Quality of Data (QoD) in Internet of Things (IOT): An Overview, State-of-the-Art, Taxonomy and Future Directions

Jameel Shehu Yalli[1], Mohd Hilmi Hasan[2], Nazleeni Samiha Haron[3], Mujeeb Ur Rehman Shaikh[4], Nafeesa Yousuf Murad[5], Abdullahi Lawal Bako[6]
Computer and Information Sciences, Universiti Teknologi PETRONAS, Perak, Malaysia[1,2,3,4,5]
Department of Computing Science, University of Aberdeen, Scotland, United Kingdom (UK)[6]

***Abstract*—The Internet of Things (IoT) data is the main component for finding the basis that allows decisions to be made intelligently and enables other services to be explored and used. Data originates from smart things that have the capabilities to connect and share data enormously with other things in the IoT ecosystem. However, the level of intelligence obtained and the type of services provided, all depend on whether the data is trusted or not. High-quality data is the most trusted;, it can be used to extract meaningful insights from an event and can also be used to provide good services to humans. Therefore, decisions based on high-quality and trusted data could be good, whereas those based on low-quality or untrusted data are not only bad but could also have severe consequences. The term Quality of Data (QoD) is used to represent data trustworthiness and is used throughout this paper. To the best of our knowledge, this work is the first to coin the term QoD. The problems that hinder QoD are identified and discussed. One if it is an outlier, it is a major feature of the data that degrades its overall quality. Several machine-learning techniques that detect outliers have been studied and presented, with few data-cleaning techniques. This paper aims to present the elements necessary to ensure QoD by presenting the overview of the IoT state-of-the-art. Then, data quality, data in IoT, and outliers are studied, and some quality assurance techniques that maintain data quality is presented. A comprehensive taxonomy is shown to provide state-of-the-art data in IoT. Open issues and future directions were suggested at the end of the paper.**

***Keywords*—Quality of Data (QoD); Internet of Things (IoT); RFID; WSN; Taxonomy; trustworthiness; outlier; anomaly; confusion matrix; QoD assurance technique**

I. Introduction

The Internet of Things (IoT) has emerged as the new evolution of internet connecting different entities, things and objects from different sources around the globe, thereby generating enormous amounts of data every time, every second. The amount of data generated by the IoT is used and consumed by the objects with which it communicates than by humans. The number of servers needed to hold information for access by users is very large, giving an insight into the number of devices that connect to the internet. Then the number of IoT-connected things and devices is ten times the number of those internet PCs [1].

The tremendous amount of data generated by different things brings out the realization of big data problems, intelligent decision-making, and the development of many IoT applications. To start with the big data problem, when these things generate all the data and share, exchange, and store it in the cloud, the cloud centers need to provide enough storage to handle this data and also enough services to manage the data. Providing these two becomes a challenge for the cloud centers [2]. Advances in the technology have exploited many capabilities of these data from the things, thereby encouraging the continuous flow of data. Thus, IoT has become a major catalyst for big data problems.

For example, the scale of IoT is continuously expanding; reports show that by the end of the year 2025, the number of IoT connections could reach 24.6 billion, with a compound annual growth rate of 13% [3]. According to the International Data Corporation (IDC), there will be more than 38 billion linked things in 2025 and reach about 50 billion by 2030. Another projection by [4] reports that connections to IoT could be about 41 billion by 2025, which could generate approximately 79 zettabytes of data. According to [5], approximately 50 billion device connections exist today, with an estimated 75.44 billion device connections by 2025.

Intelligent decision-making is achieved when enough data is obtained from things covering enough scenarios and events to compare, deduce, and reach a conclusion. However, the IoT can perform all these based on the type and quality of data received; if the data is of good quality, decisions are likely to be good, but if the data lacks quality, decisions derived would also be bad [6]. Therefore, for a reliable, trusted, and intelligent decision, the data must be trustworthy.

Users and vendors found a lot of opportunities in the prevalence of IoT. New applications are being developed for the ease and comfort of the user. Some researchers are also working on AI applications that incorporate IoT, such as smart homes, smart cities, efficient energy management and distribution, and so on. To achieve optimality in IoT applications for both driven applications and network optimization, research has used meta-heuristic and heuristic algorithms to simulate physical and biological phenomena [7].

Examples of IoT applications in smart homes includes the adjustment of blinds according to temperature and environmental changes, the opening of doors for authorized vehicles, and the ordering of medical services when there is an emergency. In the traditional home, home devices are part of existing Internet expansion, but when IoT arrives, the migration of smart things begins to the IoT network [8]. When devices get corrupted, the consequences are severe. For example, when smart locks

are hacked, anyone can access the home; when baby monitors are compromised. The homeowner can be scared; and when microwaves are hacked, it can cause fire. If the security of smart devices cannot to achieved, then smart homeowners may not want to live in their smart homes. On the contrary, they can expect to improve home safety by using intelligent surveillance services [9]. In addition, the privacy of smart homeowners must be preserved. However, the continuous incoming of data from smart devices can reveal the secrets of house owners. And this can pose serious threats to their privacy.

Some widely adopted applications include smart homes, smart cities, smart grids and smart transportation. IoT technologies have drastically changed our way of life [10]. The widespread interconnection of intelligent IoT objects distributed physically extends computational operation and communication costs to IoT objects with different specifications. These devices' sensor capabilities enable them to collect real-time data from the physical world. The analysis of such data enables us to build an intelligent world and make better decisions for its management. If these security concerns are not adequately addressed, the wide adoption of IoT applications will be severely hampered. Consider the two typical IoT application areas, Smart Home and Smart Health, where system-sensitive information and critical assets require high protection [10].

IoT will continue to affect our lives in many ways, both in our homes, offices, healthcare, cities, etc. IoT in our society can represent a symbolic capital of power [11]. The way to deal with this enormous amount of data has changed from manually entered data to autonomous devices such as RFID readers, sensor nodes, etc. Our common appliances could have embedded components to allow them to communicate and become more intelligent to ease our lives. Examples are the light bulb that warns you of its remaining life, a toaster that toasts bread and provides a weather forecast, a refrigerator, a television, a video camera, and a solar panel roof, which might all be IoT devices. Despite the comforts we enjoy when these appliances generate such data, it has posed a challenge to the servers to manage and process such a huge amount of data from around the globe, which leads to big data problems.

During the last decade, we have worried about computer protection. Last five years, we have been worried about our smartphones' protection, now we are worried about car protection; home appliances, wearables, and many other IoT devices. According to Hewlett-Packard, in 2014, 70% of the most common IoT devices were infected with serious vulnerabilities. The authors of [12] discussed various current security challenges: interoperability, resource constraints, the protection of privacy on the Internet of Things Security 297, and scalability. Thus, the security of IoT is currently the main concern and requires research attention.

IoT devices collect large amounts of data and transmit it to the network. There are many personal data in these data, such as blood pressure, pulse, electrocardiograms, place environment data, area humidity, room temperature, etc. Another authentication scenario is to consider the types of entities involved in the remote client and server scenarios [13]. Clients want to access servers' services. After the first registration, the client can have a mutual authentication with the server. After the authentication, the two can create a shared key, and the client can use this key to access the server's service. A server can provide its clients with different types of services. Servers have the responsibility to perform registration and password changes. Before these services can be provided to the client, the server must verify whether the client is registered or not [14].

We researched the quality of the data in this work and coined the term Quality (QoD), and to the best of our search and knowledge, we are the first to coin the term QoD. However, many factors contribute to the data inefficiency and lack of QoD. The first problems associated with the data and IoT devices include constraint capabilities, intermittent loss of connection, and deployment hazards [14]. Other problems come from smart things, such as node failure, faulty nodes, data loss, network congestion, architectural flaws, and so on [15]. A third-world problem is one created by humans and launched into the deployment field to gain some benefit; examples are side-channel attacks, node capture attacks, sensor impersonation, stolen verifier attacks, Sybil attacks, etc. For IoT to gain wide acceptance and embrace more deployment, the QoD needs to be ensured.

This survey first investigated the nature of the QoD, or data trustworthiness, in an IoT ecosystem. The data in IoT is explored further, from the data lifecycle to its characteristics and quality considerations in IoT, then the technology of RFID that allows the data to be shared is studied, and then down to the factors that affect the QoD in IoT. Some of the reliable techniques to ensure QoD are studied, and the techniques are presented in tabular form for ease of comparison. Data outliers, is the main component that compromises QoD, are researched, and the types and impacts of the outliers are presented for prevention and measurement. A comprehensive taxonomy that shows all the forms of data that can be used is designed for ease of understanding. Some IoT application domains, open issues, and future directions are presented as well.

The remainder of this article is presented as follows: after the introduction in Section 1, data in IoT is presented in Section II. Section III is the QoD assurance techniques and data outliers made in Section IV. A comprehensive data taxonomy is shown in Section V while some of the most common IoT application domains are made in Section VI. Open issues in QoD are presented in Section VII, and then, lastly, future directions and conclusion are in Section VIII. The paper has eight sections in all.

II. Overview of Data In IoT

Data is an important component that makes up the IoT paradigm and is the source of information and means of communication. Furthermore, QoD, or trustworthiness, is an essential requirement for any IoT ecosystem (i.e., IoT services). In the following sub-sections, we present the data life cycle in the context of the IoT. We also discuss the characteristics of IoT data. In addition, we discuss QoD in IoT. Moreover, we looked at RFID as the first technology on which the IoT is built, which allows data sharing among IoT devices. And then some of the factors that affect data quality in IoT were also discussed.

Since data is considered a valuable asset because of the insights gained about a phenomenon, it is used to provide

intelligence in our daily lives dealings. Researchers, therefore, exploit the insights and intelligence in the data using different mining techniques and algorithms [16]. Data trustworthiness is essential in QoD to have a reliable handling of the data from the data itself, data interpretation, simulation results, and any other form of data representation. Data is characterized by losing its quality when some factors, such as things constrained resources, large-scale deployment, and intermittent connections are not obtained [17].

Some of these problems can be measured from the data quality dimensions that arise as a result of hazardous elements. One way to ensure that data quality is compromised is the identification of data outliers [18]. However, some outliers appear only to describe errors, while others describe rare events, e.g. unusually high temperature in a warehouse, maybe as a result of a fire. In IoT, QoD problems need to be solved. QoD is an essential requirement for any data user (things, entities, IoT services, and IoT user applications).

A. Data Lifecycle

In the original landscape setting of the internet, data primarily originates from users using their computers, surfing the web, engaging on social media networks, and generally being utilized to offer services to these users. In contrast, the IoT sees a paradigm shift where the majority of data is generated by interconnected devices, serving both as the source and primary recipient to deliver services to individuals. The Machine-to-Machine (M2M) is a precursor to IoT, which emphasizes data as the primary communication channel [19], facilitating autonomous collaboration among IoT objects to offer innovative services. Data holds significant value in the IoT, serving as a crucial asset that provides insights into various phenomena, individuals, or entities. These insights are leveraged by applications to deliver intelligent services ubiquitously. The accuracy of data is paramount, as any inaccuracies may compromise the reliability of extracted knowledge and subsequent actions based on it. Fig. 1 presents data life cycle stages.

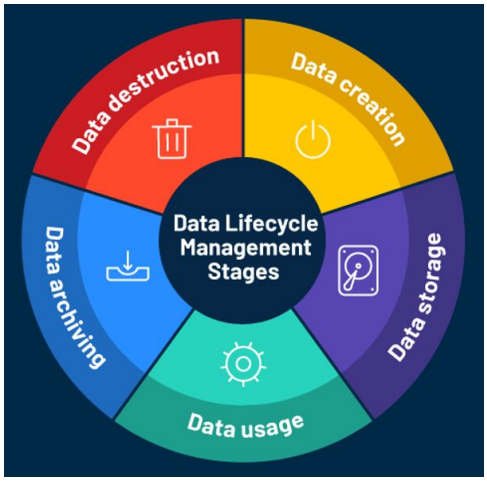

Fig. 1. Data life cycle stages.

B. Characteristics of IoT Data

The IoT device is embedded with a chip that can sense the environment and collect and share data with similar devices. The IoT devices are deployed mostly in hazardous environments, making them susceptible to natural effects such as earthquakes, rain, erosion, wind, etc. They can also be vulnerable to physical attacks and forced alteration by humans. These IoT sensors can be designed to measure variables of interest such as temperature, pressure, humidity, sleep habits, slope of a pipe, fitness level, movement positions, light intensity, and many more. However, some of the IoT characteristics are considered omnipresent, that is, erroneous, uncertain, noisy, distributed, voluminous, etc., while other characteristics can be considered dependent on the measured phenomenon, that is, continuous, smooth variation, periodicity, correlation, and Markovian behavior [20]. Some of these characteristics are:

1) Uncertain, noisy and erroneous data: uncertainty in QoD can make the data either incomplete, ignorant, ambiguous or imprecision caused by the constraint nature of the IoT nodes. Any factor that could make the data uncertain or put noise in the data, or make the data entirely wrong and have some wrong elements in it will compromise the QoD. And this can easily occur in any IoT ecosystem where the data is generated from volatile devices [21].

2) Voluminous and distributed data: In the IoT ecosystem, sensors can be deployed at any place to measure the parameters of interest. These sensors are densely deployed everywhere to gather enough data for decision-making and data management. The heterogeneous nature of IoT devices generates enormous amounts of data that are nearly impossible to manage and have challenges to manage. There is no standard IoT architecture to manage the total amount of data generated by its devices [22].

3) Smooth variation: in a continuous setting of data flow, such as the time stamp or time series, data is flown continuously at some interval. The collection and processing of such data requires some technique (like a machine learning technique) to collect the data and process it accordingly. An example is watering a tomato garden at some regular intervals [23].

4) Continuous data: The data here is similar to the smooth variation characteristic, but not necessarily that the data comes at a regular interval. The data can be random and have different patterns of arrival e.g., batch, stream, or real time. An example is to report any incident of traffic violation [24].

5) Correlation: The correlation feature exists in the IoT data set because of the heterogeneous nature of the IoT network. It consists of different sensors that measure different parameters. The data can have two correlations: spatial and temporal. When the data is correlated spatially according to the positions of the sensors, the processing of the data can give the best results in information form for certain phenomena. Whereas when the data is correlated temporally, the data might depend on its timestamp. For example, when the temperature values for the future are to be predicted, then the current temperature values can be used to make the prediction. It is also possible for the data set to have both types of correlations; the data can either be spatial (as related to memory space), temporal (as related to the time of its arrival), or both spatial and temporal [25].

6) Periodicity: This can be defined as the accuracy or age of specific data or the difference between the previous time stamp and the current time stamp as the item's punctuality or the data being sufficiently up to date for a task. Data sets that are related to scenarios may have inherently periodic patterns where the same values may occur at specific intervals [26].

7) Markovain behavior: An IoT sensor can be a function of a previous sensor, at a given time stamp of the previous sensor denoted by ti–1 [27].

C. Quality of Data in IoT

Quality of Data (QoD) in IoT can be seen as the possibility to ascertain the integrity of the data provided from its origin or the probability of the accuracy of the data [28]. Data must be clean, sensitized, and free from errors before it is transferred from a lower layer to an upper layer in the architectural stack. In other words, data must be reliable and trustworthy before it can be transmitted to preceding layers or peers for further processing. To compute trust in IoT ecosystem, it starts with the reliability of the sensors. Data from a reliable sensor could be considered trustworthy whereas data from a non-reliable sensor must be evaluated further to ascertain its correctness. However, the deployment nature of the IoT nodes in an unprotected environment makes the sensor vulnerable to attacks and unreliable [29].

Mechanisms have been developed to measure the trustworthiness in the WSN and the traditional internet; however, these mechanisms may not be suitable to ascertain the correctness of data in an IoT ecosystem due to the heterogeneous nature of the IoT, which is not the same as in WSN and the internet [29], [30]. Therefore, different mechanisms suitable for measuring QoD need to be developed, assessed, and implemented. Data is gathered massed from smart things providing ubiquitous services to users. For QoD to be ensured, the technologies used to allow the generation and sharing of the data should be addressed.

The first technologies used to allow things to connect and communicate were RFID, then WSN, which continues to evolve into other technologies up to the most widely used today, the internet (802.11 and its families). [31] assessed the quality of the data considering five dimensions, which include confidence, accuracy, timeliness, volume, and completeness. The new arising IoT applications rely on distributed and heterogeneous data for proper functioning; thereby, integration into IoT data is necessary [32]. However, maintaining such data becomes challenging due to the different sources it comes from [33], [34].

Data integrity is an essential asset in the authentication process of IoT devices. It ensures that node credentials are correct and unaltered. To ensure data integrity in an IoT distributed architecture, the MQTT protocol requires more attention because, when a connection is established between the nodes it has to transmit data to the destination, the credentials must be mutually verified to ensure that they have not been altered [35].

Assuming that the issuer "P1" connects to the broker "B1" directly, where the subscriber "S1" connects to the broker "B2" directly, the data is transferred from "P1" to "B1", and also to "B2" before it reaches the destination "S1". During this data transfer, the credentials of the data sender and the data recipient must be verified mutually. A common method of cryptography to ensure data integrity is the hash function (such as SHA-1 and SHA-2) [36].

However, ensuring QoD presumes fulfilling the criteria of accuracy, timeliness, precision, completeness, and reliability [37], [38], [39]. The authors of [40] define "timeliness as the data being current. That is, the most updated data in the most recent time". While [31], sees timeliness from two different perspectives, i.e., "an error recovery of the data and its age item, this distinguishes the recorded timestamp from current system time while the regularity of the data is with respect to its application context". Again, [32] defines "timeliness as the mean age value of the data in a source". According to [41], timeliness is defined as "the extent to which data are sufficiently up-to-date for a task".

D. Quality of Data in RFID

The first technology to be embraced by communicating entities is Radio Frequency Identification (RFID) [42]. The RFID system for authentication comprises three important tangible components: tags, readers, and data centers. The reader scans the tag to collect the necessary information and stores it in the data center. RFID can be seen as a transmitter microchip that is similar to an adhesive sticker. Active receives batteries that always emit data signals, while passive gets activated only when they are activated. The concept of radio technology was developed from RFID, where the chip does not have to view the reader in physical vision before it can communicate with it. While barcode technology requires the physical view of the reader to communicate with it [43]. RFID is also an actuator that stimulates events, an action a barcode could not do. WSN is a multi-strip wireless network connected to a dispersed sensor field that measures a specific data collection device's speed, humidity & temperature, whose values are transmitted to processing equipment. RFID is a short-range communication technology that is termed asymmetric, whereas WSN technology has a relatively long range and communication ability in a peer-to-peer fashion.

Since IoT's idea is to allow automatic connection and sharing of data among entities and any object with the ability to sense, process, transmit, and store information via the internet, it has made the network heterogeneous due to the different backgrounds of the objects. The early technologies start with RFID, then WSN, then Bluetooth, then wireless local area networks (WLANs), then wireless metropolitan area networks (WMANs), then cellular networks (LTE, 2G, 3G, 4G, 5G, and now 6G) [44], [45]. The IoT's vision is to [42] enable people and things or entities to communicate to anyone, at any time, anywhere through some sort of medium such as the internet [46]. With the rapid development of RFID technologies, Bluetooth, sensors, and smartphones, the applications and usage of IoT have increased tremendously, which directly affects daily life [47].

RFID automatically identifies entities, objects, and people. RFID's operation comes in three frequency ranges: the first is low frequency (LF), the second is high frequency (HF) and the third is ultra-high frequency (UHF). RFID devices can

be separated into two groups: active and passive [48]. Active RFID requires an energy source, while passive RFID does not require any energy to power it. The encryption levels in RFID are: the first encryption mode with no traffic encryption, the second encryption mode with the Data Encryption Standard (DES), and the third encryption mode with AES-128 bits [49]. RFID uses radio waves to communicate with data in electronic devices to identify and sense the location around them.

In recent years, WSN and RFID have gained tremendous attention in the field of IoT applications. Both technologies have different capabilities and are used in different scenarios depending on their needs [50]. RFIDs provide reading content that is used to detect and identify objects they are associated with. WSNs provide dynamic content based on the environment in which they are installed [51]. However, these technologies are used as connectors between devices and local networks or even the wider Internet. This technology is necessary to identify devices, share information with each other, verify each other's identities, and broadcast other useful information to the network. In Table I, a list of some technologies used in the WSN and the internet are presented.

E. Factors Affecting Quality of Data

Data in the IoT is itself the weakest point due to many factors that affect its quality. When data lacks quality, it cannot represent the actual scenario it is being assigned to monitor, and it could have other negative effects both on the decision being made and the operational levels of any business or organization [61]. In order to have a phenomenon of interest, some potential problems in the IoT need to be addressed. Some of the issues facing the maintenance of QoD in the IoT include but are not limited to:

1) Resource constraints: Since their inception, IoT devices have been characterized as being resource constraints in nature. Because of the limited memory available in it, the small power consumption, and the lower computational ability of the IoT devices, it becomes difficult to trust all the data that comes from them especially if the data is more than what the device can hold, and naturally, more and more data keeps coming from these devices. The IoT devices are mostly battery-powered, and resources are scarce, data collection policies and trade-offs are inherently utilized to improve the quality and cleanliness of data [62].

2) Scalability: IoT is today being deployed on a global scale, starting from organizations to homes to cities and now to the globe. Any setting of the IoT deployment generates large amounts of data, and merging any setting or integrating any application makes the data even larger, thereby increasing the chances of error occurrence in the data [63].

3) Heterogeneity: IoT devices are heterogeneous in nature, they come from different settings and backgrounds the same way their data differs from the background it comes from. It is always more challenging to manage data of the same kind with data of different kind. IoT devices can only achieve functional optimality if they integrate heterogeneous data. Therefore, the issue of heterogeneity needs to be addressed perfectly [64].

4) Sensors: When sensors are deployed, they may suffer from a lack of accuracy in reporting their readings or from a loss of calibration. Some sensors may become faulty and then report incorrect data. This is a challenge that makes it mostly difficult to find the faulty sensor, especially in a large deployment setting [65].

5) Environment: The deployment is mostly in an unprotected surrounding affected by hazards and natural effects such as rain, earthquake, erosion, the mountain's summit, wind, or the intended attack by humans [66].

6) Network: The connection often gets lost and regained again due to limited resources, bad weather, infrastructural interference, and a bad signal. IoT is an IP network with a constrained higher loss of packets than the conventional network [67].

7) Vandalism: The environment is mostly unprotected and therefore suffers from physical attacks that include damage, stealing, altering, and forceful extraction of data from it. The vandalism also extends to animals whose aim is to search for food or scatter in any setting they come across. Therefore, this factor affects the QoD [68].

8) Dead node: It often happens in many circumstances that a node is dead, but data is continuously received from the node. This has made the quality of the data untrustworthy [69].

9) Privacy: This is a major part of the acceptance of IoT globally. People's data is not guaranteed to be secure, and when data is breached (like a patient's data), the damage is too high [70].

10) Data stream: Data from the smart IoT device is received and sent continuously in the back-end pervasive applications that use them [71].

Other problems include sleeplessness habits of some nodes, unauthorized access, altering the source code & attributes, incompleteness, etc. In the the memory devices could neither send large packets nor report events frequently due to constraints; therefore, only small-sized messages could be sent, which is insufficient to report all events. Also, the scarcity of resources will cause things to go into sleep mode to save energy. However, Internet Protocols (IP) maintain the backbone of IoT connectivity and are unsuitable for sleep modes so it requires the smart things to be operational at all times, unreliable readings, multi-source data inconsistencies & alignment, data duplication, data leakage, etc.

III. Data Outliers

A data outlier is an uncertainty in an event or scenario; it is a deviation from the normal distribution setting, resulting in problems in the data set and incorrect results in the model. Outliers are the major manifestations of QoD problems. Data outliers can also be defined as phenomena with extremely small chances of occurrence. It is again defined as points in a data set that is highly unlikely to happen in a given model [72].

In other words, anomalies are defined as patterns in data that do not conform to a well-defined notion of normal behavior [73]. Data outliers belong to a class of unreliable sets or groups; they fall outside of the normal status. In most machine learning models, they are considered the unusual points in a given data set [74]. Although these points have few chances

TABLE I. THE USE OF RFID & WSN TECHNOLOGY IN IOT, PROS AND CONS

N0	AUTHORS	DOMAIN	TECHNOLOGY	OBJECTIVE	PROS	CONS
1	[3]	WSN	RFID	To provide ultra-lightweight authentication by exploiting the RFID cache reader	It achieves reduced computational cost especially when authenticating a large number of tags, It achieves security	However, it needs to provide or expand storage space for a large number of tags
2	[52]	Big Data	RFID	To provide lightweight authentication based on simpler authentication protocols	The scheme combines multiple authentication protocols and runs well	Availability is guaranteed
3	[53]	WSN	WLAN	To develop a protocol needed to incorporate the TEPANOM solution and its architecture with the EAP infrastructure	This EAP supports many authentication mechanisms by introducing lower communication overhead compared to others, it does not require any global infrastructure, thus it is scalable	It cannot integrate closely the TEPANOM solution and its architecture with the EAP infrastructure
4	[54]	Smart Grid	6LowPAN, CoaP, IEEE, 802.15.4	To provide a lightweight authentication in the smart grid application	It maintains message integrity	The scheme is only tested and proved on a small scenario field nodes
5	[55]	WSN	GWN	To introduce a novel authentication and key agreement using bio hashing to eliminate false accept rate and false reject rate	Bio hashing has some functional advantages over biometrics such as high secure operation of imposter populations and genuine zero equal error rate level	The design is inefficient with limitations to support forward secrecy and unlinkability in two factor authentication. Also lack a dynamic identity mechanism to involve nonpublic key
6	[56]	WSN	GWN SN	To develop a lightweight biometric scheme to authenticate remote users and key agreement scheme for secure IoT services	User is authenticated remotely and offline	Memory requirements need to be found in the testbed and the lightweight feature extends to real IoT devices
7	[57]	IIoT	WSN	To design a lightweight computational biometric user authentication and key agreement scheme	The protocol is lightweight and less complex authentication is achieved. Authentication, availability, and integrity of data packets are guaranteed and the protocol needs further investigation	
8	[58]	WSN	WLAN	To design a light Weight node-to-node and node-to-node authentication protocols continuously	It authenticates each data transmitted between two nodes within a pre-defined time period in the IoT ecosystem	A more accurate model needs to be designed to minimize use of battery energy consumption and to discover more dynamic device features is challenging
9	[59]	NFC	RFID	To develop an Ultralight Weight Mutual Authentication Protocol to achieve forward security by using sub key and sub index number into its key update	Computational wise, the scheme is lightweight and proved to protect against synchronization attack	The NFC authentication needs to improve performance and function while still considering the security and privacy of the system
10	[60]	NFC	NFC	To develop a novel lightweight NFC identity authentication protocol for mobile NFC IoT networks	It has three working modes in the NFC mobile phone that it can work in the tag as the card, the reader, and support peer-to-peer file sharing	Its application electronic financial services still need privacy for public trust

to occur in the whole data set, but they often occur. Another formal metric-based outlier is the distance-based outlier (DB) [75]. It is defined when an object, let's say, 'x' in a given data set 'T', and the fraction of 'x' is higher than the distance 'D' within the context of 'DB', is considered an outlier. Fig. 2 shows an example of an outlier.

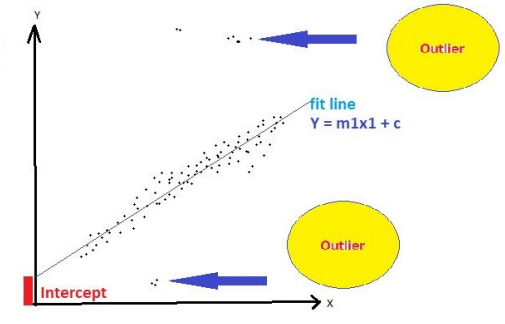

Fig. 2. Data outliers.

A. Types of Outliers

Data outliers as defined, are anomalies that do not conform with the normal behavior of the remaining data in the data set. Sometimes an outlier may represent an error, sometimes it may represent an incompatible element in a cluster, and sometimes it may even represent useful information. There are different types of data outliers in a model, but the most common ones are as follows:

1) Error: This is any value generated as a result of node failure or node malfunction. When the node's battery is drained and is not replaced and the node continues to send data to the user's access point, it is very likely that the data is wrong or that repetitive data is sent. Sometimes the node may be altered by an attacker and forced to send the wrong data [76]. Attackers often try to extract some useful information from a captured node while trying to let it continue to send information to the base station. This process has already altered the normal process, so therefore, wrong data may be sent. It can also be affected by some natural effects, such as wind displacement, i.e., when a node is deployed in an environment and configured to measure a parameter of interest within a given ratio and the node is displaced outside that ratio, then the readings sent will not represent that area of interest [77].

2) Event: This is any scenario with an associated value generated due to a change in a certain setting or phenomenon. This can be demonstrated by a natural effect of event occurrence; for example, if in a hazardous environment, there is an accident or any natural phenomenon that changes the setting of the environment suddenly, then definitely the reading from the node at that particular time follows any sudden change as well, thereby not giving the expected outcome [78].

3) Point anomaly: This is the deviated data, a group of similar data that differ in value, behavior, and attributes. When usual data occurs in a given data set, deviating from the normal distribution pattern of the remaining data and the difference is huge enough to believe that it is out of context, then that outlier is considered a point anomaly [79]. For example; a model records the card withdrawals of an employee in Asia to occur once every day and then there appears to be a withdrawal transaction in Europe three hours from the last transaction in Asia, this translates into an impossible scenario; therefore, the data point of the transaction in Europe is classified as a clear point anomaly [80].

4) Contextual anomaly: This represents a value that could be an anomaly, but it does not depend on the context. Sometimes a deviation may occur, and still, the data may not be an outlier due to the context of the data set it belongs to [72]. Many scenarios happen where unusual data becomes an outlier in one context while being considered normal in another context. For example, given two data sets A and B, where data set A is a small data set in size, let's say with 100 rows, and data set B is large let's say it has 1000 rows. The calculation of its 'variance' using the same data point may become an outlier in data set A while proving normal in data set B [80].

5) Collective anomaly: represents a set of collected values that differs largely from other values in the data set. When more than one anomaly, let's say a group of anomalies appears in the data set, forming another cluster of anomalies, then it is referred to as a collective anomaly. This type of outlier is mostly identified in clustering algorithms such as the K-means algorithm, the Naive Bayes algorithm, the Decision Tree, etc. [81]. Fig. 3 gives examples of some of these outliers.

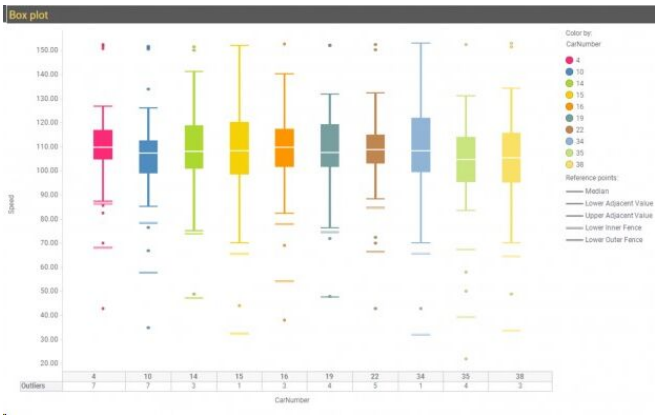

Fig. 3. Types of outliers.

B. An Outlier from the Confusion Matrix

A machine learning model is a powerful technique widely used to detect an outlier in a data set [82]. Today, there is seamless integration between the IoT domain and machine learning models. Many machine learning models are designed to solve IoT problems. One technique used to solve an IoT problem is the identification of an outlier by a model called the confusion matrix. This model has two true classes and two negative classes that are passed to the machine mode. An output is given by the model based on what it has learned. The four classes are described as follows:

1) True positive: when a true instance is passed to the model, the machine model computes the prediction and gives an output based on the data it learned. If the prediction corresponds to the actual or real event, then it is considered True Positive. For example, when it rains, the event is passed to the model, and the outcome is confirmed to be 'rain' or 'it rains'.

2) True negative: When the model receives the actual event and predicts wrongly giving an output that does not correspond to the expected outcome, then the scenario is considered True Negative. An example is when, in reality, it rains, the event is passed to the model, and the model predicts 'not to rain' or 'not raining'.

3) False positive: The other way around is to supply the model with the wrong event and expect it to learn and produce an output based on what it learns. When the outcome produced reveals the actual scenario that occurred while it was fed with the wrong event, then it is considered a False Positive. For example, when the status of an impregnated woman is passed to the model and it gives the output that the woman is pregnant then it is considered a False Positive.

4) False negative: But when the event passed to the model is the wrong event and the algorithm learns and predicts that the event is the wrong one, then it is called a False Negative. For example, if the model is fed with the woman's status as not being pregnant and produces the output as 'not pregnant'. Fig. 4 illustrates the confusion matrix in a diagram.

C. Impact of Outliers in the IOT

In IoT, environment data is obtained from the sensor as a result of measurement of parameters of interest, such as temperature, pressure, humidity, etc. and serves as an input to mine data so as to gain insights about a monitored phenomenon (e.g home, environment, health, etc. [83]. Based on these

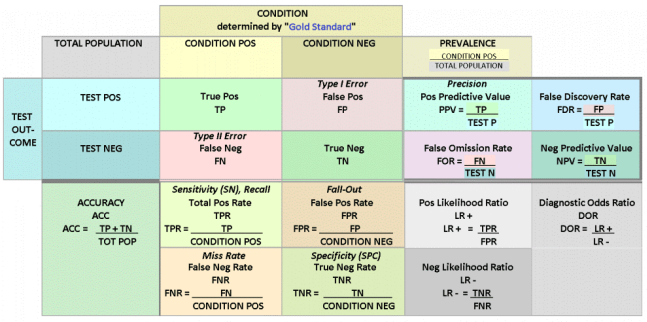

Fig. 4. The confusion matrix.

insights, decisions can be made from different angles. It is clear that the conclusions reached from an erroneous data will produce bad and unsound decisions. For example, a model giving out too many false positives and false negatives such as a scenario of a campus fire alarm, the alarm system rings many times every week while in reality there is no cause for the panic [84].

Another scenario involves monitoring forest fires to respond quickly and take appropriate measures. In addition to component monitoring applications, the data on the state of the components should be reported accurately in order to protect expensive systems and avoid damage. An inability to provide accurate data can cause damage to whole systems or even the lives of people [75]. Other examples include forest fire alerting system that requires quick action, earthquake that requires quick evacuation to safe zones, etc.

The importance of accurate and reliable data is paramount, consider the examples above and imagine if any of the system do not alert about the occurrence of the dangers happening or the model reports otherwise about the danger, the consequences will be very severe and the lives of the people and entities is as high risk. When these nodes report actual data and the model predicts right and the system alarms correct, it will help and save lives and properties while if the nodes report faulty data, the model may easily predict wrong and the whole system may not function well to give the correct outcome, this is jeopardizing lives and properties.

The trustworthiness of the data is essential to the engagement of users and to the acceptance of IoT services and is therefore important to the success of the large-scale implementation of the IoT domain [85]. Data as a component of a holistic approach to managing IoT trust collection, reliability, and accuracy are the main concerns of data perception reliability.

As experiments and simulations are a good way to demonstrate and understand IoT systems likewise machine learning models are a good way to identify and prevent outliers in any given dataset. Many IoT experimentation test beds, such as the FIT-Equipex exist [86]. the authors of [87] examined several other existing testbeds (public and private). However, in order to study more on the impact of data anomalies. There are two real-world cases that examine the impact of QoD problems on the field of electronic health applications.

The first case study by [88] examines the effect of QoD problems on electronic health monitoring applications. This work identifies QoD issues that affect QoD criteria (e.g. accuracy, precision, timeliness, accessibility, and consistency) that are critical to providing appropriate help to the patient. There are three levels of data management defined to monitor cardiac scenarios: Data Acquisition, Data Processing, and Data Discovery. For example, at the level of data collection, the problems relate mainly to the performance of body sensors, the amount of data processed, and the quality of communications [89].

The second reported case study by [90] examines the poor impact of QoD on Ambient Assisted Living systems (AAL) systems, which results from the convergence of ambient intelligence and assisted living technologies. AAL supports monitoring applications (i.e. monitoring of health and well-being) to people in their homes. This paper argues that poor QoD alters the representation of events that occur, which hinders the system from giving appropriate support to users and causes incorrect reports on the health of patients, inefficient management of environmental conditions at home, etc. [91]. The application of e-health is one of the most important IoT applications taking into account the factors affecting human life and therefore tolerating uncertainty in the QoD.

IV. Quality of Data Assurance Techniques

In conventional programming, a common rule states: "Never trust user input," while in IoT the rule can be stated as: "Never trust things". This is proven as a result of uncertainties and inconsistencies in sensor data. In order to reduce the expensive effects of low QoD, a technique is needed to prepare data and improve its quality [92]. The following are five main techniques that could promise QoD in an IoT paradigm:

A. Outlier detection

This is to find the elements that differ from the normal distribution setting or deviate from the normal behavior of the data. The ultimate goal is to highlight outliers [93]. Identifying outlier detection in a model increases its overall reliability and efficiency. In addition, detecting outliers is the first step needed to handle all the events of inconsistencies. The next is the accuracy and reliability of data processing. If these are handled, then QoD will be ensured and better decisions will be made. Note that individual data element accuracy at the level itself does not increase because it relates only to the source of data generation and processing and cannot be improved [94]. The parameters used to detect outliers pay attention to highlighting data value differences to find out the outliers (for example, values that are not consistent with an established model) [95]. However, the QoD dimension values used to evaluate data are seen as insufficient, for example, the accuracy extracted from the measurement precision class of the sensor specified by the manufacturer. When the sensor fails to clean due to any cause, the accuracy becomes unreliable and irrelevant.

B. Interpolation

Interpolation is defined as a data generation method that can hence improve the QoD dimension of the data size (i.e., add the available data elements). However, interpolation is the opposite of completeness in effect, i.e., the ratio of the

available data to non-interpolated items (i.e., both interpolated and non-interpolated) in the stream window in question. This scenario can be explained as an optimization effect to find the best compromise between these two dimensions, but nonetheless, is limited to satisfying user-defined QoD requirements [96]. Furthermore, when selecting interpolation techniques, it is important to consider the accuracy of the interpolation value [63], which is expected to meet user requirements. This technique involves the use of missing values based on a dataset. Data flows are described as missing data flow attributes or tuples (from sensor malfunctions and loss of connection) [97]. The missing data points represent gaps in the dataset that is available for a particular entity or a topic of interest. A model knowledge-generating processing of such a dataset, including those missing gaps in it would have incomplete knowledge and hence reduce QoD.

C. Data Integration

All the heterogeneous data from different landscapes must be integrated to overcome structural differences and inconsistencies and really benefit the universal service. Frameworks for Data Quality (DQ) techniques such as Resource Description Framework (RDF) and Web Ontology Language (OWL, 2009) provide standard mechanisms for data description to perform search, retrieval, and processing tasks more directly [98]. Also, linked data is a reliable approach to trigger data retrieval and integration in the IoT ecosystem. Another framework proposed by [99] integrates semantic data that uses the principles of Linked Data and semantic web technologies. The authors of [100] proposed an architectural model to integrate and incorporate all intelligent features into the smart application. Again, a Service Architecture Paradigm (SOA) model that extracts heterogeneity in intelligent objects and improves interoperability in the context of e-health applications is developed.

D. Data Deduplication

is a technique for the compression of data designed to lessen the volume of data stored by deleting duplicate data elements and replacing them with unique data references that remain unchanged. Data deduplication is a process of duplicating redundant data elements. It decreases the amount of data and affects the QoD of the data volume. [101] proposed a video duplication technique with considerations for privacy protection. The authors of [102] specify the deduplication technology for cloud-storage encrypted data. While [103] proposed a model for exploiting the deduplication capabilities together with the Hadoop framework.

E. Data Cleaning

The cleaning of data defines the life cycle of data; it starts with the selection of errors and goes down to the correction of identified errors and the identification of potential errors. It is also defined that the detection of anomalies is limited to the identification of anomalies, while data cleaning goes further to suppress the elements discovered. It has become a widely adopted technique for enterprise data management in data warehouses [104]. Data cleaning is a widespread topic of research in big data analysis [105]. It consists of three main stages: (i) determining the type of error; (ii) identifying potential errors; and (iii) correcting identified errors. It is also very common to manage enterprise data in the context of data storage. Fig. 5 presents some QoD techniques.

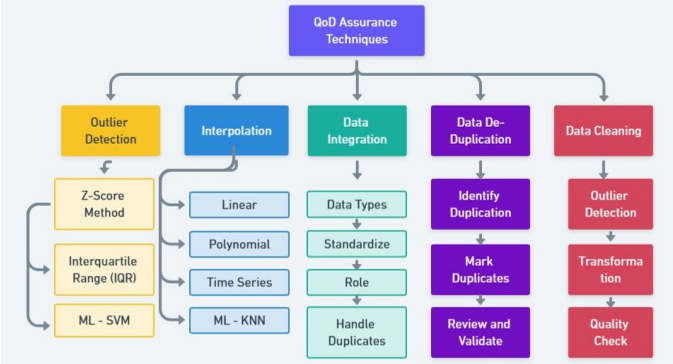

Fig. 5. QoD Assurance techniques.

V. Taxonomy for Quality of Data in IOT

The IoT and other similar domains such as WSN, have a set of attributes, features, characteristics, protocols, and technologies that are suited for their deployment and implementation. With a taxonomy, it can be seen what techniques have been used to measure data trustworthiness, what are the most important parameters to consider in ensuring QoD, what domains are integrated in the IoT data trustworthiness ecosystem, and what has been done and what needs to be done in the area.

A. Data Source

The data comes from one or more sources, and the source of the data determines whether the data can be trusted or not. Data could come from a sensor configured to provide the readings of an event, which is the source of the data, or from humans directly, especially in situations where interactive information is required by the model or the environment [106].

B. Data Processing

The first form is called stream or batch where data is usually sent at some specific time interval. The data in the same batch is mostly similar or has the same attributes. Data from the sensor could be sent in batch for processing, the data is first gathered and stored in the temporary memory of the device and then sent across the network to the server, or any base station provided. An example of batch processing can be any event whose action is not urgent [106]. Real-time processing is when the data is sensitive and needs to be processed immediately after it is obtained. In real-time, events such as fast decision-making making, banking transactions are examples of such scenarios. Another form of processing the data is computed in the near time, where the history of data is referred before making any computation. The processing of such data is not immediate, but time is also to be considered.

C. Data Type

The data obtained from the sensor could be presented to the user in the form of either Numeric, Alpha, Alpha Numeric, or even Symbols. The type of the data depends on the parameter measured and result representation [20].

D. Trust Type

The type of trust can be either direct or indirect. When data is gotten directly from a sensor it is considered as direct trust, likewise when the data is obtained from other sensors passing it to neighboring nodes, the quality of the data might degrade or the data may be intercepted and altered by natural phenomena, so therefore, the data is considered as indirect trust [20].

E. Trust Computation Location

The location of the node can be computed to determine the trustworthiness of the node. The computation can either be distributed or centralized. A cluster-based wireless sensor architecture can have the cluster head to compute the trustworthiness of the sensor node in the network, which is below a certain assigned value, where the actual data is sent to the gateway and then to the application layer. This considered as a centralized computation. Whereas in distributed trust computation, the node assesses the trustworthiness itself and send actual data to the cluster and gateway [107]

F. Trust Aggregation Method

is used to summarize trust evidence that has been gathered via nearby sensor node feedback or self-observations [108]. The majority of the literature's work use the following methods:

1) Weighted sum: This is the simplest model for aggregating trust scores. The method can summarise several factors that contribute to trust scores, the factors are multiplied by the specified weight, then adds up all results that contribute to a product that represents the trust score. Therefore, it is a commonly used technique for calculating trust score [108].

2) Bayesian inference: Due to its simplicity and solid statistical foundation, Bayesian inference is a popular confidence calculation model. This method considers trust as a random variable that follows a distribution of probability and which parameters are updated with new results.

3) Belief theory: The theory of belief, also known as the theory of evidence or Dempster-Shafer Theory (DST) provides a method for summarising confidence values from different pieces of evidence using Dempster's Rule of Combination. The rule assumes that this evidence is independent. Evidence is the confidence values computed by different sensor nodes in the network [109].

4) Regression analysis: Regression analysis is a method of aggregating confidence scores by calculating the relationships between data. The scores are calculated based on estimating the relationships of the trust factors and a number of other variables that affect the trust.

5) Fuzzy logic: This is a method that deals with estimation rather than fixed and exact conclusions, fuzzy logic also provides rules for reasoning. The confidence value determined with fuzzy logic can have a value between 0 and 1 with fuzzy measures [108].

6) Game theory: This involves making decisions between two or more decision-makers involved in certain conflicts or competitions. Game theory can be used to predict competitive rules of action with certainty. An example of using game theory models to ensure data reliability is the work of [19], which develops a defense strategy that ensures that sensor nodes are protected against attacks so that the difference between the value accepted by the sink and the true sense value is below a certain assigned value.

G. Trust Establishment

This refers to how to end a trust score from multiple properties. There are two aspects of the establishment of trust, namely single trust and multi-trust. A single trust implies that only one trust property is taken into account in the calculation of the total trust rating. On the contrary, multi-trust is a combination of trust and trust. Establishments use several trust factors to calculate the total data trust. Many proposed techniques utilize multi-trust factors to calculate trust scores, with two factors chosen on average. Among the factors used were communication, nodes' familiarity, energy, and nodes' reliability [20].

H. Trust Results

These are also called Trust Decisions, and are considered as an element of data trust calculation that deals with how the results are presented to the requestor or user. There are two options for representing the results either in binary or in a range of values or judgments. Binary represents the results as either trust or non-trust only. From this point of view, users or application layers can simply choose trusted data to process further. In terms of range, this means that the data reliability value calculated can fall within any range of possible degrees of trust. This is similar to the Likert scale, but the trust values can be determined by more than two options. As such, the requestor or user and application layers can decide accordingly on the basis of their decision logic [107].

I. Data QoS

The summation of all the packets involved in the transmission having the maximum delay (i.e., 400ms standard) and a regular jitter interval between its consecutive packets (i.e., 1ms interval) should then be able to produce the maximum number of packets received [110].

J. Node Quality

This is of two types: a resistant node which is able to prevent itself against side-channel attack and unclonability and any form of memory extraction. Most of such nodes have multiple ICs designed over one another to make it harder for an attack to recover anything from it. An unresistant node is one that is unable to protect itself from the attack mentioned [107].

K. Node State

A node can either be in a passive state where it remains ideal until it is triggered to send data, this node performs better in such a state since its battery will last longer while an active

node is the one that is continuously measuring and sending data, the problem associated with this type of node is called sleep deprivation attack [111].

L. Measurable Parameters

These nodes are deployed to measure some readings called parameters. The parameters include but are not limited to the volume of an object, the temperature of a room or place, the pressure of equipment, humidity of a place, slope position of ground, fitness level of things, sleep habits in pipelines, state of the area, movement positions of liquids, etc. [107].

M. Data Accuracy

The data from the node is considered accurate as long as it is precise, correct, and reliable. The basis for having a sound decision is from accurate data while bad data produces a severe decision that comes with consequences. When data is accurate it is said to represent the actual scenario of an event [112].

N. Data Consistency

For the data to be consistent, it should satisfy standards and integrity, and the codes with which it is burned to the device must also be consistent in producing the actual result any time it is being run.

O. Data Completeness

Data is said to be complete if its record in the database is complete, the field data is complete and the single Unique Identifier is complete [112].

P. Data Timeliness

The timeliness or the regularity of the data is measured from its time stamp as well as its real-time updates [113].

Q. Data Relevance

Data is relevant according to the user requirements it satisfies and the contextual relevance of the data.

R. Data Robustness

When data is robust it should be able to accommodate fault tolerance by being resilient as well as quality monitoring [111].

S. Data Security

There are many forms to protect the data but for the sake of this research the most commonly practiced are access control mechanisms, encryption (symmetric and asymmetric), and data masking [110].

T. Metadata Quality

Metadata is the data from which data is made up or simply some supporting data that helps in the execution of the actual data. In Fig. 6, a comprehensive taxonomy for QoD in IoT is shown.

VI. IOT Application Domains

- Smart Home: Smart homes have the vision to integrate intelligence into everyday objects such as appliances, door locks, surveillance cameras, garage doors, etc., and to communicate with existing cyber infrastructure [6]. Adding intelligence to physical objects is beneficial for improving people's lives, such as improving their comfort, convenience, security, and effective use of natural resources. For example, a smart home can adjust the blinds according to environmental changes, open garage doors automatically when an authorized vehicle approaches, or order medical services when there is an emergency. In smart homes, traditional home devices are part of existing Internet expansion. When devices are damaged, the consequences can be serious. For example, successfully hacking smart locks allows strangers to enter the house; compromising baby monitors can scare visitors away from the baby; hacking microwaves can cause a fire at home [8]. Smart homeowners may not want to live in smart homes if security is not guaranteed. On the contrary, they can expect to improve home safety by using intelligent surveillance services. In addition, the privacy of smart homeowners must be preserved. However, the continuous collection of data from smart home devices can reveal the private activities of house owners, as indicated in [9]. It poses serious threats to the privacy of owners.
- Another typical IoT application is the creation of smart networks, Smart grids are designed and implemented to improve the reliability, cost reduction, and efficiency of traditional power grid systems. It not only integrates green and renewable energy such as solar power, wind power, heat, etc. but also aims to improve the reliability and efficiency of traditional energy networks. Intelligent grid data communication networks connect many smart grid devices and play an essential role in achieving the above-mentioned objectives. It collects energy consumption data and monitors the state of smart grid systems. More applications can be developed based on smart and communication networks. For example, utilities can allocate and balance load more wisely based on energy use information collected. It can also help to design fair but scaled pricing models by taking into account unbalanced energy consumption in space and time. With a smart grid status monitoring application, you can identify faults in the grid system as quickly as possible and as well design new fault-tolerant mechanisms to better react to them. Many technologies, including Automatic Measurement Infrastructure (AMI), have been proposed to build smart network communication networks. Because so much data is moving around the mission-critical system, security is one of the most important concerns of such systems. Invading the smart grid [114] and cutting the supply of electricity to a large area can cause enormous physical and economic damage to society. Analyzing energy usage data can also reveal people's daily private activities.
- By embedding smart medical devices into the medical

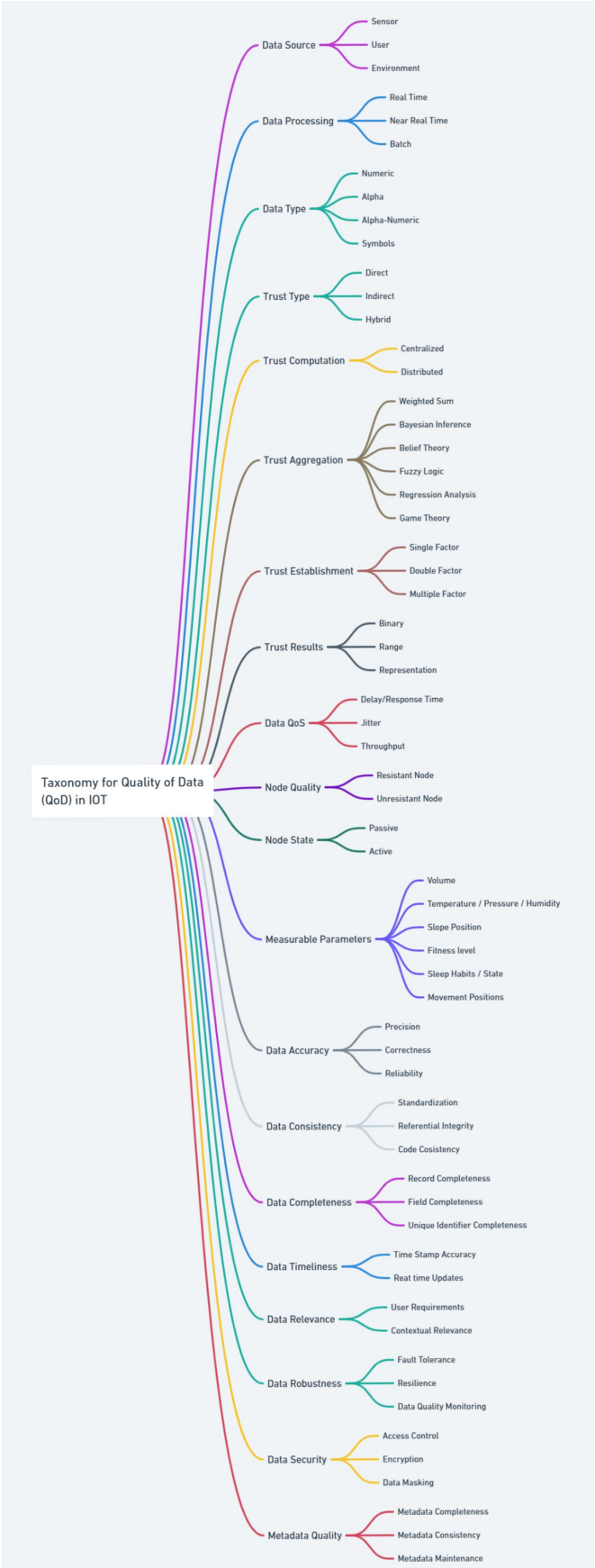

Fig. 6. Taxonomy of QoD in IOT.

infrastructure, health professionals can better monitor patients and use the data to determine those who need the most attention. In other words, healthcare professionals believe that prevention is more important and effective than treatment, so they can build proactive management systems based on collected data using the best of these networks of devices. Researchers also studied other possible techniques for implanting sensors into human bodies to monitor people's health status [115]. Using the collected data, health personnel can discover behavioral changes in the body of patients and medicines during treatment. Security is also an essential issue in Smart Connected Health. In network medical devices, data collection and monitoring of the status of the device is convenient, but there are also risks because instructions can be sent to terminate the device functions [116]. Stopping medical devices that are important to the patient's life such as heart injuries is extremely dangerous.

- Intelligent Transportation System (ITS) refers to the use of advanced technologies and data-based solutions to improve the efficiency, safety, and sustainability of transportation systems. Smart transportation uses technology and data to create more efficient, safer, and sustainable transportation systems that benefit individuals and communities. These systems play a key role in addressing urbanization, congestion, and environmental impacts in rapidly developing cities. In addition, IoT applications can manage passenger luggage at airports, and advise drivers about road conditions [117].
- The telecommunications industry includes a variety of technologies and services such as Global System for Mobile Communications (GSM) and Digital mobile network standards for voice and data communications on mobile phones and a few others. The Bluetooth technology is a short-range data exchange that is commonly used to connect headphones, speakers, and other peripheral devices. Wireless Local Area Network (WLAN): Wireless network technology allows devices to connect to the Internet or local network within a limited area. Wi-Fi called (wireless fidelity) is a service that allows voice calls and texts to be sent via Wi-Fi networks, providing coverage in areas with weak mobile signals [118]. Global Positioning System (GPS) is a satellite-based navigation system that provides accurate location information and is widely used in mobile phones, navigation devices, and vehicle tracking [119]. These technologies are essential to modern communication, connectivity, and location-based services.
- Logistics and Supply Chain Management: In logistics, RFID-embedded intelligent shelves enable real-time tracking of items, improving inventory visibility, accuracy, and efficiency [120]. This technology simplifies business, reduces errors, and helps companies optimize their supply chains to improve their performance and customer satisfaction. RFID-embedded smart shelves track items in real-time.
- Aerospace and Aviation Industry: In the aerospace and aviation industries, the Internet of Things (IoT) plays an essential role in enhancing safety, maintenance, and efficiency. Sensors and connected devices monitor aircraft components, collect performance data, allow real-time maintenance, improve reliability and safety, and reduce operational costs [121]. IoT has revolutionized the aerospace and aviation industries by providing real-time data and connectivity solutions to improve safety, reduce costs, improve passenger experience, and optimize the entire ecosystem, improve safety and security of elements.
- Automotive Industry: In the automotive industry, the IoT has changed the design, manufacture, operation, and maintenance of vehicles. It revolutionizes the automotive industry by making connected, autonomous, and safer vehicles, improving manufacturing processes, and improving consumer driving experiences. Sensors monitor and report vehicle parameters [122].

VII. Open Issues in Quality of Data for IOT

The following are a few issues that require additional research presented in four broad categories:

A. Scalability

The IoT can now be seen being deployed on a bigger scale, to say on an unprecedented scale that exceeds even the scale of the traditional Internet. Most of the solutions, however, are concentrated, and unlike distributed architectures, they do not provide sufficient flexibility and scaling for large-scale deployment [123].

B. Heterogeneity of Data Sources

The data generated in the IoT ecosystem comes from multiple types of objects, entities, sensors, RFID tags, etc. The architecture developed for IoT must be able to adapt to the heterogeneity of data origin. Furthermore, proposed technologies must be able to process different variables to meet the requirements of IoT applications. In order to meet the requirements of IoT applications, which may provide complex services based on multiple parameters such as user behavior, energy management, and home temperature relative to external temperature [124].

C. Domain-agnostic/automated verification

In IoT visions, things share data automatically with neighboring nodes based on their configuration. Domain-agnostic data cleaning methods confirm that data transmitted between "things" is uninterrupted, without human involvement, and with minimal human control, which is essential to the creation of a seamless IoT service [125].

D. Distributed Architectures

In addition to IoT scaling issues, distributed architectures also provide a platform that can adapt to faults and failure resilience. These functions are vital in the IoT perspective, because of the continuity and accessibility of data cleaning infrastructure, providing all-encompassing services even in the event of failures in ecosystems [126].

VIII. Conclusion and Future Directions

IoT offers great potential to connect millions of everyday objects and provide intelligent and ubiquitous services to help you live. The amount of data generated from the IoT infrastructure is very large. The collected data serves as the basis for obtaining insights that aid in decision-making, data management, and other services. QoD is an important interest in this scenario. Data security is linked to data quality, and the security of any model begins with data trustworthiness, which is essential to user participation and acceptance in the IoT paradigm. IoT is a promising domain, and there have been exciting results recorded in this field. In this context, QoD plays an important role. However, more research is needed to investigate how to improve QoD to ensure the widespread adoption and acceptance of IoT. Therefore, more work needs to be done to ensure effective and perfect decision-making, since data reliability is highly needed in IoT. The following are a few suggestions for future work to maintain QoD in IoT: Lightweight outlier detection Techniques, IoT network Traffic based Outlier Detection, Personalized QoD management platforms, QoD assessment-based outlier techniques, and QoD management middleware.

Acknowledgment

The authors would like to thank Yayasan Universiti Teknologi Petronas (YUTP) for providing the funding to carry out this research work. YUTP grant (015LC0-311) is granted by Universiti Teknologi PETRONAS (UTP).

References

[1] E. E. Broday and M. C. Gameiro da Silva, "The role of internet of things (iot) in the assessment and communication of indoor environmental quality (ieq) in buildings: a review," *Smart and Sustainable Built Environment*, vol. 12, no. 3, pp. 584–606, 2023.

[2] G. R. C. de Aquino and C. M. de Farias, "Asclepius: Data quality framework for iot," in *Proceedings of the Int'l ACM Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications*, 2023, pp. 69–76.

[3] K. Fan, Y. Gong, C. Liang, H. Li, and Y. Yang, "Lightweight and ultralightweight rfid mutual authentication protocol with cache in the reader for iot in 5g," *Security and Communication Networks*, vol. 9, no. 16, pp. 3095–3104, 2016.

[4] M. Saqib, B. Jasra, and A. H. Moon, "A lightweight three factor authentication framework for iot based critical applications," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 9, pp. 6925–6937, 2022.

[5] A. Kumar, R. Saha, M. Conti, G. Kumar, W. J. Buchanan, and T. H. Kim, "A comprehensive survey of authentication methods in internet-of-things and its conjunctions," *Journal of Network and Computer Applications*, vol. 204, p. 103414, 2022.

[6] A. GhaffarianHoseini, N. D. Dahlan, U. Berardi, A. GhaffarianHoseini, and N. Makaremi, "The essence of future smart houses: From embedding ict to adapting to sustainability principles," *Renewable and Sustainable Energy Reviews*, vol. 24, pp. 593–607, 2013.

[7] I. Rouf, H. Mustafa, M. Xu, W. Xu, R. Miller, and M. Gruteser, "Neighborhood watch: Security and privacy analysis of automatic meter reading systems," in *Proceedings of the 2012 ACM conference on Computer and communications security*, 2012, pp. 462–473.

[8] X. Pan, Z. Ling, A. Pingley, W. Yu, N. Zhang, and X. Fu, "How privacy leaks from bluetooth mouse?" in *Proceedings of the 2012 ACM conference on Computer and communications security*, 2012, pp. 1013–1015.

[9] K. Sha, W. Wei, T. A. Yang, Z. Wang, and W. Shi, "On security challenges and open issues in internet of things," *Future generation computer systems*, vol. 83, pp. 326–337, 2018.

[10] Z. Liu, K.-K. R. Choo, and M. Zhao, "Practical-oriented protocols for privacy-preserving outsourced big data analysis: Challenges and future research directions," *Computers & Security*, vol. 69, pp. 97–113, 2017.

[11] L. Nataliia and F. Elena, "Internet of things as a symbolic resource of power," *Procedia-Social and Behavioral Sciences*, vol. 166, pp. 521–525, 2015.

[12] R. Chetan and R. Shahabadkar, "A comprehensive survey on exiting solution approaches towards security and privacy requirements of iot," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 4, p. 2319, 2018.

[13] M. Nair, S. Dang, and M. A. Beach, "Iot device authentication using self-organizing feature map data sets," *IEEE Communications Magazine*, 2023.

[14] E. E.-D. Hemdan, Y. M. Essa, M. Shouman, A. El-Sayed, and A. N. Moustafa, "An efficient iot based smart water quality monitoring system," *Multimedia Tools and Applications*, pp. 1–25, 2023.

[15] T. C. C. Nepomuceno, V. D. H. de Carvalho, K. T. C. Nepomuceno, and A. P. C. Costa, "Exploring knowledge benchmarking using time-series directional distance functions and bibliometrics," *Expert Systems*, vol. 40, no. 1, p. e12967, 2023.

[16] S. M. Tahsien, H. Karimipour, and P. Spachos, "Machine learning based solutions for security of internet of things (iot): A survey," *Journal of Network and Computer Applications*, vol. 161, p. 102630, 2020.

[17] R. Huo, S. Zeng, Z. Wang, J. Shang, W. Chen, T. Huang, S. Wang, F. R. Yu, and Y. Liu, "A comprehensive survey on blockchain in industrial internet of things: Motivations, research progresses, and future challenges," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 88–122, 2022.

[18] N. ALBAZZAI, O. RANA, and C. PERERA, "Camera as a sensor towards augmenting anomaly detection in internet of things systems: A survey."

[19] M. Gupta, J. Gao, C. C. Aggarwal, and J. Han, "Outlier detection for temporal data: A survey," *IEEE Transactions on Knowledge and data Engineering*, vol. 26, no. 9, pp. 2250–2267, 2013.

[20] N. Haron, J. Jaafar, I. A. Aziz, M. H. Hassan, and M. I. Shapiai, "Data trustworthiness in internet of things: A taxonomy and future directions," in *2017 IEEE conference on big data and analytics (ICBDA)*. IEEE, 2017, pp. 25–30.

[21] S. Sagar, A. Mahmood, K. Wang, Q. Z. Sheng, J. K. Pabani, and W. E. Zhang, "Trust–siot: Towards trustworthy object classification in the social internet of things," *IEEE Transactions on Network and Service Management*, 2023.

[22] H. Foidl and M. Felderer, "An approach for assessing industrial iot data sources to determine their data trustworthiness," *Internet of Things*, vol. 22, p. 100735, 2023.

[23] S. Sagar, A. Mahmood, Q. Z. Sheng, J. K. Pabani, and W. E. Zhang, "Understanding the trustworthiness management in the social internet of things: a survey," *arXiv preprint arXiv:2202.03624*, 2022.

[24] M. Alabadi, A. Habbal, and X. Wei, "Industrial internet of things: Requirements, architecture, challenges, and future research directions," *IEEE Access*, 2022.

[25] L. Wei, Y. Yang, J. Wu, C. Long, and B. Li, "Trust management for internet of things: A comprehensive study," *IEEE Internet of Things Journal*, vol. 9, no. 10, pp. 7664–7679, 2022.

[26] F. M. R. Junior and C. A. Kamienski, "A survey on trustworthiness for the internet of things," *IEEE Access*, vol. 9, pp. 42 493–42 514, 2021.

[27] Z. N. Aghdam, A. M. Rahmani, and M. Hosseinzadeh, "The role of the internet of things in healthcare: Future trends and challenges," *Computer methods and programs in biomedicine*, vol. 199, p. 105903, 2021.

[28] L.-A. Tang, X. Yu, S. Kim, Q. Gu, J. Han, A. Leung, and T. La Porta, "Trustworthiness analysis of sensor data in cyber-physical systems," *Journal of Computer and System Sciences*, vol. 79, no. 3, pp. 383–401, 2013.

[29] G. Han, J. Jiang, L. Shu, J. Niu, and H.-C. Chao, "Management and applications of trust in wireless sensor networks: A survey," *Journal of Computer and System Sciences*, vol. 80, no. 3, pp. 602–617, 2014.

[30] D. Hui-hui, G. Ya-jun, Y. Zhong-qiang, and C. Hao, "A wireless sensor networks based on multi-angle trust of node," in *2009 International Forum on Information Technology and Applications*, vol. 1. IEEE, 2009, pp. 28–31.

[31] A. Klein and W. Lehner, "Representing data quality in sensor data streaming environments," *Journal of Data and Information Quality (JDIQ)*, vol. 1, no. 2, pp. 1–28, 2009.

[32] ——, "How to optimize the quality of sensor data streams," in *2009 Fourth International Multi-Conference on Computing in the Global Information Technology*. IEEE, 2009, pp. 13–19.

[33] Q. Jing, A. V. Vasilakos, J. Wan, J. Lu, and D. Qiu, "Security of the internet of things: perspectives and challenges," *Wireless Networks*, vol. 20, pp. 2481–2501, 2014.

[34] S.-M. Luo, Z.-J. Ge, Z.-W. Wang, Z.-Z. Jiang, Z.-B. Wang, Y.-C. Ouyang, Y. Hou, H. Schatten, and Q.-Y. Sun, "Unique insights into maternal mitochondrial inheritance in mice," *Proceedings of the National Academy of Sciences*, vol. 110, no. 32, pp. 13 038–13 043, 2013.

[35] K. Gupta and K. Yadav, "Data collection method to improve energy efficiency in wireless sensor network," in *International Conference of Advance Research and Innovation (ICARI-2015)*, 2015.

[36] S. Shitole and A. Gujar, "Securing broker-less publisher/subscriber systems using cryptographic technique," in *2016 International Conference on Computing Communication Control and automation (ICCUBEA)*. IEEE, 2016, pp. 1–6.

[37] J. E. Bailey and S. W. Pearson, "Development of a tool for measuring and analyzing computer user satisfaction," *Management science*, vol. 29, no. 5, pp. 530–545, 1983.

[38] C. Batini, A. Rula, M. Scannapieco, and G. Viscusi, "From data quality to big data quality," *Journal of Database Management (JDM)*, vol. 26, no. 1, pp. 60–82, 2015.

[39] W. S. Geisler, "Contributions of ideal observer theory to vision research," *Vision research*, vol. 51, no. 7, pp. 771–781, 2011.

[40] T. Dasu and T. Johnson, *Exploratory data mining and data cleaning*. John Wiley & Sons, 2003.

[41] C.-C. Lai, T.-C. Wang, C.-M. Liu, and L.-C. Wang, "Probabilistic top-^{k} dominating query monitoring over multiple uncertain iot data streams in edge computing environments," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8563–8576, 2019.

[42] X. Jia, Q. Feng, T. Fan, and Q. Lei, "Rfid technology and its applications in internet of things (iot)," in *2012 2nd international conference on consumer electronics, communications and networks (CECNet)*. IEEE, 2012, pp. 1282–1285.

[43] A. Khan, A. Ahmad, M. Ahmed, J. Sessa, and M. Anisetti, "Authorization schemes for internet of things: requirements, weaknesses, future challenges and trends," *Complex & Intelligent Systems*, vol. 8, no. 5, pp. 3919–3941, 2022.

[44] H. Chen, X. Jia, and H. Li, "A brief introduction to iot gateway," in *IET international conference on communication technology and application (ICCTA 2011)*. IET, 2011, pp. 610–613.

[45] M. Kumhar and J. Bhatia, "Emerging communication technologies for 5g-enabled internet of things applications," *Blockchain for 5G-Enabled IoT: The new wave for Industrial Automation*, pp. 133–158, 2021.

[46] S. S. Sabry, N. A. Qarabash, and H. S. Obaid, "The road to the internet of things: a survey," in *2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON)*. IEEE, 2019, pp. 290–296.

[47] S. Ijaz, M. A. Shah, A. Khan, and M. Ahmed, "Smart cities: A survey on security concerns," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 2, 2016.

[48] A. W. Nagpurkar and S. K. Jaiswal, "An overview of wsn and rfid network integration," in *2015 2nd International Conference on electronics and communication systems (ICECS)*. IEEE, 2015, pp. 497–502.

[49] K. Pal, "Challenges of using wireless sensor network-based rfid technology for industrial iot applications," *Handbook of Research on Advancements of Contactless Technology and Service Innovation in Library and Information Science*, pp. 80–100, 2023.

[50] G. Mudra, H. Cui, and M. N. Johnstone, "Survey: An overview of lightweight rfid authentication protocols suitable for the maritime internet of things," *Electronics*, vol. 12, no. 13, p. 2990, 2023.

[51] G. B. Mohammad, S. Shitharth, S. A. Syed, R. Dugyala, K. S. Rao, F. Alenezi, S. A. Althubiti, and K. Polat, "Mechanism of internet of things (iot) integrated with radio frequency identification (rfid) technology for healthcare system," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–8, 2022.

[52] A. Kumar, K. Gopal, and A. Aggarwal, "Cost and lightweight modeling analysis of rfid authentication protocols in resource constraint internet of things," *Journal of Communications Software and Systems*, vol. 10, no. 3, pp. 179–187, 2014.

[53] M. P. Pawlowski, A. J. Jara, M. J. Ogorzalek *et al.*, "Compact extensible authentication protocol for the internet of things: enabling scalable and efficient security commissioning," *Mobile Information Systems*, vol. 2015, 2015.

[54] M. M. Fouda, Z. M. Fadlullah, N. Kato, R. Lu, and X. S. Shen, "A lightweight message authentication scheme for smart grid communications," *IEEE Transactions on Smart grid*, vol. 2, no. 4, pp. 675–685, 2011.

[55] J. Srinivas, S. Mukhopadhyay, and D. Mishra, "Secure and efficient user authentication scheme for multi-gateway wireless sensor networks," *Ad Hoc Networks*, vol. 54, pp. 147–169, 2017.

[56] P. K. Dhillon and S. Kalra, "A lightweight biometrics based remote user authentication scheme for iot services," *Journal of Information Security and Applications*, vol. 34, pp. 255–270, 2017.

[57] B. Khalid, K. N. Qureshi, K. Z. Ghafoor, and G. Jeon, "An improved biometric based user authentication and key agreement scheme for intelligent sensor based wireless communication," *Microprocessors and Microsystems*, vol. 96, p. 104722, 2023.

[58] Y.-H. Chuang, N.-W. Lo, C.-Y. Yang, and S.-W. Tang, "A lightweight continuous authentication protocol for the internet of things," *Sensors*, vol. 18, no. 4, p. 1104, 2018.

[59] K. Fan, P. Song, Y. Yang *et al.*, "Ulmap: Ultralightweight nfc mutual authentication protocol with pseudonyms in the tag for iot in 5g," *Mobile Information Systems*, vol. 2017, 2017.

[60] K. Fan, C. Zhang, K. Yang, H. Li, and Y. Yang, "Lightweight nfc protocol for privacy protection in mobile iot," *Applied Sciences*, vol. 8, no. 12, p. 2506, 2018.

[61] C. Liu, P. Nitschke, S. P. Williams, and D. Zowghi, "Data quality and the internet of things," *Computing*, vol. 102, no. 2, pp. 573–599, 2020.

[62] R. Perez-Castillo, A. G. Carretero, M. Rodriguez, I. Caballero, M. Piattini, A. Mate, S. Kim, and D. Lee, "Data quality best practices in iot environments," in *2018 11th International Conference on the Quality of Information and Communications Technology (QUATIC)*. IEEE, 2018, pp. 272–275.

[63] L. Zhang, D. Jeong, and S. Lee, "Data quality management in the internet of things," *Sensors*, vol. 21, no. 17, p. 5834, 2021.

[64] R. Kollolu, "A review on wide variety and heterogeneity of iot platforms," *The International journal of analytical and experimental modal analysis, analysis*, vol. 12, pp. 3753–3760, 2020.

[65] D. Sehrawat and N. S. Gill, "Smart sensors: Analysis of different types of iot sensors," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2019, pp. 523–528.

[66] S. L. Ullo and G. R. Sinha, "Advances in smart environment monitoring systems using iot and sensors," *Sensors*, vol. 20, no. 11, p. 3113, 2020.

[67] K. Gulati, R. S. K. Boddu, D. Kapila, S. L. Bangare, N. Chandnani, and G. Saravanan, "A review paper on wireless sensor network techniques in internet of things (iot)," *Materials Today: Proceedings*, vol. 51, pp. 161–165, 2022.

[68] X. Yang, L. Shu, Y. Liu, G. P. Hancke, M. A. Ferrag, and K. Huang, "Physical security and safety of iot equipment: A survey of recent advances and opportunities," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 7, pp. 4319–4330, 2022.

[69] M. Shafiq, H. Ashraf, A. Ullah, M. Masud, M. Azeem, N. Jhanjhi, and M. Humayun, "Robust cluster-based routing protocol for iot-assisted smart devices in wsn." *Computers, Materials & Continua*, vol. 67, no. 3, 2021.

[70] P. M. Chanal and M. S. Kakkasageri, "Security and privacy in iot: a survey," *Wireless Personal Communications*, vol. 115, no. 2, pp. 1667–1693, 2020.

[71] B. M. Alencar, R. A. Rios, C. Santana, and C. Prazeres, "Fotstream: A fog platform for data stream analytics in iot," *Computer Communications*, vol. 164, pp. 77–87, 2020.

[72] M. A. Samara, I. Bennis, A. Abouaissa, and P. Lorenz, "A survey of outlier detection techniques in iot: review and classification," *Journal of Sensor and Actuator Networks*, vol. 11, no. 1, p. 4, 2022.

[73] A. Gaddam, T. Wilkin, and M. Angelova, "Anomaly detection models for detecting sensor faults and outliers in the iot-a survey," in *2019 13th International Conference on Sensing Technology (ICST)*. IEEE, 2019, pp. 1–6.

[74] M. A. Bhatti, R. Riaz, S. S. Rizvi, S. Shokat, F. Riaz, and S. J. Kwon, "Outlier detection in indoor localization and internet of things (iot) using machine learning," *Journal of Communications and Networks*, vol. 22, no. 3, pp. 236–243, 2020.

[75] H. Ghallab, H. Fahmy, and M. Nasr, "Detection outliers on internet of things using big data technology," *Egyptian Informatics Journal*, vol. 21, no. 3, pp. 131–138, 2020.

[76] M. V. Brahmam and S. Gopikrishnan, "Nodstac: Novel outlier detection technique based on spatial, temporal and attribute correlations on iot bigdata," *The Computer Journal*, p. bxad034, 2023.

[77] P. D. Rosero-Montalvo, Z. István, P. Tözün, and W. Hernandez, "Hybrid anomaly detection model on trusted iot devices," *IEEE Internet of Things Journal*, 2023.

[78] C. Karras, A. Karras, and S. Sioutas, "Pattern recognition and event detection on iot data-streams," *arXiv preprint arXiv:2203.01114*, 2022.

[79] A. Gaddam, T. Wilkin, M. Angelova, and J. Gaddam, "Detecting sensor faults, anomalies and outliers in the internet of things: A survey on the challenges and solutions," *Electronics*, vol. 9, no. 3, p. 511, 2020.

[80] E.-S. Apostol, C.-O. Truică, F. Pop, and C. Esposito, "Change point enhanced anomaly detection for iot time series data," *Water*, vol. 13, no. 12, p. 1633, 2021.

[81] A. Shahraki and Ø. Haugen, "An outlier detection method to improve gathered datasets for network behavior analysis in iot," 2019.

[82] J. Liang, "Confusion matrix: Machine learning," *POGIL Activity Clearinghouse*, vol. 3, no. 4, 2022.

[83] D. Gupta *et al.*, "Prediction of sensor faults and outliers in iot devices," in *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. IEEE, 2021, pp. 1–5.

[84] L. Boukela, G. Zhang, M. Yacoub, S. Bouzefrane, S. B. B. Ahmadi, and H. Jelodar, "A modified lof-based approach for outlier characterization in iot," *Annals of Telecommunications*, vol. 76, pp. 145–153, 2021.

[85] Z. Yan, P. Zhang, and A. V. Vasilakos, "A survey on trust management for internet of things," *Journal of network and computer applications*, vol. 42, pp. 120–134, 2014.

[86] G. Z. Papadopoulos, A. Gallais, G. Schreiner, and T. Noel, "Importance of repeatable setups for reproducible experimental results in iot," in *Proceedings of the 13th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks*, 2016, pp. 51–59.

[87] C. O'Reilly, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Anomaly detection in wireless sensor networks in a non-stationary environment," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1413–1432, 2014.

[88] S. Rodríguez-Valenzuela, J. A. Holgado-Terriza, J. M. Gutiérrez-Guerrero, and J. L. Muros-Cobos, "Distributed service-based approach for sensor data fusion in iot environments," *Sensors*, vol. 14, no. 10, pp. 19 200–19 228, 2014.

[89] J. McNaull, J. Augusto, M. Mulvenna, and P. McCullagh, "Ambient assisted living systems and technologies: a data and information quality perspective," *ACM Transactions on Data and Information Quality*, vol. 4, no. 1, pp. 1–15, 2012.

[90] J. McNaull, J. C. Augusto, M. Mulvenna, and P. McCullagh, "Data and information quality issues in ambient assisted living systems," *Journal of Data and Information Quality (JDIQ)*, vol. 4, no. 1, pp. 1–15, 2012.

[91] ——, "Multi-agent system feedback and support for ambient assisted living," in *2012 Eighth International Conference on Intelligent Environments*. IEEE, 2012, pp. 319–322.

[92] R. Togneri, G. Camponogara, J.-P. Soininen, and C. Kamienski, "Foundations of data quality assurance for iot-based smart applications," in *2019 IEEE Latin-American Conference on Communications (LATINCOM)*. IEEE, 2019, pp. 1–6.

[93] M. K. Abiodun, J. B. Awotunde, R. O. Ogundokun, E. A. Adeniyi, and M. O. Arowolo, "Security and information assurance for iot-based big data," in *Artificial Intelligence for Cyber Security: Methods, Issues and Possible Horizons or Opportunities*. Springer, 2021, pp. 189–211.

[94] M. Bures, T. Cerny, and B. S. Ahmed, "Internet of things: Current challenges in the quality assurance and testing methods," in *International conference on information science and applications*. Springer, 2018, pp. 625–634.

[95] C. A. Ardagna, E. Damiani, J. Schütte, and P. Stephanow, "A case for iot security assurance," *Internet of Everything: Algorithms, Methodologies, Technologies and Perspectives*, pp. 175–192, 2018.

[96] J. Buelvas, D. Múnera, D. P. Tobón V, J. Aguirre, and N. Gaviria, "Data quality in iot-based air quality monitoring systems: a systematic mapping study," *Water, Air, & Soil Pollution*, vol. 234, no. 4, p. 248, 2023.

[97] D. Li, L. Yan, Y. Liu, Q. Yin, S. Guo, and H. Zheng, "Data quality improvement method based on data correlation for power internet of things," in *2019 12th International Symposium on Computational Intelligence and Design (ISCID)*, vol. 2. IEEE, 2019, pp. 259–263.

[98] H. Khodkari, S. Maghrebi, and R. Branch, "Necessity of the integration internet of things and cloud services with quality of service assurance approach," *Bulletin de la Société Royale des Sciences de Liège*, vol. 85, no. 1, pp. 434–445, 2016.

[99] A. M. Nagib and H. S. Hamza, "Sighted: a framework for semantic integration of heterogeneous sensor data on the internet of things," *Procedia Computer Science*, vol. 83, pp. 529–536, 2016.

[100] R. Morabito, R. Petrolo, V. Loscrí, and N. Mitton, "Enabling a lightweight edge gateway-as-a-service for the internet of things," in *2016 7th International Conference on the Network of the Future (NOF)*. IEEE, 2016, pp. 1–5.

[101] S. Li, L. D. Xu, and S. Zhao, "The internet of things: a survey," *Information systems frontiers*, vol. 17, pp. 243–259, 2015.

[102] X. Huang, P. Craig, H. Lin, and Z. Yan, "Seciot: a security framework for the internet of things," *Security and communication networks*, vol. 9, no. 16, pp. 3083–3094, 2016.

[103] R. Sethi, B. Bhushan, N. Sharma, R. Kumar, and I. Kaushik, "Applicability of industrial iot in diversified sectors: evolution, applications and challenges," *Multimedia technologies in the Internet of Things environment*, pp. 45–67, 2021.

[104] X. Ding, H. Wang, G. Li, H. Li, Y. Li, and Y. Liu, "Iot data cleaning techniques: A survey," *Intelligent and Converged Networks*, vol. 3, no. 4, pp. 325–339, 2022.

[105] T. Wang, H. Ke, X. Zheng, K. Wang, A. K. Sangaiah, and A. Liu, "Big data cleaning based on mobile edge computing in industrial sensor-cloud," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 1321–1329, 2019.

[106] J. Guo and R. Chen, "A classification of trust computation models for service-oriented internet of things systems," in *2015 IEEE International Conference on Services Computing*. IEEE, 2015, pp. 324–331.

[107] J. S. Yalli and M. H. Hasan, "A unique puf authentication protocol based fuzzy logic categorization for internet of things (iot) devices," in *Proceedings of the 2023 12th International Conference on Software and Computer Applications*, 2023, pp. 246–252.

[108] H.-S. Lim, G. Ghinita, E. Bertino, and M. Kantarcioglu, "A game-theoretic approach for high-assurance of data trustworthiness in sensor networks," in *2012 IEEE 28th International Conference on Data Engineering*. IEEE, 2012, pp. 1192–1203.

[109] N. Mohamed, *Critical Socio-Technical issues surrounding mobile computing*. IGI Global, 2015.

[110] J. S. Yalli, S. B. Abd Latif, A. H. A. Hashim, and M. K. Alam, "An improved qos in the architecture, model and huge traffic of multi-media applications under high speed wireless campus network," 2006.

[111] J. Yalli, S. Latif, M. Masud, M. Alam, and A. Abdallah, "A comprehensive analysis of improving qos and imm traffic of high speed wireless campus network," in *2014 IEEE Symposium on Computer*

Applications and Industrial Electronics (ISCAIE). IEEE, 2014, pp. 12–17.

[112] J. S. Yalli, S. B. Abd Latif, and S. Bari, "Interactive multi-media applications: Quality of service guaranteed under huge traffic," *International Journal of Computer Applications*, vol. 105, no. 7, 2014.

[113] F. A. Garba, K. I. Kunya, Z. A. Zakari *et al.*, "A proposed novel low cost genetic-fuzzy blockchain-enabled internet of things (iot) forensics framework," *Scientific and practical cyber security journal*, 2021.

[114] M. A. Faisal, Z. Aung, J. R. Williams, and A. Sanchez, "Data-stream-based intrusion detection system for advanced metering infrastructure in smart grid: A feasibility study," *IEEE Systems journal*, vol. 9, no. 1, pp. 31–44, 2014.

[115] G. Leroy, H. Chen, and T. C. Rindflesch, "Smart and connected health [guest editors' introduction]," *IEEE Intelligent Systems*, vol. 29, no. 3, pp. 2–5, 2014.

[116] M. Rahman, B. Carbunar, and M. Banik, "Fit and vulnerable: Attacks and defenses for a health monitoring device," *arXiv preprint arXiv:1304.5672*, 2013.

[117] S. Muthuramalingam, A. Bharathi, S. Rakesh Kumar, N. Gayathri, R. Sathiyaraj, and B. Balamurugan, "Iot based intelligent transportation system (iot-its) for global perspective: A case study," *Internet of Things and Big Data Analytics for Smart Generation*, pp. 279–300, 2019.

[118] A. M. Al-Momani, M. A. Mahmoud, and M. S. Ahmad, "Factors that influence the acceptance of internet of things services by customers of telecommunication companies in jordan," *Journal of Organizational and End User Computing (JOEUC)*, vol. 30, no. 4, pp. 51–63, 2018.

[119] A. M. Luthfi, N. Karna, and R. Mayasari, "Google maps api implementation on iot platform for tracking an object using gps," in *2019 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob)*. IEEE, 2019, pp. 126–131.

[120] Y. Song, F. R. Yu, L. Zhou, X. Yang, and Z. He, "Applications of the internet of things (iot) in smart logistics: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4250–4274, 2020.

[121] G. Karakuş, E. Karşıgil, and L. Polat, "The role of iot on production of services: A research on aviation industry," in *Proceedings of the International Symposium for Production Research 2018 18*. Springer, 2019, pp. 503–511.

[122] M. A. Rahim, M. A. Rahman, M. M. Rahman, A. T. Asyhari, M. Z. A. Bhuiyan, and D. Ramasamy, "Evolution of iot-enabled connectivity and applications in automotive industry: A review," *Vehicular Communications*, vol. 27, p. 100285, 2021.

[123] C. Campolo, G. Genovese, G. Singh, and A. Molinaro, "Scalable and interoperable edge-based federated learning in iot contexts," *Computer Networks*, vol. 223, p. 109576, 2023.

[124] I. Bedhief, M. Kassar, and T. Aguili, "Empowering sdn-docker based architecture for internet of things heterogeneity," *Journal of Network and Systems Management*, vol. 31, no. 1, p. 14, 2023.

[125] J. Gaskin, A. Elmaghbub, B. Hamdaoui, and W.-K. Wong, "Deep learning model portability for domain-agnostic device fingerprinting," *IEEE Access*, 2023.

[126] F. Azzedin and T. Alhazmi, "Secure data distribution architecture in iot using mqtt," *Applied Sciences*, vol. 13, no. 4, p. 2515, 2023.