concepts, industrial automation systems, and modern IT architecture and cloud architectures are combined in the resulting software architecture. It is easy and versatile design, and the help of state-of-the-art approaches (containerisation, continuous integration (CI), continuous deployment (CD)) makes it just as suitable for cloud, fog, and edge deployment. All these features make it possible to deploy device-level services and communication protocols to make it possible for heterogeneous systems and

protocols to be transparent and automatically integrated on request [39]. The IIoT connects control systems to major business and industry innovations. However, new cybersecurity vulnerabilities are also involved in this progress. As the utility of IIoT systems is at the edge level, they can be sought by attackers. It is, therefore, of highest concern to protect physical structures at the edge by detection and identification of malicious activities based on an effective detection model. A detection model based on profound learning techniques, which can learn and test using data from Remote Telemetry Unit (RTU) gas pipeline device streams, was proposed. It uses sparse and denotes self-encoding methods to create high-level data representation through unlabelled and noisy information for unchecked learning and deep neural networks for supervised learning. The findings exhibit great success in detecting malicious activities in the proposed model [40]. The IoT continues to expand rapidly and has an increasing manifestation in previously unknown domains. These domains may impose unique constraints which make the development and implementation of IoT systems difficult. Examples of such constraints include the absence of particular protocols, limiting information types that can be obtained, obligation on providing information to the public, and monitoring the communication process. The fast and effective implementation of these projects is vital to capture, reflect, design, and reuse these limitations. For use within an industrial environment, an IoT human was modelled in the loop monitoring system. Experiences with the design and development of both the first system implementation and software architecture variance points have been identified; they account for subsequent versions and implementation in other settings [41].

When the data is generated on huge scale it is termed big data. Big data is useful; it is considered the next technology on the market, which has many advantages for many applications. Many tools have therefore been created for analysing this data to profit from it because big data can hardly be analysed using conventional tools. Big data analytics have now been one of the latest research subjects in the last decade. Big data and its properties, forms, challenges, analytical tools, and applications have been addressed in business, security, health, education, and industry [42].

## 6. Existing Approaches of IoT and Big Data in Different Research Areas

General strategies against computer network security risks have been summarized and debated in small- and medium-sized firms. There is a broad discussion of emerging new threats beyond conventional threats and modern IoT applications beyond traditional SMEs. The research contribution is to alert IT experts to the potential threats that SMEs may be facing. The dynamics as a whole are derived from Google's big data. Specific interpretations and suggestions are also given to nontechnical business owners [43]. In the automotive sector in particular, the IIoT is a paradigm change. Due to improved operating effectiveness in manufacturing processes, intelligent object identity

mechanisms, smart automation capabilities, and clock monitoring capabilities, the idea is very attractive to most of the industrial sectors. It decreases the intervention of employees in dangerous industrial circles. Some of the best areas for training and working for the IIoT are factory floors, inventory processing, installation, manufacturing procedures, finalising products, and other logistic entrance and exit tasks. The IIoT phenomenon is based on the IoT technologies, which currently guarantee effective work performance in many areas, in industry and also in commercial and social fields. IIoT concepts and meanings are debated about market drivers behind the technological growth and the progression of this phenomenon. The basic tenets of technological implementation methods in various areas and associated frameworks were also addressed. Japanese case studies have been conducted for the industry, in which procedures similar to IIoT have already been applied. This included Tsuchiya Gousei, Toyota, Hitachi, and the Zenitaka Corporation [44]. Connectivity is a one-word definition of Industry 4.0 revolution. The IoT and IIoT have grown in importance as a result of the rise of industrialisation and Industry 4.0. The massive integrated devices of the IIoT have rendered cybersecurity and user privacy critical components as new opportunities raise new challenges. The importance of industrial network intrusion detection is particularly high. It is a key factor, for example, in improving the safe operation of smart grid systems while also protecting customer privacy. Similarly, for industrial networks, data streaming is a viable option for moving research from the cloud into the fog, as it benefits from fast intrusion detection as well as buying time for intrusion mitigation [45]. The Fourth Industrial Revolution aims to improve the efficiency, flexibility, and automation of internal processes that include value chains so that companies can plan and deliver new services based on data generated by various technologies. As a result, businesses are devoting considerable resources to figuring out how Industry 4.0 technology can be used to enhance current processes and provide a more attractive business model to both existing and new consumers. The proposal report presents the findings of a study that aimed to increase market awareness of Industry 4.0, identify key contributors to the advancement of IoT and big data implementation, and recommend additional research to help Collaborative Industry 4.0 Networks expand [46].

Given various factors in the healthcare system, such as privacy and confidentiality, maintaining the reliability and accuracy of health data is difficult. EHRs are commonly used because of their diverse clinical advantages to ensure that everybody has real-time health information. EHR refers to a system of historical electronic records that includes patient health information, including demographic information, health problems, drugs, health exam results, progress in recovery, and previous medical records. In order to provide prompt care, the EHR system permits the electronic exchange of information between interested parties. Even if EHRs have contributed significantly to health recording and storage, interoperability still remains an issue. The ability to share, communicate, and use health information through

organisations to improve the quality of healthcare delivery to individuals and to the public is known as interoperability in healthcare care. A lack of interoperability prevents successful healthcare data sharing. It not only impacts health providers for health related programmes, but also limits patients' contact and access to medical records. In the medical sector, IoT has been used widely for a long time. The majority of IoT health apps are designed to recognize and monitor people and objects, collect data from patients and staff, and use sensors for specific purposes (temperature, smoke, etc.). The IoT provides a patient-care ecosystem. Many medical devices now have sensors that capture continuous health indicators for patients, such as blood pressure, blood oxygen levels, heart rate, cholesterol level, and other information. The data is then sent over a wireless network to a central computer or a mobile device for analysis and classification. IoT helps medical professionals save time and money by allowing them to monitor patients in a continuous data flow rather than conduct repetitive data collection activities. Patient data can be accessed and tracked remotely at any time with wireless IoT solutions. Collecting a patient's full health profile as a guide to treatment decisions and appropriate medications is easy thanks to a network of sensors and healthcare wearables. Doctors and nurses can take care of all vital signs and use records to prevent misdiagnosis or medical misuse. In-time RFID tag monitoring systems or IrDA technology will keep patients as well as hospital staff up to date with the real-time location and conditions. The marks may be applied to medical equipment or a patient's bracelet, to determine where the tagged items are located. Under emergency conditions, the device may help locate a patient's exact position or warn caregivers when a patient leaves the hospital without permission [1].

Fog computing, cloud computing, semantic computing, edge computing, and other innovations have all evolved in parallel with the IoT. Sensor and imaging data produced by ubiquitous IoT applications, such as healthcare, must be properly processed. In a traditional IoT framework, cloud computing confirms an optimal solution for the successful managing of vast volumes of data and provides shared services and infrastructure. Most IoT applications are extremely time sensitive, and requirements are latency bound. When data is transmitted between the cloud and the application, a delay that is not reasonable is created. Various facets and trends have been revealed to resolve the challenges existing in the conventional IoT world, in the assimilation evolving computing systems, and in the disruption technologies like edge/fog computing, big data, and IoT blockchain. In addition, a number of IoT and cloud computer framework problems exist, including the fact that any single component of the IoT architecture can serve as a point of departure that can disrupt the entire network. Trends in the bracing and administration of massive data using Internet data centres are discovered on the IoT ecosystems. The evaluation is promoted by a case study of fog/edge computing and cloud computing on waste management systems. In addition, some developments in the application of the blockchain are also examined within

the IoT ecosystem. An overview of the fundamental aspects of different computer paradigms and approaches that can help solve big data problems by building IoT ecosystems is also presented [47].

The interaction between digitalized (data), intellectual property, privacy laws, and competition law is currently triggering politicians, businesspeople, academic sector, and even the general public in the IoT scenario. The groups are concerned for a variety of reasons: businesspeople, for example, will have the opportunity to create resources; researchers will be able to easily compile, analyse, and distribute information; and everyone agrees that the processing and sharing of personal data will raise concerns regarding privacy and data protection. It is difficult to understand the interface between legal systems caused by data processing, delivery, and use, trying to dissect this interface with details, such as "the data," from its sources, to clients and ultimately to consumers in IoT environment. Data sources are diverse, but they can be grouped into three categories: First, government ("open data") collects data from public sector bodies. Second, data can be voluntarily provided by consumers, clients, or businesses through e-platforms or other IT-based service formats. Third, it may be collected by means of cookies, patient data, and ISP data. In this step an attempt was made to identify what legal frameworks apply. When computers collect data, disburse it in "the cloud," and eventually reuse it, what intellectual property law scheme can apply? Who "owns" personally identifiable information? Do data privacy laws create new rules on personal data? The research proposed when producers and users of raw or processed data should benefit from the applications of competition law, whether public or private authorities should assist with this, and whether competition law facilitates established needs. The research focused mainly on applying competition law to bodies which collect or maintain data. The question of the sector-specific regulation is posed in the data arena. Data access is a controversial topic not only under the general "competition" regulations but as regards sector-specific regulations including the directive on public sector information, e-Call, financial services, and e-platforms. Indeed, there appear to be rules on access to information (antiregulation) currently being included in industry-specific regulations which require either data exchange or open access to the data collection device. It was concluded that general competition law does not apply readily to the access of generic data, except where the set of data is indispensable for the access to business or to a market in particular, whereas sector-specific regulations tend to arise as a data processing tool owned by competitors or enterprises in general. However, at its present stage of growth, the key question under competition law in the data industry is to establish level playing field by attempting to promote the adoption of the IoT [48].

In order to collect and share data, millions of devices equipped with sensors are linked together. The IoT is defined as the phenomenon of the everyday objects which

are interconnected through an integrated system. These sensors produce a massive amount of data from a large range of equipment or items at the same time and continuously, also called large data. When time, energy, and processing capacities are constrained, handling this enormous volume of data and different variants imposes significant challenges; big data analytics are therefore increasingly difficult for the data obtained by IoT. Data processing, data interpretation, unstructured data analytics, data visualisation, interoperability, data semantics, scalability, data fusion, data integration, data quality, and knowledge discovery were all revealed and solved as part of the IoT big data project [49]. Big data (BD) and the IoT are regarded as creating large enterprises and impacting on the labour market through and beyond 2050. However, the success of BD and IoT applications still slows down, especially in view of a lack of standardisation and the challenges of the various players to cooperation. On the other hand, access to vast cloud test beds has become an asset, and enormous potential in areas such as agriculture, automotive, green energy, health, and smart cities can be optimized. The objective is to present ideas and steps for business growth that a young and intelligent entrepreneur can take to begin and speed up his successful business career, which will increase beyond 2050 [50].

Emerging technologies have developed a new wave of industrial reform in recent years. The latest industrial revolution is profoundly integrated into modern industrial and manufacturing industries to support transformation and enhancement by the new information technology generation. Smart equipment plays an important role in the reform since it is the backbone of the manufacturing industry. An innovative method for designing smart equipment has been suggested. First of all, a method's architecture and various layers for processing the data were suggested, with reference to the Cognitive Internet of Things and architecture of industrial big data. An algorithm was then used to evaluate and decide the acquired external data, together with CIoT and industry big data technology. Finally, this approach was validated by case studies as being accurate and feasible. The findings revealed that the method could significantly reduce the depth of understanding about smart equipment while also providing more valuable information about design assistance for significant impact on firm equipment design [51]. The IoT connects computers, individuals, locations, and even abstract objects like events. Smart sensors, heavy embedded microelectronics, high-speed networking, and Internet standards are on the

verge of changing today's value chains. Big data is both a product and a driver of the IoT system, with high speed, high volume, and variation of modes. Market data development poses whole new risks and opportunities. IoT requires robust modelling tools to address the technological risks associated with the interaction of "anything." Furthermore, the processing and storage of unstructured, structured, repetitive, and nonrepetitive flow of data in real time require the development of new IT systems and architectures. Only powerful analytical tools can derive "significance" from the increasing amount of data and, as a result, data science has now become a strategic advantage. The existence of IoT is largely based on technology standards that ensure the interoperability of everything. Some basic standardization exercises are outlined and methods for analytics such as large data processing approaches for real-time processing are presented. IoT is therefore a (fast) evolutionary mechanism that depends heavily on the close collaboration between standardisation organisations, open-source communities, and information technology experts to penetrate all dimensions of life [52]. Table 1 depicts the approaches of IoT and big data in different domains of research.

Various most popular libraries were searched in order to find relevant materials in the area under study. These libraries include ScienceDirect, IEEE, ACM, PubMed, and Springer. The initial search results were identified in the various libraries. Figure 1 shows the details of the libraries searched.

## 7. Analysing Literature in the Area of Research

This section discusses the overview of the analysis done from various perspectives which are visually described. Popular libraries were used for the search process of the proposed study. Figure 1 describes the details of search process in the libraries.

The initial number of papers obtained was filtered in order to further reduce the materials and to find the exactly matching papers. Figure 2 describes the final number of included papers. The figure shows high number of publications in the ScienceDirect library.

After this, each library was presented separately for its representation. The representation SD is given in Figure 3.

The representations of the ACM library with their details are given in Figure 4.

The search details of IEEE library are shown in Figure 5.

After this, the Springer library was searched, and the results are presented in Figure 6.

Lastly, the PubMed library was searched, and the details are depicted in Figure 7.

TABLE 1: Approaches of big data and IoT in different domains.

No.	Method	Year	Citation
1	Analytics big data and IoT in healthcare	2019	[1]
2	Clustering of data streams	2018	[2]
3	Industrial surveillance through fog-enabled IoT and big data	2018	[6]
4	SDN-based IIoT with edge computing	2018	[7]
5	Performance assessment of companies under IIoT architectures	2018	[11]
6	Data compression and prediction for IIoT	2018	[12]
7	Cloud of things based on linked data	2018	[13]
8	Depth estimation approach of light field imaging	2018	[14]
9	Location privacy protection for big data in IIoT	2018	[15]
10	Secure high-order CFS algorithm on clouds for IIoT	2018	[16]
11	Social big-data-based content dissemination in Internet of Vehicles	2018	[17]
12	FSO2 life extension program through IIoT and big data	2019	[18]
13	Outlier detection method towards big sensor data in IoT	2019	[19]
14	Data acquisition and storage system for designing industrial data	2019	[20]
15	Cybersecurity in an environment of IIoT	2019	[21]
16	Secure fabric blockchain-based data transmission for IIoT	2019	[22]
17	Concept drift detection and adaption in big IIoT data	2019	[23]
18	City geospatial dashboard	2019	[24]
19	IoT service platform for locally disseminating socially valuable data	2019	[25]
20	Spatiotemporal modelling and integration of manufacturing big data	2020	[26]
21	Data transmission scheme in industrial IoT	2020	[27]
22	Big data-enabled framework for energy efficient software-defined data centres in IoT	2020	[28]
23	Federated tensor mining for secure IIoT	2020	[29]
24	BIM big data storage in WebVRGIS	2020	[30]
25	Clustering big data in cognitive IIoT	2020	[31]
26	Big data cleaning based on mobile edge computing	2020	[32]
27	ADTT	2020	[32]
28	Big data driven edge-cloud collaboration architecture	2020	[33]
29	Dynamic data mining of sensor data	2020	[34]
30	Open ecosystem for future IIoT	2020	[35]
31	Deterministic virtualization in the IIoT	2019	[36]
32	IIoT based ransomware detection	2019	[37]
33	HCI and CSCW within industry settings	2018	[38]
34	Microservice architecture for the IIoT	2018	[39]
35	Intrusion detection for edge system in brownfield IIoT	2019	[40]
36	IIoT monitoring application	2017	[41]
37	Big data analytics tools and applications	2019	[42]
38	Precautions against security threats for computer networks in SMEs	2017	[43]
39	IIoT, principles, processes, and protocols	2019	[44]
40	Industrial networks and IIoT	2020	[45]
41	Awareness of Industry 4.0	2018	[46]
42	IoT and big data analytics in healthcare	2019	[1]
43	Big data & disruptive computing platforms braced IoT	2020	[47]
44	Big data, privacy regulations, open data, intellectual property, and competition law in an IoT	2018	[48]
45	Big data challenges for the IoT	2017	[49]
46	Big data and IoT	2016	[50]
47	Equipment design assisted by cognitive IoT and industrial big data	2020	[51]
48	IoT and big data	2018	[52]
49	Accurate vehicle state estimation	2019	[53]

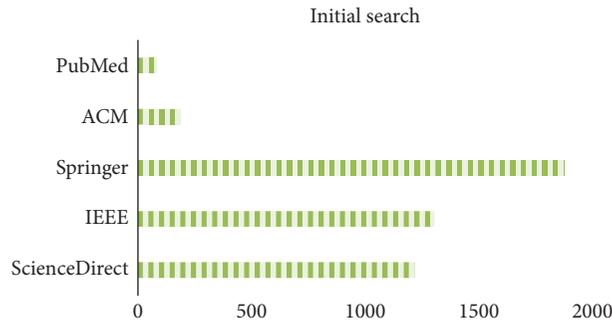


FIGURE 1: Details of search process in the libraries.

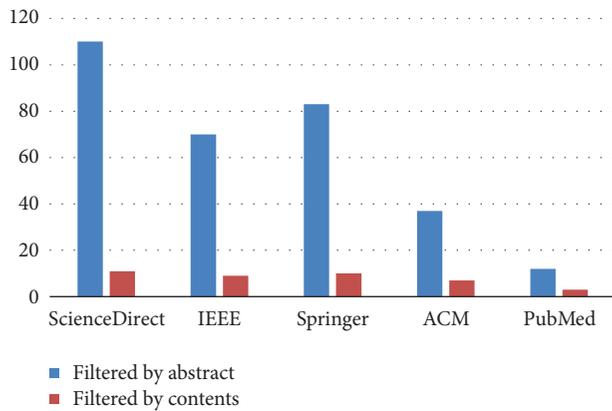


FIGURE 2: Final papers in the given libraries.

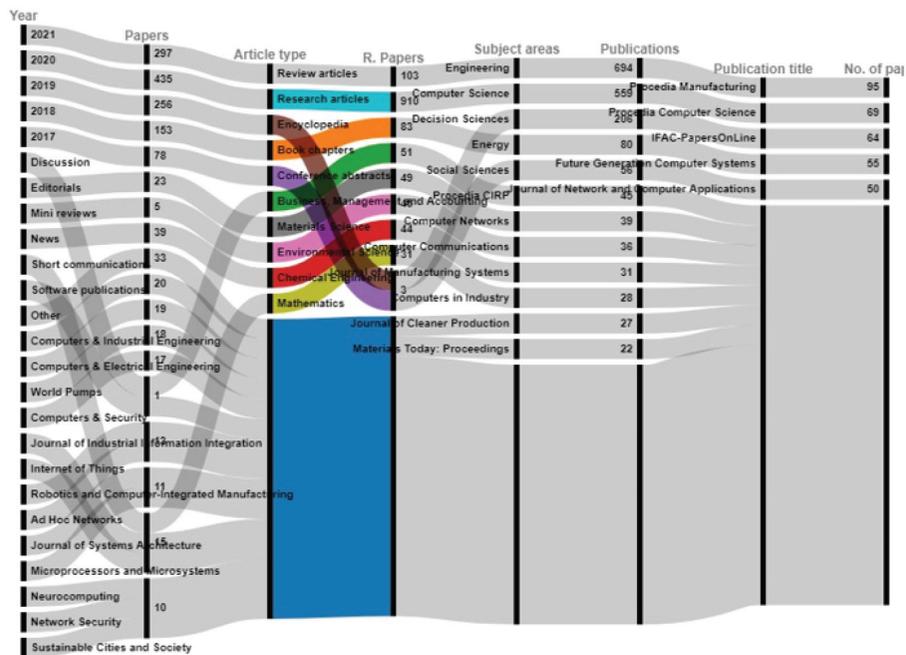


FIGURE 3: Representation of ScienceDirect library.



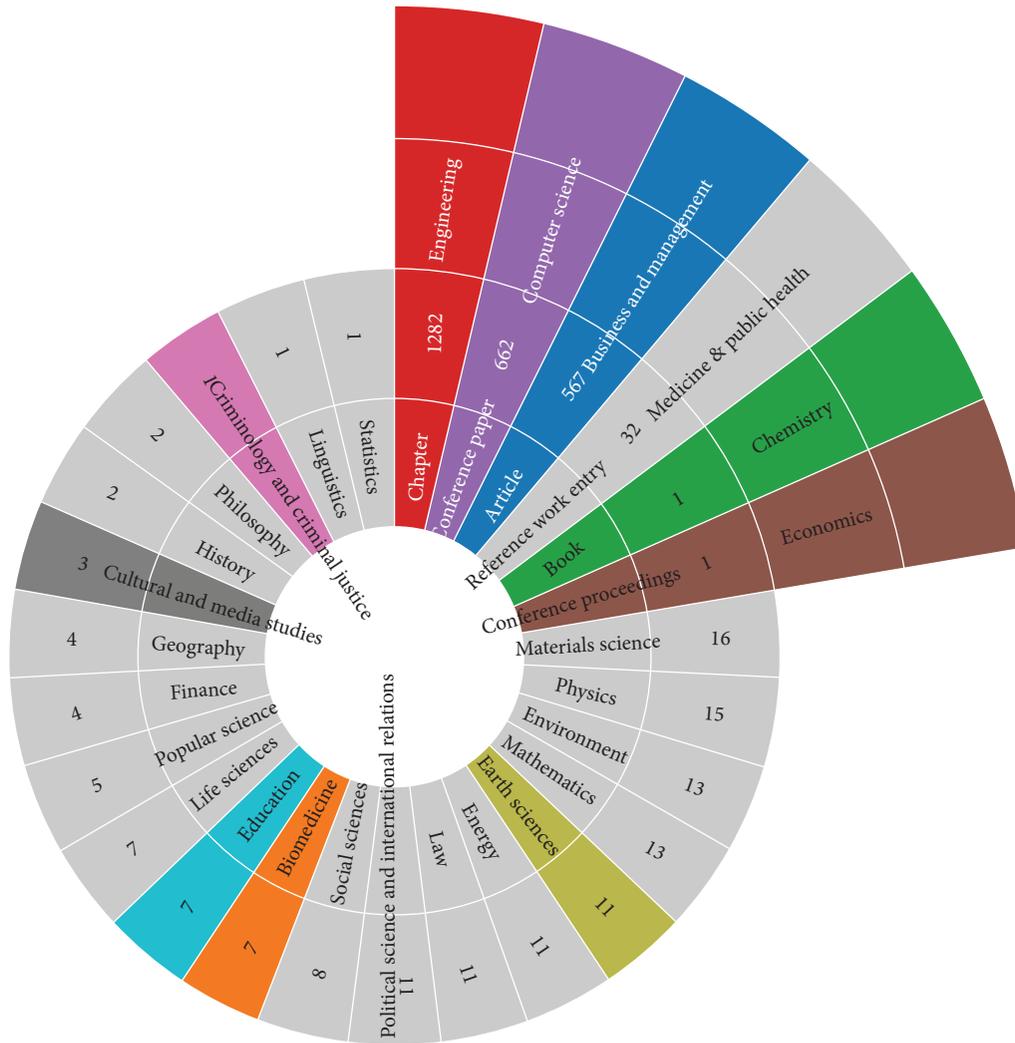


FIGURE 6: Details of Springer library.

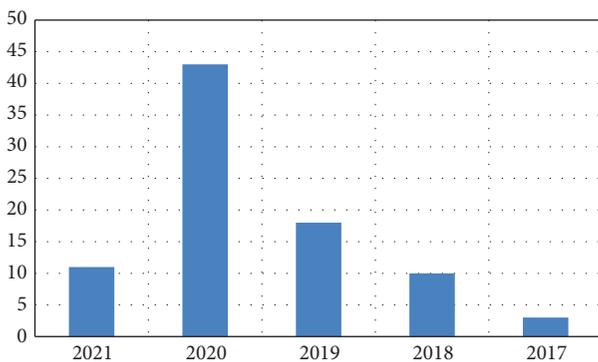


FIGURE 7: Details of PubMed library.

## 8. Conclusion

The fast and growing development of smart devices such as actuators, sensors, and wearable devices has made the IoT enable for smart and sustainable developments in the area. The IoT is one of the emergent network and information technologies comprehending automatic

operations in the network of devices connected to the IoT. The use of IoT in effective and efficient way in the area has increased effectiveness and decreased errors. Physical objects are linked with these smart devices for analysing, processing, and managing the data produced from the surroundings. Such data can then be further used for different purposes such as smart decision making, early analysis, and many other purposes. The IoT network is connected with big data through Internet for manipulating and storing huge bulk of data on cloud storage. Managing huge bulk of data in real time is a very crucial task. Various approaches are in practice for the management and recovery problems of data. Extracting correlation data became difficult specifically as the coupling degree between diverse perceptual attributes is low. The existing literature has provided comprehensive techniques, tools, and methods for various purposes of data. The proposed study has reported a wide-ranging overview on big data and its V's with IoT to describe the state-of-the-art research into the field with in-depth review of existing literature. Various popular libraries were

searched for analysing the existing literature, and comprehensive report is presented.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding this study.

## References

- [1] J. Chanchaichujit, A. Tan, F. Meng, and S. Eaimkhong, "Internet of things (IoT) and big data analytics in healthcare," in *Healthcare 4.0*, Palgrave Pivot, Singapore, 2019.
- [2] J. Diaz-Rozo, C. Bielza, and P. Larranaga, "Clustering of data streams with dynamic Gaussian mixture models: an IoT application in industrial processes," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3533–3547, 2018.
- [3] Z. Xiao, S. Xu, D. Wang, and R. Zhang, "On extracting regular travel behavior of private cars based on trajectory data analysis," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 14537–14549, 2020.
- [4] Y. Huang, Z. Xiao, D. Wang, H. Jiang, and D. Wu, "Exploring individual travel patterns across private car trajectory data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 12, pp. 5036–5050, 2020.
- [5] Y. Huang, Z. Xiao, X. Yu, D. Wang, V. Havyarimana, and J. Bai, "Road network construction with complex intersections based on sparsely-sampled private car trajectory data," *ACM Transactions on Knowledge Discovery from Data*, vol. 13, no. 35, pp. 1–28, 2019.
- [6] S. Din, A. Paul, A. Ahmad, B. B. Gupta, and S. Rho, "Service orchestration of optimizing continuous features in industrial surveillance using big data based fog-enabled internet of things," *IEEE Access*, vol. 6, pp. 21582–21591, 2018.
- [7] X. Li, D. Li, J. Wan, C. Liu, and M. Imran, "Adaptive transmission optimization in SDN-based industrial internet of things with edge computing," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1351–1360, 2018.
- [8] S. Nazir, M. Nawaz, S. Anwar et al., "Big data visualization in cardiology—a systematic review and future directions," *IEEE Access*, vol. 7, no. 1, pp. 115945–115958, 2019.
- [9] S. Nazir, M. Nawaz, A. Adnan, S. Shahzad, and S. Asadi, "Big data features, applications, and analytics in cardiology—a systematic literature review," *IEEE Access*, vol. 7, no. 1, pp. 143742–143771, 2019.
- [10] S. Nazir, S. Khan, S. Ali et al., "A comprehensive analysis of healthcare big data management, analytics and scientific programming," *IEEE Access*, vol. 8, pp. 95714–95733, 2020.
- [11] C. Mbohwa and A. K. Sahu, "Performance assessment of companies under IIoT architectures: application of grey relational analysis technique," in *Proceedings of the 2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 1350–1354, Coimbatore, India, July 2018.
- [12] J. Park, H. Park, and Y. Choi, "Data compression and prediction using machine learning for industrial IoT," in *Proceedings of the 2018 International Conference on Information Networking (ICOIN)*, pp. 818–820, Chiang Mai, Thailand, January 2018.
- [13] Y. Son and K. Lee, "Cloud of things based on linked data," in *Proceedings of the 2018 International Conference on Information Networking (ICOIN)*, pp. 447–449, Chiang Mai, Thailand, January 2018.
- [14] Y. Wu, "Research on depth estimation method of light field imaging based on big data in internet of things from camera array," *IEEE Access*, vol. 6, pp. 52308–52320, 2018.
- [15] C. Yin, J. Xi, R. Sun, and J. Wang, "Location privacy protection based on differential privacy strategy for big data in industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3628–3636, 2018.
- [16] Y. Zhao, L. T. Yang, and J. Sun, "A secure high-order CFS algorithm on clouds for industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3766–3774, 2018.
- [17] Z. Zhou, C. Gao, C. Xu, Y. Zhang, S. Mumtaz, and J. Rodriguez, "Social big-data-based content dissemination in internet of vehicles," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 768–777, 2018.
- [18] A. Bamrungwong, "Implementation of the FSO2 life extension program by using big data and IIoT," in *Proceedings of the 2019 Petroleum and Chemical Industry Conference Europe (PCIC EUROPE)*, pp. 1–8, Paris, France, May 2019.
- [19] X. Deng, P. Jiang, X. Peng, and C. Mi, "An intelligent outlier detection method with one class support tucker machine and genetic algorithm toward big sensor data in internet of things," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 6, pp. 4672–4683, 2019.
- [20] D. Geng, C. Zhang, C. Xia, X. Xia, Q. Liu, and X. Fu, "Big data-based improved data acquisition and storage system for designing industrial data platform," *IEEE Access*, vol. 7, pp. 44574–44582, 2019.
- [21] F. A. B. Juárez, "Cybersecurity in an industrial internet of things environment (IIoT) challenges for standards systems and evaluation models," in *Proceedings of the 2019 8th International Conference on Software Process Improvement (CIMPS)*, pp. 1–6, Leon, Mexico, October 2019.
- [22] W. Liang, M. Tang, J. Long, X. Peng, J. Xu, and K. Li, "A secure FaBric blockchain-based data transmission technique for industrial internet-of-things," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3582–3592, 2019.
- [23] C. Lin, D. Deng, C. Kuo, and L. Chen, "Concept drift detection and adaption in big imbalance industrial IoT data using an ensemble learning method of offline classifiers," *IEEE Access*, vol. 7, pp. 56198–56207, 2019.
- [24] K. K. Lwin, Y. Sekimoto, W. Takeuchi, and K. Zetsu, "City geospatial dashboard: IoT and big data analytics for geospatial solutions provider in disaster management," in *Proceedings of the 2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, pp. 1–4, Paris, France, December 2019.
- [25] Y. Shoji, K. Nakauchi, W. Liu, Y. Watanabe, K. Maruyama, and K. Okamoto, "A community-based IoT service platform to locally disseminate socially-valuable data: best effort local data sharing network with no conscious effort?" in *Proceedings of the 2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, pp. 724–728, Limerick, Ireland, April 2019.
- [26] W. Fang, Y. Guo, W. Liao, S. Huang, C. Yang, and K. Cui, "The spatio-temporal modeling and integration of manufacturing big data in job shop: an ontology-based approach," in *Proceedings of the 2020 IEEE 7th International Conference on Industrial Engineering And Applications (ICIEA)*, pp. 394–398, Bangkok, Thailand, April 2020.

- [27] H. Hui, C. Zhou, S. Xu, and F. Lin, "A novel secure data transmission scheme in industrial internet of things," *China Communications*, vol. 17, no. 1, pp. 73–88, 2020.
- [28] K. Kaur, S. Garg, G. Kaddoum, E. Bou-Harb, and K. R. Choo, "A big data-enabled consolidated framework for energy efficient software defined data centers in IoT setups," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2687–2697, 2020.
- [29] L. Kong, X. Liu, H. Sheng, P. Zeng, and G. Chen, "Federated tensor mining for secure industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 2144–2153, 2020.
- [30] Z. Lv, X. Li, H. Lv, and W. Xiu, "BIM big data storage in WebVRGIS," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2566–2573, 2020.
- [31] A. K. Tripathi, K. Sharma, M. Bala, A. Kumar, V. G. Menon, and A. K. Bashir, "A parallel military dog based algorithm for clustering big data in cognitive industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, 2021.
- [32] X. Wang, L. T. Yang, Y. Wang, L. Ren, and M. J. Deen, "ADTT: A highly efficient distributed tensor-train decomposition method for IIoT big data," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 1573–1582, 2020.
- [33] C. Yang, S. Lan, L. Wang, W. Shen, and G. G. Q. Huang, "Big data driven edge-cloud collaboration architecture for cloud manufacturing: a software defined perspective," *IEEE Access*, vol. 8, pp. 45938–45950, 2020.
- [34] Y. Yin, L. Long, and X. Deng, "Dynamic data mining of sensor data," *IEEE Access*, vol. 8, pp. 41637–41648, 2020.
- [35] P. Zhang, Y. Wu, and H. Zhu, "Open ecosystem for future industrial Internet of things (IIoT): architecture and application," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 1, pp. 1–11, 2020.
- [36] R. Jan and S. Wilfried, "The need for deterministic virtualization in the industrial internet of things," in *Proceedings of the Workshop on Fog Computing and the IoT*, Montreal, Canada, April 2019.
- [37] A.-H. Muna and S. Elena, "Industrial internet of things based ransomware detection using stacked variational neural network," in *Proceedings of the 3rd International Conference on Big Data and Internet of Things*, Melbourne, Australia, August 2019.
- [38] M. Henrik, C. Röcker, T. Ludwig et al., "The industrial internet of things: new perspectives on HCI and CSCW within industry settings," in *Proceedings of the Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, Jersey City, NJ, USA, November 2018.
- [39] D. Jurgen, I. Johannes, K. Michael, and K. Christian, "A microservice architecture for the industrial internet-of-things," in *Proceedings of the 23rd European Conference on Pattern Languages of Programs*, Irsee, Germany, July 2018.
- [40] A.-H. Muna, S. Elena, and H. Frank den, "An efficient intrusion detection model for edge system in brownfield industrial internet of things," in *Proceedings of the 3rd International Conference on Big Data and Internet of Things*, Melbourne, Australia, August 2019.
- [41] T. M. Ethan, K. Matthew, D. M. John, and M. Laine, "Designing for reuse in an industrial internet of things monitoring application," in *Proceedings of the 2nd Workshop on Social, Human, and Economic Aspects of Software*, Salvador, Brazil, May 2017.
- [42] Q. S. Mohammed, Y. Muneer Bani, A. Qusai, and N. Lujain, "Big data analytics tools and applications: survey," in *Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems*, Dubai, UAE, December 2019.
- [43] W. Yaoyang, F. Simon, and Z. Yan, "General precautions against security threats for computer networks in SMEs: from the perspective of big data and IOT," in *Proceedings of the International Conference on Big Data and Internet of Thing*, London, UK, December 2017.
- [44] S. Madakam and T. Uchiya, "Industrial internet of things (IIoT): principles, processes and protocols," in *The Internet of Things in the Industrial Sector. Computer Communications and Networks*, Z. Mahmood, Ed., Springer, Cham, Switzerland, 2019.
- [45] A. Sari, A. Lekidis, and I. Butun, "Industrial networks and IIoT: now and future trends," in *Industrial IoT*, I. Butun, Ed., Springer, Cham, Switzerland, 2020.
- [46] M. Flores, D. Maklin, M. Golob, A. A. Ashaab, and C. Tucci, "Awareness towards industry 4.0: key enablers and applications for internet of things and big data," in *Collaborative Networks of Cognitive Systems. PRO-VE 2018. IFIP Advances in Information and Communication Technology*, L. Camarinha-Matos, H. Afsarmanesh, and Y. Rezgui, Eds., Springer, Cham, Switzerland, 2018.
- [47] N. Jayakumar and D. P. Joshi, "Big data & disruptive computing platforms braced internet of things: facets & trends," in *Internet of Things, Smart Computing and Technology: A Roadmap Ahead. Studies in Systems, Decision and Control*, N. Dey, P. Mahalle, P. Shafi, V. Kimabahune, and A. Hassanien, Eds., Vol. 266, Springer, Cham, Switzerland, 2020.
- [48] L. Bjorn, "Big data, open data, privacy regulations, intellectual property and competition law in an internet-of-things world: the issue of accessing data," in *Personal Data in Competition, Consumer Protection and Intellectual Property Law. MPI Studies on Intellectual Property and Competition Law*, M. Bakhroum, B. Conde Gallego, MO. Mackenrodt, and G. Surblytė-Namavičienė, Eds., Vol. 28, Springer, Berlin, Germany, 2018.
- [49] W. Pornpit, K. Jaswinder, P. Vidyasagar, and D. Abhishek, "Big data challenges for the internet of things (IoT) paradigm," in *Connected Environments for the Internet of Things. Computer Communications and Networks*, Z. Mahmood, Ed., Springer, Cham, Switzerland, 2017.
- [50] G. Suci, "Big data and internet of things—challenges and opportunities for accelerated business development beyond 2050," in *Wireless World in 2050 and beyond: A Window into the Future!. Springer Series in Wireless Technology*, R. Prasad and S. Dixit, Eds., Springer, Cham, Switzerland, 2016.
- [51] J. Wan, J. Li, Q. Hua, A. Celesti, and Z. Wang, "Intelligent equipment design assisted by Cognitive Internet of Things and industrial big data," *Neural Computing & Applications*, vol. 32, no. 9, pp. 4463–4472, 2020.
- [52] N. Jesse, "Internet of things and big data: the disruption of the value chain and the rise of new software ecosystems," *Ai and Society*, vol. 33, no. 2, pp. 229–239, 2018.
- [53] Z. Xiao, D. Xiao, H. Jiang et al., "Toward accurate vehicle state estimation under non-Gaussian noises," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10652–10664, 2019.