DOI 10.12694/scpe.v20i4.1562
Scalable Computing: Practice and Experience ISSN 1895-1767
Volume 20, Number 4, pp. 607–629. http://www.scpe.org

DATA ANALYTICS IN THE INTERNET OF THINGS: A SURVEY

TAUSIFA JAN SALEEM* AND MOHAMMAD AHSAN CHISHTI†

Abstract. The plethora of sensors deployed in Internet of Things (IoT) environments generate unprecedented volumes of data, thereby creating a data deluge. Data collected from these sensors can be used to comprehend, examine and control intricate environments around us, facilitating greater intelligence, smarter decision-making, and better performance. The key challenge here is how to mine out proficient information from such immense data. Copious solutions have been put forth to obtain valuable inferences and insights, however, these solutions are still in their developing stages. Moreover, conventional procedures do not address the surging analytical demands of IoT systems. Motivated to resolve this concern, this work investigates the key enablers for performing desired data analytics in IoT applications. A comprehensive survey on the identified key enablers including their role in IoT data analytics, use-cases in which they have been applied and the performance results of the use-cases is presented. Furthermore, open research challenges and future research opportunities are also discussed. This article can be used as a basis to foster advanced research in the arena of IoT data analytics.

Key words: Internet of things, Big Data, Data Analytics, Data Mining, Machine Learning, Time Series Forecasting.

AMS subject classifications. 68M11, 97R40

1. Introduction. IoT is regarded as the vital research paradigm in the current epoch. It has dramatically revolutionized every facet of our lives. This technology is characterized by enormous amounts of smart devices that cooperate seamlessly with each other by means of a global network infrastructure, thereby facilitating a wide number of pervasive and ubiquitous applications spanning diverse fields [1]. It is an ecosystem of smart devices, i.e., devices that possess sensing and processing efficacies and can comprehend and respond to their surroundings via sensors and actuators. The blend of diverse technological advancements like near field communication, Radio Frequency Identification (RFID), real-time localization, embedded systems, and the networking expedite the conversion of day to day entities into smart entities [2]. These entities are incorporated impeccably into a web like framework so that they can communicate with one another and with other cyber agents so as to accomplish goal-oriented tasks [3]. IoT enables sensors and objects to interact coherently within smart environs and facilitates information transfer in a suitable manner. The continuum of devices in IoT are connected via several diverse access networks and communication solutions equipped with technologies such as RFID, Wireless Sensor Network (WSN), Bluetooth, Wi-Fi, ZigBee, GSM etc. [4]. Over 30 billion [5] devices ranging from smart phones, to vehicles are prophesied to be linked to the Internet by 2020. The large number of sensors deployed in IoT environments continuously generate unprecedented volumes of structured, unstructured and semi-structured data (Big Data), that cant be handled by conventional processing, storage and analytical systems [6]. Data generated from IoT is different from traditional data in following ways [7]: (i) data is generated continuously at high speed, (ii) Apart from structured data, data may be of semi-structured or unstructured nature as well, (iii) data sources are diverse and fully distributed, and (iv) integration of multi-modal data becomes complex.

The fundamental objective of IoT is to augment the standard of living. Nevertheless, this vision is based on being able to efficiently process, analyze, and comprehend the data generated by IoT devices [8]. Hence, analyzing IoT data in order to divulge trends, concealed patterns, hidden correlations, inferences, and actionable insights is crucial for dispensing elite services to IoT users. In this regard, investigating the technological advancements that can assist in analyzing unstructured and semi-structured data apart from structured data, integrating the data from heterogeneous data sources, performing real-time analytics in delay critical applications as well as in optimizing the process of data analytics becomes indispensable.

1.1. Related Surveys. To the best of our knowledge, this work is the first of this kind that investigates the key enablers for IoT data analytics and surveys their role in the IoT use-cases. There are few research works in the literature that have surveyed data analytics in IoT applications, however, the focus of those research works is different from the prime concern of this article. The main focus of this article is on the key enablers for

*National Institute of Technology Srinagar, India (tausifasaleem19@gmail.com)
†National Institute of Technology Srinagar, India

IoT data analytics, the use-cases in which they have been applied, their purpose in those use-cases, the datasets that have been used, and the performance results of the use-cases.

In [9], Siow et al. have reviewed the applications of IoT data analytics across different domains. Moreover, the article proposes taxonomy for IoT data analytics in order to guide future research in the field. The focus of the paper is the enabling infrastructure for IoT data analytics that includes data generation, data collection, data aggregation and integration, storage technologies etc. However, the article lacks the detailed survey on the enablers like data mining techniques, machine learning techniques, dimensionality reduction techniques, time series forecasting, etc.

The work in [10] by Ahmed et al. investigated the latest developments in big data analytics for IoT applications. Moreover, the opportunities created from the amalgam of big data analytics and IoT were also identified. Several research challenges in IoT data analytics were also discussed. However, the article lacks the comprehensive survey of the IoT use cases and the analytics techniques used.

1.2. Contribution and Structure of Paper. This work provides a delineation of the present state-of-the-art in the realm of data analytics in IoT. More precisely, it:

- investigates the key enablers for IoT data analytics
- surveys the role of identified key enablers in IoT data analytics
- identifies the challenges that the research community still has to face in this arena.

The rest of the paper is structured as follows. In Section 2 an introduction of IoT data analytics is provided. Section 3 presents the purpose of data analytics in IoT applications. Section 4 discusses the key enablers for IoT data analytics in detail. Furthermore, the role of the key enablers in IoT data analytics is surveyed in this section. Section 5 highlights the challenges faced by IoT data analytics. Moreover, future research opportunities in the arena of IoT data analytics are also provided in this section. Section 6 presents the concluding remarks.

2. IoT Data Analytics. With the brisk advancements in sensing, communicating, analytic and actuating technologies, the vision of intelligent recognition, real-time observation, monitoring, and management is turning into reality [11,12]. The surfeit of sensors deployed in IoT environments generate masses of structured, unstructured and semi-structured data, including health care data, logistic data, astronomical data, environmental data, etc. [13]. Data collected from these sensors can be utilized to comprehend, examine and control intricate environments around us, facilitating greater intelligence, smarter decision making, and better performance. The enormous amounts of heterogeneous and diverse data generated by millions of IoT devices (monitoring certain phenomenon) make traditional information processing solutions obsolete. This is because traditional information processing systems cannot handle such volume of data [14]. IoT data analytics refers to the analysis of every fragment of data generated from IoT devices at right time in order to extract intelligent insights [15]. It is receiving a wide range of attention from researchers and practitioners, as extracting intelligent insights from IoT data is a tricky task and demands a great deal of attention. The question of how to collect, aggregate, and analyze the data generated from IoT environs has become an important impediment that requires urgent solution [16, 17]. Data Mining and Machine learning may help in creating smarter IoT by extracting unseen patterns, hidden correlations, trends, inferences, and actionable insights, facilitating greater intelligence, smarter decision making, enhancing performance, automation, productivity, and accuracy. However, the unprecedented rise in the magnitude and intricacy of data pose novel challenges to these domains [18, 19]. Moreover, it is crucial to formulate appropriate techniques for dealing with noisy, inaccurate, uncertain and real-time data. Furthermore, in several settings, it is indispensable to merge historical data with the current sensor data so as to draw out effective insights [20].

Most of the IoT devices continuously emanate massive volumes of time series data and such data are ephemeral in nature, thereby demanding real-time action. Consequently, apart from Big IoT data analytics, IoT demands one more category of data analytics, i.e., real-time analytics, to support time stringent applications. Examples include self-driving cars, elder posture recognition, surveillance systems etc. Hence, the aforementioned applications demand fast data analytics with minimal delay. In such applications, transferring data to cloud for analysis is not feasible. The finest remedy for such time stringent IoT applications is to bring analytics closer to IoT data source in order to remove needless delays. However, bringing analytics closer to IoT data source puts forth a new set of challenges, including limitation of power, storage and computing resources [21].

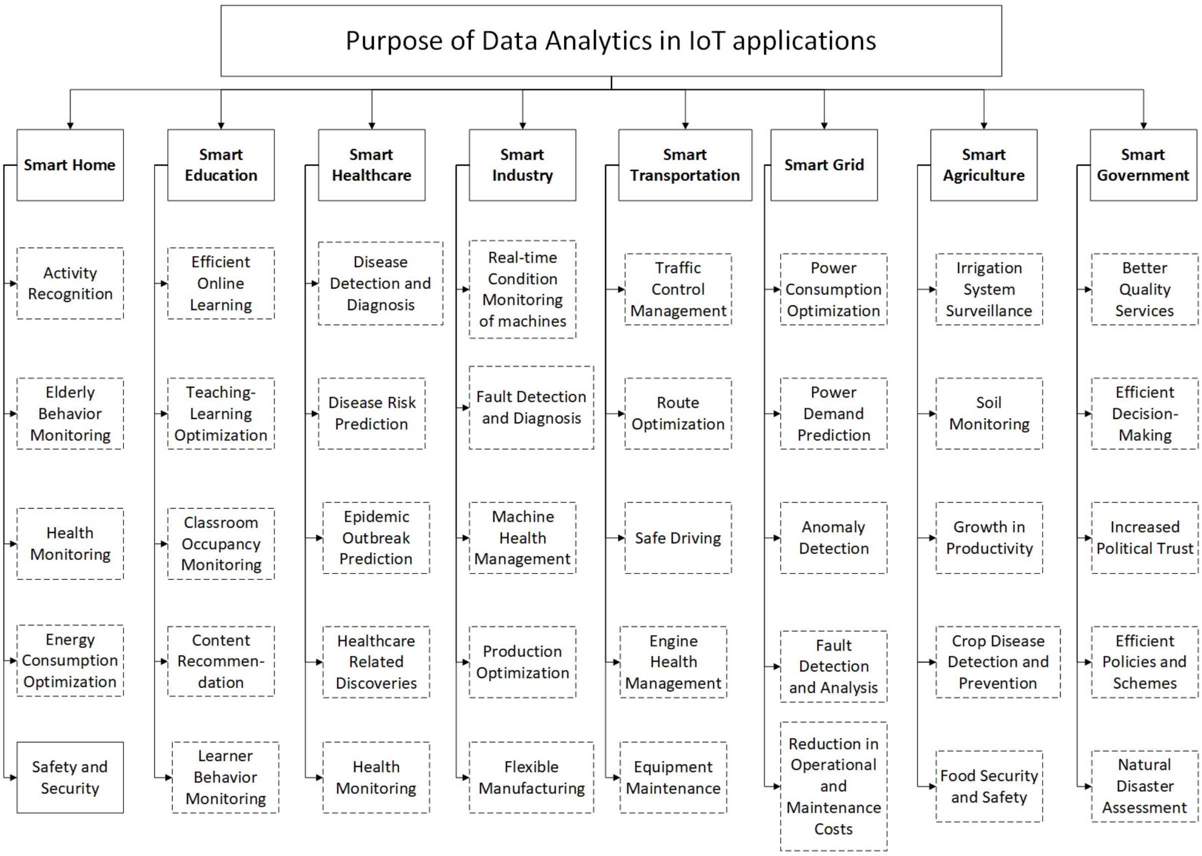

FIG. 3.1. *Purpose of Data Analytics in IoT applications*

IoT data analytics can be categorized into three groups, descriptive analytics, predictive analytics and prescriptive analytics [10]. Descriptive analytics delineates what has occurred and what is going on. It assists in perceiving novel business challenges and opportunities by utilizing data aggregation and data mining techniques. Descriptive analytics use-cases include energy consumption [22], urban designing [23], etc. Predictive analytics describes what will happen and why. It envisages future conditions and states precisely with the aid of statistical models and prediction techniques. Predictive analytics use-cases include disease prediction [24], predicting energy usage [25], machine failure prediction [26], anomaly prediction [27], etc. Prescriptive analytics characterizes what to do and why it needs to be done. It employs decision support systems to explore diverse possibilities and provides recommendations for decision-making using optimization and simulation algorithms. Prescriptive analytics use-cases include failure risk management in industrial IoT [28], clinical process design and optimization in healthcare [29], etc.

3. Purpose of Data Analytics in IoT Applications. IoT has brought colossal value to our lives by facilitating the growth of a myriad of business-specific and user-oriented applications in different sectors. These applications have triumphed in providing massive benefits to the users. Data analytics has a remarkable part in the development and success of IoT applications. It is used to extract meaningful inferences from IoT data and these inferences are generally in the form of intelligent control decisions, patterns, and statistics that assist IoT applications in powerful decision-making. Hence, utilization of data analytics in IoT applications bring immense benefits including better services, improved productivity, automation, and smarter decision-making. Fig 3.1 presents the purpose of data analytics in IoT applications. Following presents a brief discussion on the IoT applications and the role played by data analytics in these applications.

3.1. Smart Home. A smart home is an important development of IoT in which the dwellings are embodied with intelligence to provide smart services like user comfort, healthcare, security, remote monitoring and control of devices, energy conservation, etc. [30]. Smart homes provide a better standard of living by incorporating automation in the device access, control, and monitoring. The purpose of data analytics in smart home is to render intelligence in order to produce an interactive environment by utilizing foundational services like physiological and psychological state detection, image recognition, voice recognition etc. Analytics of smart home data help in tracking daily activities of the inhabitants, monitoring elderly behavior, optimization of energy consumption, ensuring security, health monitoring, etc.

3.2. Smart Healthcare. Increase in the number of long-term illness cases and regularly aging population is putting a consequential burden on nowadays healthcare organizations. Consequently, there is a dire need to alleviate the stress on healthcare organizations whilst continuing to dispense exorbitant healthcare services to patients. IoT has been recognized as a prospective panacea to reduce the stress on healthcare organizations, thereby transforming healthcare into smart healthcare [31]. The purpose of data analytics in smart healthcare is to provide intelligence in order to ensure remote health monitoring, assist in early disease diagnostics, make novel findings in disease trends, etc. by utilizing foundational services like physiological and psychological state detection, image recognition, voice recognition, etc.

3.3. Smart Industry. with the brisk advancement in communication, computing and manufacturing technologies, production in industrial organizations is being shifted from digital to intelligent [32]. Smart Industry is a smart manufacturing system which integrates production and services together to meet the industrial requirements. The data generated from smart industry typically consist of data pertinent to machine logs and manufacturing processes. Analytics of such data results in services like condition monitoring of machines, fault detection and analysis, machine health management, production optimization, flexible manufacturing, etc. [32,33].

3.4. Smart Transportation. This system aims to exploit sturdy and leading sensing, computational and communication technologies in order to facilitate smart recognition, tracking, and monitoring of vehicles [4]. These technologies will capacitate vehicle-to-vehicle communication in a meticulous way without human arbitration. Moreover, incorporating IoT in transportation systems will provide smart services like traffic congestion management, route optimization, safe driving, etc. Furthermore, real-time information about the availability of parking slots, weather condition, road condition, engine health, equipment maintenance will also be provided [34].

3.5. Smart Grid. Smart Grid is another consequential advancement of IoT for administering and disseminating electricity between suppliers and consumers in order to ameliorate efficiency, safety, reliability with real-time tracking and control [4]. Integrating IoT with electrical systems will facilitate services like optimization of power system performance, fault detection and analysis, security, reduction in operational and maintenance costs [35]. Sensors deployed in smart Grid continuously emanate data pertinent to control loops and security and the data generated demands fast analytics in order to optimize power consumption, predict future power supply needs, detect anomalies, etc.

3.6. Smart Agriculture. Factors like increasing population and dwindling of cultivable land as a result of urbanization, demand extraction of the most out of available resources. Smart agriculture is a novel approach of accomplishing farming tasks by mitigating human endeavor and by efficient utilization of farming resources. Smart agriculture employs advanced sensing, communication, computing and actuating technologies in order to facilitate services like climate control based on harvesting requisites, growth in productivity, automatic irrigation system monitoring, crop disease detection and prevention, soil monitoring, livestock monitoring, etc. [36]. Sensors deployed in Smart agriculture generate data pertinent to moisture content of soil, diameter of the trunk of plants, climatic conditions, humidity conditions, etc. and the generated data demand real-time analytics in order to facilitate aforementioned services.

3.7. Smart Government. Governments can attain a number of benefits from the amalgam of IoT and data analytics. Almost, all the tasks pertaining to government administration demand accurate analysis and prediction. Incorporating IoT and data analytics in the government functionalities will lead to better quality

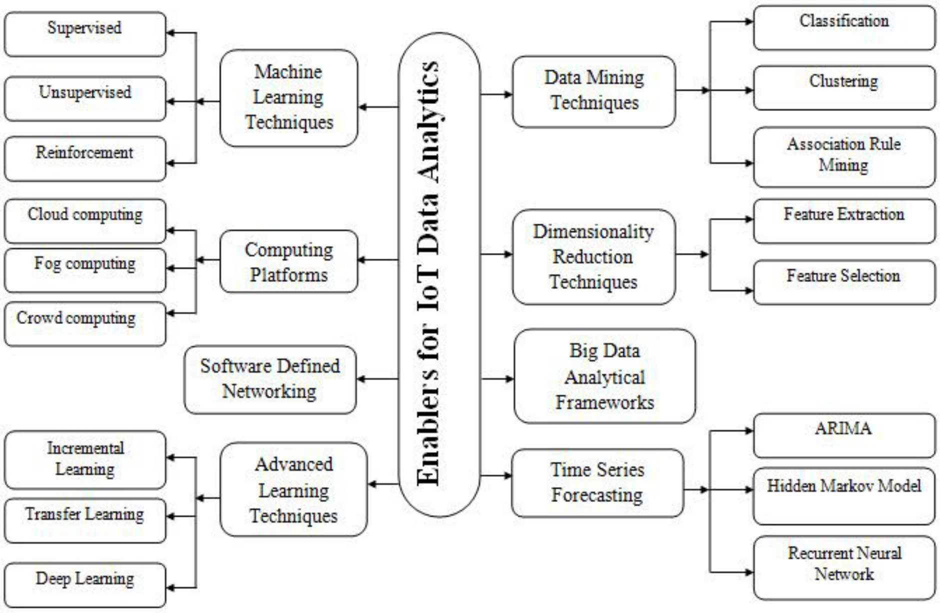

Fig. 4.1. *Key Enablers for IoT data analytics*

services, efficient decision-making, cost optimization, efficient policies and schemes, increase political trust, environmental monitoring, prediction and assessment of natural disasters, assessment of public demands, etc. [37].

3.8. Smart Education. IoT and data analytics contribute to the competence of education systems to a greater extent by enabling services like efficient online learning, teaching-learning optimization, classroom occupancy monitoring, content recommendation, learner behavior monitoring etc [38,39]. Moreover, Integrating IoT with education systems helps in motivating students, identifying weak and struggling students, learners progress assessment, and hence makes learning process efficient.

4. Key Enablers for IoT Data Analytics. From the discussion on IoT data analytics presented in section 2, it is apparent that recognizing and extracting the hidden information from IoT data is a pressing chore that surpasses the potential of conventional information processing and analyzing strategies. However, recent advancements in computational intelligence, data mining, and machine learning approaches are paving way for requisite data analytics in IoT. Fig 4.1 presents the key enablers for IoT data analytics. In the following subsections, we present an elucidation of these techniques in the realm of data analytics.

4.1. Data Mining Techniques. Data mining is utilized to discover concealed patterns and information from the data generated by IoT devices. The main objective of the data mining procedure is to reveal implicit knowledge from the data and mutate it into a valuable shape. Data mining techniques are of three types: classification, clustering and association rule mining.

Classification is a supervised learning procedure that uses a set of labeled data for training purposes to categorize data items into pre-defined classes [4]. The prime goal of utilizing classification in IoT is to predict a class for every instance of input data (unlabelled data). The set of labeled data is utilized for training to build the classification model while as unlabelled data is classified by the classification model. The objective of classification is to develop a classifier that learns the distribution of patterns in the set of labeled data. Classification has been used in numerous IoT use-cases including real-time ECG monitoring [40], twitter sentiment analysis [41], ebola virus outbreak control [42], real-time monitoring of breast cancer patients [43], automatic people counter in stores [44], real-time fall detection system for elderly people [45], defect detection in machines [46], cardiac arrest prediction [47], video surveillance [48], rice disease monitor and control [49], real-time condition monitoring of electric machines [50], etc.

Clustering is an unsupervised learning procedure that groups data items with similar characteristics together into the same cluster [18]. In other words, data items in the same cluster have identical traits and data items in different clusters have highly disparate traits. Examples of IoT use-cases that utilized clustering include activity recognition [51], heart disease survival prediction [52], electricity load prediction [53], behaviour visualization of Sybil attacker [54], Type 2 diabetes monitoring [55], wormhole attack detection [56], weather data analysis [57], safe driving [58], gesture recognition [59], etc.

Association rule mining includes recognition of frequently occurring attribute-value relationships. It assists in the creation of more qualitative information for effective decision-making [18]. Association rule mining focuses on discovering all the frequently occurring associations from a set of data items. It has been used in diverse IoT use-cases including data mining in medical applications [60], human activity recognition [61], extraction of usage patterns of devices [62], etc.

Table 4.1 presents the purpose of data mining in the IoT use-cases mentioned in this sub-section.

4.2. Machine Learning Techniques. Machine learning offers the ability to systems to automatically learn and improve from experience without demanding the obligation of adhering to static program directions. Machines learning approaches craft an effective correlation among input data instances and the output actions and are competent of accomplishing forecasting and decision-making tasks in IoT applications [20]. These approaches are generally divided into three categories: supervised, unsupervised and reinforcement learning.

Supervised learning techniques model dependencies and associations between the target prediction outcome and the input attributes so that outputs for upcoming data instances are forecasted depending on the associations it learned from the dataset [63,64]. Techniques in this category include Linear Regression, Decision tree, Random Forests, Naive Bayes, K-Nearest Neighbour (KNN), Support Vector Machine (SVM) and Artificial Neural Network (ANN).

Regression is a supervised learning algorithm that is used to forecast a real-valued output from the correlations learned from the training data. Linear regression presumes a linear correlation between the input predictors and the target output. Example of IoT use-case that utilized linear regression is energy consumption prediction in digital manufacturing systems [65].

Decision tree follows a greedy strategy to classify data items by arranging them based on attribute values. Example of IoT use case that utilized decision tree is activity and movement recognition [66].

Random Forest is a supervised learning technique in which a myriad of decision trees are trained on different subsets of training set chosen randomly. Example of IoT use-case that utilized random forest is diagnosis and prediction of diseases [67].

Naive Bayes is a supervised learning technique for performing multi-class classification. It uses Bayes theorem for determining the probability of a class given a data item. Example of IoT use-case that utilized naive bayes is device problem detection [68].

KNN is a supervised learning technique in which outputs for new data instances are predicted by exploring K identical data instances in the dataset and taking the mode of their output values as the predicted output for the new data instance. Example of IoT use-case that utilized KNN is appliance recognition in power management systems [69].

SVM, a supervised learning technique is based on the concept of augmenting the margin, i.e., each of the two sides of a hyperplane that splits the linearly separable input variable space into two classes. Example of IoT use-case that implemented SVM is indoor acoustic surveillance [70].

Artificial Neuron is the elementary computational unit in an ANN. It accepts one or more inputs and performs their weighted sum, which is then passed as an input to a non-linear function called as activation function. Example of IoT use-case that implemented ANN is intelligent intrusion detection [71].

Unsupervised learning applies techniques on the input data instances to mine useful information, detect patterns and group the data instances so that valuable insights are obtained [63,72]. These techniques include K-Means Clustering, Apriori and FP Growth.

K-Means clustering is an unsupervised learning technique that is utilized in scenarios with unlabeled data. The objective of this algorithm is to group data items into a K number of clusters. Example of IoT use-case

Table 4.1
Purpose of Data Mining in IoT use-cases.

Work	IoT Use-Case	Data Mining Method	Purpose of Data Mining	Dataset	Performance Results
[40]	Real-time ECG monitoring	Classification	To classify ECG data into different cardiovascular conditions	Data obtained from ECG sensor	-
[41]	Twitter sentiment analysis	Classification	To categorize tweets into two classes, positive and negative	Gold standard dataset from SemEval 2017	Accuracy: 99.2 percent
[42]	Ebola virus outbreak control	Classification	To assess the intensity of infection in a user based on the symptoms	EVD database	Accuracy: 94 percent
[43]	Real-time monitoring of breast cancer patients	Classification	To categorize breast cancer into two classes, benign and malignant	Breast cancer dataset from UCI repository	Accuracy: 95.6 percent
[44]	Automatic people counter in stores	Classification	To categorize people into adults and children based on their height	-	Accuracy: 91 percent
[45]	Real-time fall detection system for elderly people	Classification	To classify images into two types; standing state and falling state	Dataset consisting of fall/non fall events	Accuracy: 95.5 percent
[46]	Defect detection in machines	Classification	To categorize products into defected and non-defected classes	-	-
[47]	Cardiac arrest prediction	Classification	To classify ECG signal patterns into two types; normal and abnormal	Data collected from subjects with different age groups and heights	-
[48]	Video surveillance	Classification	To categorize traffic into five classes: non-critical traffic, little critical traffic, rather critical traffic, critical traffic, very critical traffic	Network traffic	Accuracy: 77 percent
[49]	Rice disease monitor and control	Classification	To classify rice diseases into four categories; rice bacterial blight, rice blast, rice brown spot and rice sheath rot	Images of infected rice leaves	Accuracy: 89.23 percent
[50]	Real-time condition monitoring of electric machines	Classification	To formulate condition monitoring decisions for electric machines based on the vibration patterns of the shaft	Data is gathered from the vibration analysis of the shaft	-
[51]	Activity recognition	Clustering	To categorize the activity patterns of the user into different clusters	Data from Washington State University (WSU) CASAS smart home project	Accuracy: 88 percent
[52]	Heart disease survival prediction	Clustering	To group data items into two clusters based on the attribute value similarity	Heart Disease Dataset	-
[53]	Electricity load prediction	Clustering	To categorize the massive dataset into small clusters	Electric load data from power industry	MAPE: 3.0554
[54]	Behavior visualization of Sybil attacker in IoT	Clustering	To group compromised identities and deploy the sybil node for corresponding identities without violating the set of adjacent nodes	Network Traffic	Coverage: 48.7 percent
[55]	Type-2 diabetes monitoring	Clustering	To categorize data into different clusters	Data of individuals with Type-2 Diabetes	-
[56]	Wormhole attack detection in IoT	Clustering	To divide the nodes into various clusters based on their location from the root node	Data from RPL network in IoT	Accuracy: 93 percent
[57]	Analysis of weather data and sensor fault detection	Clustering	To categorize the regions with different weather data characteristics	Linked Sensor Data and Linked Observation Data	-
[58]	Safe driving	Clustering	To identify accident-prone areas	Data collected using accelerometer, and GPS sensor	-
[59]	Gesture recognition	Clustering	To detect the presence of an event	-	Accuracy: 100 percent
[60]	Data Mining in medical applications	Association Rule Mining	To find similar items in the dataset	Medical Data	Number of scans: 122
[61]	Human activity recognition	Association Rule Mining	To mine frequent patterns	Data collected using wearable sensors	Accuracy: 95.16 percent
[62]	Extraction of usage patterns of IoT devices	Association Rule Mining	To extract device co-usage patterns	Data gathered from 201 residential broadband subscribers of a large European ISP	Confidence: 0.78

Table 4.2
Purpose of Machine Learning in IoT use-cases.

Work	IoT Use-Case	Machine Learning Technique	Purpose of Machine Learning techniques	Dataset	Performance Results
[65]	Predicting energy consumption of digital manufacturing systems	Linear Regression	To Predict the power consumption	Data obtained from SLS manufacturing system (EOS P700)	Accuracy: 96.1 percent
[66]	Activity and movement recognition	Decision Tree	To recognize the activities and movements of the patient	Data obtained from smart phone	Accuracy: 76.83 percent
[67]	Diagnosis and prediction of diseases	Random Forests	To predict the risk of chronic heart disease for the stroke affected patients	Data obtained from patients body	Accuracy: 93 percent
[68]	Device problem detection	Naive Bayes	To predict the problem in a device	-	-
[69]	Appliance recognition in Power Management systems	K Nearest Neighbour	To recognize an appliance	Appliance signature database	Accuracy: 92.73 percent
[70]	Indoor acoustic surveillance	Support Vector Machine	To identify high stress speech signals	Surveillance of Waterloo International Airport	Accuracy: 89.67 percent
[71]	Intelligent intrusion detection	Artificial Neural Network	To identify benign and malicious network traffic	Malicious shellcode data	Accuracy: 98 percent
[73]	Optimization of real-time traffic network assignment	K-Means Cluster	To cluster the similar data points	GIS data	-
[74]	Human sequential Movement Prediction	Apriori	To predict the human movement sequence patterns	Data collected using GPS device	F-measure: 0.687
[75]	Early detection of liver cancer	FP Growth	To discover patterns from liver cancer dataset for early detection	Data obtained from the British Columbia Cancer (BC) Agency	-
[77]	Predictive analytics	Q-learning	To forward a query to a proper query processor	-	-

that implemented K-Means clustering is optimization of real-time traffic network assignment [73].

Apriori is an extensively used algorithm for association rule mining. It is used for recognizing frequently occurring attribute-value relationships in the dataset. Example of IoT use-case that utilized apriori is human sequential movement prediction [74].

Another procedure for association rule mining is FP Growth. Apriori utilizes a breadth-first search approach to determine the set of frequently occurring data items and hence is quite expensive in terms of memory usage. While as FP Growth algorithm utilizes a depth-first search approach. Example of IoT use-case that utilized FP Growth is early diagnosis of liver cancer [75], etc.

Reinforcement learning algorithms learn incessantly from the experience of the environment in an iterative manner until they inspect the full range of feasible states [63, 76] e.g., Q-Learning.

Q-learning is a reinforcement learning algorithm that is based on value instead of policy. It is an easy method for agents to comprehend how to proceed efficiently in controlled environments. It operates by continuously advancing its evaluation measures of the quality of specific actions at specific states. Example of IoT use-case that utilized reinforcement learning is predictive analytics in smart cities [77].

Table 4.2 presents the purpose of machine learning in the IoT use-cases mentioned in this sub-section.

4.3. Advanced Machine Learning Techniques. Apart from traditional machine learning approaches, various advanced learning approaches like deep learning, incremental learning, and transfer learning are also used to dig out valuable knowledge from IoT data. Deep learning is appropriate for modeling complex behaviours of diverse data sets and transfer learning is mostly useful for scenarios with limited data sets while as incremental learning means real-time learning. It is appropriate for the scenarios where data arrive over time in a sequential fashion.

Deep learning is a representation learning approach that utilizes a hierarchical learning process to mine representations from data by making use of several hidden layers with non-linear transformations [78]. It offers an exemplary solution for various classification and recognition tasks as it encapsulates various levels of abstraction. It is appropriate for modeling complex behaviours of diverse datasets. Deep learning consists of diverse architectures including Restricted Boltzmann Machine (RBM), Deep Belief Network (DBN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), Auto-Encoder (AE) etc. RBM and DBN capture high-level representations of input data in an unsupervised manner. CNN works exceptionally well with image data. RNN and LSTM are utilized for time series forecasting. AEs are utilized for dimensionality reduction of high dimensional data. Deep learning models have been utilized in numerous IoT use-cases including transportation analysis [79], localization [80], air quality prediction [81], human activity detection [82], malware detection [83], traffic sign detection [84], crop recognition [85], fault diagnosis [86], plant classification [87], pose detection [88], etc.

Incremental learning means real-time learning. It is appropriate for the scenarios where data arrive over time in a sequential fashion [20, 89]. By means of their sequential treatment, these learning settings offer an elegant inferencing scheme for processing big data. To make upcoming learning and data analytics effective and beneficial, data-rigorous use cases demand that the learning algorithms should have the ability of performing incremental learning so that knowledge base is built over time [90]. Examples of IoT use-cases that utilized incremental learning include fire detection [91], self learning [92], outlier detection [93], etc.

Transfer learning is mostly useful for scenarios with limited datasets. It is a machine learning approach in which the learning parameters of a modeled predictive task are exploited to improve generalization in a different but related problem with limited data [94, 95, 96]. Transfer learning ensures better performance by saving time while modeling a predictive problem. Given the massive resource requirements of deep learning models on large and challenging datasets, transfer learning is admired in deep learning. Transfer learning involves the following steps:

1. Select a related source task: A related predictive modeling problem with ample amount of data is chosen.
2. Develop a model for the chosen source task.
3. The model developed for the source task is then used as a starting point for developing a model on the actual task.
4. Tune model.

Examples of IoT use-cases that harnessed transfer learning include human activity recognition [97], microscopic image classification [98], acceleration of neural network model execution [99], etc.

Table 4.3 presents the purpose of advanced machine learning techniques in the IoT use-cases mentioned in this sub-section.

4.4. Dimensionality Reduction Techniques. Data pre-processing is a vital step for effectual machine learning and data mining. Most machine learning, time series forecasting, and data mining techniques may not be effective for high dimensional data. Dimensionality means the number of attributes in the input data instances of a dataset. When the number of attributes in the input data instances is very huge as opposed to the number of instances in the dataset, certain algorithms struggle to train effective and efficient models. This anomaly is known as the Curse of Dimensionality [100]. To combat this curse of Dimensionality phenomenon, data downsizing techniques have been designed. These techniques are broadly classified into two types: Feature Selection and Feature Extraction.

Feature extraction techniques create a new, smaller set of features that are able to capture most of the useful information [101]. These techniques consist of Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and AEs.

PCA generates linear combination of the original attributes. The new attributes formed are arranged according to their explained variance. Examples of IoT use-cases that utilized PCA include include soil moisture retrieval [102], face recognition [103], intrusion detection [104], structural health monitoring [105], network

Table 4.3
Purpose of Advanced Machine Learning in IoT use-cases.

Work	IoT Use-Case	Advanced Learning Technique	Purpose of Advanced Machine Learning Techniques	Dataset	Performance Results
[79]	Transportation analysis	Deep Learning/ RBM and RNN	To forecast congestion of traffic	GPS data	Accuracy: 88 percent
[80]	Localization	Deep Learning	To predict indoor positioning based on indoor fingerprinting	CSI values gathered from three antennas	Mean Error: 0.9425
[81]	Air quality prediction	Deep Learning/ LSTM	To predict air quality	Pollution dataset from City Pulse EU FP7 Project	Precision: 98 percent
[82]	Human activity detection	Deep Learning/ CNN and LSTM	To predict activities based on data from multimodal wearable sensors	Opportunity dataset	F1 score: 95.8 percent
[83]	Malware Detection	Deep Learning/ DBN	To detect android malwares in smart phone	Android applications	Accuracy: 96 percent
[84]	Traffic sign detection	Deep Learning/ CNN	To detect traffic signs	-	-
[85]	Crop recognition	Deep Learning/ CNN	To distinguish summer crop types	Data collected from Landsat-8 and Sentinel-1A RS satellites in Ukraine	Accuracy: 85 percent
[86]	Fault diagnosis	Deep Learning/ AE	To learn the useful fault features and carry out fault diagnosis	CWRU bearing data	Accuracy: 94.11 percent
[87]	Plant Classification	Deep Learning/ CNN	To classify images of plants based on their types	Data collected from TARBIL project in Turkey	Accuracy: 97.47 percent
[88]	Pose detection	Deep Learning	To detect human poses	Image Parse dataset	Percentage of correct parts: 69 percent
[91]	Fire detection	Incremental Learning	To become accustomed to the timely changes in the data	Dataset from the Metropolitan Fire Brigade from a state in Australia	-
[92]	Self-learning	Incremental Learning	To enable self learning in IoT environments	-	-
[93]	Outlier detection	Incremental Learning	To detect outliers	-	-
[97]	Human Activity recognition	Transfer Learning	To utilize a pre-trained Autoencoder based activity model for unseen human activity recognition with unlabeled data	Data collected from accelerometer sensor	Accuracy: 98 percent
[98]	Microscopic image classification	Transfer Learning	To utilize features extracted from pre-trained Convolutional Neural Network models	2D-Hela and PAP-smear datasets	Accuracy (2D-Hela): 92.57 percent Accuracy (PAP-smear): 92.63 percent
[99]	Acceleration of neural network model execution	Transfer Learning	To make the deployment of Deep learning architectures possible on edge devices	-	-

anomaly detection [106], machine health management [107], etc.

LDA also generates linear combinations of original attributes. However, contrary to PCA, LDA doesn't maximize the explained variance. Rather, it augments the separability between classes. Examples of IoT use-cases that implemented LDA include ECG classification [108], event prediction [109], irrigation system surveillance [110], online activity recognition [111], intrusion detection [112], etc.

AEs are neural networks that are trained to regenerate their original inputs. The idea is to structure the hidden layer to have lesser neurons than the input/output layers. With the result, the hidden layer learns to build a smaller representation of the input. Examples of IoT use-cases that utilized AEs include human activity recognition [113], privacy preservation in sensor data analytics [114], prediction performance improvement in sensor and wearable systems [115], botnet traffic detection [116], fault diagnosis [117], etc.

Feature selection techniques filter irrelevant or redundant features from the dataset. These techniques include Genetic Algorithms. Genetic Algorithm accomplishes supervised feature selection. It efficiently selects features from high dimensional data sets where exhaustive search is not feasible. Examples of IoT use-cases that have utilized genetic algorithm include intrusion detection [118], medical image feature extraction and selection [119], pattern recognition [120], building energy optimization [121], gait analysis [122], etc.

Table 4.4 presents the purpose of dimensionality reduction techniques in IoT use-cases mentioned in this sub-section.

4.5. Time Series Forecasting. Most of the data produced by IoT devices are time-indexed [123]. And analyzing such data to extract relevant features, predict future instances, and to explore the relationship between multiple data streams is the main aim of time series modeling [124,125]. Time series data exhibit the property of autocorrelation i.e., the current value in the time series is correlated with the past values. In linear models, the current value depends linearly on the past observations while as in nonlinear models, the current value is a nonlinear function of past values. If the properties of a stochastic process fluctuate with time, it is hard to forecast the future values from its observed time series, this phenomenon is known as non-stationarity. Time series modeling techniques include Auto-Regressive Integrated Moving Average (ARIMA), Hidden Markov Model (HMM) and Recurrent Neural Network (RNN).

ARIMA is an extension of Auto-Regressive Moving Average [126]. Both of these techniques are used to forecast future instances in the series. However, ARMA cannot be applied in scenarios where data exhibit non-stationarity. In order to combat this problem, ARIMA was proposed. Since ARIMA is inherently linear, it is not able to model complex data patterns as opposed to approaches like HMM and RNN. ARIMA has been applied in various IoT use-cases including failure prediction in machines [127], weather forecasting [128], occupancy prediction in smart buildings [129], load prediction [130], building energy consumption forecasting [131], etc.

HMM is eminent for its competence in modeling short-term dependencies between adjoining observations. However, it is not suitable for scenarios with long-term dependencies [132]. Examples of IoT use-cases that have employed HMM include anomaly detection [133], physical activity recognition [134], traffic control management [135], health monitoring [136], prediction of user mobility [137], detection of sitting posture activities [138], etc.

RNN and its variants are highly effective in modeling sequences with complex structures because of the following reasons:

- They can extract patterns in time series data with long time lags.
- They are Robust to noise and can perform prediction in the presence of missing values.
- They are inherently non-linear which makes them suitable for modeling complex data patterns.
- They provide support for multi-variate and multi-step forecasting.

RNN suffers from vanishing/exploding gradient problem [139] due to which its performance gets degraded significantly while modeling long input sequences. To overcome this problem Long Short-Term Memory (LSTM), a variant of RNN was designed, that works exceptionally well for long input sequences. Examples of IoT use-cases that have used RNN include activity recognition based on multi-sensor data [140], network traffic classification [141], real-time deterministic control [142], weather forecasting [143], etc.

Table 4.5 provides the purpose of time series forecasting techniques in the IoT use-cases mentioned in this sub-section.

Table 4.4
Purpose of Dimensionality Reduction in IoT use-cases.

Work	IoT Use-Case	Dimensionality Reduction Technique	Purpose of Dimensionality Reduction Techniques	Dataset	Performance Results
[102]	Soil moisture retrieval	Principal Component Analysis	To reduce the number of dimensions in the feature set	Data collected from UWB radar sensor (P410)	Accuracy: 95.96 percent
[103]	Face recognition	Principal Component Analysis	To diminish the dimension of face feature so as to improve the computational efficiency	Cohn-Kanade face database	Accuracy: 95 percent
[104]	Intrusion detection	Principal Component Analysis	To decrease the number of features so as to decrease the processing time	Mobile network traffic data	F-measure weighted average: 0.834
[105]	Structural health monitoring	Principal Component Analysis	To get rid of environment interferences from the sensor data	Data is gathered from sensors attached to architectural structure	Accuracy: 94 percent
[106]	Network anomaly detection	Principal Component Analysis	To alleviate the dimensionality of the dataset	-	-
[107]	Machine health management	Principal Component Analysis	To convert high-dimensional data to low-dimensional space	-	-
[108]	ECG classification	Linear Discriminant Analysis	To decrease the number of features in the ECG signal	-	-
[109]	Event prediction	Linear Discriminant Analysis	To reduce the number of features in the dataset for improving the event prediction performance of SVM	Data obtained from IoT devices	Precision: 87.17 percent
[110]	Irrigation system surveillance	Linear Discriminant Analysis	To retrieve the colour of plants and soil images	Data collected from sensors deployed in agricultural fields	-
[111]	Online Activity recognition	Linear Discriminant Analysis	Feature Extraction	WSU Cairo ADL dataset	Accuracy: 98.36 percent
[112]	Intrusion detection	Linear Discriminant Analysis	Classification for intrusion detection	Network traffic data	Accuracy: 99.44 percent
[113]	Human activity recognition	Autoencoders	Feature Extraction	-	-
[114]	Privacy preservation in sensor data analytics	Autoencoders	To convert sensitive discriminative features of data into non-sensitive features in order to guard privacy of the users	Opportunity, Skoda, and Hand-Gesture datasets	F1 Score (Opportunity dataset): 97.36 percent F1 Score (Skoda dataset): 94.94 percent F1 Score (Hand-Gesture dataset): 75.43 percent
[115]	Prediction performance improvement in mobile and wearable systems	Autoencoders	To enquire about unfamiliar features in an efficient manner	HAPT dataset	Accuracy: 91 percent
[116]	Botnet traffic detection	Autoencoders	To extract new set of features in order to distinguish malicious and benign network traffic	Network traffics from ISCX	True Positive Rate: 91 percent
[117]	Fault Diagnosis	Autoencoders	Gearbox fault diagnosis	-	-
[118]	Intrusion detection	Genetic Algorithm	To carry out the attribute reduction of the feature sets	KDD-CUP99 dataset	Accuracy: 96.8 percent
[119]	Medical image feature extraction and selection	Genetic Algorithm	To choose the reduced set of features	-	-
[120]	Pattern recognition	Genetic Algorithm	Feature selection	Leaf shape image dataset	-
[121]	Building energy optimization	Genetic Algorithm	To curtail the expenditure of the energy consumption	Data obtained from a building in Cardiff, UK	Energy saving: 25 percent
[122]	Gait analysis	Genetic Algorithm	To select meaningful features	Data Collected using the 8 camera ELITE stereo-photogrammetric system	Accuracy: 97 percent

TABLE 4.5
Purpose of Time Series Forecasting in IoT use-cases.

Work	IoT Use-Case	Time Series Forecasting Technique	Purpose of Time Series Forecasting Techniques	Dataset	Performance Results
[127]	Failure Prediction in machines	Auto Regressive Integrated Moving Average	To predict failure in machines	Data collected from the sensors attached to a slitting machine	Accuracy: 98.69 percent
[128]	Weather forecasting	Auto Regressive Integrated Moving Average	Time series based weather forecasting	Data collected from sensors that measure following parameters: Temperature, Humidity, Station Barometric pressure, Wind speed, and Wind direction	Root Mean Squared Error: 0.003867201
[129]	Occupancy prediction in smart buildings	Auto Regressive Integrated Moving Average	To predict the number of residents in a smart building at a given location and time	Wifi dataset collected from University of Houston main campus	-
[130]	Load prediction	Auto Regressive Integrated Moving Average	To predict load behaviour in IoT	Data collected from the IoT devices	Average Response Time: 49.79 milliseconds
[131]	Building energy consumption forecasting	Auto Regressive Integrated Moving Average	To predict building energy consumption	Data collected from sensors	MAPE: 1.05-2.59
[133]	Anomaly detection	Hidden Markov Model	To detect the device anomaly	Data collected from the IoT devices	Accuracy: 98 percent
[134]	Physical activity recognition	Hidden Markov Model	To recognize physical activities	Data collected from 10 subjects	Precision: 82.51 percent
[135]	Traffic control management	Hidden Markov Model	To learn the profile information of traffic in less time	Data collected from different traffic profiles	Accuracy: 95 percent
[136]	Health monitoring	Hidden Markov Model	Real time monitoring of cardio vascular patients	Data collected from patients body	-
[137]	Prediction of user mobility	Hidden Markov Model	To estimate the next location	27 day traffic data of mobile network	Prediction time: 1.39 seconds
[138]	Detection of sitting posture activities	Hidden Markov Model	To identify sitting posture activities	Kinect and Smartwatch based 42 dimensional data	Accuracy: 64.88 percent
[140]	Activity recognition based on multi-sensor data	Recurrent Neural Network	To predict future activities of a resident	MIT dataset for activity	Accuracy: 90.85 percent
[141]	Classification of network traffic	Recurrent Neural Network	To classify the traffic flowing in a network	Dataset from RedIRIS	Accuracy: 99.59 percent
[142]	Real time deterministic control	Recurrent Neural Network	Knowledge discovery	Nottingham and CMU datasets	Accuracy for Nottingham: 93.9 percent Accuracy for CMU: 82.3 percent
[143]	Weather forecasting	Recurrent Neural Network	To predict weather	Dataset obtained from Valley weather station in Anglesey (North Wales, UK)	Mean Absolute Error: 0.0476

4.6. Computing Platforms. Cloud computing [144,145,146] and Fog computing [20] are two important models for managing the enormous volume of data produced from IoT environs. With the brisk growth of the IoT, the traditional cloud computing is facing stern issues, like undesirable network latency, and spectral inefficiency which does not make it suitable for scenarios requiring minimal latency, real-time treatment, and mobility support. Determined to resolve these issues, new paradigm transfers the functioning of cloud computing

Table 4.6
Distinction between Cloud Computing and Fog Computing

Features	Cloud computing	Fog computing	References
Size	Server extremely large in size	Servers small in size	[151]
Computational Capacity	Huge	Limited	[18]
Applications	Appropriate for delay-tolerant and computationally exhaustive implementations	Appropriate for delay-critical applications requiring minimal latency, real-time treatment	[18,151]
Communication Overhead	High, as devices are connected to the internet during the whole period	Low, because devices can acquire cached contents straight from edge gateway	[18]
Deployment	Demands composite installation planning	Require ad-hoc installation with no or slight drafting	[18,151]
Operation	Operate in environments fully guarded by cloud operators	Usually, operate in scenarios that are mainly determined by requirements of customers	[151]
Location	Centralized	Distributed over the large geographical area	[151]

closer to the data source. This technology is referred to as Fog computing [147]. With the result network congestion is reduced and decision-making becomes fast. However, these fog devices generally do not have adequate storage and computational resources. Table 4.6 provides the comparison of these computing platforms.

Examples of IoT use-cases that utilized cloud computing as the computing platform include industrial IoT big learning [148], hybrid systems for smart agriculture [149], disease diagnosis [150], disease prevention in precision agriculture [151], temperature control systems [152] etc.

Examples of IoT use-cases that harnessed fog computing as the computing platform include preventive healthcare and assisted living in smart ambient [153], video surveillance [154], asset provisioning for crowd sensing applications in IoT [155], crime assistance [156], data analytics in smart cities using big data [157], etc. [155] proposed a fog based computing scheme known as Mist computing that provides cost-effective resource provisioning for IoT crowd sensing applications.

Apart from cloud computing and fog computing, another emerging computing paradigm known as crowd computing can be utilized for managing the data produced by IoT systems. In crowd computing, IoT devices deployed close to each other and with related interests can share computing and power resources so as to optimize the performance of the IoT systems [158]. Crowd computing has got a massive potential in IoT applications.

4.7. Big Data Analytical Frameworks. This section explores the big data analytical frameworks that can be utilized for analyzing humongous volumes of data produced from IoT environments. Applying the right data analytical framework is fundamental for the successful development of an IoT application. Depending on the analytical requirements of IoT application, data analytics can be performed either in Cloud or near the IoT data source (Fog nodes). Table 4.7 summarizes big data analytical frameworks.

4.8. Software Defined Networking. Software Defined Networking (SDN) is a novel technology that simplifies network administration by segregating the control plane from the data plane, thereby centralizing the network intelligence [174]. It enables virtualization within the network and enhances networking capability. SDN fulfils following fundamental requisites of IoT applications:

4.8.1. Network Management. Network management is a vital consideration in IoT for supervising the tremendous number of devices and the massive volume of data produced by them [175]. SDN enables programmatically efficient control mechanism and hides the complexities of network management from end users. It optimizes network management functionalities such as efficient utilization of bandwidth, minimization of latency and load balancing.

Table 4.7
Big Data Analytical Frameworks

Framework	Description	Benefits	Limitations	Used in IoT application
Hadoop [159]	Apache Hadoop is a software platform that carries out batch processing of massive datasets in a distributed manner using clusters of computing devices	Less susceptible to failure Scalable Provides reliable storage	Slow Lacks security functionalities Lacks support for stream processing	Smart City development and Urban planning [164] Defining human behaviours in social IoT [165], etc.
Spark [160]	Apache Spark is a platform for conducting analytics of huge datasets using distributed computing. It endorses in-memory data processing so as to augment the proficiency of data analytical applications	Supports stream processing Relatively faster because of In-memory computation Fault-tolerant	Expensive Requires manual optimization	Smart Building system [166] Cloud based data analytics for smart cities [167], etc.
Storm [161]	Apache Storm is a platform that carries out data processing in real-time in a distributed manner and generates the result promptly with minimal delay	Scalable Fault-tolerant Less latency Supports stream processing	No flow control	Real time monitoring system in Automotive Manufacturing [168] Intelligent data processing on edge devices [169], etc.
Flink [162]	Apache Flink offers immense potential to perform real-time data processing in a fault tolerant manner at a rate of millions of events per second	Faster Better memory management High throughput Requires less configuration	Not common	Analytics in Industrial Environments [170] Real-time analysis of social networks [171], etc.
Azure Stream Analytics [163]	An event processing engine that analyzes massive volumes of streaming data in order to extract inferences, recognize patterns etc.	Real-time processing capabilities Scalable Data aggregation capabilities	Lacks Job management	Multimedia analytics [172] Monitoring and performance analysis of power plants in real-time [173], etc.

4.8.2. Efficient resource utilization. Efficient resource utilization is fundamental for improving the performance of the network [175]. SDN relieves the simpler edge devices from accomplishing the multifaceted networking tasks and utilizes the available resources efficiently [174].

4.8.3. Energy Management. Massive number of data centers are deployed to process the humongous magnitude of data sensed by IoT devices. Consequently, large quantity of energy is utilized to power these data centers. SDN plays an important role in optimizing the energy usage as it maps the traffic efficiently to the suitable servers and switches off the other unnecessary devices in the data center [175].

4.8.4. Security and Privacy. The utilization of flow-rule-based traffic forwarding concept in SDN facilitates secure control of flows between the various devices in the network, which in turn improves the security and privacy of the data generated by IoT devices [175].

SDN has been applied in numerous IoT use-cases including efficient traffic management for emergency situations [176], intrusion detection [177], service delivery [178], traffic congestion avoidance [179], dynamic distribution of IoT analytics and effective utilization of network resources [180], low latency anomaly detection in smart city [181], etc.

5. Vision and Open Challenges. Data analytics has brought considerable benefits to IoT applications. However, in order to leverage the full potential of data analytics in IoT applications, following major challenges need to be addressed:

5.1. Data Pre-processing. In data pre-processing, noisy data are smoothened, ambiguities in the data are removed, and missing values are filled, thus making it suitable for further processing. Because of the constrained nature of IoT sensors and intermittent loss of connectivity, the massive scale of IoT data contains more irregularities and uncertainties, thereby complicating data pre-processing. Moreover, IoT data may contain missing and incomplete values that lead to poor data quality. Ensuring completeness in IoT data is vital for its data quality. Efficient pre-processing techniques that can remove irregularities and uncertainties in the data and make it suitable for further processing should be researched.

5.2. Data compression and redundancy reduction. Not all the data generated from IoT environments are useful. Also, there exists a high level of redundancy in IoT data. The closely deployed sensor nodes in IoT tend to capture similar information that leads to redundancy in IoT data. Redundant data not only lead to energy wastage but also dissipate the storage space. Moreover, it also affects the feature extraction process. Hence, removing redundancy in IoT data and ensuring its uniqueness is crucial for its data quality. Effective data compression and redundancy reduction techniques need to be employed so as to alleviate the burden of storage and analytics in such systems.

5.3. Data Integration. How to integrate and analyze heterogeneous data arriving from distributed and diverse sources so that a unified view of these diverse data formats is created is an impediment to IoT data analytics. Deep learning is quite effective in analyzing heterogeneous data. However, the severe resource requirements of deep learning algorithm limit its use in IoT applications. Hence, investigating the ways that will reduce the computational requirements of deep learning models becomes crucial. However, this should be done while preserving the accuracy of deep learning models.

5.4. Visualization. Data visualization aims to make data more meaningful for further analysis and interpretation. However, inappropriate data visualization will diminish the significance of the original data and may even thwart efficient data analysis. Orchestrating visualization in IoT data is complex because of its massive scale. Moreover, visualization in case of highly heterogeneous and diverse IoT environments is a challenging task. Given the importance of appropriate data visualization, devising visualization techniques that are well-suited for the representation of highly complex IoT data becomes crucial.

5.5. Expandability and scalability. The sharply growing IoT data bring in the challenges of expandability and scalability for the IoT analytical systems. Analytical paradigms that are proficient enough to deal with progressively growing complex datasets are highly required. Running analytical techniques on distributed systems with parallel processing is the potential solution for this problem.

5.6. Energy management. With the exponential growth of IoT data, transmission, storage, processing and analysis of such enormous data will certainly dissipate more energy. Energy consumption control and management solutions should be designed for such systems.

5.7. Security and Privacy. Data generated from IoT environs are susceptible to external intervention. Hence, Authentication, authorization, and encryption techniques should be utilized to ensure security of IoT systems. However, conventional data protection solutions are not applicable to the IoT data because of its massive scale and highly diverse nature. To this purpose, design of novel Security and Privacy solutions for such systems becomes inevitable.

6. Conclusion. Ample volumes of data have been generated since the previous decade with the escalation in the number of smart devices. Analyzing this voluminous magnitude of data so as to explore novel knowledge, forecast potential insights and to formulate management decisions is a vital procedure that makes IoT a laudable technology for enhancing the standard of our lives. The prime requisite for most of the IoT applications is an intelligent analytical mechanism that can carry out tasks like classification, clustering, association rule mining, or time series analysis. However, conventional analytical procedures do not tackle the surging analytical needs of IoT systems. To this purpose, this paper identifies the key enablers for IoT data analytics and surveys their role in data analytics. Furthermore, several challenges faced by IoT data analytics were identified so as to stimulate research directions in this arena. At last, it is worthwhile to state that data analytics has brought immense benefits to IoT applications, however, a number of challenges still remain unaddressed, the list of which has been discussed. To the best of our knowledge, this work is the first of this kind and we hope that this survey will be beneficial for the researchers in the field of Data Analytics to understand the key enablers and lead them to the direction of possible future research in this field.

REFERENCES

[1] M. Ge, H. Bangui, and B. Buhnova, *Big data for internet of things: A survey*, Future Generation Computer Systems, Elsevier, 2018.

[2] Gerd Kortuem, Daniel Fitton, and Vasughi Sundramoorth, *Smart Objects as Building Blocks for the Internet of Things*, IEEE Internet Computing, January/February, pp.44-51, 2010.

[3] Yunchuan Sun, Houbing Song, Antonio J. Jara, and Rongfang Bie , *Internet of Things and Big Data Analytics for Smart and Connected Communities*, IEEE Access, Vol. 14, No. 8, 2015.

[4] Mohsen Marjani, Fariza Nasaruddin, Abdullah Gani, Ahmad Karim, Ibrahim Abaker Targio Hashem, and Aisha Siddiqa,Ibrar Yaqoob, *Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges*, IEEE Access, Vol. 5, pp. 5247-5261, 2016.

[5] Amy Nordrum, *Popular Internet of Things Forecast of 50 Billion Devices by 2020 Is Outdated*, https://spectrum.ieee.org/tech-talk/telecom/internet/popular-internet-of-things-forecast-of-50-billion-devices-by-2020-is-outdated, 2016.

[6] S. Yin, and O. Kaynak, *Big data for modern industry: Challenges and trends*, Proceedings of IEEE, Vol. 103, No. 2, pp. 143-146, 2015.

[7] Han Hu, Yonggang Wen, Tat-Seng Chua, and Xuelong Li, *Toward scalable systems for big data analytics: A technology tutorial*, IEEE Access, Vol. 2, pp. 652687, 2014.

[8] Simmhan Y., and Perera S., *Big Data Analytics Platforms for Real-Time Applications in IoT*, Big Data Analytics, Springer, New Delhi, pp. 115-135, 2016.

[9] Eugene Siow, Thanassis Tiropanis, and Wendy Hall., *Analytics for the Internet of Things: A survey*, ACM Computing Surveys (CSUR) Vol. 1, No. 1, 2018.

[10] Ejaz Ahmed, Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, Imran Khan, Abdelmuttlib Ibrahim Abdalla Ahmed, Muhammad Imran, and Athanasios V. Vasilakos, *The role of big data analytics in Internet of Things*, Computer Networks, Elsevier, Vol. 129, pp. 459-471, 2017.

[11] Meng Ma, Ping Wang, and Chao-Hsien Chu, *Data Management for Internet of Things: Challenges, Approaches and Opportunities*, International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, IEEE, pp. 1144-1151, 2013.

[12] Tingli Li, Yang Liu, Ye Tian, Shuo Shen, and Wei Mao, *A Storage Solution for Massive IoT Data Based on NoSQL*, International Conference on Green Computing and Communications, Conference on Internet of Things, and Conference on Cyber, Physical and Social Computing, IEEE, Besancon, France , pp. 50-57, 2012.

[13] M. Chen, S. Mao, and Y. Liu, *Big Data: A Survey*, Mobile Networks and applications, Springer, Vol. 19, pp. 171-209, 2014.

[14] Mervat Abu-Elkheir, Mohammad Hayajneh, and Najah Abu Ali, *Data Management for the Internet of Things: Design Primitives and Solution*, Sensors, Vol. 13, pp. 15582-15612, 2013.

[15] Shikhar Verma, Yuichi Kawamoto, Zubair Md. Fadlullah, Hiroki Nishiyama, and Nei Kato, *A Survey on Network Methodologies for Real-Time Analytics of Massive IoT Data and Open Research Issues*, IEEE Communications Surveys and Tutorials, Vol. 19, No. 3, pp. 1457-1477, 2016.

[16] Sai Xie, and Zhe Chen, *Anomaly Detection and Redundancy Elimination of Big Sensor Data in Internet of Things*, https://arxiv.org/abs/1703.03225, 2017.

[17] Hongming Cai, Boyi Xu, Lihong Jiang, and Athanasios V. Vasilakos, *IoT-based Big Data Storage Systems in Cloud Computing: Perspectives and Challenges*, IEEE Internet of Things Journal, Vol. 4, No. 1, pp. 75-87, 2016.

[18] Chun-Wei Tsai, Chin-Feng Lai, Ming-Chao Chiang, and Laurence T. Yang, *Data Mining for Internet of Things: A Survey*, IEEE Communications Surveys and Tutorials, Vol. 16, No. 1, pp. 77-97, 2014.

[19] Furqan Alam, Rashid Mehmood, Iyad Katib, and Aiiad Albeshri, *Analysis of Eight Data Mining Algorithms for Smarter Internet of Things (IoT)*, International Workshop on Data Mining in IoT Systems, Procedia Computer Science, Elsevier, 2016.

[20] Shree Krishna Sharma, and Xianbin Wang, *Live Data Analytics With Collaborative Edge and Cloud Processing in Wireless IoT Networks*, IEEE Access, Vol. 5, pp. 4621-4635, 2017.

[21] Mehdi Mohammadi, Ala Al-Fuqaha, Sameh Sorour, and Mohsen Guizani, *Deep Learning for IoT Big Data and Streaming Analytics: A Survey*, IEEE Communications Surveys and Tutorials, 2018.

[22] Albert Malama, Lilias Makashini, Henry Abanda, Austine Ngombe, and Priscilla Mudenda, *A Comparative Analysis of Energy Usage and Energy Efficiency Behavior in Low- and High-Income Households: The Case of Kitwe, Zambia*, Resources, Vol. 4, pp. 871-902, 2015.

[23] Lszl Lengyel, Pter Ekler, Tams Ujj, Tams Balogh, and Hassan Charaf, *SensorHUB: An IoT Driver Framework for Supporting Sensor Networks and Data Analysis*, International Journal of Distributed Sensor Networks, 2015.

[24] Min Chen, Yixue Hao, Kai Hwang, Lu Wang, and Lin Wang, *Disease Prediction by Machine Learning Over Big Data From Healthcare Communities*, IEEE Access, Vol. 5, pp. 8869-8879, 2017.

[25] Antorweep Chakravorty, Chunming Rong, Pal Evensen, and Tomasz Wiktor Wlodarczyk, *A Distributed Gaussian-Means Clustering Algorithm for Forecasting Domestic Energy Usage*, proceedings of SMARTCOMP, IEEE, China, 2014.

[26] J.C. O'Brien, J.R. Leech, C.C. Wright, C.R. Reeves, N.C. Steele, and C.Y. Choi, *Neural Networks for early prediction of machine failure*, Colloquium on Advanced Vibration Measurements, Techniques and Instrumentation for the early prediction of failure, IEEE, UK, 1992.

[27] Aaron Waibel, Abdullah Ali Alshehri, Soundararajan Ezekiel, *Anomaly Prediction in Seismic Signals Using Neural Networks*, Proceedings of Applied Imagery Pattern Recognition Workshop, IEEE, USA, 2013.

[28] Davorin Kuchan, *Prescriptive Analytics for Industrial IoT Failure Risk Management*, https://www.sparklinglogic.com/prescriptive-analytics-industrial-iot/, 2016.

[29] River Logic, *10 Use Cases for Prescriptive Analytics in Healthcare*, https://blog.riverlogic.com/infographic-10-use-cases-prescriptive-analytics-in-healthcare, 2016.

[30] Muhammad Raisul Alam, Mamun Bin Ibne Reaz, and Mohd Alauddin Mohd Ali, *A Review of Smart HomesPast, Present, and Future*, IEEE Transactions On Systems, Man, And CyberneticsPart C: Applications And Reviews, Vol. 42, No. 6, pp. 1190-1203, 2012.

[31] Stephanie B. Baker, Wei Xiang, and Ian Atkinson, *Internet of Things for Smart Healthcare: Technologies, Challenges, and Opportunities*, IEEE Access, Vol. 5, pp. 26521-26544, 2017.

[32] Baotong Chen, Jiafu Wan, Lei Shu, Peng Li, Mithun Mukherjee, and Boxing Yin, *Smart Factory of Industry 4.0: Key Technologies, Application Case, and Challenges*, IEEE Access, Vol. 6, pp. 6505-6519, 2017.

[33] Hyoung Seok Kang, Ju Yeon Lee, SangSu Choi, Hyun Kim, Jun Hee Park, Ji Yeon Son, Bo Hyun Kim, and Sang Do Noh, *Smart Manufacturing: Past Research, Present Findings, and Future Directions*, International Journal Of Precision Engineering And Manufacturing-Green Technology Vol. 3, No. 1, pp. 111-128, 2016.

[34] Luo Qi, *Research on Intelligent Transportation System Technologies and Applications*, Workshop on Power Electronics and Intelligent Transportation System, IEEE, Guangzhou, China, 2018.

[35] Rabab Hassan, and Ghadir Radman, *Survey on Smart Grid*, International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS), IEEE, Pudukkottai, India, 2010.

[36] Aqeel-ur-Rehman, and Zubair A. Shaikh, *Smart Agriculture*, Applications of Modern High Performance Networks, Bentham Science Publishers, pp. 120-129, 2009.

[37] Feng Chen, Pan Deng, Jiafu Wan, Daqiang Zhang, Athanasios V.Vasilakos, and Xiaohui Rong , *Data Mining for the Internet of Things: Literature Review and Challenges*, International Journal of Distributed Sensor Networks, 2015.

[38] M.B.Ibanez, A.Di Serio, D.Villaran, and C.D.Kloos, *Experimenting with electromagnetism using augmented reality Impact on flow student experience and educational effectiveness* , Computers and Education, Vol.71, pp. 1-13, 2014.

[39] L.-f. Kwok, *A vision for the development of i-campus*, Smart Learning Environments, vol. 2, pp. 112, 2015.

[40] Granados J., Westerlund T., Zheng L., and Zou Z., *IoT Platform for Real-Time Multichannel ECG Monitoring and Classification with Neural Networks*, Research and Practical Issues of Enterprise Information Systems, CONFENIS 2017, Lecture Notes in Business Information Processing, Springer, Cham, Vol. 310, 2018.

[41] Salha M. Alzahrani, *Development of IoT Mining Machine for Twitter Sentiment Analysis: Mining in the Cloud and Results on the Mirror*, 15th Learning and Technology Conference, Jeddah, Saudi Arabia, IEEE, 2018.

[42] Sanjay Sareen, Sandeep K. Sood, and Sunil Kumar Gupta, *IoT-based cloud framework to control Ebola virus outbreak*, J Ambient Intell Human Comput., Springer, Vol. 9, pp. 459-476, 2016.

[43] A. Suresh, R. Udendhran, M. Balamurgan, and R. Varatharajan, *A Novel Internet of Things Framework Integrated with Real Time Monitoring for Intelligent Healthcare Environment*, Journal of Medical Systems, Springer, Vol. 43, 2019.

[44] Viriyavisuthisakul S., Sanguansat P., Toriumi S., Hayashi M., and Yamasaki T., *An Automatic People Counter in Stores Using a Low-Cost IoT Sensing Platform*, Advances in Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP 2017, Smart Innovation, Systems and Technologies, Springer, Cham, Vol. 81, 2018.

[45] Xiangbo Kong, Zelin Meng, Naoto Nojiri, Yuji Iwahori, Lin Meng, and Hiroyuki Tomiyama, *A HOG-SVM based fall detection IoT system for elderly persons using deep sensor*, Procedia Computer Science, Vol. 147, pp. 276282, 2019.

[46] Der-Chen Huang, Chun-Fu Lin, Chih-Yen Chen, and Jyh-Rou Sze, *The Internet Technology for Defect Detection System with Deep Learning Method in Smart Factory*, 4th International Conference on Information Management, IEEE, Oxford, UK, 2018.

[47] AKM Jahangir Alam Majumder, Yosuf Amr ElSaadany, Roger Young Jr., and Donald R.Ucci, *An Energy Efficient Wearable Smart IoT System to Predict Cardiac Arrest*, Advances in Human-Computer Interaction, Hindawi, 2019.

[48] Albert Rego, Alejandro Canovas, Jose M. Jimnez, and Jaime Lloret, *An Intelligent System for Video Surveillance in IoT Environments*, IEEE Access, Vol. 6, pp. 31580-31598, 2018.

[49] Amrita A. Joshi, and B.D.Jadhav, *Monitoring and Controlling Rice Diseases Using Image Processing Techniques*, International Conference on Computing, Analytics and Security Trends (CAST), IEEE, 2016.

[50] D.Ganga, and V. Ramachandran, *IoT based Vibration Analytics of Electrical Machines*, IEEE Internet of Things Journal, 2017.

[51] Serge Thomas, Mickala Bourobou, and Younghwan Yoo, *User Activity Recognition in Smart Homes Using Pattern Clustering Applied to Temporal ANN Algorithm*, Sensors, Vol. 15, pp. 11953-11971, 2015.

[52] Manish Joshi, Bramah Hazela, and Vineet Singh, *An application of IoT on Hungarian database using Data mining Techniques: A collaborative approach*, 3rd International Conference on Advances in Computing, Communication and Automation, IEEE, Dehradun, India, 2017.

[53] Xishuang Dong, Lijun Qian, Lei Huang, *Short-Term Load Forecasting in Smart Grid: A Combined CNN and K-Means Clustering Approach*, International Conference on Big Data and Smart Computing (BigComp), IEEE, 2017.

[54] Alekha Kumar Mishra, Asis Kumar Tripathy, Deepak Puthal, and Laurence T. Yang, *Analytical Model for Sybil Attack Phases in Internet of Things*, IEEE, Internet of Things Journal, 2018.

[55] John M Dennis, Beverley M Shields, William E Henley, Angus G Jones, and Andrew T Hattersley, *Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data*, Lancet Diabetes Endocrinol, 2019.

[56] Prachi Shukla, *ML-IDS: A Machine Learning Approach to Detect Wormhole Attacks in Internet of Thing*, Intelligent Systems Conference, IEEE, London, UK, 2017.

[57] Aras Can Onal, Omer Berat Sezer, Murat Ozbayoglu, and Erdogan Dogdu,*Weather Data Analysis and Sensor Fault Detection Using An Extended IoT Framework with Semantics, Big Data, and Machine Learning*, International Conference on Big Data (BIGDATA), IEEE, Boston, MA, USA., 2017.

[58] Mohd Abdullah Al Mamun, Jinat Afroj Puspo, and Amit Kumar Das, *An Intelligent Smartphone Based Approach Using IoT for Ensuring Safe Driving*, International Conference on Electrical Engineering and Computer Science (ICECOS), IEEE, Palembang, Indonesia, 2017.

[59] Raul parada, Kamruddin Nur, Joan Melia-Segui, and Rafael Pous, *Smart Surface: RFID-Based Gesture Recognition Using k-Means Algorithm*, 12th International Conference on Intelligent Environments, IEEE, London, UK, 2016.

[60] Mohiuddin Ali Khan, Sateesh Kumar Pradhan, and Huda Fatima, *An Efficient Technique for Apriori Algorithm in Medical Data Mining*, Innovations in Computer Science and Engineering, Springer, 2019.

[61] Martin Atzmueller, Naveed Hayat, Matthias Trojahn, and Dennis Kroll, *Explicative Human Activity Recognition using Adaptive Association Rule-Based Classification*, International Conference on future IoT technologies (Future IoT), IEEE, Eger, Hungary, 2018.

[62] Gevorg Poghosyan, Ioannis Pefkianakis, Pascal Le Guyadec, and Vassilis Christophides, *Extracting Usage Patterns of Home loT Devices*, Symposium on Computers and Communications (ISCC), IEEE, 2017.

[63] David Fumo, *Types of Machine Learning Algorithms You Should Know*, https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861, 2017.

[64] Amanpreet Singh, Narina Thakur, and Aakanksha Sharma, *A Review of Supervised Machine Learning Algorithms*, 3rd International Conference on Computing for Sustainable Global Development, IEEE, New Delhi, India, 2016.

[65] Jian Qin, Ying Liu, and Roger Grosvenor, *Data analytics for energy consumption of digital manufacturing systems using Internet of Things method*, 13th Conference on Automation Science and Engineering (CASE), IEEE, Xi'an, China, 2017.

[66] Igor Bisio, Alessandro Delfino, Fabio Lavagetto, and Andrea Sciarrone, *Enabling IoT for In Home rehabilitation Accelerometer signals classification methods for activity and movement recognition*, IEEE Internet of Things Journal, 2016.

[67] Ani R, Krishna S, Anju N, Sona Aslam M, and O.S Deepa, *IoT Based Patient Monitoring and Diagnostic Prediction Tool using Ensemble Classifier*, International Conference on Advances in Computing, Communications, and Informatics (ICACCI), IEEE, Udupi, India, 2017.

[68] Pooja A. Dhobi, and Niraj Tevar, *IoT Based Home Appliances Control*, Proceedings of the International Conference on Computing Methodologies and Communication (ICCMC), IEEE, Erode, India, 2017.

[69] Lester James V. Miranda, Marian Joice S. Gutierrez, Samuel Matthew G. Dumlao, and Rosula SJ Reyes, *Appliance Recognition using Hall Effect Sensors and K-Nearest Neighbors for Power Management Systems*, Region 10 Conference (TENCON): IEEE, Singapore, 2016.

[70] Allaa R.Hilal, Aya Sayedelahl, Arash Tabibiazar, Mohamed S.Kamel, and Otman A.Basir, *A distributed sensor management for large-scale IoT indoor acoustic surveillance*, Future Generation Computer Systems, Elsevier, 2018.

[71] Alex Shenfield, David Day, and Aladdin Ayesh, *Intelligent intrusion detection systems using artificial neural networks*, ICT express, Elsevier, Vol.4, pp. 95-99, 2018.

[72] Greene D., Cunningham P., and Mayer R., *Unsupervised Learning and Clustering*, Machine Learning Techniques for Multimedia. Cognitive Technologies. Springer, Berlin, Heidelberg, 2018.

[73] Jiachen Yang, Yurong Han, Yafang Wang, Bin Jiang, Zhihan Lv, and Houbing Song, *Optimization of Real Time Traffic Network Assignment Based on IoT Data Using DBN and Clustering Model in Smart City*, Future Generation Computer Systems, Elsevier, 2017.

[74] S. Sridhar Raj, and M. Nandhini, *Ensemble human movement sequence prediction model with apriori based probability tree classifier (APTC) and Bagged J48 on machine learning*, Journal of King Saud University-Computer and Information Sciences, Elsevier, 2018.

[75] Pinheiro, Fabiola, Kuo, Mu-Hsing, Thomo, Alex, Barnett, and Jeff, *Extracting association rules from liver cancer data using the FP-growth algorithm*, 3rd International Conference on Computational Advances in Bio and Medical Sciences (ICCABS), IEEE, New Orleans, LA, USA, 2013.

[76] Leslie Pack Kaelbling, Michael L. Littman, and Andrew W. Moore, *Reinforcement Learning: A survey*, Journal of Artificial Intelligence Research, Vol. 4, pp. 237-285, 1996.

[77] Kostas Kolomvatsos, and Christos Anagnostopoulos, *Reinforcement Learning for Predictive Analytics in Smart Cities*, Informatics, MDPI, Vol. 3, No. 16, 2017.

[78] Alexandra Lheureux, Katarina Grolinger, Hany F. Elyamany, and Miriam A. M. Capretz, *Machine Learning With Big Data: Challenges and Approaches*, IEEE Access, Vol. 5, pp. 7776-7797, 2017.

[79] X. Ma, H. Yu, Y. Wang, and Y. Wang, *Large-scale transportation network congestion evolution prediction using deep learning theory*, PloS one, Vol. 10, No. 3, 2015.

[80] X. Wang, L. Gao, S. Mao, and S. Pandey, *Deepfi: Deep learning for indoor fingerprinting using channel state information*, Wireless Communications and Networking Conference (WCNC), IEEE, New Orleans, LA, USA, 2015.

[81] brahim KK, Mehmet Ulvi MEK, and Suat ZDEMR, *A deep learning model for air quality prediction in smart cities*, International Conference on Big Data (BIGDATA), IEEE, Boston, MA, USA, 2017.

[82] F. J. Ordonez, and D. Roggen, *Deep convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition*, Sensors, Vol. 16, 2016.

[83] Z. Yuan, Y. Lu, Z. Wang, and Y. Xue, *Droid-sec: deep learning in android malware detection*, Computer Communication Review, SIGCOMM, ACM, Vol. 44, No. 4, pp. 371372, 2014.

[84] K. Lim, Y. Hong, Y. Choi, and H. Byun, *Real-time traffic sign recognition based on a general purpose gpu and deep-learning*, PLoS one, Vol. 12, No. 3, 2017.

[85] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, *Deep learning classification of land cover and crop types using remote sensing data*, Geoscience and Remote Sensing Letters, IEEE, 2017.

[86] H. Shao, H. Jiang, F. Wang, and H. Zhao, *An enhancement deep feature fusion method for rotating machinery fault diagnosis*, Knowledge Based Systems, Elsevier, Vol. 119, pp. 200220, 2017.

[87] Salar Razavi, and Hulya Yalcin, *Plant classification using group of features*, 24th Signal Processing and Communication Application Conference (SIU), IEEE, 2016.
[88] A. Toshev, and C. Szegedy, *Deeppose: Human pose estimation via deep neural networks*, Conference on Computer Vision and Pattern Recognition, IEEE, Columbus, OH, USA, 2014.
[89] Alexander Gepperth, and Barbara Hammer, *Incremental learning algorithms and applications*, proceedings of European Symposium on Artificial Neural Networks, Computational Intelligence, and Machine Learning, Belgium, 2016.
[90] Haibo He, Sheng Chen, Kang Li, and Xin Xu, *Incremental Learning from Stream Data*, IEEE Transactions on Neural Networks, Vol. 22, No. 12, pp. 1901-1914, 2011.
[91] Rashmika Nawaratne, Damminda Alahakoon, Daswin De Silva, Prem Chhetri, and Naveen Chilamkurti, *Self-evolving intelligent algorithms for facilitating data interoperability in IoT environments*, Future Generation Computer Systems, Elsevier, 2018.
[92] Arun Kishore Ramakrishnan, Davy Preuveneers, and Yolande Berbers, *Enabling self learning in Dynamic and open IoT environments*, Procedia Computer Science, Elsevier, Vol. No. 32, pp. 207-214.
[93] Alexander Gepperth, and Barbara Hammer, *Incremental learning algorithms and applications*, Proceedings of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), Bruges, Belgium, 2016.
[94] Lisa Torrey, Jude Shavlik, *Transfer Learning Handbook of Research on Machine Learning Applications*, IGI Global, 2009.
[95] Jason Brownlee, *A Gentle Introduction to Transfer Learning for Deep Learning*, https://machinelearningmastery.com/transfer-learning-for-deep-learning/, 2017.
[96] Pranoy Radhakrishnan, *What is transfer Learning?*, https://towardsdatascience.com/what-is-transfer-learning-8b1a0fa42b4, 2017.
[97] Md Abdullah Al Hafiz Khan, and Nirmalaya Roy, *UnTran Recognizing Unseen Activities with Unlabeled data using Transfer Learning*, Third International Conference on Internet of Things Design and Implementation, IEEE/ACM, Orlando, FL, USA, 2018.
[98] Long D. Nguyen, Dongyun Lin, Zhiping Lin, and Jiuwen Cao, *Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation*, International Symposium on Circuits and Systems (ISCAS), IEEE, Florence, Italy, 2018.
[99] Chang-Jiun Chen, Kai-Chun Chen, and May-chen Martin-Kuo, *Acceleration of Neural Network Model Execution on Embedded Systems*, International Symposium on VLSI Design, Automation and Test (VLSI-DAT), IEEE, Hsinchu, Taiwan, 2018.
[100] Lei Yu, Jieping Ye, and Huan Liu, *Dimensionality Reduction for Data Mining - Techniques, Applications and Trends*, http://www.cs.binghamton.edu/ lyu/SDM07/DR-SDM07.pdf.
[101] *Dimensionality Reduction algorithms: Strengths and weaknesses*, https://elitedatascience.com/dimensionality-reduction-algorithms.
[102] Jing Liang, Xiaoxu Liu, and Kuo Liao, *Soil Moisture Retrieval using UWB Echoes via Fuzzy Logic and Machine Learning*, IEEE Internet of Things Journal, Vol. 14, No. 8, 2015.
[103] Yong-Ping Chen, Qi-Hui Chen, Kuan-Yu Chou, and Ren-Hau Wu, *Low-Cost Face Recognition System Based on Extended Local Binary Pattern*, International Automatic Control Conference (CACS), IEEE, Taiwan, 2016.
[104] Mohammed Faisal Elrawy, Ali Ismail Awad, and Hesham F. A. Hamed, *Flow based Features for a Robust Intrusion Detection System Targeting Mobile Traffic*, 23rd International Conference on Telecommunications (ICT), IEEE, Thessaloniki, Greece, 2016.
[105] Hongyang Zhang, Junqi Guo, Xiaobo Xie, Rongfang Bie, and Yunchuan Sun, *Environmental Effect Removal Based Structural Health Monitoring in the Internet of Things*, Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, IEEE, Taichung, Taiwan, 2013.
[106] Gilberto Fernandes Jr., Joel J. P. C. Rodrigues, Luiz Fernando Carvalho, Jalal F. Al-Muhtadi, and MarioLemes Proena Jr., *A comprehensive survey on network anomaly detection*, Telecommunication Systems, Springer, 2018.
[107] Gil-Yong Lee et al., *Machine health management in smart factory: A review*, Journal of Mechanical Science and Technology, Springer, Vol. 32, No. 3, pp. 987-1009, 2018.
[108] R. Varatharajan, Gunasekaran Manogaran, and M. K. Priyan, *A big data classification approach using LDA with an enhanced SVM method for ECG signals in cloud computing*, Multimed Tools Appl, Springer.
[109] Sina Dami, and Mahtab Yahaghizadeh, *Efficient Event Prediction in an IOT Environment Based On LDA Model and Support Vector Machine*, 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), IEEE, Kerman, Iran, 2018.
[110] Kabilan N, and Dr. M. Senthamil Selvi, *Surveillance and Steering of Irrigation System in Cloud using Wireless Sensor Network and Wi-Fi Module*, Fifth International Conference on Recent Trends in Information Technology, IEEE, Chennai, India, 2016.
[111] Z. Huang, K.-J. Lin, B.-L. Tsai, S. Yan, and C.-S. Shih, *Building edge intelligence for online activity recognition in service-oriented IoT systems*, Future Generation Computer Systems, Elsevier, 2018.
[112] Eduardo Viegas, Altair Santin, Luiz Oliveira, Andre Franca, Ricardo Jasinski, and Volnei Pedroni, *A Reliable and Energy Efficient Classifier Combination Scheme for Intrusion Detection in Embedded Systems*, Computers and Security, Elsevier, Vol. 78, pp. 16-32, 2018.
[113] Henry Friday Nweke, Ying Wah Teh, Mohammed Ali Algaradi, and Uzoma Rita Alo, *Deep Learning Algorithms for Human Activity Recognition using Mobile and Wearable Sensor Networks State of the Art and Research Challenges*, Expert Systems With Applications, Elsevier, Vol. 105, pp. 233-261, 2018.
[114] Mohammad Malekzadeh, Richard G. Clegg, and Hamed Haddadi, *Replacement AutoEncoder: A Privacy-Preserving Algorithm for Sensory Data Analysis*, Third International Conference on Internet-of-Things Design and Implementation,

IEEE, IEEE/ACM, Orlando, FL, USA, 2018.

[115] Mohammad Kachuee, Anahita Hosseini, Babak Moatamed, Sajad Darabi, Majid Sarrafzadeh, *Context-Aware Feature Query To Improve The Prediction Performance*, Global Conference on Signal and Information Processing (GlobalSIP), IEEE, Montreal, QC, Canada, 2017.

[116] Homayoun S., Ahmadzadeh M., Hashemi S., Dehghantanha A., and Khayami R., *BoTShark: A Deep Learning Approach for Botnet Traffic Detection*, Cyber Threat Intelligence, Advances in Information Security, Vol 70. Springer, Cham.

[117] Guifang Liu, Huaiqian Bao, and Baokun Han, *A Stacked Autoencoder-Based Deep Neural Network for Achieving Gearbox Fault Diagnosis*, Mathematical Problems in Engineering, Hindawi.

[118] Lianbing Deng, Daming Li, Xiang Yao, David Cox, and Haoxiang Wang, *Mobile network intrusion detection for IoT system based on transfer learning algorithm*, Cluster Computing, Springer, 2018.

[119] G.Nagarajan, R.I.Minu, B Muthukumar, V.Vedanarayanan, and S.D.Sundarsingh, *Hybrid Genetic Algorithm for Medical Image Feature Extraction and selection*, International Conference on Computational Modeling and Security (CMS), Procedia Computer Science, Elsevier, Vol. 85, pp. 455-462, 2016.

[120] Sourabh S. Patil, and Dr. Mrs. A. S. Bhalchandra, *Pattern Recognition Using Genetic Algorithm*, International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), IEEE, Palladam, India.

[121] Jonathan Reynolds, Yacine Rezgui, Alan Kwan, Solne Piriou, and A ZoneLevel, *Building Energy Optimisation Combining an Artificial Neural Network, a Genetic Algorithm, and Model Predictive Control*, Energy, Elsevier, Vol. 151, pp. 729-739, 2018.

[122] Rosa Altilio, Luca Liparulo, Andrea Proietti, Marco Paoloni, and Massimo Panella, *A Genetic Algorithm for Feature Selection in Gait Analysis*, IEEE Congress on Evolutionary Computation (CEC), IEEE, Vancouver, BC, Canada, 2016.

[123] Sameh Ben Fredj, *Analyzing IoT Data: Introduction to Time Series Forecasting with Python*, http://iot-ee.com/en/2017/08/07/analysing-iot-data-introduction-time-series-forecasting-python/, 2017.

[124] Ratnadip Adhikari, and R. K. Agrawal, *An Introductory Study on Time Series Modeling and forecasting*, Available: https://arxiv.org/ftp/arxiv/papers/1302/1302.6613.pdf.

[125] Siddhartha Bhandari, Neil Bergmann, Raja Jurdak, and Branislav Kusy, *Time Series Data Analysis of Wireless Sensor Network Measurements of Temperature*, Sensors, Vol. 17, 2017.

[126] Jeff Morrison, *Autoregressive Integrated Moving Average Models (ARIMA)*, http://www.forecastingsolutions.com/arima.html

[127] Ameeth Kanawaday, and Aditya Sane, *Machine Learning for Predictive Maintenance of Industrial Machines using IoT Sensor Data*, 8th International Conference on Software Engineering and Service Science, IEEE, Beijing, China, 2017.

[128] Gaurav Chavan, and Dr. Bashirahamad Momin, *An Integrated approach for Weather Forecasting over Internet of Things: A Brief Review*, International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), IEEE, Palladam, India, 2017.

[129] Basheer Qolomany, Ala Al-Fuqaha, Driss Benhaddou, and Ajay Gupta, *Role of Deep LSTM Neural Networks And Wi-Fi Networks in Support of Occupancy Prediction in Smart Buildings*, 19th International Conference on High Performance Computing and Communications; 15th International Conference on Smart City; 3rd International Conference on Data Science and Systems, IEEE, Bangkok, Thailand, 2017.

[130] Rodrigoda Rosa Righi, Everton Correa, Mrcio Migu el Gomes, and Cristiano Andrda Costa, *Enhancing performance of IoT applications with load prediction and cloud elasticity*, Future Generation Computer Systems, Elsevier, 2018.

[131] Chirag Deb, Fan Zhang, Junjing Yang, Siew Eang Lee, and Kwok Wei Shah, *A review on time series forecasting techniques for building energy consumption*, Renewable and Sustainable Energy Reviews, Elsevier, Vol. 74, pp. 902-924, 2017.

[132] Bilal Esmael, Arghad Arnaout, RudolfK. Fruhwirth, and Gerhard Thonhauser, *Improving Time Series Classification Using Hidden Markov Models*, 12th International Conference on Hybrid Intelligent Systems, IEEE, Pune, India, 2012.

[133] Palani kumar, and Meenakshi DSouza, *Design a Power Aware Methodology in IoT based on Hidden Markov Model*, 9th International Conference on Communication Systems and Networks (COMSNETS), IEEE, Banglore, India.

[134] Jun Qi, Po Yang, Martin Hanneghan, Stephen Tang, Bo Zhou, *A Hybrid Hierarchical Framework for Gym Physical Activity Recognition and Measurement Using Wearable Sensors*, IEEE Internet of Things Journal, 2018.

[135] Ramkumar Eswaraprasad,and Linesh Raja, *Improved Intelligent Transport System for Reliable Traffic Control Management by Adapting Internet of Things*, International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions), IEEE, Dubai, United Arab Emirates, 2017.

[136] Maryem Neyja, Shahid Mumtaz, Kazi Mohammed Saidul Huq, Sherif Adeshina Busari, Jonathan Rodriguez, and Zhenyu Zhou, *An IoT-Based E-Health Monitoring System Using ECG Signal*, Global Communications Conference, IEEE, Singapore, 2017.

[137] Yihang Cheng, Yuanyuan Qiao, and Jie Yang, *An Improved Markov Method for Prediction of User Mobility*,IFIP, 2016.

[138] M. Tariq, H. Majeed, M.O. Beg, F.A. Khan, and A. Derhab, *Accurate detection of sitting posture activities in a secure IoT based assisted living environment*,Future Generation Computer Systems, Elsevier, 2018.

[139] Qingchen Zhang, Laurence T. Yang, Zhikui Chen , and Peng Li, *A survey on deep learning for big data*, Information Fusion, Elsevier, Vol. 42, pp. 146-157, 2018.

[140] Jiho Park, Kiyoung Jang, and Sung-Bong Yang, *Deep Neural Networks for Activity Recognition with Multi-Sensor Data in a Smart Home*, 4th World Forum on Internet of Things (WF-IoT), Singapore, 2018.

[141] Manuel Lopez-Martin, Belen Carro, Antonio Sanchez-Esguevillas, and Jaime Lloret, *Network Traffic Classifier*

With Convolutional and Recurrent Neural Networks for Internet of Things, IEEE Access, Vol. 5, pp. 18042-18050, 2017.
[142] Shin Kamada, and Takumi Ichimura, *Knowledge Extracted from Recurrent Deep Belief Network for Real Time Deterministic Control*, International Conference on Systems, Man, and Cybernetics (SMC), IEEE, Banff, Canada, 2017.
[143] A.J. Hussain et al., *A dynamic neural network architecture with immunology inspired optimization for weather data forecasting*, Big Data Research, Elsevier, 2018.
[144] Singh, Parminder, Pooja Gupta, and Kiran Jyoti, *Tasm: technocrat arima and svr model for workload prediction of web applications in cloud*, Cluster Computing Vol. 22, No. 2, pp. 619-633, 2019.
[145] Singh, Parminder, Pooja Gupta, Kiran Jyoti, and Anand Nayyar, *Research on Auto-Scaling of Web Applications in Cloud: Survey, Trends and Future Directions*, Scalable Computing: Practice and Experience Vol. 20, No. 2, pp. 399-432, 2019.
[146] Singh, Parminder, Pooja Gupta, and Kiran Jyoti, *Triangulation Resource Provisioning for Web Applications in Cloud Computing: A Profit-Aware Approach*,Scalable Computing: Practice and Experience Vol. 20, No. 2, pp. 207-222, 2019.
[147] Mung Chiang, and Tao Zhang, *Fog and IoT: An Overview of Research Opportunities*, IEEE Internet of Things Journal, Vol. 3, No. 6, pp. 854-864, 2016.
[148] Qingchen Zhang, Laurence T. Yang, Zhikui Chen, Peng Li, and Fanyu Bu, *An Adaptive Dropout Deep Computation Model for Industrial IoT Big Data Learning with Crowdsourcing to Cloud Computing*, IEEE Transactions On Industrial Informatics, 2017.
[149] Sahitya Roy, Dr Rajarshi Ray, Aishwarya Roy, Subhajit Sinha, Gourab Mukherjee, Supratik Pyne, Sayantan Mitra, Sounak Basu, and Subhadip Hazra, *IoT, Big Data Science and Analytics, Cloud Computing and Mobile App based Hybrid System for Smart Agriculture*, 8th Industrial Automation and Electromechanical Conference (IEMECON), IEEE, Bangkok, Thailand, 2017.
[150] P. Verma, and S.K. Sood, *Cloud-centric IoT based disease diagnosis healthcare framework*, Journal of Parallel and Distributed Computing, Elsevier, Vol. 116, pp. 27-38, 2017.
[151] Karim Foughali, Karim Fathallah, and Ali Frihida, *Using Cloud IOT for disease prevention in precision agriculture*, Procedia Computer Science, Elsevier, Vol. 130, pp. 575-582, 2018.
[152] Pablo Palacios, and Andres Cordova, *Approximation and Temperature Control System via an Actuator and a Cloud An Application Based on the IoT for Smart Houses*, International Conference on eDemocracy and eGovernment (ICEDEG), IEEE, Ambato, Ecuador, 2018.
[153] Luca Cerina, Sara Notargiacomo, Matteo Greco, Luca Paccani, and Marco Domenico Santambrogio, *A Fog-Computing architecture for Preventive Healthcare and Assisted Living in Smart Ambients*, 3rd International Forum on Research and Technologies for Society and Industry (RTSI), IEEE, Modena, Italy, 2017.
[154] Ning Chen, Yu Chen, Yang You, Haibin Ling, Pengpeng Liang, and Roger Zimmermann, *Dynamic Urban Surveillance Video Stream Processing Using Fog Computing*, Second International Conference on Multimedia Big Data (BigMM), IEEE, Taipei, Taiwan, 2016.
[155] Hamid Reza Arkian, Abolfazl Diyanat, and Atefe Pourkhalili, *MIST: Fog-based Data Analytics Scheme with Cost-Efficient Resource Provisioning for IoT Crowdsensing Applications*, Journal of Network and Computer Applications, Elsevier, Vol. 82, pp. 152-165, 2017.
[156] Augusto J. V. Neto, Zhongliang Zhao, Joel J. P. C. Rodrigues, Hugo B. Camboim, and Torsten Braun, *Fog-based Crime-Assistance in Smart IoT Transportation System*, IEEE Access, 2018.
[157] Bo Tang, Zhen Chen, Gerald Hefferman, Shuyi Pei, Tao Wei, Haibo He, and Qing Yang, *Incorporating Intelligence in Fog Computing for Big Data Analysis in Smart Cities*, IEEE Transactions on Industrial Informatics, 2016.
[158] S.Bi, R. Zhang, Z. Ding, and S. Cui, *Wireless communications in the era of big data*, IEEE Commun. Mag., Vol.53, No.10, pp.190199, Oct.2015.
[159] *Apache Hadoop*, https://hadoop.apache.org/
[160] *Apache Spark*, http://spark.apache.org/
[161] *Apache Storm*, http://storm.apache.org
[162] *Apache Flink*, https://flink.apache.org/
[163] *Azure Stream Analytics*, https://azure.microsoft.com/en-in/services/stream-analytics/
[164] Muhammad Mazhar Ullah Rathore, Awais Ahmad, Anand Paul, and Dr. Seungmin Rho, *Urban planning and building smart cities based on the Internet of Things*, Computer Networks, Vol. 101, pp. 63-80, 2016.
[165] Anand Paul, Awais Ahmad, M. Mazhar Rathore, and Sohail Jabbar, *Smartbuddy: Defining Human Behaviors Using Big Data Analytics In Social Internet Of Things*, Wireless Communications, IEEE, pp. 68-74, 2016.
[166] Muhammad Rizwan Bashir, and Asif Qumer Gill, *Towards an IoT Big Data Analytics Framework: Smart Buildings Systems*, 18th International Conference on High Performance Computing and Communications, IEEE, Sydney, NSW, Australia, 2016.
[167] Zaheer Khan, Ashiq Anjum, Kamran Soomro, and Muhammad Atif Tahir, *Towards cloud based big data analytics for smart future cities*, Journal of Cloud Computing Advances, Systems and Applications, Vol. 4, No. 2, 2015.
[168] Muhammad Syafrudin, Ganjar Alfian, Norma Latif Fitriyani, and Jongtae Rhee, *Performance Analysis of IoT Based Sensor, Big Data Processing, and Machine Learning Model for Real Time Monitoring System in Automotive Manufacturing*, Sensors, MDPI, 2018.
[169] Roger Young, Sheila Fallon, and Paul Jacob, *An Architecture for Intelligent Data Processing on IoT Edge Devices*, 19th International Conference on Modelling and Simulation, IEEE, Cambridge, UK, 2017.
[170] Michael Zimmermann, Felix W. Baumann, Michael Falkenthal, Frank Leymann, and Ulrich Odefey, *Automating the Provisioning and Integration of Analytics Tools with Data Resources in Industrial Environments using OpenTOSCA*, 21st International Enterprise Distributed Object Computing Conference Workshops, IEEE, Quebec City, QC, Canada,

2017.
[171] Giacomo Marciani, Marco Piu, Michele Porretta, and Matteo Nardelli, *Grand Challenge: Real-time Analysis of Social Networks Leveraging the Flink Framework*, DEBS, ACM, Irvine, CA, USA, 2016.
[172] Aleksandr Farseev, Ivan Samborskii, and Tat-Seng Chua, *bBridge: A Big Data Platform for Social Multimedia Analytics*, ACM, Amsterdam, The Netherlands, 2016.
[173] Patric Boscolo, *Real time monitoring and performance analysis of power plants at BaxEnergy*, https://microsoft.github.io/techcasestudies/iot/2017/06/30/baxenergy.html, 2017.
[174] Ahmet Cihat Baktir, Atay Ozgovde, and Cem Ersoy, *How Can Edge Computing Benefit from Software-Defined Networking: A Survey, Use Cases and Future Directions*, IEEE Communications Surveys and Tutorials, Vol. 19, pp. 2359-2391, 2017.
[175] Samaresh Bera, Sudip Misra, and Athanasios V. Vasilakos, *Software-Defined Networking for Internet of Things: A Survey*, IEEE Internet of Things Journal, Vol. 4, No. 6, pp. 1994-2008, 2017.
[176] A.Rego, L.Garcia, S.Sendra, and J.Lloret, *Software Defined Network-based control system for an efficient traffic management for emergency situations in smart cities*, Future Generation Computer Systems, Elsevier, Vol. 88, pp. 243-253, 2018.
[177] Wani A., and Revathi S, *Analyzing Threats of IoT Networks Using SDN Based Intrusion Detection System (SDIoT-IDS)*, Smart and Innovative Trends in Next Generation Computing Technologies, NGCT, Communications in Computer and Information Science, Vol. 828, Springer, Singapore, 2018.
[178] Arbiza L.M.R., Tarouco L.M.R., Bertholdo L.M., and Granville L.Z.S, *SDN-Based Service Delivery in Smart Environments*, Intelligent Distributed Computing IX, Studies in Computational Intelligence, Vol. 616, Springer, Cham, 2016.
[179] Raul Muoz, Ricard Vilalta, Noboru Yoshikane, Ramon Casellas, Ricardo Martnez, Takehiro Tsuritani, and Itsuro Morita, *IoT-aware Multi-layer Transport SDN and Cloud Architecture for Traffic Congestion Avoidance Through Dynamic Distribution of IoT Analytics*, European Conference on Optical Communication (ECOC), IEEE, Gothenburg, Sweden, 2017.
[180] Raul Muoz, Ricard Vilalta, Noboru Yoshikane, Ramon Casellas, Ricardo Martnez, Takehiro Tsuritani, and Itsuro Morita, *Integration of IoT, Transport SDN and Edge/Cloud computing for Dynamic Distribution of IoT Analytics and Efficient Use of Network Resource*, Journal of Lightwave Technology, Vol. 36, No. 7, pp. 1420-1428, 2018.
[181] Jose Santos, Philip Leroux, Tim Wauters, Bruno Volckaert, F.D Turck, *Anomaly detection for Smart City applications over 5G Low Power Wide Area Network*, Network Operations and Management Symposium, Taiwan, 2018.

Edited by: Ashish Kumar Luhach
Received: May 21, 2019
Accepted: Oct 25, 2019