

Received November 25, 2021, accepted December 26, 2021, date of publication January 4, 2022,
date of current version January 14, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3140409

The Internet of Things and Architectures of Big Data Analytics: Challenges of Intersection at Different Domains

DINA FAWZY, SHERIN M. MOUSSA[✉], AND NAGWA L. BADR

Department of Information Systems, Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt

Corresponding author: Sherin M. Moussa (sherinmoussa@cis.asu.edu.eg)

ABSTRACT The current exponential advancements in the Internet of Things (IoT) technologies pave a vast intelligent computing platform by integrating smart objects with sensing, processing and communication capabilities. The core element of IoT is the complex big data generated from different interconnected sources at real-time, presenting divergent processing and analysis challenges. Best practices in software engineering have been continuously addressed in IoT technologies to handle such big data efficiently at different domains. Despite of the massive studies dedicated for IoT, no explicit processing architecture is proposed based on real investigation of software engineering concepts and big data analytics characteristics in IoT. This paper provides a systematic literature review for the current state-of-the-art of IoT systems in different domains. The study investigates the current techniques and technologies that serve IoT systems from the big data analytics and software engineering perspectives, revealing a matrix for the specific IoT data features and their encountered challenges and gaps for each domain. The review deduces a proposed domain-independent software architecture for big IoT data analytics, maintaining various IoT data processing challenges, including data scalability, timeliness, heterogeneity, inconsistency, confidentiality and correlations. Finally, the main research gaps are emphasized for future considerations.

INDEX TERMS Big data, data mining, Internet of Things, IoT, software engineering.

I. INTRODUCTION

The power of recent Internet technologies has invaded all devices, facilitating the control of innumerable autonomous gadgets to introduce the Internet of Things (IoT) [1]. IoT smartly interconnects smart objects and actuators in order to offer a smart interactive environment with human [2]. IoT are now representing a major source of data, as billions of connected objects gather, analyze, share and transmit data in an intelligent way [3]. These data are generated continuously in huge amounts, with different types of heterogeneity and unreliability at real time from various IoT devices to fulfil their responsibilities [4]. Hence, understanding and processing these data in a powerful way are major challenges, as the current conventional approaches are not capable of manipulating such data. This creates a crucial need for new practices in software engineering (SWE) in conjunction with big data analytics (BDA) and IoT [5], [6] to optimally process and analyze IoT data. From the BDA perspective, IoT

data should be maintained regarding the main ten big data Vs [7]–[10]. As from the SWE perspective, IoT data require a novel processing model that manages their challenges, such as data concurrency, data consistency and availability that significantly complicate any associated development [11], [12]. Thus, the merge of a powerful SWE model and BDA offers the intelligent service of IoT [13].

IoT technologies have made tremendous economic and environmental impacts that affect society in many domains [14]–[16]. Figure 1 shows a domain-based taxonomy for the main IoT-based systems from the BDA and SWE perspectives. We have investigated several research studies in various IoT domains, such as: (1) smart environments, (2) human analytics, (3) network analytics, (4) energy analytics, and (5) environmental analytics to explore the most common IoT-based SWE and BDA challenges.

Although different survey studies have been conducted for IoT-based systems, many research gaps are still vague, summarized as follows:

- 1) SWE concepts have not been thoroughly investigated for seamless adoption to the complex IoT-specific data

The associate editor coordinating the review of this manuscript and approving it for publication was M. Sabarimalai Manikandan[✉].

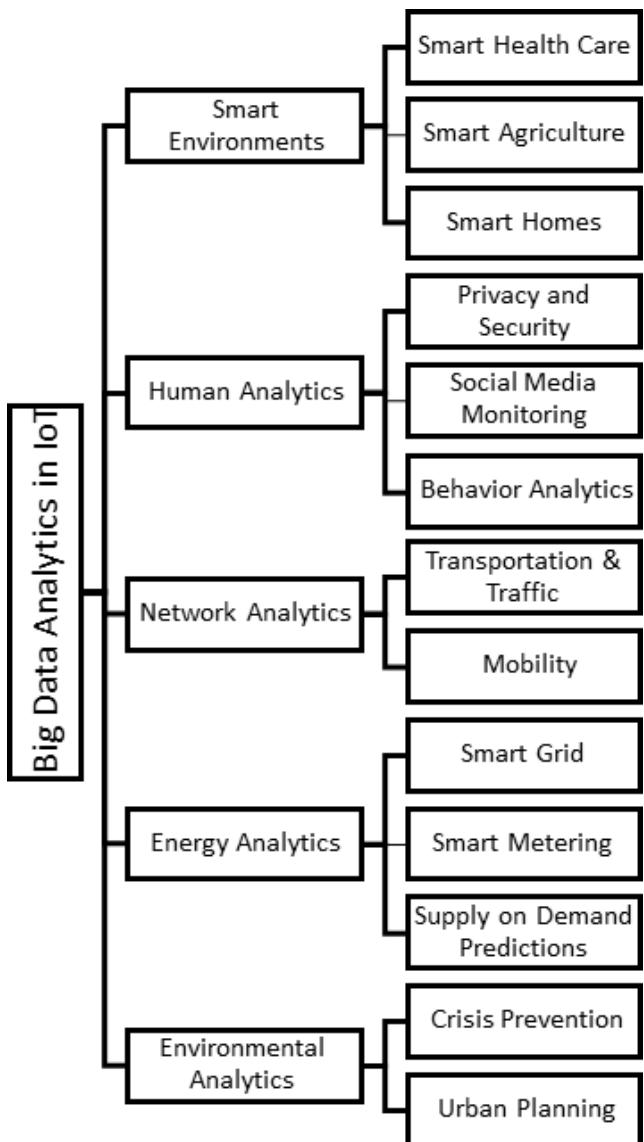


FIGURE 1. A domain-based taxonomy of big data analytics systems in IoT.

nature, reflecting different IoT data processing and performance challenges, such as system scalability, reliability, etc.

- 2) Big data characteristics have not been enough evaluated with respect to various IoT data analytics challenges, such as IoT analytics accuracy and consistency, urging for serious improvements to cope with such IoT data.
- 3) No IoT-specific data processing architecture is proposed to consider IoT data processing and data analytics challenges that maps the IoT data nature to both the SWE concepts and BDA characteristics.

Therefore, the scope of this study is to present a comprehensive systematic literature review by analyzing the current techniques and technologies used in IoT-based systems from the SWE and BDA perspectives at different

domains. It explores the IoT-based data nature and thoroughly investigates the associated most critical challenges in IoT applications with respect to both SWE concepts and BDA characteristics. Thus, this study demonstrates the mapping between the current state of IoT technologies, SWE development practices and BDA details with respect to the IoT-specific data features for the main IoT-based systems to highlight the present research gaps and the future directions that can be adopted.

The evaluation metrics of this study in order to analyze the presented IoT-based systems include the BDA target, approach, technology and the challenges from the BDA perspective that would reflect processing concerns, as well as the applied evaluation criteria. The metrics also include the applied software (SW) architecture and design, whether model-driven engineering (MDE), separation of concerns, system validation and verification (V&V) are considered and the challenges from the SWE perspective that would affect the SW Quality of Service (QoS). To the best of our knowledge, this study is the first analytical study that addresses BDA architectures in IoT-based systems at several domains then evaluates them from the SWE perspective to highlight the main research gaps in this area.

Thus, the main contributions of this research study are summarized as follows:

- 1) Explore the key features of IoT data by analyzing the main IoT-based systems at different domains.
- 2) Identify the SWE concepts that are concerned with such IoT data features.
- 3) Diagnose the SWE challenges for IoT-based systems that would arise if such SWE concepts were disregarded.
- 4) Determine the big data characteristics concerned with the IoT data features.
- 5) Explore the BDA challenges resulted from the identified IoT data features.
- 6) Propose a domain independent BDA-based IoT processing architecture that handles the highlighted challenges from both the SWE and BDA perspectives.

The rest of the paper is organized as follows. Section II explores the related survey studies that have considered IoT-based systems from different perspectives. Section III discusses the research methodology adopted to conduct this study. Sections IV, V, VI, VII and VIII present our systematic literature review of the main IoT-based systems in the fields of smart environments, and the analytical applications related to humans, networks, energy and environment respectively to evaluate them from both the SWE and BDA perspectives. Section IX investigates the deduced IoT data features and the associated challenges in SWE and BDA with respect to the current IoT-based systems and discusses the main persistent research gaps. In Section X, we propose a recommended SW architecture for big IoT data analytics that addresses most of the highlighted challenges. Finally, section XI concludes our findings and directs future research to the missing gaps.

II. RELATED REVIEWS STUDIES

Previous survey studies have been conducted for data analytics in IoT applications at different areas, i.e. network management, data mining, data fusion, etc. However, the focus of these reviews differs from the prime concern of this review study. This study focuses on investigating the mapping of IoT-specific data features with both SWE and BDA paradigms to explore the possible IoT data processing and data analytics challenges at various IoT domains. The networks protocols were investigated in [17] among different BDA applications for IoT in the smart transportation, E-healthcare, safety and security and smart energy metering domains to contribute the most appropriate protocol per each domain and reveal its limitations for IoT. This study devised a taxonomy to bring forth a generic overview of the IoT paradigm for smart cities, their opportunities and major requirements. Thus, it focused on the use of IoT in the context of smart cities only.

Authors in [18] provided a survey for deep learning techniques only in IoT by investigating the current types of deep learning approaches applied in many IoT applications and evaluating their processing time to formulate the deep learning processing challenges and research gaps in IoT systems.

Another perspective was adopted in [19], in which BDA approaches were grouped per applied domain, specifying the analytical function for each presented approach, such as cleansing, visualization, etc. to highlight the future directions and research gaps with respect to BDA functions for IoT applications. A study in [20] considered a variant perspective that enclosed different types of BDA approaches for IoT systems, as offline analytics, business analytics memory-level, real-time and massive analytics, with their corresponding processing platforms to indicate their research gaps.

Authors in [21] investigated BDA for IoT applications from the context awareness perspective, where each approach was evaluated with respect to its context model and applied data mining technique to devote the context awareness analytical challenges for IoT systems. Data fusion in IoT systems was addressed in [22] by presenting a survey of the applied mathematical models and data fusion analytical approaches that were evaluated from the probabilistic parameters perspective. From the processing type perspective, many BDA techniques were applied on several current scalable processing platforms in [23], such as fog computing and cloud computing to evaluate these platforms' performance and deduce their gaps regarding IoT applications.

Authors in [24] presented a comprehensive survey for the data collection schemes in healthcare. The study analyzed different data collection and secure transmission schemes to provide a suggested reliable and secured data transmission taxonomy with low computational cost and compression ratio. From the data management perspective, authors in [25] reviewed the various solutions for managing IoT data at diverse IoT fields, presenting different types of IoT data and introducing new basic concepts of data management for IoT

in order to deduce the open issues and lessons learnt in data management for IoT.

III. ADOPTED SYSTEMATIC REVIEW METHODOLOGY

This survey adopts the evidence-based systematic review approach [26], as this type of review follows a specific methodology that aims to comprehensively identify all relevant studies on a specific topic and to select the appropriate studies based on explicit criteria. The main research questions driving this study can be formulated as:

- (RQ1) How are IoT data currently being analyzed at the different domains?
- (RQ2) What are the key features encountered for IoT data that would require deep consideration from the SWE and BDA perspectives?
- (RQ3) What are the challenges faced by the IoT-based systems with respect to SWE and BDA?
- (RQ4) What should an appropriate architecture for IoT-based system be composed of in order to efficiently process IoT data and maintain the resultant challenges from both SWE and BDA perspectives?

Relevant papers were identified by searching on Scopus and Google Scholar that index journals and conferences across multiple databases. The inclusion set of search criteria included "big data", "data mining", "big data analytics", "Software engineering" keywords, filtered by "internet of things". The articles were screened manually using the following inclusion criteria: 1) published papers, 2) described the actual designs, implementations and results in enough detail, 3) concerned with analytics and 4) papers after 2012.

The most relevant papers were selected from 13 domains. Figure 2 presents detailed statistics for the selected papers with respect to the IoT domains and the publication year. Previous surveys considering BDA in IoT systems were first investigated to confirm our study's uniqueness and singularity as discussed in section II. RQ1 is addressed in sections IV, V, VI, VII and VIII, evaluating the current state-of-the-art in IoT-based systems from both SWE and BDA perspectives. RQ2 and RQ3 are handled in section IX, exploring the key features of IoT data and their associated challenges from SWE and BDA perspectives, whereas a detailed discussion is conducted in section X to satisfy RQ4 by proposing a recommended SWE model for big IoT data analytics that manipulates the identified challenges.

IV. IoT SYSTEMS FOR SMART ENVIRONMENT

Previous IoT facilitates the surrounding objects to become intelligent, which enables the creation of a smart life [27]. This section presents the main studies of IoT-based approaches that serve the environment to be more comfortable for people. Most of the SW architectures in this domain ignored system V&V, MDE and separation of concerns concepts, which caused serious challenges in system reliability, availability, security, scalability, maintainability and reusability.

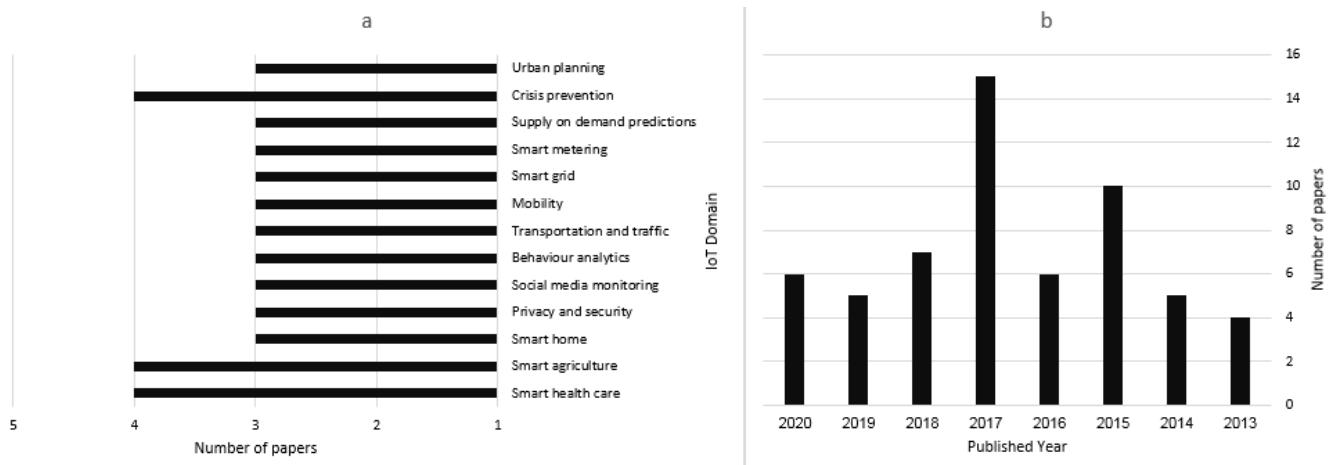


FIGURE 2. The statistics of the considered papers for IoT-based systems, (a) Number of papers vs. the IoT domains, (b) Number of papers vs. the publication year.

A. SMART HEALTH CARE

A new era of medical healthcare systems has emerged based on IoT, in which different analysis types on different SWE architectures were considered to provide various smart services in healthcare [28]. In [29], authors proposed an IoT-based system through a layered architecture for patients' data authentication on the cloud using Ethereum blockchain model, which processed the data generated from sensors attached to a human body. The architecture consisted of four layers; (1) application layer, (2) middleware layer, (3) network layer and (4) perception layer. Their proposed system ignored the massive data amount and data quality that affected the accuracy of the resultant analytics and degraded the huge data analytics support.

Another IoT-based system based on a component-based architecture for diseases detection was proposed based on ontology in [30]. All diseases' details such as symptoms, causes and effects were used to construct an ontology in a centralized processing server that monitored public health using body attached sensors in order to facilitate decision making in case of emergencies. The architecture applied separation of concerns on the detailed diseases data using ontological classes and their actions, i.e., cause of, symptom of, etc. The ontological classes were driven from a medical activity model. The analytical model did not manage the timeliness, valuable big data streams of sensors' readings and diseases details on the distributed environment, which affected the analytics accuracy and caused latency.

In [31], a distributed component-based architecture implemented on a cloud-IoT-based platform was presented for improving health services through optimizing the usage of cloud resources, data storage and retrieval using three different versions of optimization algorithms and inspected on the CloudSim dataset. The architecture aimed to get bests selection of Virtual Machines (VMs) on the cloud, helping patients reduce execution time. It applied separation of concerns by building the architecture in four components:

(1) users' devices, (2) users' requests, (3) cloud broker, (4) network administrator. However, it did not consider the sensors' valuable context data for VMs allocation, which required anywhere spatial analytics operations, affecting the analytics accuracy. Although the proposed architecture targeted cloud resources management, it ignored the limited number of processors when existed.

A Smart Personal Health Advisor (SPHA) was introduced in [32] as another health monitoring system based on a three-layer architecture: (1) sensors infrastructure, (2) network management, (3) deep learning analysis on Hadoop. The system monitored the patient's physiological and physical cases to recognize his medical factors using Principal Component Analysis (PCA) for feature extraction, in which the patient health status was diagnosed using the Convolutional Neural Network (CNN) for deep learning. The system was verified using developed testbeds that collected different categories of personal data. Yet, the presented analytics did not consider maintaining the inconsistent body sensors' data, which implied an inconsistent analytical model.

B. SMART AGRICULTURE

Agriculture is a key worldwide economic demand. The agricultural crops could be damaged for diverse reasons, causing huge losses [33]. Hence, authors in [34] proposed a model for predicting the best crop sequence by analyzing the farmland crops history and the soil information given. The model was built on a layered architecture consisted of four layers, which were (1) IoT data sources, (2) transmission management, (3) data analytics accomplished on Hadoop distributed environment and (4) analytics interpretation. The model used different classification algorithms, like decision tree algorithm (C4.5), SVM, REPTree, as well as association rule mining using FP-Growth to explore soil characteristics needed for classification. Overlooking the management of massive and inconsistent data without reduction methods would cause a

challenge regarding the analytics consistency, as well as a difficulty in huge data analytics support.

In [35], authors introduced an IoT-based real-time monitoring model for smart irrigation and water utilization. It depended on thermal imaging analysis captured by sensors and transferred to cloud server to analyze several factors, such as humidity, light intensity and temperature measurements to avoid unnecessary waste of water. The model was developed on a multi-layered architecture that collected crops sensors data and transferred them via Internet to central data processing services on the cloud. Not considering spatially correlated and heterogeneous data during analysis was the main BDA challenge, which indicated poor analysis anywhere and low accuracy.

Authors in [36] proposed a model for smart irrigation by monitoring water amount quantity and comparing it to other environmental parameters. The model was built on a layered architecture consisted of three layers; (1) Application layer, (2) Network layer management and (3) Perception layer. The model used statistical equations for comparisons. Overlooking the management of massive and inconsistent data without reduction methods would cause a challenge regarding the analytics consistency, as well as a difficulty in huge data analytics support.

In [37], an IoT platform was introduced for pest and disease monitoring. The model predicted pests' occurrences using probabilistic methods over the collected environmental sensors data, such as weather, humidity and temperature. The platform was implemented on a multi-layered architecture consisting of four layers, which were (1) sensor nodes, (2) data acquisition, (3) wireless communication technology and (4) intelligent data processing on the cloud. However, the traded data were not managed during transmission to the cloud server processor to perform analysis, resulting in poor analytics confidentiality.

C. SMART HOME

Use IoT frameworks emphasized the intelligent phenomenon of smart homes, which serve to process home sensors data using home resources and intelligently identify home activities [38]. In [39], authors proposed a real-time IoT-based approach to identify homes' actions using home deployed sensors on a three-layer architecture as follows: (1) different types of sensors to sense different human activities, (2) sensors communication manager and (3) the recognition layer, where various analytics were accomplished on stream processing engines. Different dynamic tests were applied to verify the approach responses to human activities. The approach aimed to analyze the real time sensors' usage readings within home using statistical analytics to predict how healthy is an individual. The prediction was based on the sensors' readings time and order without data reduction operations for massive data and data inconsistencies handling, which affected the support of huge data and analytics consistency in the model.

Another IoT-based approach was proposed in [40] to predict home activities using home deployed sensors and a mixture technique of clustering and classification, as K-patterns followed by an artificial neural network (ANN), on a cloud distributed environment following a client-server model. The K-pattern clustering was used to group related activity models, aiming to extract the ANN features to be learned and predicted from the resultant clustered activity models. The ANN predicted the activity based on time and pattern. The clustering step was used to ensure the prediction model's accuracy but managing data inconsistencies before analysis was not tackled, which heavily reduced the analytics results' consistency and quality.

Another IoT-based approach was introduced in [41] as a home decision support system through a three-tier architecture as follows: (1) client node layer, where home sensors data were acquired (2) mobile edge node layer and (3) fog computing layer, where a deep machine learning (ML) technique using NNs was applied. Yet, accomplishing the complex operations included in the NN without managing the timeliness valuable home credential data affected the analytical model latency. Table 1 summarizes our analysis and evaluation for all presented IoT-based studies in smart environments from both BDA and SWE perspectives.

V. IoT SYSTEMS FOR HUMAN ANALYTICS

IoT links digital information to human world for smarter online lives [42]. This section discusses the main studies that considered human-related data analytics in terms of their processes and information flow between humans and physical objects. Most of the SW architectures proposed in this domain have challenges in system reliability, availability, security, scalability, maintainability and reusability due to missing main SWE concepts, such as system V&V, MDE and separation of concerns.

A. PRIVACY AND SECURITY

The continuous streams of data generated from IoT usually hold important information regarding time, context and user credentials. These data should be fed to several software management and control systems connected to autonomous devices [43]. This section discusses the main approaches for data security and privacy related to the IoT-based systems. A generic BDA privacy model was proposed in [44] as a policies recommender applied in the public health domain, driven out of the requirements engineering offered by the General data Protection Regulation (GDPR) organization. The policy making process is going through three steps: (1) Case recognition, which determines privacy gaps to address by the health policy using regression discontinuity techniques, (2) Action strategy, which sets the procedures considering the available resources by medical experts and (3) Plan monitoring for policy life cycle to evaluate the initiated policy efficiency. The model was implemented on a multi-layered architecture consisting of four layers, which were (1) data ingestion, (2) management layer, (3) data analysis on Hadoop

TABLE 1. IoT frameworks for smart environments from BDA and SWE perspectives: an overview on the associated IoT data features and their encountered challenges.

Ref #	BDA Target	BDA Approach	BDA Technology	Evaluation Criteria	SW Architecture and Design	Applied Model-driven Engineering?	Separation of Concerns	System Validation & Verification	Challenges from BDA Perspective	Challenges from SWE Perspective	Causes from IoT-Data Perspective
29	Authenticate patients data	Ethereum blockchain model	Cloud processing	-	Distributed Multi-Layered Architecture	-	-	Verified by dynamic tests	Massive data support and Analytics Consistency	Reliability, Availability, Security, Scalability, Maintainability and Reusability	Massive and Inconsistent Data
30	Disease detection and remote health monitoring	Ontology Analytics	Centralized Processing	Analytics Accuracy	Component-Based Architecture	Activity Model	Separated Ontology Classes	-	Analytics Accuracy, Latency and Massive data	Reliability, Security, Scalability and Availability	Timeliness, Valuable and Massive Data
31	Optimizing patients request execution time in healthcare	Optimization techniques like Genetic Algorithm	Cloud processing	Processing time	Distributed Component-Based Architecture	-	Conventional View Components	-	Analytics Anywhere and Accuracy	Reliability, Scalability, Reusability, Availability, Security and Maintainability	Spatially Correlated and Valuable Data
32	Remote health monitoring	Feature extraction: PCA classification: CNN	Hadoop & Map Reduce	Processing time	Distributed Multi-Layered Architecture	-	-	Verified by dynamic testbeds	Analytics Consistency	Scalability, Security, Reusability and Maintainability	Massive and Inconsistent Data
34	Agricultural DSS for predicting crop yields	Classification: C4.5, SVM and FP-Growth	Hadoop & Map Reduce	Analytics Accuracy	Distributed Multi-Layered Architecture	-	-	-	Massive data support and Analytics Consistency	Reliability, Scalability, Availability, Security, Maintainability and Reusability	Massive and Inconsistent Data
35	Efficient irrigation	Real-time image processing	Cloud processing	Analytics Accuracy	Distributed Multi-Layered Architecture	-	-	-	Analytics Anywhere and Accuracy	Reliability, Availability, Security, Scalability, Maintainability and Reusability	Heterogenous and Spatially correlated Data
36	Water quantity monitoring for smart irrigation	Statistical Analysis	Cloud processing	Analytics Accuracy	Distributed Multi-Layered Architecture	-	-	-	Massive data support and Analytics Consistency	Reliability, Availability, Security, Scalability, Maintainability and Reusability	Massive and Inconsistent Data
37	Pest prediction	Statistical Analysis	Cloud processing	Analytics Accuracy	Distributed Multi-Layered Architecture	-	-	-	Analytics Confidentiality	Reliability, Availability, Security, Scalability, Maintainability and Reusability	Massive and Traded Data
39	Home activities' recognition	Real-time statistical analysis	-	Analytics & throughput	Distributed Multi-Layered Architecture	-	Dynamic tests to recognize human activity	Massive data and Analytics Consistency	Scalability, Security, Availability, Maintainability and Reusability	Massive and Inconsistent Data	
40	Home activities' recognition	K-pattern Clustering and ANN	Cloud processing	Analytics Accuracy	Distributed Activity Model	-	-	Analytics Consistency	Reliability, Security, Scalability, Availability, Maintainability and Reusability	Inconsistent Data	
41	Decision Support System for home events	Neural Networks	Fog computing	-	Distributed Multi-Layered Architecture	-	-	Analytics Latency	Reliability, Availability, Security, Scalability Maintainability and Reusability	Timeliness and Valuable Data	

distributed environment and (4) data visualization. The model considered the big data processing aspects, but it was a limited security model that considered security concerns related to the public health domain only, ignoring other valuable data that caused a challenge in the analytics accuracy.

Another approach for improving security in IoT-based healthcare systems was proposed in [45] to prevent patients' data attacks using an anomaly detection technique for untrusted parties' and audiences' access identification. The approach was implemented on a cloud distributed environment in a client-server model that used a cryptographic service to reach a higher level of trusted access data security. However, it did not filter noise data, like disconnected access, which downgraded the analysis accuracy, consistency and interpretation.

In [46], authors presented a client-server-based approach for preserving privacy data aggregation using a distributed fog computing environment. The approach employed the hash chain technique to facilitate fog devices to filter untrusted data to prevent external attackers. The main faced obstacle was the real-time processing of the timeliness data and the complex mathematical methods used in the hash chain technique, which affected the analytics latency to respond to any unsecured attack.

B. SOCIAL MEDIA MONITORING

The huge amounts of data provided by smartphones and social networks help identifying individuals' activities and actions. The convergence of IoT and social networks has evolved to present the Social IoT (SIoT), which enables each object to identify services by using its relationships [47]. In this context, authors in [48] proposed a three-layer architectural model (data, network and application layers) to detect telegram data infusion through classifying telegram notifications' speeds using the genetic algorithm. The architecture did not handle spatial data and data timeliness that affected analytics latency and anywhere analytics.

Another generic social IoT-based three-layered architecture for social analytics was proposed in [49] as (1) data producers layer, (2) context data management layer and (3) data consumer layer. Both humans and things equally participated to generate data, in which the Internet allowed the interaction between them and the devices. An intelligent system was included to orchestrate and manage data, but due to the offline computing of timeliness, traded data and the unsecure data transfer resulted in additional challenges in analytics confidentiality and latency.

Furthermore, in [50], authors proposed a system to predict human activities dealing with social networks, in which all objects were formed by interconnected devices through the Internet to integrate IoT and social networks with a cloud server. The system was implemented using a three-layer architecture consisted of (1) objects layer, (2) SIoT server on the cloud and (3) application layer. The used classification techniques were used to predict human activities and were tested on different Twitter datasets to validate the system

throughput, but the heterogenous data was managed before analysis. This implied a poor analytical model to support different data types.

C. BEHAVIOUR ANALYTICS

There are many studies for exploring the interconnection between devices and humans in order to identify human activities. IoT has introduced a new connection theme through human-device, rather than the traditional way of the Internet through human-human [51]. Accordingly, an IoT-based dynamic client-server model was introduced in [52] to predict human location by adapting body sensors relocation. It developed the k-nearest neighbor classifier using trained versions with various sensor contexts without labelled training data on a cloud distributed environment. However, spatial information was not considered in the analysis phase that impacted the analytics accuracy and anywhere analytics.

Another IoT-based client-server approach was introduced in [53] for Human Activity Recognition (HAR) using cloud distributed processing with body deployed sensors to monitor patients with heart diseases. The activity recognition was accomplished using C4.5 decision tree classifier within different situations. Nevertheless, big data veracity was not considered due to the absence of any data pre-processing, which highly affects the analytic model consistency and accuracy.

In [54], a smart sensing IoT-based approach based on a client-server model was proposed for Human Computer Interaction (HCI) applications to act like a human responding to an activity. The approach was applied in a simulated game environment based on imagery analysis on a cloud distributed environment, in which the application responded to the player's actions in real environments. The experiments were run on KTH dataset with real environment images captured using human devices, then 3D Convolutional NNs were applied to identify human actions. The CNN features were driven from an activity model. However, securing user data was not addressed, leading to poor analytics confidentiality. Table 2 presents the detailed evaluation for all discussed IoT-based studies in human analytics.

VI. IoT SYSTEMS FOR NETWORKS ANALYTICS

The internet-connected devices currently surrounding day-to-day activities emit enormous real-time mobile data [55]. Analyzing such data allow people to enhance their mobility and to provide better transportation services as discussed in this section. Nevertheless, most of these works did not handle separation of concerns, MDE approaches and system V&V.

A. TRANSPORTATION AND TRAFFIC

With the exponential increase of populations, the demand for real-time services and smart infrastructures has also increased. For instance, IoT-based systems can provide online traffic guidance using road deployed sensors [56]. Authors in [57] proposed a parallel processing graph-oriented approach for a smart transportation system that used traffic sensors readings and vehicular network to identify

TABLE 2. IoT frameworks for human analytics from BDA and SWE perspectives: an overview on the associated IoT data features and their encountered challenges.

Ref #	BDA Target	BDA Approach	BDA Technology	Evaluation Criteria	SW Architecture and Design	Applied Model-driven Engineering?	Separation of Concerns	System Validation and Verification	Challenges from SWE Perspective		Causes from IoT-Data Perspective
									Challenges from BDA Perspective	Challenges from SWE Perspective	
44	Privacy policies recommender	Situation analytics	Hadoop & Map Reduce	-	Distributed Multi-layered Architecture	Requirements Engineering	-	-	Analytics Accuracy	Reusability, Availability, Reliability, Security, Scalability and Maintainability	Valuable Data
45	Data attacks prevention	Anomaly detection	Cloud processing	-	Distributed Processing on a client server model	-	-	-	Analytics Accuracy, Consistency and Interpretation	Reliability, Availability, Security, Scalability, Reusability and Maintainability	Inconsistent Data
46	Private data aggregation	Hash Chain Technique	Fog computing	Communication overheads	Distributed Processing on a client server model	-	-	-	Analytics Latency	Reliability, Availability, Security, Scalability, Reusability and Maintainability	Data Timeliness
48	Social Data infusion detection	Genetic algorithm	-	Analytics Accuracy	Distributed Multi-layered Architecture	-	-	-	Analytics Anywhere and Latency	Reliability, Availability, Security, Scalability, Reusability and Maintainability	Spatially Correlated and Timeliness
49	Social Data Management	Data management processing	-	-	Distributed Multi-layered Architecture	-	-	-	Analytics Confidentiality and Latency	Reliability, Availability, Security, Scalability, Reusability and Maintainability	Timeliness and Traded Data
50	Human activities prediction	Classification	Cloud processing	Throughput	Distributed Multi-layered Architecture	-	-	-	Testing different Data Models Support	Scalability, Reusability and Maintainability	Heterogenous Data
								categories to validate system throughput			
52	Human location prediction	Classification with K nearest neighbor	Cloud processing	-	Distributed Processing on a client server model	-	-	-	Analytics Anywhere and Accuracy	Reliability, Availability, Security, Scalability, Reusability and Maintainability	Spatially Correlated
53	Human activity recognition	Classification using C4.5 decision tree	Cloud processing	Processing time	Distributed Processing on a client server model	-	-	-	Analytics Consistency and Accuracy	Reliability, Availability, Security, Scalability, Reusability and Maintainability	Inconsistent Data
54	Human computer interaction	Convolutional Neural Networks	Cloud processing	Processing time	Distributed Processing on a client server model	Activity Model	-	-	Analytics Confidentiality	Reliability, Availability, Security, Scalability, Reusability and Maintainability	Traded Data

vehicles' instant location and speed. The proposed approach was implemented on a seven-layer architecture as follows: (1) data sources, (2) communication, (3) graph building, (4) processing on Hadoop, (5) results, (6) interpretation and (7) application. The system analyzed the graphs of sensors' big data using Giraph tool, without considering the huge sensors' data that require special reduction techniques. This impacted the analytics latency and their ability to cope with huge data.

In [58], an IoT-based approach was introduced using fog computing as an online transportation assistant to provide continuous traffic support such as driving assistant, collision detection and hazard prevention. It followed a three-layer architecture consisted of (1) intelligent processing layer on a fog distributed environment, (2) real-time BDA layer and (3) data sources layer. The approach analyzed data in parallel, employing data prioritization methods to ensure the availability to end-users. Nevertheless, it did not handle the heterogenous data formats before processing, which affected the analytical model to deal with variant data types.

Another IoT vehicular data analytic approach on cloud was proposed in [59] to recommend intelligent parking and avoid congestion based on a three-layer architecture with: (1) IoT devices layer, (2) communication layer and (3) data analytics on a cloud distributed environment. The approach collected geographic location information and analyzed parking availability information using Logistic Regression (LR). Many challenges were faced by the analytical model, such as its ability to support huge data, as no data reduction techniques were applied.

B. MOBILITY

Smart objects in IoT-based environments generate massive amounts of location data that allow identifying people capacities and optimizing routes through smart cities [60]. An IoT-based approach was introduced in [61] for mobility analytics using cloud processing deployed on a three-layer architecture as follows: (1) devices layer, (2) network layer and (3) data analytics layer on a cloud distributed processing. It predicted people movements within a building to help directing them in emergencies and validated using automatic tests that generated various locations datasets from ETSIIT dataset at different times. The analysis applied the autoregressive and moving average (ARIMA) times series prediction. However, it did not handle the variant sensors while performing data aggregation, which affected the analysis consistency.

Another approach for location identification services improvement using mobile IoT sensors data recognition was presented in [62]. The approach was developed on a three-layer architecture as follows: (1) physical layer, (2) network layer and (3) utility layer on a cloud environment. The approach applied a utility-driven service model based on sensing requests from the physical layer. Ontology analytics were used to dynamically apply cloud applications without addressing real-time processing of timeliness data, which affected the model latency.

A general service-oriented architecture for context-aware mobile IoT applications was introduced in [63], utilizing cloud resources and IoT technologies as a service. It consisted of a marketplace layer for the user interface, a runtime module and interconnected smart objects. The runtime engine managed the mobile IoT data heterogeneity, aggregation and business analytics. Nevertheless, the analytics faced a challenge in its ability to support huge data, as no data reduction techniques were provided. Table 3 summarizes the evaluation for all presented IoT-based studies concerning network analytics.

VII. IoT SYSTEMS FOR ENERGY ANALYTICS

IoT devices connected to data centers, running on wireless networks allowed their operation in many domains, especially in the energy field [64]. Several challenges were encountered, such as managing supply on demands, monitoring power consumption, energy pricing, energy storage and energy theft prevention, in which many studies have been conducted as discussed herein. From SWE perspective, most of the proposed architectures in this domain face challenges in system reliability, availability, security, scalability, maintainability and reusability as a result of ignoring system V&V, MDE and separation of concerns concepts.

A. SMART GRID

IoT facilitates the communication between the energy measurement components of the smart grid, producing huge data for valuable analytics [65]. In [66], authors proposed an online scalable smart grid management system on a cloud service-oriented architecture to report power usage at different time intervals. The system applied different data-driven forecasting models for big energy data forecasting, like regression tree and ARIMA techniques to predict the energy consumed at different spatial and temporal granularities. The system was tested on the Los-Angeles Smart Grid dataset. Nevertheless, it did not ensure data quality in terms of managing data inconsistencies before the prediction process, which affected both the analytics model's accuracy and consistency.

Another smart grid real-time distributed processing analytic approach based on a client-server model for real time energy pricing was proposed in [67]. It used regression analysis via Hadoop-Spark ML libraries, which managed big data velocity and volume, but it ignored managing big data variety in terms of the heterogeneity of sensors data that affected the model efficiency by not supporting diverse data models, as well as it ignored ensuring grid cyber security that affected system security.

Authors in [68] presented a smart grid management tool using BDA on a cloud implemented on a four-layer architecture as follows: (1) data resource layer, (2) data transmission layer, (3) data storage layer and (4) data analysis layer. The third and fourth layers were implemented on a cloud distributed environment. It planned the energy storage scheme of all devices deployed at a consumer's home based on historical energy consumption data using genetic algorithms. Besides, it presented an energy scheduling scheme based on the Nash

TABLE 3. IoT frameworks for network analytics from BDA and SWE perspectives: an overview on the associated IoT data features and their encountered challenges.

Ref #	BDA Target	BDA Approach	BDA Technology	Evaluation Criteria	SW Architecture and Design	Applied Model-driven Engineering?	Separation of Concerns	System Validation and Verification	Challenges from BDA Perspective	Causes from IoT-Data Perspective
57	Transportation monitoring	Graph-oriented Analysis	Hadoop & Map Reduce	Processing time	Distributed Multi-layered Architecture	-	-	-	Massive data support and Analytics Latency	Reliability, Availability, Security, Scalability, Reusability and Maintainability
58	Transportation routing assistant	Real-time data prioritization methods	Fog computing	-	Distributed Multi-layered Architecture	-	-	-	Different Data Models Support	Reliability, Availability, Security, Scalability, Reusability and Maintainability
59	Smart parking guide	Logistic Regression	Cloud processing	-	Distributed Multi-layered Architecture	-	-	-	Massive data support	Reliability, Availability, Security, Scalability, Reusability and Maintainability
61	People movement prediction	ARIMA	Cloud processing	Processing time	Distributed Multi-layered Architecture	-	-	Validated by automated tests	Analytics Consistency	Scalability, Reusability and Maintainability
62	Location recognition	Ontology analytics	Cloud processing	-	Distributed Multi-layered Architecture	Utility Driven Service Model	-	-	Analytics Latency	Reusability, Availability, Reliability, Security, Scalability and Maintainability
63	General mobile application	context-aware analytics	Cloud processing	-	Distributed Service-Oriented Architecture	-	-	-	Massive data support	Reliability, Availability, Security, Scalability, Reusability and Maintainability

equilibrium game theory for the daily energy usage to reduce energy billing. Yet, using such techniques with complex mathematical operations without considering any data reduction technique for massive data negatively affected the analytic model latency and massive data support.

B. SMART METERING

Smart meters become more intelligent by the power of IoT, which facilitate deploying them at customers' environments to broadcast power consumption data to the service providers at regular interval of time [69]. An approach for real-time smart meter data analytics on Hadoop distributed environment based on a client-server model was presented in [70] to examine the weather impact on energy consumption using the Periodic Auto-Regression (PAR) technique for time series data prediction and tested on a home power consumption dataset. The main contradiction was the short-term power forecast used to cope with real-time processing, ignoring the valuable data insights from long-term forecasting. This affected the regression model training, implying poor analytic model accuracy. Another approach based on a client-server model using Hadoop was presented in [71] to filter smart meters noisy data caused by unpredicted events, or data communication faults. It determines the relationship between the data at different time periods using autoregression and ANN techniques. The approach was validated using rolling and moving cross validation approaches. However, the approach ignored stream data handling for timeliness data, which caused poor analytic model latency.

Authors in [72] proposed an approach for energy theft recognition using smart meter data analytics, implemented on a distributed processing of a centralized server and distributed nodes. It utilized clustering-based feature extraction on the load data to address the imbalanced data, then applied the SVM classifier to detect the theft. Running the approach in parallel supported big data velocity but managing the data volume was still a challenge, as no data reduction was applied.

C. SUPPLY ON DEMAND PREDICTIONS

The fast spreading of IoT deployments prevented energy harvesting for managing supply on demand [73]. In [74], an IoT-based time series analytics model was proposed to analyze and forecast energy consumption patterns to manage supply on demand, implemented on a central server. The analysis was associated with several time intervals ranges from hour to whole year. The proposed model used the unsupervised data clustering K-means to group energy time intervals followed by Bayesian technique to predict energy usage and tested on a domestic appliance level electricity dataset. The main drawbacks were ignoring the valuable data of environmental factors affecting energy production, as well as the reduction of massive data, causing deterioration to both massive data support and analytics accuracy.

An IoT-based smart Energy Management System was presented in [75] for energy consumption monitoring, implemented on a distributed three-layer architecture as follows:

(1) sensors and actuators layer, (2) data management layer and (3) data processing on a central server. The data acquisition module in the second layer is connected to each network device for acquiring and transmitting consumption data to a centralized processing server. The Benchmarking was used for energy data analysis through a generic business intelligence tool that generated user interactive charts and reports for annual power consumptions. The system was validated by validating the incremental systems prototypes. However, the massive and timeliness data and the unsecured valuable energy data transmission were the major challenges that seriously impacted the massive data support of the analytical model with respect to latency and confidentiality.

Another application was presented in [76] to manage energy consumption and balance between supply and actual demand in smart cities using online support vector regression (SVR), applied on a power consumption data that were processed on a central server. The proposed application followed unscalable centralized processing and applied energy short term forecasting using data kept in memory that ensured big data velocity, but it did not consider the SVR cold start and the data volume by reduction methods, which affected the model efficiency to support huge data. Table 4 summarizes the evaluation from the two considered SWE and BDA perspectives for all presented IoT-based studies in energy analytics.

VIII. IoT SYSTEMS FOR ENVIRONMENTAL ANALYTICS

In relation to data analytics, IoT-based systems can ensure human civilization by offering safe surroundings and improving modern lifestyles [77]. Herein, IoT-related studies that addressed different environmental analytics are investigated. However, important SWE concepts, such as separation of concerns and MDE approaches were ignored, which leads to major challenges in system scalability, security, reusability, maintainability, interoperability and availability.

A. CRISIS PREVENTION

One of the IoT blessings is context and timing identification that serve in natural disasters management [78]. Thus, IoT enhances the effectiveness of disaster response and quick decision making. Authors in [79] proposed a real-time monitoring approach to respond intelligently in crisis situations, deployed on a cloud distributed environment with a three-layer architecture as follows: (1) social media data source layer, (2) data crawling layer and (3) data analysis and interpretation layer. It gathered social media information using human language interpreting and analyzed them online using sentiment analysis. However, working on such valuable data without securing them was the main concern in the analytic model confidentiality.

Another approach was introduced in [80] as a decision support in crisis relief situations on central processing architecture as per various forms of both structured and unstructured social data. The approach was validated using both convergent validity and discriminant validity. The analysis

TABLE 4. IoT frameworks for network analytics from BDA and SWE perspectives: an overview on the associated IoT data features and their encountered challenges.

Ref #	BDA Target	BDA Approach	BDA Technology	Evaluation Criteria	SW Architecture and Design	Applied Model-driven Engineering?	Separation of Concerns	System Validation and Verification	Challenges from BDA Perspective	Challenges from SWE Perspective	Causes from IoT-Data Perspective
66	Smart grid power prediction	Regression Tree and ARIMA	Cloud processing	Processing time	Distributed Service-Oriented Architecture	Data-Driven Analytic Models	-	-	Analytics Accuracy and Consistency	Reliability, Availability, Reliability, Security, Scalability and Maintainability	Inconsistent Data
67	Real-Time energy pricing	Real-time regression analysis	Hadoop & Map-Reduce	Processing time	Distributed Processing on a client server model	-	-	-	Different Data Models Support	Reliability, Availability, Security, Scalability, Reusability and Maintainability	Heterogeneous Data
68	Energy storage management	Genetic Algorithms and Nash equilibrium	Cloud processing	-	Distributed Multi-layered Architecture	-	-	-	Massive data support and Analytics Latency	Reliability, Availability, Security, Scalability, Reusability and Maintainability	Massive Data
70	Power consumption predictions	Real-time periodic Auto-Regression	Hadoop & Map-Reduce	Processing time	Distributed Processing on a client server model	-	-	-	Analytics Accuracy	Reliability, Availability, Security, Scalability, Reusability and Maintainability	Valuable Data
71	Smart metering data filtering	autoregression and artificial neural network	Hadoop & Map-Reduce	-	Distributed Processing on a client server model	-	-	Automated cross validation	Analytics Latency	Scalability, Reusability and Maintainability	Timeliness Data
72	Energy theft detection	Data clustering followed by SVM classifier	Centralized server and distributed nodes	-	Distributed Processing on a client server model	-	-	-	Massive data support	Reliability, Availability, Security, Scalability, Reusability and Maintainability	Massive Data
74	Energy consumption forecasting	Clustering followed by Bayesian network prediction	Centralized Processing	Processing time	Centralized Processing	-	-	-	Massive Data Support and Analytics Accuracy	Scalability, Reliability, Security, Reusability, and Maintainability, Interoperability and Availability	Massive and Valuable Data
75	Energy consumption monitoring	Business intelligence software tool	-	-	Distributed Multi-layered Architecture	-	-	Validation using prototypes	Massive Data Support, Analytics Latency and Confidentiality	Scalability, Reusability and Maintainability	Timeliness, Massive and Traded Data
76	Supply on demand prediction	Online support vector regression	Centralized Processing	-	Centralized Processing	-	-	-	Massive Data Support	Scalability, Reliability, Security, Reusability, Interoperability, Availability	Massive Data

determined disaster relief activities using the partial least square regression (PLSR) technique. Although the approach managed big data variety, but it revealed poor massive data support and analytics latency, as no data stream handling or data reduction were applied.

Authors in [81] introduced an approach that aggregated sensors' and users' data to establish spatial risk manager, implemented on a cloud-based three-layer architecture as follows: (1) acquisition layer, (2) integration layer and (3) decision support layer. The manager was developed using statistical analysis on WSN for flood prevention. Yet, the approach did not consider real-time processing for timeliness data, which damaged model latency.

In [82], an integrated real-time processing system combining satellite images and social media data was presented for flood monitoring and hazard detection on a central processing server. The satellite imagery was used for monitoring floods using image processing techniques, while the social media data helped in the forecasting process that applied Bayesian-based relevance ranking strategy. However, managing data inconsistencies before both imagery and social analysis was not addressed, resulting in analytics inconsistency.

B. URBAN PLANNING

IoT infrastructures support construction engineers to better make their land-use plans based on understanding nature rules and human needs [83]. A proposed IoT-based monitoring system based on collected noise maps was presented in [84] to identify noise and support noise prevention rules in important areas, implemented on a client-server architecture over a cloud distributed environment. The system was validated using cross validation on algemesí dataset at different locations. The system used statistical spatial-temporal prediction to detect noise levels, but it faced many challenges regarding spatial data analytics, like using a spatial database for data storage and supporting big data variety, in terms of supporting different data types rather than images.

Another smart IoT-based architecture for managing water consumption in smart cities was introduced in [85] based on a three-layer architecture as follows: (1) IoT data sources, (2) communication management and (3) data analysis on Hadoop. It analyzed rainwater amounts to manage water consumption and prevent floods using Global Active Archive of Large Flood Events dataset. Yet, ignoring massive data reduction was a critical concern, resulted in poor model support of huge data.

Authors in [86] proposed an IoT-based water supply management system implemented on a client-server model over a cloud distributed environment to connect water sources deployed sensors via WSN to massive water consumers. The system optimizes the water consumption process using SCADA and metering analysis without managing data inconsistencies, affecting the model's consistency. Table 5 presents the evaluation summary of all environmental analytics studies.

IX. THE EXPLORATION OF IoT DATA FEATURES CAUSING SWE AND BDA CHALLENGES

Upon investigating all presented studies in sections IV, V, VII, VIII and IX, this section introduces the key features of IoT data that we explored as our first contribution in this study. The explored IoT data features are then mapped to the SWE challenges that would arise, based on the SWE concepts concerned with each IoT data feature as our second contribution. The BDA challenges, resultant from mapping these explored IoT data features to the current big data characteristics, are then investigated as our third contribution.

The three contributions are mapped into the inspected studies and presented in Tables 1, 2, 3, 4 and 5. These tables provide a comprehensive comparative summary for the presented IoT-based studies with respect to the main metrics considered for comparison as highlighted in section I, which included the targeted analysis type, the applied analytical approach and used technology, the considered evaluation criteria to evaluate the associated results, as well as the SWE concepts that are used as evaluation metrics. The tables also include our inferred challenges from both SWE and BDA perspectives, in addition to the causes from the IoT data features perspective. As shown in these tables, ignoring some SWE concepts, like separation of concerns, MDE approaches and system V&V while processing dynamic, valuable and traded IoT data in each domain, caused poor system reliability, security, scalability, availability, interoperability, maintainability and reusability.

Hence, the concepts of SWE and BDA are coupled; SWE seeks to maintain generalizable, powerful and reusable processing architectures for data analytics, while BDA append the smart characteristic by offering effective, continuous, reliable and secured techniques for manipulating the hidden patterns in data. Thus, the main objective of this evaluation is to reveal the trade-off relation between BDA, SWE and IoT, as presented, as well as to disclose the relation between the challenges of SWE and BDA to support the most optimum processing platform for IoT-based systems.

A. IoT DATA FEATURES

The IoT era that interconnects several devices presents a different scheme of data called IoT data. Presenting IoT data nature helps IoT system analysts, SW engineers and developers to understand these data and to improve their traditional approaches to cope with them [87]. Studying various IoT-based systems, IoT data features can be compiled as follows.

- 1) *Multidimensional Massive Data*: IoT connected devices generate gigantic amounts of data, which enforce the scaling of traditional approaches to store, process and analyze them.
- 2) *Data Timeliness*: IoT data are dynamically changing over time, which require real time processing with high speed at regular basis.
- 3) *Heterogeneous Data*: IoT data come from many diversified sources with heterogeneous forms.

TABLE 5. IoT frameworks for environmental analytics from BDA and SWE perspectives: an overview on the associated IoT data features and their encountered challenges.

Ref #	BDA Target	BDA Approach	BDA Technology	Evaluation Criteria	SW Architecture and Design	Applied Model-driven Engineering?	Separation of Concerns	System Validation and Verification	Challenges from BDA Perspective	Causes from IoT-Data Perspective
79	Crisis decision support system	Real-time natural language processing	Cloud processing	-	Distributed Multi-layered Architecture	-	-	-	Analytics Confidentiality	Reliability, Availability, Scalability, Security,
80	Disaster relief activities prediction	Partial least square regression	Centralized Processing	-	Centralized Processing	-	Convergent validity and discriminant validity	Massive Data Support and Analytics Latency	Maintainability, Security, Reusability, Maintainability, Interoperability and Availability, Reliability, Availability, Security,	Massive and Timeliness Data
81	Risk manager	Statistical analysis	Cloud processing	-	Distributed Multi-layered Architecture	-	-	Analytics Latency	Scalability, Reusability and Maintainability, Scalability, Reliability, Security, Reusability, Maintainability, Interoperability and Availability, Reliability, Availability, Security,	Timeliness Data
82	Hazard detection	Real-time image processing and social media analysis	Centralized Processing	-	Centralized Processing	-	-	Analytics Consistency	Scalability, Reusability, Inconsistent Data	
84	Noise abatement recommender	Spatial-temporal prediction	Cloud processing	Processing time	Distributed processing on a client server model	-	Cross validation	Different Data Models Support	Scalability, Reusability and Maintainability, Interoperability and Availability, Scalability, Reusability and Maintainability	Heterogeneous Data
85	Water consumption management	Regression analysis	Hadoop& Map-Reduce	-	Distributed Multi-layered Architecture	-	-	Massive Data Support	Reliability, Availability, Security, Scalability, Reusability and Maintainability, Reliability, Availability, Security,	Massive Data
86	Water supply management	SCADA and metering analysis	Cloud processing	-	Distributed processing on a client server model	-	-	Analytics Consistency	Scalability, Reusability and Maintainability	Inconsistent Data

This represents a complexity concern on such data to be interpreted and analyzed.

- 4) *Inconsistent Data*: The completeness and quality of the collected IoT data may vary, which may include uncertainties, irregularities, or noise.
- 5) *Traded Data*: IoT data are continuously in use and easily accessible from different agents. Thus, ensuring privacy concerns on these data is a challenge.
- 6) *Valuable Data*: IoT sensors are interconnected in different ways autonomously, generating data of high business and social value that should be analyzed accurately.
- 7) *Spatially Correlated*: Excessive different sensors are deployed within the same area to generate diverse data. These data are spatially correlated, making data aggregation processes more complex.

B. SWE CHALLENGES

With more devices being connected, the IoT presents a new data scheme that traditional platforms cannot directly process [88]. Analyzing the considered SWE metrics for the presented studies, we deduce the resultant SWE challenges that would be provoked because of specific IoT data features while disregarding some SWE concepts as shown in Figure 3. As per our analysis, the SWE challenges that have emerged from the IoT perspective include:

- 1) *System Availability*: The IoT-based platform should confront the diversity of fast IoT data for storage, access, processing, analysis and visualization using distributed processing themes. The lack of distributed environments threatens the availability of SW systems, as central processing systems are more likely to fail since they have a single point of failure. Furthermore, any dereliction in SW V&V tests related to system failures detection would cause poor system availability.
- 2) *System Reliability*: The IoT-based platform should be fault tolerant and smoothly operating despite of the untrusted IoT data acquired by unreliable sensors, unclear imagery, or imperfect natural human language using different event handlers. Powerful V&V are responsible for ensuring accepted system performance and detecting system faults. Overlooking enough testing results in unreliable SW systems.
- 3) *System Security*: The lack of control while acquiring, processing and visualizing different and opened access IoT data sources affects the overall system security. Thus, preserving IoT data security at all platform layers should be addressed to ensure system security. Inefficient V&V tests that would examine system resilience at different levels of access put the system security at risk. In addition, the reliance on central processing systems makes the system subject to various attacks. Besides, non-separated SW architectures result in poor system security.
- 4) *System Scalability*: The IoT-based platform should accommodate the massive IoT data, provided the limited hardware resources and used bandwidth, which cause a bottleneck in processing and communication.
- 5) *System Reusability*: Designing reusable systems would support IoT-based applications in order to cope with the rapidly dynamic IoT data timeliness. In addition, cross-cutting concerns would restrain such reusability goal. Thus, neglecting the separation of concerns concept, or any organized architectures, as in multi-layered, or modular architectures, would complicate the extension of functionalities in such systems. Moreover, the abandonment of the MDE development approaches, i.e. Unified Modeling Language (UML) diagrams, as well as domain-specific languages, would make it difficult to consider reusability. According to our inspection, none of the considered studies has mentioned whether they addressed any domain-specific languages.
- 6) *System Maintainability*: Maintaining a SW system for IoT is costly, in which the system's functionalities are forced to cope with different dynamic faulty devices and IoT data timeliness. Ignoring the concept of separation of concerns, MDE development approaches or the usage of domain-specific languages that facilitate offering basic domain-based functionalities, would complicate the maintenance activities of such SW systems.
- 7) *System Interoperability*: The spectrum of IoT includes heterogeneous data sources using various protocols and interfaces for communication and database types, which needs to connect and exchange information without restriction. The processing platform should allow seamless transition and managing connections among divergent platforms. This challenge occurs as a result of the absence of organized designs, such as modular, component-based, or service-oriented architectures, where no defined interfaces are available to communicate with different systems.
- 8) *Functional Suitability*: Developing a SW system that correctly covers all functional requirements and user objectives is a challenge for IoT, because of the heterogeneous and inconsistent data sources with correlated and domain dependent IoT data. Thus, ignoring the consideration of domain-specific languages that facilitate covering domain functional requirements would endanger functional suitability. Unfortunately, domain-specific languages were not addressed in any of the presented studies.

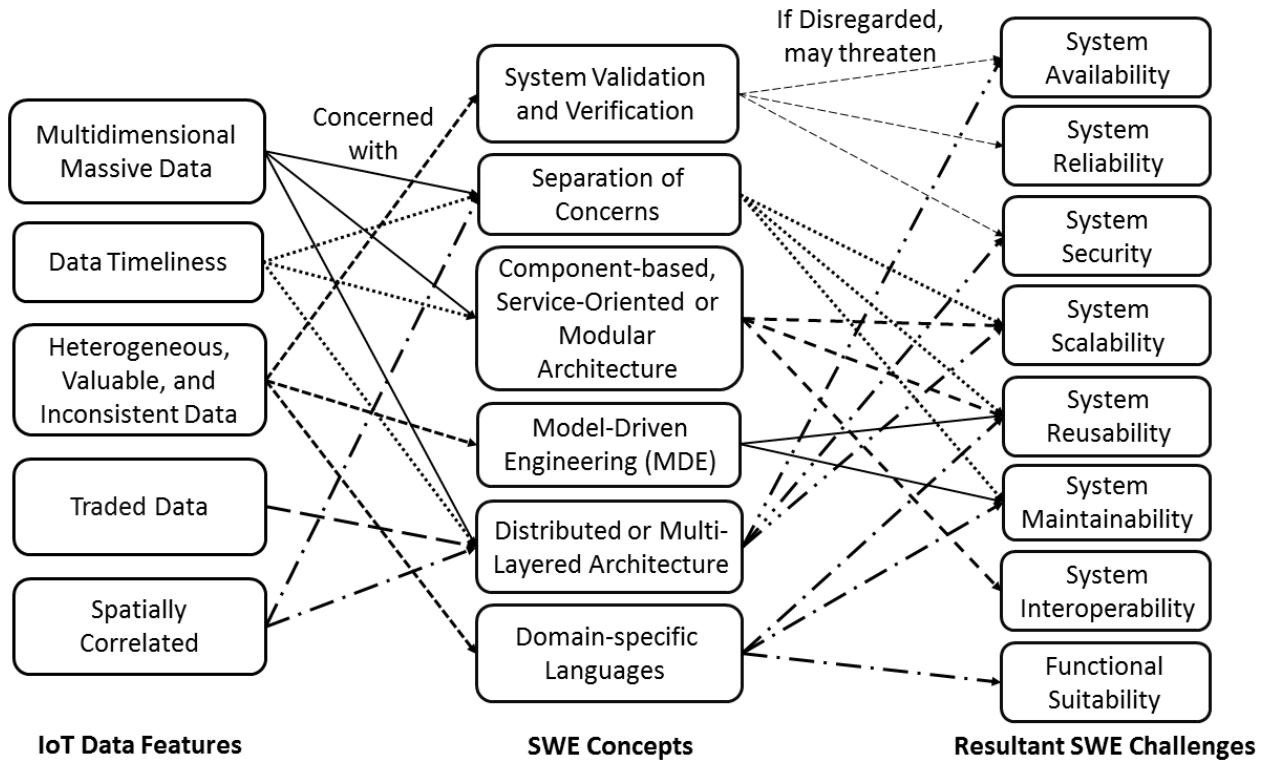


FIGURE 3. The explored SWE challenges resultant from IoT data features.

C. BDA CHALLENGES

The magic of IoT extends the communication from human-machine and machine-human to machine-machine context, where data analytics are addressed in terms of the 10Vs as volume, velocity, variety, veracity, value, variability, volatility, validity, visualization and vulnerability [6], [7], [89]. Therefore, BDA has a great effect on analyzing and interpreting these data to append the intelligent feature of IoT. This section aims to reveal the BDA challenges through investigating the effect of IoT data features on the big data characteristics as shown in Figure 4. Thus, the main challenges that should be considered in such IoT-BDA model include:

- 1) *Massive Data Support*: The high dimensionality reduction of IoT data should be considered in the IoT-BDA model. This challenge appears by ignoring the data volume in the IoT-BDA model. Several data reduction approaches could be applied on the IoT data based on their data nature, such as compression, filtering, sampling, etc.
- 2) *Analytics Latency*: The IoT-BDA model should fit the timeliness nature of IoT data. This challenge occurs by not considering big data velocity and volatility while constructing the IoT-BDA model. IoT-BDA model should force analytics to be accomplished in minimal time by avoiding techniques that would consume much time for training or mathematical operations.
- 3) *Different Data Models Support*: The IoT-BDA model should support heterogenous data, which sometimes are spatially correlated. Neglecting big data variety in IoT-BDA model would cause such challenge. Traditionally, this challenge is addressed by various data

integration forms, which require the perception of sources scheme to define proper data adapters.

- 4) *Analytics Accuracy*: The IoT-BDA model should ensure accurate results despite of the massive, heterogeneous, inconsistent and valuable IoT data nature. Therefore, this challenge occurs when ignoring big data volume, variety, veracity, validity, or vulnerability. Accuracy can be granted using different testing approaches and efficient BDA model training.
- 5) *Analytics Consistency*: The heterogeneous, spatially correlated and inconsistent nature of IoT data due to data noise. Thus, without handling big data variety and veracity in IoT-BDA model would cause this challenge. Therefore, the IoT-BDA model should maintain these data before analysis through several data cleansing techniques that ensure the truthfulness of data and thus, analysis accuracy.
- 6) *Analytics Confidentiality*: The IoT-BDA model should consider the valuable, traded, heterogeneous and timeliness IoT data while processing. These features concern the variety, value, variability and vulnerability characteristics of big data. Ignoring managing these characteristics would cause a serious concern regarding the privacy of IoT data, leading to poor model confidentiality.
- 7) *Analytics Anywhere*: The spatially correlated and timeliness IoT data are concerned with the variability and volatility of big data. Thus, the IoT-BDA model should consider geographic location correlations and time information during analysis.
- 8) *Analytics Interpretation*: The massive, timeliness, heterogeneous and inconsistent IoT data are concerned with

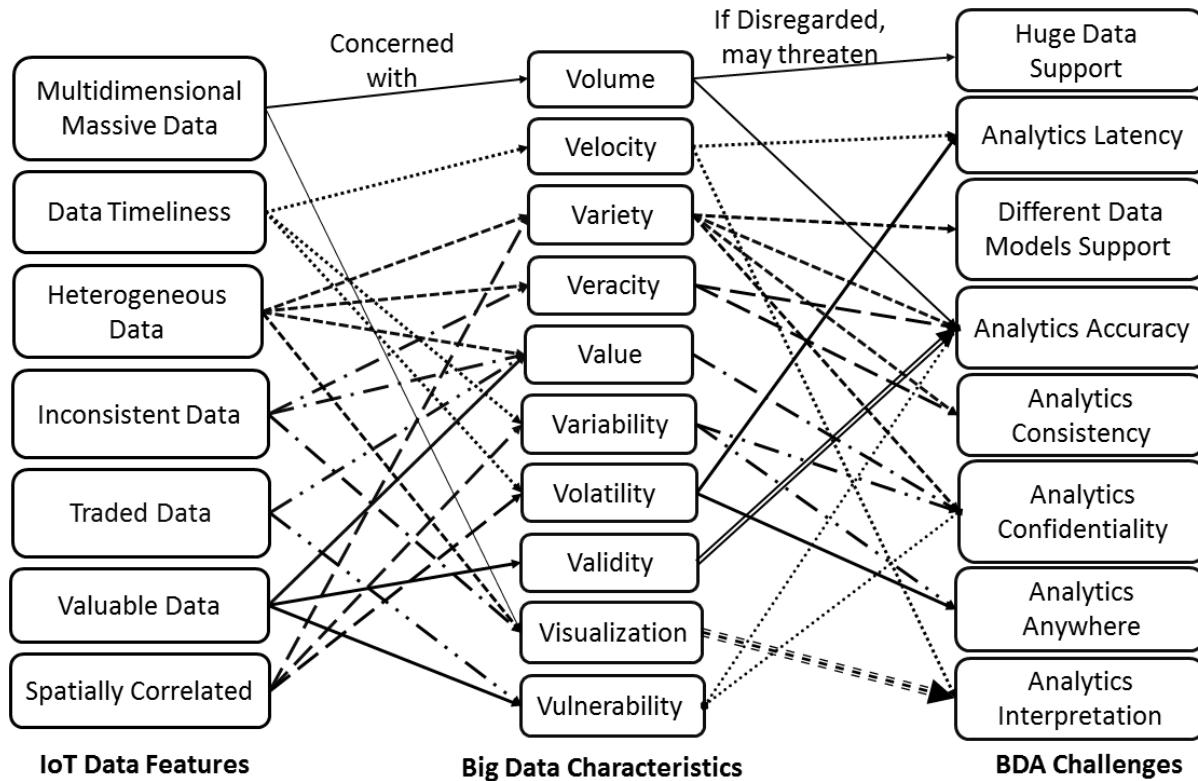


FIGURE 4. The explored BDA challenges resultant from IoT data features.

big data visualization and velocity. Ignoring these characteristics in the IoT-BDA model would cause a critical challenge while presenting the power of analysis, leading to poor analysis interpretation.

Table 6 relates each SWE challenge to its associated BDA challenge(s) that should be considered in IoT processing platforms as per our deduced effects of the IoT data features. It presents the inferred impact of each IoT data feature and the resultant challenges from both SWE and BDA perspectives.

X. THE PROPOSED DOMAIN INDEPENDENT BDA-BASED IoT ARCHITECTURE

Considering the comprehensive analysis for the state of art in the current IoT-based systems provided in sections IV through VIII, we propose a domain-independent BDA-based IoT architecture, as shown in Figure 5. The proposed architecture considers the discussed BDA and SWE challenges resultant from the associated IoT data features in most of the current IoT systems in section IX. To the best of our knowledge, the proposed architecture exclusively introduces a comprehensive solution for online-based data analysis in IoT-based systems that considers both SWE and BDA perspectives.

Avoiding the presented challenges of central processing, the proposed architecture can be deployed using different online computing models, such as edge computing, fog computing, or cloud computing models, in which their scalable, mobile and reliable infrastructure can be tailored as per the business needs [90]. The proposed architecture

follows the main SWE design concepts, including model-driven engineering, separation of concerns, system V&V and domain-specific language. It is composed of six layers: (1) Data Manager, (2) System Resources Controller, (3) System Recovery Manager, (4) BDA Handler, (5) SWE Handler and (6) Security Manager.

The “Business-Logic Software Architecture” layer represents the specific domain application logic and business system requirements needed from the IoT-based application to fulfil, on which our proposed layers are expected to interact. The “Raw Data Providers” represent the application dependent data sources. The architecture separates between IoT data management procedures and business logic functionalities. Regarding SW modularity, each layer consists of several modules to tackle diverse challenges as detailed below.

A. DATA MANAGER

This layer receives the raw data generated from the physical IoT objects sources, like sensors, social media interactions, location-based services, smart devices and feeds the whole architecture. It is responsible for managing the connections and interactions of the various data providers, as well as the continuous acquisition of their data. In order to check the consistency of such acquired raw data before processing, this layer consists of the following modules to ensure QoS:

- 1) *Network and Communication Manager:* This module identifies the physical objects and manages the communication protocols used to interconnect them before data are obtained. Several IoT technologies are used

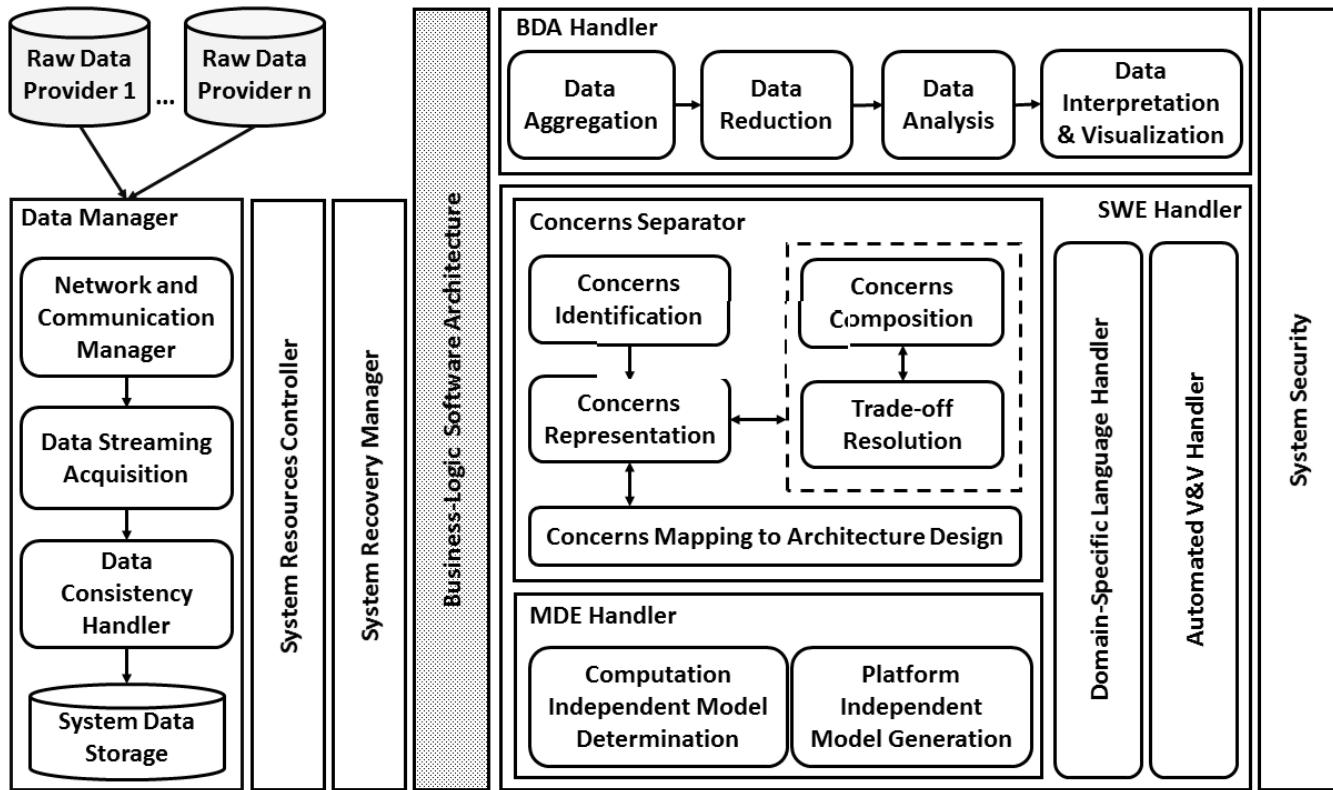


FIGURE 5. The proposed domain-independent software architecture for BDA-based IoT system.

TABLE 6. The IoT data features effect on SWE and BDA.

IoT Data Feature	Resultant SWE Challenge	Resultant BDA Challenge
Multi-dimensional Massive Data	System Scalability and Reusability	Massive Data Support
Data Timeliness	System Availability	Analytics Latency and Interpretation
Heterogeneous Data	System Interoperability, Functional Suitability and Maintainability	Different Data models Support, Analytics Consistency and Analytics Interpretation
Inconsistent Data	System Reliability	Analytics Consistency, Accuracy and Interpretation
Traded Data	System Security	Analytics Confidentiality
Valuable Data	System Security	Analytics Accuracy
Spatially Correlated Data	System Reliability and Interoperability	Analytics Anywhere

depending on the application nature. For instance, wireless body area networks (WBAN) can be used for body deployed sensors in healthcare applications. Network throughput in terms of the successful data transfer rate over time, latency in terms of the delay between sending and processing data units, and availability in terms of the efficient and uprunning network that are ensured by “System Resources Controller” module, are the main metrics that can be considered to evaluate this module [91]. The efficiency of these

evaluation metrics to measure system performance is proven in [92].

- 2) *Data Streaming Acquisition:* A suitable stream processing engine is needed to ensure the real time processing of timeliness IoT data. This module can be evaluated by measuring the response time defined as the time taken between sending a request to the server and the task completion [93]. Response time is the most appropriate evaluation metric to evaluate analytics latency and system availability [94].

- 3) *Data Consistency Handler*: In this module, diverse data manipulation procedures are applied on the inconsistent IoT data, such as noisy data cleansing and filtering, depending on the data type and system purpose. In addition, data transformation techniques can be applied if special data formats are required for processing. Handling such procedures enriches analytics accuracy and consistency, as well as system reliability and availability. A data integration technique can be applied on the maintained data units to evaluate the approaches utilized in this module [95], as a data integration technique can efficiently represent the correlation between raw data and maintained data versions [96].
- 4) *System Data Storage*: The heterogeneity and massive features of IoT data enforce the usage of scalable and different database management systems to store the huge amounts of structured and unstructured data. Several database management systems can be deployed in this module, based on the structure of data units in the IoT domain in hand [96].

B. SYSTEM RESOURCES CONTROLLER

This layer ensures the reliable and scalable processing environment for IoT data. It applies data mining optimization techniques to recommend efficient allocation and distribution policies of the virtual and physical processing resources, in which the processing tasks can cope with the anytime huge data processing, access and storage [97]. This module can be evaluated through monitoring resource utilization, which is the percentage of resources consumed by the incoming workload to present how busy the CPU is [98].

C. SYSTEM RECOVERY MANAGER

System reliability is directly related to providing efficient data replication schemes and appropriate system backups solutions. This layer exploits the power of data analytics by deploying predictive analytics over the acquired data to forecast the growth of big IoT data, compared to the inspections achieved at the “System Resources Controller”, hence determine the storage space needed for future growing data. Recovery time objective, which determines how quickly systems must be recovered, can be used to evaluate the recovery methodologies of this module [99].

D. SWE HANDLER

This layer achieves the main concepts of SWE to ensure the overall system QoS. This layer is composed of the following modules:

- 1) *Concerns Separator*: This module follows the aspect-oriented approach for the separation of concerns that facilitates system scalability, reusability and maintainability. It starts by ‘Concerns Identification’, which recognizes the core and crosscutting concerns of the IoT system by dynamically exposing join points during the execution of the system. Join points include the execution of methods, creation of objects, or throwing of exceptions. “Concerns Composition” and “Trade-off

Resolution” then manage the composition of all related concerns by selecting a specific set of join points that satisfies the concerns criteria based on different logical operators. The inconsistencies resulted from this composition process are then eliminated by applying some filtering techniques. “Concerns Representation” represents the resultant concerns by constructing their equivalent connected graph, whereas “Concerns Mapping to Architecture Design” maps it further to the software architecture design. This module can be validated using correlation matrices between the concerns and the generated classes [100].

- 2) *MDE Handler*: This module ensures the overall system reusability and maintainability using model driven engineering approaches that provide a direct process of design, construction and maintenance of software by determining the Computation Independent Model (CIM) and Platform Independent Model (PIM) of the IoT system. The CIM maps the functional requirements to the domain requirements by analyzing the domain feature model of the system, expressing the reusable elements in that specific domain. The functional requirements and domain features are analyzed by constructing UML diagrams and a dictionary of services and techniques used for implementation respectively. Thus, CIM is developed by analyzing the user requirements and finding a matching set of features from the domain features dictionary. The PIM is then generated from the CIM using the domain patterns, which represent the reusable structures of the common problems across domains. Design pattern techniques are utilized to confirm that all functionalities in the CIM are reflected in the PIM. Several graphical code inspections can be conducted to evaluate this module [101]. System defects are best found by code inspections, as they reveal functionalities’ defects of a service.
- 3) *Automated V&V Handler*: System reliability, SW functional suitability and availability are ensured via this module. It presents an IoT-based validation and verification model that dynamically generates test cases from UML models and establishes a testing simulation environment to test the system’s ability to be integrated, configured, installed, executed, upgraded, scaled-up, or scaled-out, saving costs and risks of testing the real system [102]. Further, it could support self-monitoring architectural models for high fault tolerance. Tracking testing progress technique can be applied to validate the testing process and ensure tests completeness [103].
- 4) *Domain-Specific Language Handler*: As per our study, the SW functional suitability has become one of the major SWE challenges that has not been addressed yet in IoT-based systems to the best of our knowledge. This module ensures functional suitability by handling domain-specific mini-languages built on the

top of the hosting language to provide concepts and behaviors in a specific context. This presents an interface for diverse service types in a specific domain, and facilitates converting, transforming and combining data from multiple services to meet the operations needs and satisfy both business and technical requirements. Different user interfaces validation techniques from various domains can be considered to evaluate this module, which can automatically validate knowledge from heterogeneous systems [104].

E. BDA HANDLER

In this layer, data analytics are performed to maintain the big data characteristics of the collected IoT data per each of the following modules:

- 1) *Data Aggregation*: Data aggregation techniques are needed to ensure data fusion from multiple and heterogeneous IoT data sources into one data repository. This module presents a data aggregation model that supports different network protocols for different data models with a minimum time between data generation and aggregation. This supports analytics anywhere, optimum latency and accurate analytics. Confirming data features stability before and after aggregation can be used to validate this module [105].
- 2) *Data Reduction*: This module provides different approaches to support scalability despite of the huge data volume and to minimize their processing time. Data reduction techniques are applied to reduce the multidimensional massive IoT data. Data could be reduced in (1) amount, using dataset sampling or filtering, or (2) dimension, using dimensionality reduction techniques that maintain data correlations or both. This module can be evaluated via the system latency and response time before and after data reduction, in addition to the applied reduction technique accuracy metric [106].
- 3) *Data Analysis*: Upon maintaining IoT data features and considering IoT data processing challenges, this module performs the required business analytics depending on the application purpose. The Analytics accuracy and latency are managed through several accuracy metrics and multi-threading libraries [92].
- 4) *Data Interpretation & Visualization*: Information become easily interpreted when effective visualization techniques are utilized. This module supports different representation forms for the analytical results, based on the application type. For instance, charts are best suitable for numeric results, maps for spatial outcomes, ... etc. [107].

F. SYSTEM SECURITY

As secured processing is a main challenge for IoT data, this layer intersects with all layers of the architecture to ensure system security. Its responsibilities vary among the layers. For instance, it identifies attacks and authenticates

external accesses and data storage access all over the system layers, whereas at the BDA Handler and SWE Handler layers, it assigns security rules for the system users to accomplish each task. At the Data Manager layer, it validates data providers based on legal licenses of data usage and networks information encryption. This module can be evaluated by initiating several attacks from variant networks and different profiles to ensure system security and privacy [108].

G. IMPLEMENTATION CHALLENGES

This section highlights the challenges, as well as the research and development (R&D) efforts that are expected to draw the attention of the research community regarding the implementation of the proposed domain-independent BDA-based IoT architecture. Each module in the proposed architecture is subject to face the following challenges:

- 1) *Data Manager*: Supporting heterogeneous data is the main processing concern that would endure this module to cope with the different network protocols, ensuring data consistency and data storage.
- 2) *System Resources Controller*: The massive amount of data is a serious challenge to manage the system resources.
- 3) *System Recovery Manager*: Appropriate innovative data replication schemes and backup solutions should be adopted. Thus, initiating such solutions with real-time data is a major challenge.
- 4) *SWE Handler*: Considering data heterogeneity, inconsistency and freshness are the main processing challenges to maintain SWE concepts, such as V&V, MDE and separation of concerns.
- 5) *BDA Handler*: BDA accuracy is directly affected by the different IoT data features, such as the massive data amounts, data heterogeneity and inconsistency, which negatively impacts on IoT data analytics.
- 6) *System Security*: The massive amount of open-accessed data introduces a real challenge to secure IoT-based systems.

XI. DISCUSSION AND CONCLUSION

Nowadays, the Internet of Things (IoT) phenomenon produces a complicated type of data characterized by their huge amount, high speed, unstructured format, valuable and credential meaning, inconsistencies and spatial correlations. Thus, the emergence of IoT systems brings new challenges from both the big data analytics (BDA) and software engineering (SWE) perspectives to process and analyze such data. Despite of conducting many surveys for IoT-based systems from different perspectives, but the SWE concepts and BDA characteristics for such systems are still ambiguous and overlooked with respect to the IoT-specific data features. This raises different challenges and concerns for the data processing and analytics at these IoT systems. Yet, there is no adapted architecture or taxonomy dedicated to process and analyze IoT data, irrespective of the IoT domain.

To the best of our knowledge, this paper exclusively reveals the trade-off relation between BDA, SWE and IoT,

and discloses the relation between the challenges of SWE and BDA to support an optimum processing platform for IoT-based systems. It presents a systematic literature review of IoT-based studies at different domains, which are smart environments, human-related analytics and network-related analytics, energy-related analytics and environmental-related analytics. The study investigates the challenges of these systems from the BDA perspective based on the current big data 10 Vs, as well as the processing difficulties from the SWE perspective of the main IoT systems, highlighting many research opportunities for further consideration to provide secure, available, reliable, scalable, interoperable, maintainable, reusable and functional suitable real-time data processing for IoT data.

In addition, the study inspects that these SWE challenges have emerged due to neglecting specific essential SWE concepts while processing IoT data. The main abandoned SWE concepts in most of the IoT-based systems at various domains can be inferred to be domain-specific languages for functional suitability, separation of concerns for maintainability, model-driven engineering (MDE) approaches for reusability and systematic validation and verification for reliability. Regarding the IoT data challenges with respect to BDA, they are implied to include the huge volume of different data models support, as well as the analytics latency, accuracy, interpretation, consistency and confidentiality, addressing the current 10 Vs of big data. A matrix that presents a concrete overview on the specific IoT data features and their encountered challenges and gaps for each IoT domain is provided.

Accordingly, we propose a domain-independent software architecture for big IoT data analytics. The proposed architecture exclusively introduces a comprehensive solution for online-based data analysis in IoT-based systems that maintains the investigated challenges from both SWE and BDA perspectives through six main layers interacting with the business-logic layer of the IoT application and its raw data providers. A discussion is further presented to highlight the main challenges and R&D efforts that would face the research community at the implementation of the proposed architecture.

Hence, the contribution of this study can be summarized as follows: (1) deducing key IoT data features, (2) identifying the SWE concepts and big data characteristics concerned with such IoT data features, (3) inferring the resultant BDA and SWE challenges from the IoT data features, and (4) proposing a domain independent BDA-based IoT processing architecture that handles the highlighted challenges from both the SWE and BDA perspectives, as well as drawing the attention to its implementation challenges and the associated implications.

Our future research efforts focus on investigating the optimum security and privacy preservation approaches that would fit the deduced IoT data features, while maintaining system performance and timely results. In addition, future considerations are intended for better resources utilization and power

saving approaches that ideally suit IoT-based systems in various domains.

APPENDIX

Abbreviation	Explanation
ANN	Artificial Neural Network.
BDA	Big Data Analytics.
CNN	Convolutional Neural Network.
HAR	Human Activity Recognition.
HCI	Human Computer Interaction.
IoT	Internet of Things.
LR	Logistic Regression.
MDE	Model-Driven Engineering.
ML	Machine Learning.
NNs	Neural Networks.
PAR	Periodic Auto-Regression.
PCA	Principal Component Analysis.
PLSR	Partial Least Square Regression.
QoS	Quality of Service.
R&D	Research and Development.
SIoT	Social IoT.
SVM	Support Vector Machine.
SVR	Support Vector Regression.
SW	Software.
SWE	Software Engineering.
UML	Unified Modeling Language.
V&V	Validation and Verification.
VMs	Virtual Machines.
WBAN	Wireless Body Area Networks.
WSN	Wireless Sensor Network.

REFERENCES

- [1] D. Fawzy, S. Moussa, and N. Badr, "The spatiotemporal data fusion (STDF) approach: IoT-based data fusion using big data analytics," *Sensors*, vol. 21, no. 21, p. 7035, Oct. 2021, doi: [10.3390/s21217035](https://doi.org/10.3390/s21217035).
- [2] D. F. Mahmoud, S. M. Moussa, and N. L. Badr, "The spatio-temporal data reduction (STDR): An adaptive IoT-based data reduction approach," in *Proc. IEEE 10th Int. Conf. Intell. Comput. Inf. Syst. (ICICIS)*, Dec. 2021, pp. 355–360.
- [3] A. Nauman, Y. A. Qadri, M. Amjad, Y. B. Zikria, M. K. Afzal, and S. W. Kim, "Multimedia Internet of Things: A comprehensive survey," *IEEE Access*, vol. 8, pp. 8202–8250, 2020.
- [4] M. Stoyanova, Y. Nikoloudakis, S. Panagiotakis, E. Pallis, and E. K. Markakis, "A survey on the Internet of Things (IoT) forensics: Challenges, approaches, and open issues," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 1191–1221, 2nd Quart., 2020.
- [5] B. Diène, J. J. P. C. Rodrigues, O. Diallo, E. H. M. Ndoye, and V. V. Korotaev, "Data management techniques for Internet of Things," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106564.
- [6] D. Fawzy, S. Moussa, and N. Badr, "The evolution of data mining techniques to big data analytics: An extensive study with application to renewable energy data analytics," *Asian J. Appl. Sci.*, vol. 4, no. 3, pp. 756–766, 2016.
- [7] S. Saide and M. L. Sheng, "Toward business process innovation in the big data era: A mediating role of big data knowledge management," *Big Data*, vol. 8, no. 6, pp. 464–477, Dec. 2020.
- [8] D. Fawzy, S. Moussa, and N. Badr, "Trio-V wind analyzer: A generic integral system for wind farm suitability design and power prediction using big data analytics," *J. Energy Resour. Technol.*, vol. 140, no. 5, May 2018, Art. no. 051202.
- [9] D. Fawzy, S. Moussa, and N. Badr, "An approach for erosion and power loss prediction of wind turbines using big data analytics," in *Proc. 5th Int. Workshop Data Anal. Renew. Energy Integr.*, 2017, pp. 33–46.

- [10] D. Fawzy, S. Moussa, and N. Badr, "WFEC: Wind farms economic classifier using big data analytics," in *Proc. 8th Int. Conf. Intell. Comput. Inf. Syst. (ICICIS)*, Dec. 2017, pp. 154–159.
- [11] X. Larrucea, A. Combelles, J. Favaro, and K. Taneja, "Software engineering for the Internet of Things," *IEEE Softw.*, vol. 34, no. 1, pp. 24–28, Jan./Feb. 2017.
- [12] A. Bröring, S. Schmid, C.-K. Schindhelm, A. Khelil, S. Käbisch, D. Kramer, D. L. Phuoc, J. Mitic, D. Anicic, and E. Teniente, "Enabling IoT ecosystems through platform interoperability," *IEEE Softw.*, vol. 34, no. 1, pp. 54–61, Jan./Feb. 2017.
- [13] B. Morin, N. Harrand, and F. Fleurey, "Model-based software engineering to tame the IoT jungle," *IEEE Softw.*, vol. 34, no. 1, pp. 30–36, Jan. 2017.
- [14] A. Taivalsaari and T. Mikkonen, "A roadmap to the programmable world: Software challenges in the IoT era," *IEEE Softw.*, vol. 34, no. 1, pp. 72–80, Jan. 2017.
- [15] M. Strohbach, H. Ziekow, V. Gazis, and N. Akiva, "Towards a big data analytics framework for IoT and smart city applications," in *Modeling and Processing for Next-Generation Big-Data Technologies (Modeling and Optimization in Science and Technologies)*, vol. 4. Cham, Switzerland: Springer, 2015, pp. 257–282.
- [16] D. P. Acharya and K. Ahmed, "A survey on big data analytics: Challenges, open research issues and tools," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 2, pp. 511–518, 2016.
- [17] Y. Mehmood, F. Ahmad, I. Yaqoob, A. Adnane, M. Imran, and S. Guizani, "Internet-of-Things-based smart cities: Recent advances and challenges," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 16–24, Sep. 2017.
- [18] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, "Deep learning for IoT big data and streaming analytics: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2923–2960, 4th Quart., 2018.
- [19] M. Z. Ge, H. Bangui, and B. Buhnova, "Big data for Internet of Things: A survey," *Future Gener. Comput. Syst.*, vol. 87, pp. 601–614, Oct. 2018.
- [20] M. Marjani, F. Nasaruddin, A. Gani, A. Karim, I. A. T. Hashem, A. Siddiqi, and I. Yaqoob, "Big IoT data analytics: Architecture, opportunities, and open research challenges," *IEEE Access*, vol. 5, pp. 5247–5261, 2017.
- [21] O. B. Sezer, E. Dogdu, and A. M. Ozbayoglu, "Context-aware computing, learning, and big data in Internet of Things: A survey," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 1–27, Feb. 2018.
- [22] F. Alam, R. Mehmood, I. Katib, N. N. Albogami, and A. Albeshri, "Data fusion and IoT for smart ubiquitous environments: A survey," *IEEE Access*, vol. 5, pp. 9533–9554, 2017.
- [23] E. Ahmed, I. Yaqoob, I. A. T. Hashem, I. Khan, A. I. A. Ahmed, M. Imran, and A. V. Vasilakos, "The role of big data analytics in Internet of Things," *Comput. Netw.*, vol. 129, pp. 459–471, Dec. 2017.
- [24] A. Ullah, M. Azeem, H. Ashraf, A. A. Alaboudi, M. Humayun, and N. Jhanji, "Secure healthcare data aggregation and transmission in IoT—A survey," *IEEE Access*, vol. 9, pp. 16849–16865, 2021.
- [25] B. Diéne, J. J. P. C. Rodrigues, O. Diallo, E. H. M. Ndoye, and V. V. Korotaev, "Data management techniques for Internet of Things," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106564.
- [26] K. S. Khan, R. Kunz, J. Kleijnen, and G. Antes, "Five steps to conducting a systematic review," *J. Roy. Soc. Med.*, vol. 96, no. 3, pp. 118–121, 2003.
- [27] W. Yu, F. Liang, X. He, W. G. Hatcher, C. Lu, J. Lin, and X. Yang, "A survey on the edge computing for the Internet of Things," *IEEE Access*, vol. 6, pp. 6900–6919, 2017.
- [28] Y. Wang, Y. Zhao, W. Dang, J. Zheng, and H. Dong, "The evolution of publication hotspots in electronic health records from 1957 to 2016 and differences among six countries," *Big Data*, vol. 8, no. 2, pp. 89–106, Apr. 2020.
- [29] E. M. Abou-Nassar, A. M. Iliyasu, P. M. El-Kafrawy, O.-Y. Song, A. K. Bashir, and A. A. A. El-Latif, "DITrust chain: Towards blockchain-based trust models for sustainable healthcare IoT systems," *IEEE Access*, vol. 8, pp. 111223–111238, 2020.
- [30] V. Kumar, "Ontology based public healthcare system in internet of things (IoT)," *Procedia Comput. Sci.*, vol. 50, pp. 99–102, 2015.
- [31] M. Elhoseny, A. Abdelaziz, A. S. Salama, A. M. Riad, K. Muhammad, and A. K. Sangaiah, "A hybrid model of Internet of Things and cloud computing to manage big data in health services applications," *Future Gener. Comput. Syst.*, vol. 86, pp. 1383–1394, Sep. 2018.
- [32] M. Chen, Y. Zhang, M. Qiu, N. Guizani, and Y. Hao, "SPHA: Smart personal health advisor based on deep analytics," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 164–169, Mar. 2018.
- [33] O. Elijah, T. A. Rahman, I. Orikumhi, C. Y. Leow, and M. H. D. N. Hindia, "An overview of Internet of Things (IoT) and data analytics in agriculture: Benefits and challenges," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3758–3773, Oct. 2018.
- [34] S. Rajeswari, K. Suthendran, and K. Rajakumar, "A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics," in *Proc. Int. Conf. Intell. Comput. Control (IC2C)*, Jun. 2017, pp. 1–5.
- [35] M. Roopaei, P. Rad, and K.-K. R. Choo, "Cloud of things in smart agriculture: Intelligent irrigation monitoring by thermal imaging," *IEEE Cloud Comput.*, vol. 4, no. 1, pp. 10–15, Jan./Feb. 2017.
- [36] L. García, L. Parra, J. M. Jimenez, J. Lloret, and P. Lorenz, "IoT-based smart irrigation systems: An overview on the recent trends on sensors and IoT systems for irrigation in precision agriculture," *Sensors*, vol. 20, no. 4, p. 1042, Feb. 2020.
- [37] Y. Shi, Z. Wang, X. Wang, and S. Zhang, "Internet of Things application to monitoring plant disease and insect pests," in *Proc. Int. Conf. Appl. Sci. Eng. Innov.*, 2015, pp. 31–34, doi: 10.2991/asei-15.2015.7.
- [38] B. L. R. Stojkoska and K. V. Trivodaliev, "A review of Internet of Things for smart home: Challenges and solutions," *J. Cleaner Prod.*, vol. 140, no. 3, pp. 1454–1464, Jan. 2017.
- [39] H. Ghayat, S. Mukhopadhyay, X. Gui, and N. Suryadevara, "WSN- and IoT-based smart homes and their extension to smart buildings," *Sensors*, vol. 15, no. 5, pp. 10350–10379, May 2015.
- [40] S. Thomas, M. Bourbou, and Y. Yoo, "User activity recognition in smart homes using pattern clustering applied to temporal ANN algorithm," *Sensors*, vol. 15, no. 5, pp. 11953–11971, May 2015.
- [41] M. A. Rahman, M. M. Rashid, M. S. Hossain, E. Hassanain, M. F. Alhamid, and M. Guizani, "Blockchain and IoT-based cognitive edge framework for sharing economy services in a smart city," *IEEE Access*, vol. 7, pp. 18611–18621, 2019.
- [42] R. K. Chahal, N. Kumar, and S. Batra, "Trust management in social Internet of Things: A taxonomy, open issues, and challenges," *Comput. Commun.*, vol. 150, pp. 13–46, Jan. 2020.
- [43] F. Li, H. Li, C. Wang, K. Ren, and E. Bertino, "Guest editorial special issue on security and privacy protection for big data and IoT," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1446–1449, Apr. 2019.
- [44] M. Anisetti, C. Ardagna, V. Bellandi, M. Cremonini, F. Frati, and E. Damiani, "Privacy-aware big data analytics as a service for public health policies in smart cities," *Sustain. Cities Soc.*, vol. 39, pp. 68–77, May 2018.
- [45] S. Sharma, K. Chen, and A. Sheth, "Toward practical privacy-preserving analytics for IoT and cloud-based healthcare systems," *IEEE Internet Comput.*, vol. 22, no. 2, pp. 42–51, Mar. 2018.
- [46] R. Lu, K. Heung, A. H. Lashkari, and A. A. A. Ghorbani, "A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT," *IEEE Access*, vol. 5, pp. 3302–3312, 2017.
- [47] B. AlBanna, M. Sakr, S. Moussa, and I. Moawad, "Interest aware location-based recommender system using geo-tagged social media," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 12, p. 245, Dec. 2016, doi: 10.3390/ijgi5120245.
- [48] R. A. Candra, D. S. Saputra, D. N. Ilham, H. Setiawan, and H. Hardisal, "The infusion of notification design with an application of social media based on a Internet of Things (IoT)," *Sinkron, Jurnal Penelitian Teknik Informatika*, vol. 5, no. 1, pp. 129–137, 2020.
- [49] A. M. Ortiz, D. Hussein, S. Park, S. N. Han, and N. Crespi, "The cluster between Internet of Things and social networks: Review and research challenges," *IEEE Internet Things J.*, vol. 1, no. 3, pp. 206–215, Jun. 2014.
- [50] A. Ahmad, M. M. Rathore, A. Paul, and S. Rho, "Defining human behaviors using big data analytics in social Internet of Things," in *Proc. IEEE 30th Int. Conf. Adv. Inf. Netw. Appl. (AINA)*, Mar. 2016, pp. 1101–1107.
- [51] B. Guo, D. Zhang, Z. Wang, Z. Yu, and X. Zhou, "Opportunistic IoT: Exploring the harmonious interaction between human and the Internet of Things," *J. Netw. Comput. Appl.*, vol. 36, no. 6, pp. 1531–1539, Nov. 2013.
- [52] S. A. Rokni and H. Ghasemzadeh, "Autonomous sensor-context learning in dynamic human-centered Internet-of-Things environments," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design*, Nov. 2016, pp. 1–6.
- [53] D. Castro, W. Coral, C. Rodriguez, J. Cabra, and J. Colorado, "Wearable-based human activity recognition using an IoT approach," *J. Sens. Actuator Netw.*, vol. 6, no. 4, pp. 1–28, 2017.
- [54] A. G. D'Sa and B. G. Prasad, "An IoT based framework for activity recognition using deep learning technique," 2019, arXiv:1906.07247.

- [55] K. Yang, S. Liu, L. Cai, Y. Yilmaz, P.-Y. Chen, and A. Walid, "Guest editorial special issue on AI enabled cognitive communication and networking for IoT," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1906–1910, Apr. 2019.
- [56] Y. Meidan, M. Bohadana, A. Shabtai, J. D. Guarnizo, M. Ochoa, N. O. Tippenhauer, and Y. Elovici, "ProfiloIoT: A machine learning approach for IoT device identification based on network traffic analysis," in *Proc. ACM Symp.*, 2017, pp. 506–509.
- [57] M. M. Rathore, A. Ahmad, A. Paul, and G. Jeon, "Efficient graph-oriented smart transportation using Internet of Things generated big data," in *Proc. 11th Int. Conf. Signal-Image Technol. Internet-Based Syst. (SITIS)*, Nov. 2015, pp. 512–519.
- [58] T. S. J. Darwish and K. A. Bakar, "Fog based intelligent transportation big data analytics in the internet of vehicles environment: Motivations, architecture, challenges, and critical issues," *IEEE Access*, vol. 6, pp. 15679–15701, 2018.
- [59] W. He, G. Yan, and L. D. Xu, "Developing vehicular data cloud services in the IoT environment," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 1587–1595, May 2014.
- [60] T. Maksymyuk, S. Dumych, M. Brych, D. Satria, and M. Jo, "An IoT based monitoring framework for software defined 5G mobile networks," in *Proc. 11th Int. Conf. Ubiquitous Inf. Manage. Commun. (IMCOM)*, Jan. 2017, pp. 1–4.
- [61] A. Fernández-Ares, A. M. Mora, M. G. Arenas, P. García-Sánchez, G. Romero, V. Rivas, P. A. Castillo, and J. J. Merelo, "Studying real traffic and mobility scenarios for a smart city using a new monitoring and tracking system," *Futur. Gener. Comput. Syst.*, vol. 76, pp. 163–179, Nov. 2017.
- [62] I. P. Žarko, K. Pripužić, M. Serrano, and M. Hauswirth, "IoT data management methods and optimisation algorithms for mobile publish/subscribe services in cloud environments," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun. 2014, pp. 1–5.
- [63] C. Doukas and F. Antonelli, "COMPOSE: Building smart & context-aware mobile applications utilizing IoT technologies," in *Proc. Global Inf. Infrastruct. Symp.*, Oct. 2013, pp. 1–6.
- [64] G. Escobedo, N. Jacome, and G. Arroyo-Figueroa, "Big data & analytics to support the renewable energy integration of smart grids—Case study: Power solar generation," in *Proc. 2nd Int. Conf. Internet Things, Big Data Secur. (IoTBDS)*, 2017, pp. 267–275.
- [65] N. H. Motlagh, M. Mohammadrezaei, J. Hunt, and B. Zakeri, "Internet of Things (IoT) and the energy sector," *Energies*, vol. 13, no. 2, p. 494, 2020.
- [66] Y. Simmhan, S. Aman, A. Kumbhare, R. Liu, S. Stevens, Q. Zhou, and V. Prasanna, "Cloud-based software platform for big data analytics in smart grids," *Comput. Sci. Eng.*, vol. 15, no. 4, pp. 38–47, 2013, doi: [10.1109/MCSE.2013.39](https://doi.org/10.1109/MCSE.2013.39).
- [67] R. Shyam, H. B. B. Ganesh, S. S. Kumar, P. Poornachandran, and K. P. Soman, "Apache spark a big data analytics platform for smart grid," *Proc. Technol.*, vol. 21, pp. 171–178, Jan. 2015.
- [68] K. Wang, W. Yunqi, X. Hu, Y. Sun, D. J. Deng, A. Vinel, and Y. Zhang, "Wireless big data computing in smart grid," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 58–64, Apr. 2017.
- [69] O. Ardkanian, N. Koochakzadeh, R. P. Singh, L. Golab, and S. Keshav, "Computing electricity consumption profiles from household smart meter data," in *Proc. EDBT/ICDT Workshops*, vol. 14, 2014, pp. 140–147.
- [70] X. Liu, L. Golab, W. Golab, and I. F. Ilyas, "Benchmarking smart meter data analytics," in *Proc. 18th Int. Conf. Extending Database Technol. (EDBT)*, 2015, pp. 385–396.
- [71] H. N. Akouemo and R. J. Povinelli, "Data improving in time series using ARX and ANN models," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3352–3359, Sep. 2017.
- [72] S. S. S. R. Depuru, L. Wang, V. Devabhaktuni, and R. C. Green, "High performance computing for detection of electricity theft," *Int. J. Electr. Power Energy Syst.*, vol. 47, pp. 21–30, May 2013.
- [73] S. Kazmi, N. Javaid, M. J. Mughal, M. Akbar, S. H. Ahmed, and N. Alrajeh, "Towards optimization of metaheuristic algorithms for IoT enabled smart homes targeting balanced demand and supply of energy," *IEEE Access*, vol. 7, pp. 24267–24281, 2017.
- [74] S. Singh and A. Yassine, "Big data mining of energy time series for behavioral analytics and energy consumption forecasting," *Energies*, vol. 11, no. 2, p. 452, Feb. 2018.
- [75] A. R. Al-Ali, I. A. Zualkernan, M. Rashid, R. Gupta, and M. Alikarar, "A smart home energy management system using IoT and big data analytics approach," *IEEE Trans. Consum. Electron.*, vol. 63, no. 4, pp. 426–434, Nov. 2017.
- [76] P. Vrablecová, A. B. Ezzeddine, V. Rozinajová, S. Šárik, and A. K. Sangaiah, "Smart grid load forecasting using online support vector regression," *Comput. Electr. Eng.*, vol. 65, pp. 102–117, Jan. 2018.
- [77] J. Zhao, X. Liu, R. Dong, and G. Shao, "Landscape ecology and ecological planning toward sustainable development," *Int. J. Sustain. Develop. World Ecol.*, vol. 23, no. 4, pp. 293–297, Jul. 2016.
- [78] S. A. Shah, D. Z. Seker, S. Hameed, and D. Draheim, "The rising role of big data analytics and IoT in disaster management: Recent advances, taxonomy and prospects," *IEEE Access*, vol. 7, pp. 54595–54614, 2019.
- [79] H. Gibson, S. Andrews, K. Domdouzis, L. Hirsch, and B. Akhgar, "Combining big social media data and FCA for crisis response," in *Proc. IEEE/ACM 7th Int. Conf. Utility Cloud Comput.*, Dec. 2014, pp. 690–695.
- [80] T. Papadopoulos, A. Gunasekaran, R. Dubey, N. Altay, S. J. Childe, and S. Fosso-Wamba, "The role of big data in explaining disaster resilience in supply chains for sustainability," *J. Cleaner Prod.*, vol. 142, pp. 1108–1118, Jan. 2017.
- [81] F. E. A. Horita, J. P. D. Albuquerque, L. C. Degrossi, E. M. Mendiondo, and J. Ueyama, "Development of a spatial decision support system for flood risk management in Brazil that combines volunteered geographic information with wireless sensor networks," *Comput. Geosci.*, vol. 80, pp. 84–94, Jul. 2015.
- [82] B. Jongman, J. Wagelmaker, B. Romero, and E. de Perez, "Early flood detection for rapid humanitarian response: Harnessing near real-time satellite and Twitter signals," *ISPRS Int. J. Geo-Inf.*, vol. 4, no. 4, pp. 2246–2266, 2015.
- [83] H. Zhang, M. Babar, M. U. Tariq, M. A. Jan, V. G. Menon, and X. Li, "SafeCity: Toward safe and secured data management design for IoT-enabled smart city planning," *IEEE Access*, vol. 8, pp. 145256–145267, 2020.
- [84] J. S. Garcia, J. P. Solano, M. C. Serrano, E. N. Camba, S. F. Castell, A. S. Asensi, and F. M. Suay, "Spatial statistical analysis of urban noise data from a WASN gathered by an IoT system: Application to a small city," *Appl. Sci.*, vol. 6, no. 12, p. 380, Nov. 2016.
- [85] M. M. Rathore, A. Ahmad, A. Paul, and S. Rho, "Urban planning and building smart cities based on the Internet of Things using big data analytics," *Comput. Netw.*, vol. 101, no. 4, pp. 63–80, Jun. 2016.
- [86] D. Koo, K. Piratla, and C. J. Matthews, "Towards sustainable water supply: Schematic development of big data collection using Internet of Things (IoT)," *Proc. Eng.*, vol. 118, pp. 489–497, Jan. 2015.
- [87] M. H. Rehman, I. Yaqoob, S. Salah, I. Muhammad, P. Jayaraman, and C. Perera, "The role of big data analytics in industrial Internet of Things," *Future Gener. Comput. Syst.*, vol. 99, pp. 247–259, Oct. 2019.
- [88] R. C. Motta, K. M. de Oliveira, and G. H. Travassos, "On challenges in engineering IoT software systems," in *Proc. XXXII Brazilian Symp. Softw. Eng.*, 2018, pp. 42–51.
- [89] J. A. Patel and P. Sharma, "Online analytical processing for business intelligence in big data," *Big Data*, vol. 8, no. 6, pp. 501–518, Dec. 2020.
- [90] M. Fahmideh and D. Zowghi, "An exploration of IoT platform development," *Inf. Syst.*, vol. 87, Jan. 2020, Art. no. 101409.
- [91] P. J. Basford, F. M. J. Bulot, M. Apetroaie-Cristea, S. J. Cox, and S. J. Ossont, "LoRaWAN for smart city IoT deployments: A long term evaluation," *Sensors*, vol. 20, no. 3, p. 648, Jan. 2020.
- [92] M. S. Aslanpour, S. S. Gill, and A. N. Toosi, "Performance evaluation metrics for cloud, fog and edge computing: A review, taxonomy, benchmarks and standards for future research," *Internet Things*, vol. 12, Dec. 2020, Art. no. 100273.
- [93] S. Henning and W. Hasselbring, "Scalable and reliable multi-dimensional sensor data aggregation in data streaming architectures," *Data-Enabled Discovery Appl.*, vol. 4, no. 1, pp. 1–12, Dec. 2020.
- [94] A. M. Gómez-Orellana, J. C. Fernández, M. Dorado-Moreno, P. A. Gutiérrez, and C. Hervás-Martínez, "Building suitable datasets for soft computing and machine learning techniques from meteorological data integration: A case study for predicting significant wave height and energy flux," *Energies*, vol. 14, no. 2, p. 468, Jan. 2021.
- [95] S. Raghavan, B. Y. L. Simon, Y. L. Lee, W. L. Tan, and K. K. Kee, "Data integration for smart cities: Opportunities and challenges," in *Computational Science and Technology*. Singapore: Springer, 2020, pp. 393–403.
- [96] Q.-T. Doan, A. S. M. Kayes, W. Rahayu, and K. Nguyen, "Integration of IoT streaming data with efficient indexing and storage optimization," *IEEE Access*, vol. 8, pp. 47456–47467, 2020.
- [97] D. Fernández-Cerero, F. J. Ortega, A. Jakóbik, and A. Fernández-Montes, "DISCERNER: Dynamic selection of resource manager in hyper-scale cloud-computing data centres," *Future Gener. Comput. Syst.*, vol. 116, pp. 190–199, Mar. 2021.
- [98] S. M. Moussa and G. A. Agha, "Integrating encrypted mobile agents with smart spaces in a multi-agent simulator for resource management," *J. Softw.*, vol. 5, no. 6, pp. 630–636, Jun. 2010, doi: [10.4304/jsw.5.6.630-636](https://doi.org/10.4304/jsw.5.6.630-636).

- [99] M. S. Laser, N. Medvidovic, D. M. Le, and J. Garcia, "ARCADE: An extensible workbench for architecture recovery, change, and decay evaluation," in *Proc. 28th ACM Joint Meeting Eur. Softw. Eng. Conf. Symp. Found. Softw. Eng.*, Nov. 2020, pp. 1546–1550.
- [100] K. Cemus, F. Klimes, O. Kratochvil, and T. Cerny, "Separation of concerns for distributed cross-platform context-aware user interfaces," *Cluster Comput.*, vol. 20, no. 3, pp. 2355–2362, Sep. 2017.
- [101] F. Santos, I. Nunes, and A. L. C. Bazzan, "Model-driven agent-based simulation development: A modeling language and empirical evaluation in the adaptive traffic signal control domain," *Simul. Model. Pract. Theory*, vol. 83, pp. 162–187, Apr. 2018.
- [102] J. Kocić and H. Axmann, "A novel solution for an ECU simulator as a key component for automated testing and verification of a vehicle diagnostic device," in *Proc. 25th Telecommun. Forum (TELFOR)*, Nov. 2017, pp. 1–4.
- [103] N. Medhat, S. M. Moussa, N. L. Badr, and M. F. Tolba, "A framework for continuous regression and integration testing in IoT systems based on deep learning and search-based techniques," *IEEE Access*, vol. 8, pp. 215716–215726, 2020, doi: [10.1109/ACCESS.2020.3039931](https://doi.org/10.1109/ACCESS.2020.3039931).
- [104] F. Bock, C. Sippl, A. Heinz, C. Lauer, and R. German, "Advantageous usage of textual domain-specific languages for scenario-driven development of automated driving functions," in *Proc. IEEE Int. Syst. Conf. (SysCon)*, Apr. 2019, pp. 1–8.
- [105] R. K. Yadav and M. Gupta, "Data aggregation algorithms in IoT: An organized evaluation of the literature," in *Proc. 3rd Int. Conf. Smart Syst. Inventive Technol. (ICSSIT)*, Aug. 2020, pp. 300–304.
- [106] Y. Fathy, P. Barnaghi, and R. Tafazolli, "An adaptive method for data reduction in the Internet of Things," in *Proc. IEEE 4th World Forum Internet Things (WF-IoT)*, Feb. 2018, pp. 729–735.
- [107] T. Rahman, A. Khandakar, M. A. Kadir, K. R. Islam, K. F. Islam, R. Mazhar, T. Hamid, M. T. Islam, S. Kashem, Z. B. Mahbub, M. A. Ayari, and M. E. H. Chowdhury, "Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization," *IEEE Access*, vol. 8, pp. 191586–191601, 2020.
- [108] T. Martin, D. Geneiatakis, I. Kounelis, S. Kerckhof, and I. N. Fovino, "Towards a formal IoT security model," *Symmetry*, vol. 12, no. 8, p. 1305, Aug. 2020.



DINA FAWZY received the B.S. and M.S. degrees in computer and information sciences from the Faculty of Computer and Information Sciences (FCIS), Ain Shams University (ASU), Cairo, Egypt, in 2013 and 2018, respectively, where she is currently pursuing the Ph.D. degree in computer and information sciences. From 2013 to 2018, she was a Teaching Assistant with the Information Systems Department, FCIS, ASU. Since 2019, she has been an Assistant Lecturer. Her research interests include big data analytics, data mining, software engineering, the IoT data analytics, machine learning, and renewable energy. She was a recipient of the Best M.S. Thesis Award from FCIS, ASU, in April 2019.



SHERIN M. MOUSSA received the B.S. and M.S. degrees in computer and information sciences from the Faculty of Computer and Information Sciences (FCIS), Ain Shams University (ASU), Cairo, Egypt, in 2000 and 2004, respectively, and the Ph.D. degree from ASU in conjunction with the University of Illinois at Urbana-Champaign, USA, in 2010.

From 2010 to 2016, she was an Assistant Professor at FCIS, ASU, where she was an Associate Professor, from 2016 to 2020. Since 2021, she has been a Full Professor with the Information Systems Department, FCIS, ASU. She was the Director of the Quality and Accreditation Unit, FCIS, ASU; a Foreign Affairs Consultant at the Research and Development Projects Center; and the Director of Grants and International Collaboration Office at ASU. In addition to her teaching, research contributions, and international publications, she has more than 21 years of experience in IT multiple/complex national/regional projects in different governmental and private organizations. Her research interests include big data analytics, data mining, mobile data management, data streaming, GIS, software engineering, and big data privacy and publishing.



NAGWA L. BADR received the M.S. degree in computer science and the Ph.D. degree in software engineering and distributed systems from Liverpool John Moores University, U.K., in 1996 and 2003, respectively.

She had done her postdoctoral studies at Glasgow University, U.K. She is the Head of committee that contributed in research projects funded by national and international grants in information systems, bioinformatics, business analytics, and health informatics for the last few years (<http://www.heal.us.ac.uk/>). She is currently a Professor and the Dean of the Faculty of Computer and Information Sciences (FCIS), Ain Shams University (ASU). Her research interests include software engineering, cloud computing, big data analytics, social networking, Arabic search engines, and bioinformatics.