A Literature Review on Outlier Detection in Wireless Sensor Networks

Julio C. García [1,2,]*, Luis A. Rivera [1,3], and Jonny Perez [2]

[1] Faculty of Systems and Computer Engineering, Universidad Nacional Mayor de San Marcos, Lima, Perú
[2] Faculty of Communication Science, Universidad Laica Eloy Alfaro de Manabí, Manta, Ecuador
[3] Mathematical Sciences Laboratory, Universidade Estadual doNorte Fluminense, Rio de Janeiro, Brazil
Email: jcgarcia1805@gmail.com (J.C.G.); rivera@uenf.br (L.R.); jonnyperezv@hotmail.com (J.P.V.)
*Corresponding author

Abstract*—Wireless sensor networks have become an important element of technologies such as the Internet of Things due to their ability to obtain sensory data from the physical world in tracking and monitoring applications. However, such networks are susceptible to the presence of outliers mainly due to errors or failures in the sensor nodes or the presence of events that alter the reading patterns. To address this problem, many researchers have turned their efforts to the development of outlier detection techniques that achieve maximum detection rate with the highest possible efficiency, given the limited resources typical of this type of networks. In this study, 33 papers on outlier detection techniques in wireless sensor networks between 2018 and 2023 were analyzed with the aim of describing the characteristics of these techniques, their metrics and test conditions, application areas, and possible limitations. The results showed mostly hybrid, distributed, online and multivariate sensing proposals in addition to the exploitation of spatiotemporal correlations of the data. In terms of efficiency, almost all of them reported detection rates above 85% and in several cases up to 100% but in specific conditions; with application areas especially related to environmental monitoring and care. Finally, the most relevant limitations encountered include high computational complexity and high resource consumption, sensitivity to parameters, lack of scalability, and dependence on specific assumptions about data distribution.

***Keywords*—Wireless Sensor Network (WSN), outlier detection, Outlier Detection Techniques (ODTs), fault detection, event detection, distributed detection**

I. Introduction

In emerging areas such as the Internet of Things (IoT), the tracking and monitoring capabilities of Wireless Sensor Networks (WSNs) are integrated with Internet-based services and applications, thus facilitating informed decision-making in controlled environments [1]. These WSNs are composed of devices called sensor nodes. These nodes are deployed for the purpose of collecting environmental data, providing accurate representation of monitored phenomena or tracked targets, functioning as the digital skin of the IoT by delivering real-world information [2]. Specifically, WSNs have been widely used in applications related to personal, industrial, commercial, and military domains [3], such as environmental and habitat monitoring [4], structural monitoring [5], precision agriculture [6], medical and health monitoring [7]; and military applications such as defense, survival, target tracking, among many other fields [8].

Low-cost sensor nodes present serious memory, energy, bandwidth, and computational capacity limitations [9], which make them susceptible to producing abnormal readings known as anomalies or outliers. Of the many definitions in the literature for the term "anomaly", one of the most common is "patterns in the data that do not conform to a well-defined notion of normal behavior" [10]. In WSN, anomalies can be defined as "measurements in the detected data that significantly deviate from the normal data detection profile" [2]. In the same context, Jurdak and Wang *et al.* [11] identifies three types of anomalies: (1) network anomalies, (2) node anomalies, and (3) data anomalies. Data anomalies refer to an observation or a subset of observations that, compared to the rest of the dataset, appear to be inconsistent [12]; hence, they are often called "outliers" [13], although the terms anomaly and outlier are commonly used interchangeably in the literature [14]. However, the latter (outlier), in the context of WSN, serves to identify unusual behavior compared to most sensor readings [15]; that is, measurements that significantly differ from the normal pattern of the detected data [16].

In datasets from WSNs, outliers are common. This prevalence is largely due to the harsh and unattended environments in which these networks operate. Two primary reasons contribute to this: (1) the propensity of sensor nodes to fail, and (2) the influence of noisy wireless signals and malicious attacks [17]. Additionally, unusual phenomena within the monitored area's reach are added to these reasons [18]. In this regard, many studies such as [1, 9, 15, 19] address three main sources of outliers in WSNs: (1) noise or errors, (2) events, and (3) malicious attacks. It is important to clarify that, as treaties in [3, 20], this work, which consisted of a comprehensive review of outlier detection techniques in WSNs whose

Manuscript received August 10, 2023; revised August 30, 2023; accepted October 25, 2023; published March 14, 2024.

doi: 10.12720/jait.15.3.372-388

methodology is explained in more detail in Section II, only considers outliers caused by errors and events, since malicious attacks are related to network security, a field of study that is beyond this study.

As a result, the problem is that the observations collected by sensor nodes are often of low quality and unreliable. This limitation hampers real situational awareness for decision-making and motivates the need for efficient outlier detection techniques in WSNs that guarantee the quality of sensor data [21]. Outlier detection in the context of WSNs has been extensively researched in various disciplines, such as statistics, data mining, and machine learning [3]. Chandola *et al.* [10] refers to the problem of finding patterns that do not agree well with known and expected behavior. In the field of WSNs, it refers to the problem of finding data observations that deviate significantly from normal measurements over a specific period [22].

The purpose of using a WSN extends beyond merely collecting data from the field of implementation. More importantly, the analysis of this data at the opportune time that allows making some significant decisions, which is why data quality is the main concern in a WSN application [2]. In this context, outliers greatly influence the quality of the collected data, so they are usually more interesting than normal data [15]. For example, forest fires, earthquakes, or chemical spills cannot be accurately detected using inaccurate and incomplete data; therefore, ensuring the reliability and accuracy of sensor data is extremely important [3] and crucial for decision making [15].

Therefore, outliers hold significant potential value because they can represent changes in the monitoring objects or environments [23]. The importance of detecting them accurately lies in the fact that outliers translate into meaningful and often critical actionable information across application domains [10], which has greatly driven the efforts of previous studies to develop Outlier Detection Techniques (ODTs), both to provide reliability and quality to the data, and to report events of interest in the monitored area [2]. As a result, many techniques, methods, frameworks, and algorithms based on statistics, classification, and clustering among other approaches have been proposed in the literature. These techniques optimize the quality of sensor measurements and provide the best information to end users while maintaining low power consumption [15].

The aim of this paper is to describe the characteristics, test conditions, metrics used, application areas, and limitations of the outlier detection proposals for WSNs addressed in the literature. To this end, as mentioned previously, Section II outlines the questions that will guide the study and provides a synthesis of the review process undertaken. Section III provides extensive information on WSNs and discusses their applications, outliers, approaches, and taxonomy of existing detection techniques; all with the objective of providing the necessary basis to facilitate the analysis that follows. Section IV details the characteristics, evaluation conditions, metrics, application areas, and limitations of the selected detection proposals. Sections V and VI present a discussion of the results, conclusions, and future directions, respectively.

II. Review Method

Although there are many literature reviews related to anomaly detection in WSN, there are very few specifically related to outlier detection [2, 3, 14, 15]. Moreover, there are almost no current reviews that concentrate on outliers caused by faults or events and explore the details of such detection proposals. In fact, the most recent and comprehensive review was conducted in [9] covering the period between 2004 and 2018. Therefore, the purpose of this work is to provide a current review of the frameworks, methods, techniques, and algorithms proposed for outlier detection in WSNs with respect to the following questions:

- Q1: What are the characteristics of the proposals for outlier detection in WSNs?
- Q2: What conditions and metrics are used for evaluating outlier detection methods in WSNs?
- Q3: What are the application areas of outlier detection in WSNs?
- Q4: What limitations do these proposals for outlier detection in WSNs present?

For this, relevant articles obtained from the ACM Digital Library, Ebsco host, IEEE Explore, Science Direct, Scopus, Springer Link, and Web of Science were analyzed; from 2018 to June 2023 using the following search string: “detection AND (outlier OR anomaly) AND (wsn OR “wireless sensor network”)”. Of the 1688 articles obtained, after a selection process applying inclusion and exclusion criteria, 33 studies were selected and analyzed in detail based on the posed questions. In addition, we incorporated and analyzed other surveys and systematic reviews that, together with the background of the selected papers, allowed us to establish the theoretical foundations for the subsequent analysis.

III. Foundations of Outlier Detection in WSN

A. *Wireless Sensor Networks*

One of the most important elements of the IoT paradigm is constituted by WSNs, because they act as a digital perception layer that provides a means to access information from the physical world, which can be exploited by any computational system [2]. A WSN is a network composed of small nodes, with limited capabilities of energy, memory, computation, and communication bandwidth [3], self-organized and generally deployed in large numbers [24] in harsh and unattended environments [14]. A WSN can be made up of hundreds or even thousands of these low-cost sensor nodes distributed over a wide area [15].

Smart sensor nodes can employ various types of sensors—such as biological, mechanical, chemical, optical, thermal, and magnetic sensors—to measure environmental properties [25]. This combination of multiple sensors allows us to observe various characteristics of the same phenomenon [7] simultaneously. In fact, in the context of outlier detection, this feature of sensor nodes motivates univariate, bivariate, or multivariate detection approaches.

1) Network structure

The two main network structures or topologies used in a WSN implementation are: (1) flat, and (2) hierarchical or cluster based. In a flat base structure, all nodes are treated equally and are given the same functionality [2]. In the hierarchical network structure, the sensor node network is divided into clusters and each cluster has a cluster head or header node (CH) [22]. Typically, the sensor nodes at the lowest level are responsible for collecting data from the physical world and transmitting it to their cluster heads (parent nodes), which in turn send it to the main gateway node or to a base station [18]. The communication can be categorized as single-hop or multi-hop, depending on the number of hops sensor nodes use to transmit data to other nodes of equal or higher hierarchy [8].

2) WSN applications

WSNs stand out in applications related to personal, industrial, commercial, and military domains [3]. Examples include home automation in personal applications, sales tracking in commercial settings, architecture and control in industrial contexts, and monitoring and tracking of enemy targets in military operations [2].

These WSN applications can be summarized in two large groups: (1) tracking applications: human, animal, traffic, cars and busses and enemy tracking in military applications; (2) monitoring applications: inventory, animal, structural, manufacturing, engine, chemical, patient, and environmental monitoring, which includes monitoring and recognition of environmental phenomena [9]. These capabilities to monitor large areas, react in real time, access remote and hostile places, and their relative ease of use have provided scientists with a whole world of possibilities for new applications [15].

On the other hand, the limited resources of the WSN and the harsh implementation conditions cause the data generated by the sensor to be contaminated with noise, obvious errors, missing data, duplicate values, and contradictory information [15]. In a real-world WSN application, data quality is the main concern [2]. This quality is affected both by internal and external factors, making it unsuitable for decision-making processes in real events, and it should be noted that outliers are one of the most influential factors affecting data quality [3].

B. Outliers in WSN

In Ref. [26], outliers are defined as “patterns in the data that do not conform to a well-defined notion of normal behavior”. In different application domains, these patterns are often referred to as anomalies, outliers, discordant observations, exceptions, aberrations, surprises, peculiarities or contaminants. Of all of them, two of the most used terms in the literature, sometimes interchangeably, are: (1) anomalies and (2) outliers [14]. However, in the context of WSNs, in this work we adopt the term “outliers” to refer to anomalies in the data, that is, an observation or a subset of observations that, compared to the rest of the dataset, appears to be inconsistent [12].

1) Definition

Even though the terms anomaly and outlier are used interchangeably in the literature [14], outlier is useful to identify unusual behavior compared to most sensor readings [15]. However, the specific definition can vary depending on the context and the methodologies upon which ODTs are based [3, 16]. Among the most used definitions in the literature is [27] “An observation, which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism” [4, 20, 25, 28, 29]; and a little more recently that of [16] “Those measurements that significantly deviate from the normal pattern of sensed data” [21, 25, 29–31].

Finally, based on the concepts analyzed and without intending to be exhaustive, this paper proposes the following definition for outliers:

“An outlier is an observation or set of observations obtained from one or more sensor nodes, which turn out to be inconsistent: (1) with other attributes perceived by other sensors of the same node at the same time; (2) with immediately preceding readings from the same sensor node; (3) with readings obtained from neighboring nodes; or (4) with a well-defined pattern of normal behavior”.

In this same sense, outliers can be of two dimensions based on the number of attributes that each data instance integrates: (1) univariate, when a data point has a single attribute and this can be detected as an outlier in relation to other data; and (2) multivariate, when a data point has multiple attributes and this can be identified as an outlier if some of its attributes together have anomalous values with respect to other data [15].

2) Sources of outliers in WSN

In a WSN, anomalies or outliers can originate from various sources, including variability in data collection devices, limitations in energy resources, accumulation of errors in numerous sensor nodes, and malicious attacks [10]. These outliers are primarily generated by noise or errors, events, and malicious attacks [3]. Noise or errors in the data are usually associated with defective sensor nodes and can affect data quality, requiring their removal or correction [1, 14, 15]. Events, representing changes in the real-world conditions [3], can yield outlier data that is crucial for decision making [9]. Malicious attacks that threaten the network’s security, can take control of sensor nodes and inject false data [15].

3) Types of outliers in WSN

Three types of anomalies are described in [11]: network, node, and data anomalies. Network anomalies involve unexpected variations in the number of packets traversing the network. Node anomalies originate from hardware or software defects, mainly caused by energy degradation. Data anomalies appear as unrealistic variations in the data captured by sensors. These latter can be classified into three types: temporal, spatial, and spatiotemporal [32, 33]. Temporal ones arise when comparing consecutive readings from the same node, spatial ones when comparing a node with its neighbors, and spatiotemporal ones combine both.

Other works like Chandola *et al.* [10] who classify anomalous data by their complexity into point anomalies (a single data instance is anomalous), contextual (a data instance is anomalous in one context but not in another), and collective (a set of data instances is anomalous). Rajasegarar *et al.* [34] categorizes them based on their degree of impact as first-order anomalies (where some observations on a node are anomalous), second-order (where all data from a node are anomalous in comparison to its neighbors), and third-order (where a set of nodes is anomalous compared to their neighbors). Likewise, in relation to the cause of anomaly on a local node [14, 16] identifies anomaly types 1, 2, 3, and 4; as well as local outliers (detected using only the data from a single sensor node) and global outliers (detected from a global perspective considering a group of sensor nodes), depending on the scope of detection.

C. Outlier Detection in WSNs

The detection problem refers to the process of searching for patterns in the data that deviate from the expected behavior [10], a definition cited in [2, 14, 35]. In the context of WSNs, the problem lies in detecting any abnormal behavior in the sensor data flows [15] or identifying data instances that deviate from the rest of the data patterns based on certain measurements [9]. In various ODTs, especially those based on classification, the expected normal behavior is initially modeled from a historical set of collected and previously labeled data instances.

Another important aspect of detection is the way of reporting outliers. It can be binary (normal or outlier) known as labels/scalar; another more commonly used way is the outlier scores, for which it is necessary to define a threshold for anomalies [3, 8].

Depending on the source, outlier detection may include: (1) fault detection regarding noise or errors; (2) event detection concerning events; and (3) intrusion detection pertaining to malicious attacks [3]. In this work, outlier detection in WSNs is addressed both in the context of fault detection and event detection.

1) Applications of outlier detection in WSNs

The use of outlier detection in WSN is closely related to the use of WSN in real-world implementations, as reviewed in the “WSN Applications” section of this paper. These applications primarily involve monitoring and tracking tasks. Specifically, outlier detection has been used in the following applications [3, 9, 15]: environmental monitoring, habitat monitoring, medical and health monitoring, industrial monitoring, target tracking, and structural monitoring. Other applications such as credit card fraud detection [10], intrusion detection [34], and smart cities [9] are also mentioned in the literature.

2) The availability of pre-labeled data

The availability or lack of pre-labeled data is one of the determining factors for selecting an appropriate detection technique. In fact, some techniques use pre-labeled data to train or validate an initial model that allows classifying new data instances as normal or outliers [34]. Therefore, the use of pre-labeled data divides ODTs into three categories [16]: (1) Supervised techniques that build classifiers requiring pre-labeled data to learn a normal and an abnormal model, and then classify a new data point as normal or outlier according to the model in which the data point fits. Supervised techniques are often closely related to multiclass classification techniques because they model both the normal and the outlier classes [15]; (2) Semi-supervised techniques that do not require data instances labeled as outliers but do require instances labeled as normal. These techniques are quite linked with one-class classification techniques as they model only one class, usually the normal class [15]; (3) Unsupervised techniques that do not require any pre-labeled data and use other criteria to identify outliers, most of the time a similarity measure between a point and its nearest neighbors.

3) Correlations

In the context of anomaly detection in WSNs, four important types of correlations are recognized: attribute correlations, temporal correlations, spatial correlations, and spatiotemporal correlations [2, 3, 14, 15]. The attribute correlation involves a relationship between different measurements from the same sensor, which can enhance the efficiency of detection models by reducing the dimensionality of the data. Temporal and spatial correlations refer to the predictive relationship between sensor readings at consecutive times and between geographically close sensors, respectively. On the other hand, spatiotemporal correlations combine the latter two aspects, highlighting predictive relationships between data collected at different nodes and at different times. Although the use of these correlations can enhance the effectiveness and efficiency of anomaly detection in WSNs [2], their incorporation in existing work is still limited, especially about attribute correlations.

4) Challenges and requirements for outlier detection in WSNs

Designing an ODT for WSNs presents various challenges related mostly to the unique characteristics of these types of networks. In fact, even though a variety of outlier detection solutions exist for traditional (wired) networks, these solutions cannot be directly transferred to WSNs [2]. Table I summarizes several challenges compiled from Ref. [2, 3, 15] that need to be considered when designing a suitable outlier detection solution for WSNs.

TABLE I. CHALLENGES TO OUTLIER DETECTION IN WSNS

Challenge	References		
	[3]	[2]	[15]
Resource constraints	√	√	√
High communication cost	√	√	√
Distributed streaming data	√		
Dynamic streaming datax	√	√	√
Dynamic network topology	√	√	√
Network heterogeneity	√	√	√
Large-scale deployment and network scalability	√	√	
Identifying outlier sources	√		√
High-dimension data		√	√

Likewise, based on the reviewed challenges, Table II condenses the requirements that according to the referred works should be integrated into an optimal outlier detection solution for WSNs.

TABLE II. THE PROPOSED REQUIREMENTS FOR AN OPTIMAL OUTLIER DETECTION SOLUTION FOR WSNs

Requirement	**Brief description**	**References**			
		[2]	**[14]**	**[15]**	**[9]**
Distributed structure	This requirement adopts inter-node collaboration for the detection process as opposed to the centralized structure.	√			
Online detection	Conducts data analysis in real-time, either in a continuous manner or in packets.	√		√	
Detection effectiveness	Exhibits a high detection rate and low false alarm rate.	√	√	√	√
Unsupervised techniques	Not dependent on labeled data or training phases.				√
Nonparametric methods	Does not assume an a priori data distribution.				√
Adaptability to dynamic data changes	Considers the dynamic nature of the data and its nonstationary distribution.	√	√	√	
Multivariate and high-dimensional data	Ensures capability of handling data instances with multiple attributes or even high dimensionality.	√		√	√
Dimension reduction	Incorporates dimension reduction subprocesses for the previous case.	√			
Energy efficient	Low power consumption in the detection process, low computational and communication complexity, and adequate memory usage.	√	√	√	√
Autoconfiguration with respect to the network topology	Facilitates adaptation to a dynamic network topology, which means that it must be robust to possible communication failures.		√	√	√
High scalability	It works well in WSNs with few sensor nodes or in dense deployments.				√
Use of correlations	It exploits the different types of correlations between network data, allowing it to effectively distinguish between errors and events.	√	√	√	√
The dynamic update of decision threshold	In models that return an outlier score, the decision threshold is flexible with respect to the dynamic nature of the data.			√	
Automatic parameter adjustment	Minimizes human intervention. by allowing automatic parameter adjustments.	√	√		

D. Outlier Detection in WSNs

Since their beginnings in the statistical community in the 19th century, various research communities have developed a variety of techniques for detecting outliers or anomalies in data [10]. Many of the algorithms developed are used with large datasets and assume substantial processing capabilities, which is why they cannot be directly applied to WSNs, as they are too computationally complex to run on sensor nodes [34]. Subsequently, many ODTs specifically designed for WSNs have emerged from various fields of study [3], which have examined the problem from multiple perspectives and strategies.

1) Approaches

Outlier detection approaches for WSNs are classified according to the structure of the model and the mode of operation. Structural approaches include centralized approaches in which data are sent to a central location for processing, allowing more complex detection algorithms but generating high energy consumption and communication overheads [1, 2, 14, 24]. Distributed approaches, on the other hand, perform detection at each sensor node, favoring real-time detection and efficient resource utilization, although they can limit the complexity of the algorithms and thus potentially limit detection accuracy [1, 2, 14, 15, 24]. Hybrid approaches seek to combine the advantages of both [33].

In terms of mode of operation, online approaches identify outliers in real time or near-real time, although they may have a higher rate of false alarms [1, 2, 13, 14, 16]. Offline approaches, on the other hand, collect observations over long periods of time before identifying outliers, taking advantage of historical data and more powerful methods, but they can be unsuitable for WSNs that require online processing [1, 13, 16]. Hybrid offline/online techniques integrate both approaches, using offline processing for initial model training and then online processing for real-time detection [22]. Centralized approaches are less common than distributed ones, and only a few of these distributed techniques perform online detection [2, 15].

2) Taxonomy

Different taxonomies are proposed for ODTs in WSN [1, 3, 9, 15], and although in [1] it is mentioned as a taxonomy for IoT, in practical terms WSN is implied. Table III compares the referred classifications.

We can observe that the main categories are techniques based on statistics, clustering, nearest neighbors, classification, and spectral decomposition. Other less referred are based on artificial intelligence, theoretical information, spectral techniques, and hybrid techniques.

Statistical-based techniques, in general, can be parametric (based on Gaussian or non-Gaussian distributions) or nonparametric (based on kernel or histogram). Hybrid statistical approaches are less common, as are parametric ones based on regression or mixed parametric distributions. Clustering-based techniques and nearest neighbor techniques commonly do not have subcategories, except for one taxonomy that divides the latter into distance-based and density-based. On the other hand, classification-based techniques appear richer in subcategories. In general, they are divided into SVM-based (One-class) and Bayesian Network-based (multiclass) with their various variants. Less common are rule-based classification techniques or single-class KPCA-based (Kernel PCA) techniques. Likewise, spectral decomposition-based techniques only include Principal Component Analysis (PCA) as a subcategory. Another relatively common category is that based on AI, which includes the subcategories Fuzzy Logic and NN.

The latter is also included as a subcategory of classification-based techniques. Finally, the categories based on theoretical information, spectral techniques, and hybrid approaches are the least common.

TABLE III. TAXONOMIES OF ODTS FOR WSNS

Main categories	Subcategories	Sub-Subcategories	References			
			[3]	[15]	[9]	[1]
Statistical-based	Parametric	Gaussian-based	√	√	√	√
		Non-gaussian-based	√	√	√	√
		Regression			√	
		A mixture of parametric distribution-based			√	
	Non-Parametric	Kernel-based	√	√	√	√
		Histogram-based	√	√	√	√
	Hybrid Approach				√	
Nearest Neighbor-based			√	√		√
	Distance				√	
	Density				√	
Clustering-based			√	√	√	√
Classification-based	Support Vector Machine (SVM)		√		√	√
	Bayesian Network					√
		Native Bayesian Network-based	√		√	√
		Bayesian Belief Network-based	√			
		Dynamic Bayesian Network-based	√		√	
	Neural Network (NN)				√	
	Rule				√	
	One-class	SVM		√		
		KPCA		√		
	Multi-class	Bayesian Network		√		
Spectral Decomposition-Based	Principal Component Analysis (PCA)		√		√	√
Artificial intelligence (AI)-based	Fuzzi logic			√		√
	NN			√		√
Information Theoretic					√	
Spectral Technique					√	
Hybrid-based						√

3) Pros and cons of detection techniques

Several pros and cons for each category of ODTs in WSNs have been identified in the literature.

a) Statistical-based

Pros: They are mathematically justified and can effectively identify outliers if a correct probability distribution model is acquired [1, 9, 15, 16]. After constructing the model, they do not require the actual data on which the model is based [1, 9, 16]. Nonparametric techniques are more flexible and autonomous because they do not make any assumptions about the distribution characteristics of the data [9, 16]. They use temporal correlations to determine the presence of an outlier based on a sudden change in the data distribution [1, 15]. Cons: Parametric techniques require prior knowledge that is often not available or costly to compute in many real-life WSN applications [1, 2, 9, 15, 16]. Nonparametric statistical models are not as suitable for real-time applications, and the computational cost of handling multivariate data is higher [1, 15]. Histogram-based techniques are efficient for univariate data and relatively simple to implement, but they cannot capture the interactions between different attributes of multivariate data, and it is not easy to determine an optimal bin size for constructing the histogram [2, 9, 15, 16]. Kernel function-based techniques can scale well on multivariate data, but they have potentially quadratic time complexity in terms of the data size [9, 16]. The selection of the threshold depends on the application and is a difficult task, especially for a continually changing dynamic environment [2].

b) Nearest neighbor-based

Pros: They do not require assumptions about the data distribution [1, 9, 15, 16]. They are unsupervised techniques, meaning that no pre-labelled data are needed [15, 16]. They are effective in detecting outliers in data with nonlinear patterns [9, 16]. Their application in different types of data is simple and mainly requires defining a suitable distance measure [1, 15]. The use of Euclidean distance is a good option for univariate and multivariate constant features [9]. Cons: The proper choice of input parameters may be a difficult task [1, 9, 16]. The threshold value used in this technique is critical and must be carefully chosen to avoid a high false negative rate [1, 9, 15]. Defining distance measures between data instances can be challenging in sensor data [16]. Density-based techniques are not entirely efficient because some approaches may be sensitive to variations in the local density of data points, which may result in less accurate outlier detection in datasets with nonuniform densities [15]. In some multivariate datasets, it may be computationally expensive to compute the distance between data instances, which may affect the scalability of these techniques [1, 2, 9, 15, 16]. They may require high communication overhead and significant energy consumption [9]. Some strategies may only identify outliers with severe divergence and possibly overlook less extreme outliers [9].

c) Clustering-based

Pros: Clustering-based techniques do not require prior knowledge about the data distribution [16]. They can be used in an incremental model, allowing for the incorporation of new data instances and the detection of outliers [9, 15, 16]. They are unsupervised techniques, making them flexible for different datasets [9, 15, 16]. They are suitable techniques for anomaly detection from temporal data [9, 15]. The test phase is agile because the number of clusters with which each test instance must be compared is a small constant [9, 15]. Cons: The effectiveness of these techniques largely depends on their ability to capture the cluster structure of normal instances [9, 16]. Like nearest neighbor-based techniques, computing the distance between multivariate data instances using clustering-based techniques is computationally costly [2, 9, 16]. Most methods are byproducts of clustering and are not optimized to detect outliers [9]. Some clustering techniques are not suitable for WSN applications due to their dependence on the choice of cluster width [2]. Clustering techniques are not effective in adapting to ongoing changes in data streams over time. Recent models have attempted to solve this problem through incremental learning methods, but the computational cost of these methods is very high [2]. Some clustering algorithms force each instance to be assigned to some group, which can lead to outliers being assigned to a large cluster and being considered as normal instances [9].

d) Classification-based

Pros: By constructing a classification model, they provide optimal and maximum outlier identification [1, 15, 16]. In particular, multiclass approaches apply powerful algorithms that can differentiate instances of various classes [9]. The test phase is fast since each test instance only needs to be compared with the precalculated model [9, 16]. It does not require a statistical model or estimated parameters [1, 15]. It solves the problem of multidimensional data [15]. Some recent one-class techniques are unsupervised, so they do not require labeled training data [1]. Cons: SVM-based techniques have high computational complexity due to quadratic optimization and the choice of suitable kernel functions, and are inefficient for online approaches [2, 15, 16]. Some classification techniques, especially SVM, require parameter selection, which can limit the capacity of the solution and increase the need for human intervention [2]. Techniques based on Bayesian networks have difficulties in learning an accurate classification model when there are a larger number of variables [16]. Bayesian networks are effective in detecting correlations and dependencies in sensor data and attributes, but their ability to handle large multivariate datasets is limited [2]. As new data arrive, the model needs to be updated [15]. Multiclass classification techniques assume the availability of accurate labels for varied normal classes, which is difficult to obtain [9]. Labeling each test instance becomes a drawback if instead an outlier score is desired [9].

e) Spectral decomposition-based

Pros: PCA-based techniques can be applied to datasets with many dimensions [1, 9, 16]. They can operate in unsupervised mode [9, 16]. They capture the normal pattern of the data and detect outliers optimally [16]. The dimensionality reduction can be applied as a preprocessing step before applying outlier detection methods in the transformed space [9]. Cons: The selection of the right principal components to estimate the correlation matrix of normal patterns is computationally expensive [1, 9, 16]. They are useful only if outliers and normal instances are highly distinguishable in the reduced space [9, 16].

f) Artificial intelligence-based

Pros: They are rule-based techniques, providing a clear understanding of the detection process, flexibility to adapt to different contexts, and the possibility of incorporating expert knowledge [1]. They can generalize despite having limited, noisy or fragmented data [1, 15]. There is no need to retrain the system when new data or rules are added [1, 15]. Cons: Developing a model from a fuzzy system requires fine tuning and prior simulation before being operational [1, 15]. Storing the rule base can require large amounts of memory because the number of rules increases exponentially with the number of variables; also, constantly traversing them can slow down the detection process [1, 15]. Adding spatial and temporal correlation to the decision-making process further increases the number of rules [1, 15].

4) Evaluation conditions of outlier detection techniques

The main conditions related to the evaluation of an ODT in WSN are dataset selection, performance metrics, and complexity analysis.

a) Datasets

They are fundamental for training and evaluating the performance of the technique. These datasets contain a collection of measurements or events collected from real-world implementations [21], although in other cases they are also synthetically generated [8]. In Ref. [14], two datasets are proposed that meet the typical nonstationary conditions of the data generated in a WSN application for environmental condition monitoring: Intel Berkeley (IBRL) from Intel Berkeley Research Lab and Grand-St-Bernard (GSB) from SensorScope Project. These real datasets are among the most used in the literature, either natively or modified, partially or completely, for the training and evaluation of anomaly detection proposals, as shown by a review conducted in [35].

b) Performance metrics

The performance metrics of an ODT for WSN seek to demonstrate the effectiveness of said technique [2]. These metrics, generally, are built based on four possible outcomes in a binary classification: the outliers (TP) and normal cases (TN) that are correctly classified and the outliers (FN) and normal cases (FP) that are not [7]. Maintaining a high detection rate (DR) and a low False

Alarm Rate (FAR) or False Positive Rate (FPR) while consuming the least number of resources is the essence of the evaluation of a detection technique [3]. The DR, also known as the True Positive Rate (TPR) [14], represents the proportion of outliers that are correctly identified; while the FPR indicates the proportion of normal data incorrectly identified as outliers [15]. On the other hand, to evaluate the trade-off between FPR and TPR, the ROC (Receiver operating characteristic) curve is used, a 2D graph where the best performance is achieved with a TPR of 1 and an FPR of 0; besides, the greater the area under the ROC curve (AUC: area under ROC curve), the better the performance of the ODT, considering that an AUC value of 1 indicates 100% accuracy and an AUC value of less than 0.5 indicates a performance worse than the random assignment of labels [14].

Three other commonly used metrics are accuracy, precision, and recall; accuracy (ACC) is the ratio between correctly predicted observations (both atypical and normal) (TP+TN) and the entire dataset; precision is the proportion of true outliers (TP) present in all values detected as such (TP+FP), and recall, which is equivalent to DR [17]. Metrics such as sensitivity, specificity, and F1-score [36] are also mentioned in the literature.

c) *Complexity analysis*

The complexity analysis of an ODT seeks to demonstrate the efficiency of said technique [2]. Due to resource limitations in a WSN, it is crucial to analyze the detection algorithms to determine their complexity in computational, memory, and communication aspects when the number n of instances varies. Big O notation is a commonly used method for this evaluation, whose purpose is to identify the upper limit of the algorithm's complexity as O(n), also allowing for the study of how it evolves as the number (and potentially the dimension) of the data vectors used in the elaboration of models for anomaly detection in WSN increases [14]. Recent works include other analyses such as space complexity [30], time complexity [18], and asymptotic complexity [8].

IV. The analysis of Outlier Detection Proposals for WSN

The detailed analysis was conducted on 33 selected studies, spanning from 2018 to 2023, which focused on ODTs in WSNs; the analysis adhered to the questions raised in Section II.

A. *Characteristics of Proposals for Outlier Detection in WSNs*

The diversity of solutions found is evident: algorithms, techniques, and frameworks based on threshold, boundary, clustering, statistical, classification, artificial intelligence, nearest neighbor, isolation, and hybrid approaches.

The Decision Support System (DSS) methodology [6] and the Omnibus technique [37] are threshold-based approaches. The DSS methodology uses an integrated detection module to inform irrigation decisions in agriculture, while the Omnibus technique enhances the Transform-Based Contextual (TACO) framework by allowing definitions of unidimensional and multidimensional outliers, employing varied transmission window models, using Locality-Sensitive Hashing (LSH) to ensure efficiency and predictable accuracy, and incorporating different measures of similarity.

Boundary-based techniques employ Support Vector Data Description (SVDD) to tackle the detection problem. The TSVDD technique [21] applies a Toeplitz matrix and a model selection strategy to reduce algorithm complexity and avoid underfitting and overfitting. Meanwhile, the Novel Spatiotemporal and Attribute SVDD (N-STASVDD) technique [18] considers independent and identically distributed attributes and uses coresets to decrease computational complexity and energy consumption, surpassing the centralized approach. Lastly, the Improved density-compensated SVDD (ID-SVDD) technique [23] enhances the original SVDD by using the data density distribution and the Parzen-window algorithm to efficiently map data from sparse to high-density areas, thereby achieving more effective outlier detection.

The Peak-Searching Algorithm (PSA) [17] and Fault detection based on Participation Degree (FDP) [30] are clustering-based techniques. PSA uses Bayesian optimization to identify probability peaks in data and use them as initial points in EM (expectation maximization) and k-means algorithms, which enhances the precision of clustering and reduces the number of necessary iterations. Conversely, FDP employs the degree of participation in hierarchical clustering to establishes relationships between instances, eliminating the need for training with labeled data. The FDP technique uses the Agglomerative Hierarchical Clustering (AHC) algorithm and the Nearest Neighbor Boundary (NNB) to create a dendrogram tree and enhance the algorithm's efficiency in global outlier detection.

Statistical-based approaches like Copula-Based Probabilistic Multivariate (CBPM) [20] and Multivariate Outlier Detection (MOD) [38] have also been proposed. The CBPM technique considers the dependence between measurements surpassing existing statistical methods and is implemented in three stages: estimation, distributed detection, and outlier classification. On the other hand, the MOD technique focuses on improving accuracy in data aggregation in forest monitoring WSNs, using statistical analysis to identify and eliminate outliers before aggregation, resulting in a more precise dataset.

Distributed online OCSVM (doOCSVM) and Sparce doOCSVM [39], Distributed Outlier Detection Scheme (DODS) [29], and ST-CE-CKDOT (improved CKODT based on spatial-temporal technique and centered ellipsoidal scheme) [40] are classification-based approaches. DoOCSVM and Sparce doOCSVM use an approximate random function and apply stochastic gradient descent to minimize cost functions, allowing decentralized implementation. DODS uses Bayesian classifiers at each node, considering multiple types of data, temporal correlation, and remaining energy, using the Maximum a Posteriori (MAP) concept to determine optimal classes. Finally, the ST-CE-CKDOT technique identifies and locates damage in water pipe systems using a single-class classification technique, a centered ellipsoidal technique, and spatial-temporal correlations to distinguish between events, noise, and faulty sensors.

Some solutions with Artificial Intelligence, Nearest Neighbors, and isolation-based approaches were also found. A technique based on Artificial Neural Networks (ANN) to detect and correct outliers in temperature measurements in smart building wireless sensors is proposed in [41] using predictions from the ANN model trained with historical data. In Ref. [42], the Outlierness Factor based on Neighbourhood (OFN) technique is proposed to detect outliers in WSN using spatio-temporal correlation. By calculating distances, assigning weights, and determining the outlierness factor, OFN distinguishes between sensor errors and genuine events. Meanwhile, box plot-sampled iForest (BS-iForest) [43] combines the box plot method and the Isolation Forest algorithm to detect anomalies in WSN data, selecting the best isolation trees according to their fitness and using similar data points to evaluate anomalies, improving stability and detection performance.

Finally, hybrid approaches were also part of the findings. Fault-Tolerant Anomaly Detection method (FTAD) [8] is a fault-tolerant anomaly detection method that uses spatial-temporal correlation through statistical methods such as the Pauta Criterion Method (PCM) and the use of thresholds. In contrast, the distributed and real-time model proposed in [22] is based on OCPCC (One class principal component classifier) using spatial correlations and Candid Covariance-Free Incremental PCA (CCIPCA) to improve efficiency and reduce computational complexity. Combined Kernelized Outliers Detection Technique (CKODT) [5], a hybrid model, merges KFDA and One Class SVM (OCSVM) for water pipe monitoring in WSNs, while [44] presents an approach based on Optimum-Path Forest (OPF) and meta-heuristics. Techniques such as Temporal Outlier Detection (TOD) and Spacial Outlier Detection (SOD) [45], which employ statistical and graph-based approaches, enhance the detection of temporal and spatial outliers. In Ref. [46], an isolation-based nearest-neighbor ensembles (iNNE) framework is proposed using local detectors and weighted voting. In-Network Contextual Outlier Detection on Edge (INCODE) [31] is a framework for contextual detection in WSNs, using Edge Computing and Google PageRank. Online Linear Weighted Projection Regression (OLWPR) [36] detects anomalies in three phases: data compression, prediction, and anomaly detection. Deep belief network online quarter-sphere SVM (DBN-OQSSVM) [47] combines deep belief networks and online quarter-sphere support vector machines for anomaly detection in WSNs. The Local Outlier Detection Algorithm (LODA) [25] is a decentralized technique based on time series and adaptive Bayesian networks. In Ref. [7], an ensemble learning approach is proposed, combining Decision Tree, Naive Bayes, and K-Nearest Neighbor through a Random Forest. Hypergrid based Adaptive Detection of Faults (HADF) [48], a distributed method that uses hypergrid and statistical analysis to identify sensor data faults. The approach in [49] uses an outlier detection framework based on collaboration between mobile edge and cloud, including Fast angle-based outlier detection algorithm (FastABOD) and f-SVDD (SVDD + fuzzy theory). In Ref. [50], anomaly detection in physiological data is addressed with an integrated system combining correlation coefficient, random forest, dynamic threshold, and majority voting. CESVM-DR (CESVM + Dimension Reduction) [51] is a lightweight approach that combines CCIPCA and OCSVM based on CESVM (Centred-Ellipsoid SVM), seeking to reduce computational complexity and enhance detection accuracy. PiForest [4] based on iForest is a technique for anomaly detection in environments with limited resources and streaming data. In Ref. [28], a technique is presented that combines time series analysis, entropy, and classification using random forests. Finally, the technique in [52] uses Generative Adversarial Networks (GANs), an unsupervised learning approach, to detect outliers in WSNs, implementing two neural networks and autoencoders trained through the Adam optimizer.

Table S1 proposals, considering name, base algorithm on which it is founded, group or taxonomy to which it belongs, approach, number of variables or dimensions experimented with, correlations that were considered in the proposal, and sources of outliers addressed. In this table, many hybrid techniques stand out, that is, techniques based on two or more methods and their derivations. From this descriptive table, Table S2 was generated for the quantification of the frequency of the characteristics observed in the analyzed detection proposals. Similarly, Table IV synthesizes the proportion of techniques according to the taxonomy. On the other hand, based on the same information, Table V quantifies the number of times the base algorithms and methods were used for different techniques collected.

Table IV shows that hybrid approaches (54.55%) are the preferred solutions, followed by classification and boundary-based approaches, with 9.09% in each case. Likewise, Table V shows that the most utilized algorithms, regardless of the previous, are SVDD and OCSVM, boundary and classification-based algorithms, respectively; followed by random forest and isolation forest. On the other hand, as illustrated in Table S2, solutions based on distributed approaches (60.6%) prevail over centralized solutions (39.4%). Similarly, online (66.7%) and multivariate (63.6%) techniques are preferred over offline approaches or those using univariate or bivariate data.

The use of some type of correlation has been an important component for more than 80% of the proposed solutions. 36.4% exploit correlations between attributes, especially in multivariate cases. To a lesser extent, some techniques take advantage of spatial (12.1%), temporal (6.1%) and spatiotemporal (18.2%) correlations. Only three studies used all possible correlations. Regarding the source of the outliers, 21 studies (63.6%) exclusively focus on outliers stemming from faults, while the remainder incorporate the detection of outliers triggered by events (event detection).

Finally, based on the previous data, the main characteristics of the outlier detection proposals for WSNs are hybrid techniques based on algorithms that take advantage of classification or that learn to distinguish normal classes, with a distributed and online approach, that exploit the spatiotemporal correlations of the data and, in addition to detecting faults in the data, also detect events of interest in the monitored area.

TABLE IV. THE DISTRIBUTION OF ODTS IN WSNS BY TAXONOMY

Taxonomy	References	Papers	
		Freq.	**Percentage**
Hybrid	[4, 5, 7, 8, 22, 25, 28, 31, 36, 44–52]	18	54.55%
Classification-based	[29, 39, 40]	3	9.09%
Boundary-based	[18, 21, 23]	3	9.09%
Clustering-based	[17, 30]	2	6.06%
Threshold-based	[6, 37]	2	6.06%
Statistical-based	[20, 38]	2	6.06%
Nearest Neighbor-based	[42]	1	3.03%
Isolation-based	[43]	1	3.03%
AI-based	[41]	1	3.03%
Total		**33**	**100%**

TABLE V. METHODS USED AS BASIS FOR OUTLIER DETECTION PROPOSALS IN WSNS

ID	Method	Uses	F	ID	Method	Uses	F	ID	Method	Uses	F
1	ABOD	[49]	1	**16**	HGDB	[48]	1	**31**	OCPCC	[22]	1
2	AHC	[30]	1	**17**	HIA	[45]	1	**32**	OCSVM	[5, 39, 40, 51]	4
3	BAYES CLASSIFER	[29]	1	**18**	INTERVAL	[8]	1	**33**	OPFC	[44]	1
4	BAYESIAN NETWORK	[25]	1	**19**	ISOLATION FOREST	[4, 46, 43]	3	**34**	PAGERANK	[31]	1
5	CCIPCA	[22], [51]	2	**20**	KFDA	[5]	1	**35**	PCA	[4, 36]	2
6	CKDOT	[40]	1	**21**	K-MEANS	[17]	1	**36**	PCM	[8]	1
7	COEFFCIENT CORRELATION	[50]	1	**22**	KNN	[46, 7]	2	**37**	PSA	[17]	1
8	COPULA-F	[20]	1	**23**	LSH	[37]	1	**38**	PWD	[23]	1
9	DBN	[47]	1	**24**	LWPR	[36]	1	**39**	QSSVM	[47]	1
10	DECISION TREE	[7]	1	**25**	MAHALANOBIS DIST.	[23]	1	**40**	RANDOM FOREST	[7, 28, 50]	3
11	DYNAMIC THRESHOLD	[50]	1	**26**	MAJORITY VOTING	[50]	1	**41**	STATISTICAL DETECTOR	[48]	1
12	EM	[17]	1	**27**	MAP CONCEPT	[29]	1	**42**	SVDD	[21, 18, 23, 49]	4
13	ENTROPY	[28]	1	**28**	NAIVE BAYES	[7]	1	**43**	TIMES SERIES ANALYSIS	[28]	1
14	FUZZY THEORY	[49]	1	**29**	NEURAL NETWORK	[41]	1	**44**	TRFFM	[21]	1
15	GPM	[45]	1	**30**	NNB	[30]	1	**45**	N/D	[6, 42, 38, 52]	4

B. Conditions and Metrics in the Evaluation of Outlier Detection Proposals in WSN

As mentioned in Section III.D.4, datasets, are essential requirements for training and evaluating the performance of the technique, determine the effectiveness of the proposal in the detection process, and quantify its efficiency in the use of computational resources. In this context, the review of the 33 reference sources highlights the datasets and performance metrics used in the development of experiences as well as the complexity measures for their efficiency in resource-limited WSN environments.

1) Datasets

The outlier detection proposals found in the selected papers used several datasets. Table VI shows a larger number of works using IBRL, followed by GSB; this is primarily because these datasets are public, but more importantly, they were extracted from real-world implementations. Papers such as [4, 17, 18, 30, 39, 42, 45] combine the use of real datasets with synthetic datasets. This latter, in most cases, is generated entirely based on certain statistical distributions, combining the injection of artificial anomalies, manual labeling, and even normalization processes. Cases such as [29] generate synthetic datasets based on real datasets, whereas in [38] only the former type is used. In Ref. [6], static datasets are combined with others generated in real time. Finally, some studies either collect data [40] or generate it [8] through computer simulations.

TABLE VI. DATASETS USED IN THE PERFORMANCE EVALUATION OF SELECTED PROPOSALS

Dataset	Papers	Freq.
Intel Berkeley (IBRL)	[17, 20–22, 25, 28, 29, 31, 36, 37, 42, 44–46, 48, 49, 51]	17
Grand-St-Bernard (GSB)	[18, 20–23, 44, 51, 52]	7
Multiple Intelligent Monitoring in Intensive Care (MIMIC)	[50]	1
Activity Recognition based on a Multi-sensor data fusion (AReM)	[7]	1
Intelligent Sensors, Sensor Networks & Information Processing (ISSNIP)	[46]	1
Lausanne Urban Canopy Experiment (LUCE)	[48, 51]	2
Wireless Indoor Positioning Data Set (WILDS)	[30]	1
Patrouille des Glaciers (PDG)	[50]	1
Networked Aquatic Microbial Observing System (NAMOS)	[50]	1
Weather Data of University of Washington 2002	[37]	1
Italian Industrial Production Index from tsoutlier R package	[45]	1
Campus Climate and Resilience Study (Campus-CRS)	[43]	1
Breast Cancer Wisconsin (BCW)	[43]	1
Other various	[4, 6]	2

2) *Metrics and complexity analysis*

The acronyms detailed in Table VII are used to refer to performance metrics and complexity analysis of detection proposals. Further details related to the evaluation of such works and the efficiency and effectiveness results obtained are shown in Table S3.

As we can observe in Table VII, the analyzed papers use up to 20 different performance metrics. Among them, Accuracy is the most used (48.5% of the papers), followed by Area Under the Curve, F1-score, False positive rate, Receiver Operating Characteristic, Precision, Recall, True Positive Rate, Detection rate, and False Alarm Rate. Likewise, in terms of complexity analysis, 36.4% of the studies considered computational time as an important issue, followed by computational and memory complexities.

TABLE VII. ACRONYMS OF THE PERFORMANCE METRICS AND COMPLEXITY ANALYSIS USED AND THEIR FREQUENCY

Performance metrics		Freq.	Complexity analysis		Freq.
Acronym	Description		Acronym	Description	
ACC	Accuracy	16	CT	Computational time	12
AUC	Area Under the Curve	9	CPC	Computational complexity	8
F1	F1-score or F-measure	8	MC	Memory complexity	5
FPR	False positive rate	8	TC	Time complexity	4
ROC	Receiver Operating Characteristic	7	CMC	Communication complexity	4
PRE	Precision	7	SC	Space Complexity	2
REC	Recall	7	PC	Power consumption	1
TPR	True Positive Rate	6	ET	Execution time	1
DR	Detection rate	5			
FAR	False Alarm Rate	4			
FNR	False negative rate	3			
DA/DAR	Detection Accuracy / Detection Accuracy Rate	3			
PE	Percentage Error	1			
RMSE	Root Mean Square Error	1			
SEN	Sensitivity	1			
SPE	Specifcity	1			
GME	G-mean	1			
TNR	True negative rate	1			
	Forecast ACC	1			
	Aggregated Data Accuracy	1			

Table S3 shows that regarding the effectiveness of the proposed techniques, detection capabilities in most cases range between 85%–100%. It can also be observed [41, 38] experiments, experimentation was not precisely aimed at determining the effectiveness of outlier detection but at measuring other aspects related to the proposed technique. In studies [40] and [50] no quantitative results of any kind are reported.

In relation to the complexity analysis, the predominant use of Big O notation, as referred to in Section III.D.4, is confirmed. However, it is also evident that about half of the proposals excluded any type of analysis in this regard. Most of them being of hybrid type.

In addition to the metrics, the conditions under which the experimental processes occurred are of interest. In fact, Table VIII shows that 84% of outlier detection proposals for WSN were tested in simulated environments. For this, the use of software was vital; in fact, many studies made explicit the hardware and software resources that allowed such simulations, with Matlab (36.4%) and Python (24.2%) being the most mentioned, especially when it came to hybrid techniques (see Table IX). Likewise, 42.4% of the detection proposals assumed a hierarchical cluster-based network structure, discussed in Section III.A.1 (see Table X).

TABLE VIII. FREQUENCY OF EVALUATION TYPES BY TAXONOMY

Taxonomy	Testbed	Simulation	Real implementation	Numerical analysis	Simulation and Real Implementation
Hybrid		15	1	1	1
Classification-based		3			
Boundary-based		3			
Clustering-based		2			
Threshold-based	1	1			
Statistical-based		2			
Nearest Neighbor-based		1			
Isolation-based		1			
AI-based					1
Total	1	28	1	1	2
%	3.0%	84.8%	3.0%	3.0%	6.1%

TABLE IX. FREQUENCY OF SOFTWARE USED BY TAXONOMY

Taxonomy	Matlab	Python	Others	None or N/D
Hybrid	5	7	1	5
Classification-based	2		1	
Boundary-based	3			
Clustering-based				2
Threshold-based				2
Statistical-based	1		1	
Nearest Neighbor-based		1		
Isolation-based			1	
AI-based	1			
Total	12	8	4	9
%	36.4%	24.2%	12.1%	27.3%

TABLE X. FREQUENCY OF NETWORKS STRUCTURE USED BY TAXONOMY

Taxonomy	Flat	Hierarchical Cluster-based	Hierarchical Group and community based	Hierarchical architecture with three layers	N/D
Hybrid	5	6	1	1	5
Classification-based	1	2			
Boundary-based		1			2
Clustering-based		1			1
Threshold-based	1	1			
Statistical-based		2			
Nearest Neighbor-based		1			
Isolation-based					1
AI-based	1				
Total	8	14	1	1	9
%	24.2%	42.4%	3.0%	3.0%	27.3%

C. Areas of Application for Outlier Detection in WSN

In terms of application areas, very few works explicitly refer to this in the context of their proposed solution. In many studies, the application area was deduced from the nature of the dataset used, such as those that used datasets with environmental or human activity information; however, applying this logic was not possible in all cases. Interestingly, monitoring tasks served as the common denominator across most of the studies. Details are shown in Table XI, which indicates that environmental monitoring applications are of the greatest interest, at least based on the data used for testing, including agricultural monitoring, water quality, and forest fires; while structural and health monitoring applications are represented on a much smaller scale.

D. Detection Techniques Limitations

Outlier detection in WSN is a constantly evolving area of research, which is evidenced in the diversity of techniques proposed in the reviewed works. However, in most cases, we could infer some common limitations, including sensitivity to parameters, computational complexity, scalability, and dependence on specific assumptions.

The DSS methodology [6] could be sensitive to datasets containing noise that manifests as multiple values. Moreover, its centralized architecture may be less scalable and susceptible to external factors such as extreme weather conditions. Some potential limitations of TSVDD [21] include the need for a representative training dataset and sensitivity to model parameters. Also, the use of the Toeplitz matrix to reduce time and space complexity in a real application may impact the algorithm's accuracy and efficiency.

In PSEM and PSk-means [17], PSA uses BO, and this dependence implies that a prior design and a Gaussian distribution are critical for its efficiency. Furthermore, the PSA is centralized, making it unsuitable for highly distributed WSNs where data analyses are performed at each sensor node. The CKODT technique [5] combining KFDA and OCSVM may face challenges in WSN environments with limited resources, as it increases computational complexity and requires more computing capabilities, energy, and storage, which could affect the battery life of sensor nodes. Although the N-STASVDD method [18] reduces computational complexity and energy consumption compared to other methods, there may still be a computational and energy cost associated with real-time anomaly detection. Possible limitations of the Copula-based method [20] are the selection of an appropriate threshold and the determination of the dependence between captured measures. For OPF [44], a possible limitation is that it requires fine tuning of parameters such as the anomaly threshold. Additionally, although it is compared with other methods, no information is offered about the proposal's performance and effectiveness in WSN scenarios.

TABLE XI. AREAS OF APPLICATION OF THE DETECTION PROPOSALS

Application area	Papers	Freq.
Agricultural Monitoring (Precision Agriculture)	[6]	1
Monitoring of water pipes	[5, 40]	2
Water quality monitoring	[23]	1
Structural monitoring in buildings	[41]	1
Health monitoring	[50]	1
Forest fires	[4, 38]	2
Based on the data: Environmental monitoring	[17, 18, 20–22, 28, 29, 36, 37, 42, 44–46, 48, 49, 52]	16
Based on the data: Human Activity Monitoring	[7]	1
N/D	[8, 25, 30, 31, 39, 43, 47, 51]	8

Selecting suitable parameters for the approximate random function and the stochastic gradient descent may be a challenge and can affect the performance of the doOCSVM and Sparse doOCSVM techniques proposed in [39]. The application of TOD and SOD [45] to multivariate data could reduce its effectiveness and efficiency due to difficulties in establishing an appropriate detection threshold. ID-SVDD [23] may not be suitable for real-world applications due to its centralized approach, and there is also a lack of information regarding its efficiency in using computational resources.

In the DODS approach [29], a possible limitation could be its focus on detecting outliers at the level of individual nodes without considering the network's contextual or global information. On the other hand, no details are provided about its efficiency, considering the use of a Bayesian classifier and the MAP concept that might demand significant computational resources. For the FDP-based approach [30], appropriate parameter selection and adaptation to different data distributions can be challenging. In the case of OFN [42], the selection of suitable parameters and sensitivity to variations in the density function can also affect its performance and effectiveness. The MOD approach [38] not only has a centralized architecture that can affect scalability and efficiency, but also requires the selection of appropriate parameters. The ANN-based technique [41] could face problems of computational complexity and energy consumption due to the nature of neural networks. The HADF method [48] may also have limitations in terms of computational complexity and resource requirements, which could affect the battery life of sensor nodes.

In the case of ST-CE-CKDOT [40], some limitations in terms of accuracy and reliability of detection may be due to factors such as sensor quality and complexity of the piping system. Furthermore, the implementation and maintenance of this network can be costly and require careful planning. On the other hand, the INCODE technique [31] has limitations due to its dependence on community formation and consensus mechanism, influencing the quality of outlier detection. It is also necessary that the community nodes have sufficient processing and memory capacity to perform data summarization and information transmission. Likewise, the omission of temporal correlations in the analysis could restrict the effectiveness of INCODE in outlier detection in certain contexts.

The OODS framework [37] provides a solid approach to outlier detection, but defining the appropriate window size could be a difficulty and there is a lack of information about the efficient use of computational resources. The OLWPR technique [36] may require a significant amount of computational resources to run in real time at each node, which could generate scalability difficulties. The DBN-OQSSVM method [47] could require significant computing capability to scale to a distributed approach. Since LODA [25] does not consider spatial or temporal correlations with neighbors to detect outliers, it might not be able to detect certain types of outliers that could be detected through the use of correlations. The ERF algorithm [7] relies on an ensemble learning approach that combines multiple base classifiers, which may increase computational complexity and execution time compared with a simpler approach.

Some potential limitations or challenges that might arise in the outlier detection framework based on collaboration between the mobile edge and the cloud [49] include the demand for periodic updates and optimizations of the detection model to preserve its accuracy and reliability, as well as the need to maintain a balance between performance and energy consumption at the edge nodes. The technique proposed in [50] is based on the analysis of historical data gathered from multiple medical sensors to identify anomalies and dynamically adjust the threshold value. However, this dependence could limit its ability to promptly adapt to changes in the patient's medical conditions or the surrounding environment. In addition, the technique relies on a centralized approach, which could restrict its scalability. Lastly, the effectiveness and efficiency of the proposed technique are not sufficiently demonstrated.

Determining the optimal number of principal components when applying the CCIPCA algorithm for data dimensionality reduction can represent a challenge and simultaneously influence the accuracy of outlier detection. Also, the lack of leveraging of spatiotemporal correlations in the data is an additional limitation for the proposed CESVM-DR scheme [51]. In PiForest [4], the use of PCA can have an impact on the accuracy of outlier detection in datasets with complex structures or nonlinear correlations among features. Furthermore, although the proposed method is designed to handle real-time data, it may still have difficulties handling large volumes of data or quick data flows.

The technique proposed in [28] integrates multiple approaches and techniques, which could increase its complexity and computational requirements. In addition, the proposed technique does not yet address the problem of detecting malicious or faulty nodes, which can negatively affect the accuracy of outlier detection. The BS-iForest algorithm [43] can be sensitive to the choice of parameters, such as the number of trees and the maximum tree depth. Furthermore, the process could be more complex than the traditional isolation forest algorithm.

Finally, potential limitations of the proposed GAN method [52] could include the computational complexity involved in implementing the GAN, the processing time required, and the possibility of obtaining false positives or negatives in outlier detection.

A more detailed analysis of these limitations, along with the advantages identified for each of the analyzed proposals, is shown in Table S4.

In an adjacent line of research, contemporary studies employing bio-inspired algorithms suggest viable solutions to multiple limitations encountered in reviewed outlier detection techniques, including high computational complexity, sensitivity to parameter settings, and lack of scalability. For instance, the Prairie Dog Optimization (PDO) algorithm [53] effectively balances both accuracy and scalability and demonstrates robustness in dynamically adapting to dataset variations through online and incremental learning. Similarly, the Modified Elite Opposition-Based Artificial Hummingbird Algorithm (m-AHA) [54] stands out for its ability to self-tune operational parameters, a key element for efficacy in WSNs. Likewise, the Dwarf Mongoose Optimization Algorithm (DMO) [55] introduces a metaheuristic that is particularly beneficial in scenarios with a scarcity of historical data, addressing another common limitation.

V. Discussion

Certainly, outlier detection in WSNs has garnered considerable attention in the literature due to its critical importance for ensuring data quality and reliability in mission-critical applications. Various approaches have emerged to tackle this issue, yet significant challenges remain both in terms of theoretical development and practical implementation.

Regarding the proposed techniques, a trend towards the adoption of hybrid models is observed (54.5%), with a focus on distributed (60.6%) and online (66.7%) approaches aimed at detecting anomalies locally at sensor nodes, thus minimizing communication overhead. However, this localized approach is unable to capitalize on global correlations, which could enhance detection accuracy. Notably, only 18.2% of the proposed studies leveraged spatiotemporal data correlations, and a mere 9.1% also exploited correlations between attributes, particularly in multivariate solutions (63.6%). Hence, there arises a need to incorporate collaborative elements among neighboring nodes or with central nodes to optimize the model's efficacy.

At the algorithmic level, different strategies have been employed to reduce computational complexity, such as dimensionality reduction and the use of semi-supervised learning over unsupervised methods, which often suffer from higher rates of false positives. Alternatively, supervised techniques, although potentially more accurate, present the challenge of obtaining suitable labeled data. Moreover, most existing methods do not explicitly address adaptability to abrupt data changes.

Concurrently, the practical implementation of these techniques faces significant challenges. A paramount issue is the seamless integration with the specific infrastructure and topology of each sensor network. Centralized techniques (39.4%), for instance, although effective in certain scenarios, encounter scalability issues in more extensive networks. This demands meticulous mapping of distributed methods to pre-existing node clusters and roles. Additionally, energy efficiency remains a critical aspect; the application of processing or communication-intensive algorithms could rapidly deplete batteries, requiring a careful balance between detection accuracy and energy sustainability. Other challenges encompass the efficient training and updating of models, as well as adaptable parameter configuration. Moreover, deployment in real-world environments introduces environmental variables that may reveal limitations not previously identified in earlier research or simulations.

Overall, although numerous valuable contributions have been made, a comprehensive solution that satisfactorily meets all key requirements for optimal outlier detection in WSNs remains elusive from both a theoretical and applied standpoint. Further research is needed on models that effectively balance accuracy, distribution, scalability, and energy efficiency.

VI. Conclusions and Future Directions

In this work, 33 papers on ODTs in WSN from 2018 to 2023 were analyzed. These detection proposals are based on classification, boundary, clustering, thresholds, statistics and, in greater numbers, hybrid techniques. Likewise, distributed solutions prevailed over centralized ones, with online detection processes, multivariate, and in many cases taking advantage of the spatial, temporal, or spatiotemporal correlations of the detected data.

On the other hand, the reviewed works were mostly tested using simulation and specialized software such as Matlab and Python, with a network structure based on hierarchy and clusters, and using public data such as the datasets from IBRL and GSB. The performance of the techniques was measured using metrics such as ACC, AUC, F1-score, and FPR. In most of the studies, detection rates above 85% were reported; the best performances were from centralized techniques that do not consider spatiotemporal correlations of the data and operate in offline mode, which could explain these high detection rates. Likewise, the few proposals that achieved 100% detection rates did so under specific conditions, such as the use of synthetic datasets, with well-defined outliers, synthetic anomalies, or controlled noise levels; conditions that may not accurately reflect the complexity and noise of sensor readings in real-world WSN applications. On the other hand, not all the proposals reviewed incorporated an analysis of the complexity of their techniques, especially in the case of hybrid approaches.

Regarding the application area of the proposed techniques, most of the papers did not specify this aspect. However, as mentioned above, they used public datasets containing perceived observations from the environment in real-world implementations. This, along with research related to agricultural monitoring, water quality, and forest fires, shows the importance of these technologies in relation to environmental monitoring and care. This focus on environmental applications may also be influenced by the limited availability of datasets in other areas.

Finally, the main limitations found include high computational complexity and high resource consumption (energy, storage, computing capacity), sensitivity to parameters, lack of scalability, the use of a centralized approach that can be problematic in distributed WSNs, the dependence on specific assumptions about data distribution, or the presence of a representative training set. Other limitations include the lack of consideration of spatial or temporal correlations, reliance on historical data, and high cost of implementation and maintenance.

Outlier detection in WSNs still requires further research to address key limitations. A promising direction is the development of hybrid and layered models that integrate local node-based detection with global correlation and centralized optimization, balancing accuracy, and scalability. Another relevant area is dynamic adaptability to data changes through incremental and online learning. Furthermore, modern techniques for modeling the complex correlations in sensor data should be explored, along with the development of automated parameter selection mechanisms. On a practical level, more evaluation in real-world applications, theoretical analysis of energy feasibility and scalability, and the establishment of standardized benchmarks for objective assessment are needed.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

The first and second authors wrote the papers, and third author analyzed a data and made the final revision. All authors had approved the final version.

References

[1] M. A. Samara, I. Bennis, A. Abouaissa, and P. Lorenz, "A survey of outlier detection techniques in IoT: Review and classification," *J. Sens. Actuator Netw.*, vol. 11, no. 1, 4, Jan. 2022. doi: 10.3390/jsan11010004

[2] M. Rassam, A. Zainal, and M. Maarof, "Advancements of data anomaly detection research in wireless sensor networks: A survey and open issues," *Sensors*, vol. 13, no. 8, pp. 10087–10122, Aug. 2013. doi: 10.3390/s130810087

[3] Y. Zhang, N. Meratnia, and P. Havinga, "Outlier detection techniques for wireless sensor networks: A survey," *IEEE Commun. Surv. Tutor.*, vol. 12, no. 2, pp. 159–170, 2010. doi: 10.1109/SURV.2010.021510.00088

[4] P. Jain, S. Jain, O. R. Zaiane, and A. Srivastava, "Anomaly detection in resource constrained environments with streaming data," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 6, no. 3, pp. 649–659, Jun. 2022. doi: 10.1109/TETCI.2021.3070660

[5] A. Ayadi, O. Ghorbel, M. S. BenSalah, and M. Abid, "Kernelized technique for outliers detection to monitoring water pipeline based on WSNs," *Comput. Netw.*, vol. 150, pp. 179–189, Feb. 2019. doi: 10.1016/j.comnet.2019.01.004

[6] R. Khan, I. Ali, M. Zakarya, M. Ahmad, M. Imran, and M. Shoaib, "Technology-assisted decision support system for efficient water utilization: A real-time testbed for irrigation using wireless sensor networks," *IEEE Access*, vol. 6, pp. 25686–25697, 2018. doi: 10.1109/ACCESS.2018.2836185

[7] P. Biswas and T. Samanta, "Anomaly detection using ensemble random forest in wireless sensor network," *Int. J. Inf. Technol.*, vol. 13, no. 5, pp. 2043–2052, Oct. 2021. doi: 10.1007/s41870-021-00717-8

[8] N. Peng, W. Zhang, H. Ling, Y. Zhang, and L. Zheng, "Fault-tolerant anomaly detection method in wireless sensor networks," *Information*, vol. 9, no. 9, 236, Sep. 2018. doi: 10.3390/info9090236

[9] M. Safaei *et al.*, "A systematic literature review on outlier detection in wireless sensor networks," *Symmetry*, vol. 12, no. 3, 328, Feb. 2020. doi: 10.3390/sym12030328

[10] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, Jul. 2009. doi: 10.1145/1541880.1541882

[11] R. Jurdak, X. R. Wang, O. Obst, and P. Valencia, "Wireless sensor network anomalies: diagnosis and detection strategies," in *Proc. Intelligence-Based Systems Engineering*, A. Tolk and L. C. Jain, Eds., Intelligent Systems Reference Library, vol. 10. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 309–325. doi: 10.1007/978-3-642-17931-0_12

[12] V. Barnett and T. Lewis, *Outliers in Statistical Data*, 3rd ed. John Wiley Sons, 1994.

[13] D. Widhalm, K. M. Goeschka, and W. Kastner, "SoK: A taxonomy for anomaly detection in wireless sensor networks focused on node-level techniques," in *Proc. the 15th International Conference on Availability, Reliability and Security*, Aug. 2020, pp. 1–10. doi: 10.1145/3407023.3407027

[14] C. OReilly, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Anomaly detection in wireless sensor networks in a non-

stationary environment," *IEEE Commun. Surv. Tutor.*, vol. 16, no. 3, pp. 1413–1432, 2014. doi: 10.1109/SURV.2013.112813.00168

[15] A. Ayadi, O. Ghorbel, A. M. Obeid, and M. Abid, "Outlier detection approaches for wireless sensor networks: A survey," *Comput. Netw.*, vol. 129, pp. 319–333, Dec. 2017. doi: 10.1016/j.comnet.2017.10.007

[16] Y. Zhang, "Observing the unobservable: Distributed online outlier detection in wireless sensor networks," PhD. dissertation, University of Twente, Enschede, The Netherlands, 2010. doi: 10.3990/1.9789036530583

[17] T. Zhang, Q. Zhao, K. Shin, and Y. Nakamoto, "Bayesian-optimization-based peak searching algorithm for clustering in wireless sensor networks," *J. Sens. Actuator Netw.*, vol. 7, no. 1, 2, Jan. 2018. doi: 10.3390/jsan7010002

[18] Y. Chen and S. Li, "A lightweight anomaly detection method based on SVDD for wireless sensor networks," *Wirel. Pers. Commun.*, vol. 105, no. 4, pp. 1235–1256, Apr. 2019. doi: 10.1007/s11277-019-06143-1

[19] N. Ghosh, K. Maity, R. Paul, and S. Maity, "Outlier detection in sensor data using machine learning techniques for IoT framework and wireless sensor networks: A brief study," in *Proc. 2019 International Conference on Applied Machine Learning (ICAML)*, Bhubaneswar, India, May 2019, pp. 187–190. doi: 10.1109/ICAML48257.2019.00043

[20] S. K. Ghalem, B. Kechar, A. Bounceur, and R. Euler, "A probabilistic multivariate copula-based technique for faulty node diagnosis in wireless sensor networks," *J. Netw. Comput. Appl.*, vol. 127, pp. 9–25, Feb. 2019. doi: 10.1016/j.jnca.2018.11.009

[21] Z. Huan, C. Wei, and G.-H. Li, "Outlier detection in wireless sensor networks using model selection-based support vector data descriptions," *Sensors*, vol. 18, no. 12, 4328, Dec. 2018. doi: 10.3390/s18124328

[22] M. A. Rassam, M. A. Maarof, and A. Zainal, "A distributed anomaly detection model for wireless sensor networks based on the one-class principal component classifier," *Int. J. Sens. Netw.*, vol. 27, no. 3, 200, 2018. doi: 10.1504/IJSNET.2018.093126

[23] P. Shi, G. Li, Y. Yuan, and L. Kuang, "Outlier detection using improved support vector data description in wireless sensor networks," *Sensors*, vol. 19, no. 21, 4712, Oct. 2019. doi: 10.3390/s19214712

[24] D. McDonald, S. Sanchez, S. Madria, and F. Ercal, "A survey of methods for finding outliers in wireless sensor networks," *J. Netw. Syst. Manag.*, vol. 23, no. 1, pp. 163–182, Jan. 2015. doi: 10.1007/s10922-013-9287-z

[25] M. Safaei *et al.*, "Standalone noise and anomaly detection in wireless sensor networks: A novel time-series and adaptive Bayesian-network-based approach," *Softw. Pract. Exp.*, vol. 50, no. 4, pp. 428–446, Apr. 2020. doi: 10.1002/spe.2785

[26] V. Chandola, A. Banerjee, and V. Kumar, "Outlier detection: A survey," Dissertation, University of Minnesota, 2007.

[27] D. M. Hawkins, *Identification of Outliers*, Dordrecht: Springer Netherlands, 1980.

[28] M. Safaei, M. Driss, W. Boulila, E. A. Sundararajan, and M. Safaei, "Global outliers detection in wireless sensor networks: A novel approach integrating time-series analysis, entropy, and random forest-based classification," *Softw. Pract. Exp.*, vol. 52, no. 1, pp. 277–295, Jan. 2022. doi: 10.1002/spe.3020

[29] C. Titouna, F. Naït-Abdesselam, and A. Khokhar, "DODS: A distributed outlier detection scheme for wireless sensor networks," *Comput. Netw.*, vol. 161, pp. 93–101, Oct. 2019. doi: 10.1016/j.comnet.2019.06.014

[30] W. Zhang, G. Zhang, X. Chen, X. Zhou, Y. Liu, and J. Zhou, "A participation degree-based fault detection method for wireless sensor networks," *Sensors*, vol. 19, no. 7, 1522, Mar. 2019. doi: 10.3390/s19071522

[31] S. Bharti, K. K. Pattanaik, and A. Pandey, "Contextual outlier detection for wireless sensor networks," *J. Ambient Intell. Humaniz. Comput.*, vol. 11, no. 4, pp. 1511–1530, Apr. 2020. doi: 10.1007/s12652-019-01194-5

[32] A. Chirayil, R. Maharjan, and C.-S. Wu, "Survey on anomaly detection in Wireless Sensor Networks (WSNs)," in *Proc. 2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, Toyama, Japan, Jul. 2019, pp. 150–157. doi:.10.1109/SNPD.2019.8935827

[33] H. Ayadi, A. Zouinkhi, B. Boussaid, and M. N. Abdelkrim, "A machine learning methods: Outlier detection in WSN," in *Proc. 2015 16th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, Monastir, Tunisia, 2015, pp. 722–727. doi: 10.1109/STA.2015.7505190

[34] S. Rajasegarar, C. Leckie, and M. Palaniswami, "Detecting data anomalies in wireless sensor networks," in *Security in Ad Hoc and Sensor Networks*, World Scientific, 2009.

[35] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine learning for anomaly detection: A systematic review," *IEEE Access*, vol. 9, pp. 78658–78700, 2021. doi: 10.1109/ACCESS.2021.3083060

[36] I. G. A. Poornima and B. Paramasivan, "Anomaly detection in wireless sensor network using machine learning algorithm," *Comput. Commun.*, vol. 151, pp. 331–337, Feb. 2020. doi: 10.1016/j.comcom.2020.01.005

[37] N. Giatrakos, A. Deligiannakis, M. Garofalakis, and Y. Kotidis, "Omnibus outlier detection in sensor networks using windowed locality sensitive hashing," *Future Gener. Comput. Syst.*, vol. 110, pp. 587–609, Sep. 2020. doi: 10.1016/j.future.2018.04.046

[38] A. A. A. Alkhatib and Q. Abed-Al, "Multivariate outlier detection for forest fire data aggregation accuracy," *Intell. Autom. Soft Comput.*, vol. 31, no. 2, pp. 1071–1087, 2022. doi: 10.32604/iasc.2022.020461

[39] X. Miao, Y. Liu, H. Zhao, and C. Li, "Distributed online one-class support vector machine for anomaly detection over networks," *IEEE Trans. Cybern.*, vol. 49, no. 4, pp. 1475–1488, Apr. 2019. doi: 10.1109/TCYB.2018.2804940

[40] A. Ayadi, O. Ghorbel, M. S. BenSalah, and M. Abid, "Spatio-temporal correlations for damages identification and localization in water pipeline systems based on WSNs," *Comput. Netw.*, vol. 171, 107134, Apr. 2020. doi: 10.1016/j.comnet.2020.107134

[41] K. Zhang, K. Yang, S. Li, D. Jing, and H.-B. Chen, "ANN-based outlier detection for wireless sensor networks in smart buildings," *IEEE Access*, vol. 7, pp. 95987–95997, 2019. doi: 10.1109/ACCESS.2019.2929550

[42] U. Gupta, V. Bhattacharjee, and P. S. Bishnu, "Outlier detection in wireless sensor networks based on neighbourhood," *Wirel. Pers. Commun.*, vol. 116, no. 1, pp. 443–454, Jan. 2021. doi: 10.1007/s11277-020-07722-3

[43] J. Chen, J. Zhang, R. Qian, J. Yuan, and Y. Ren, "An anomaly detection method for wireless sensor networks based on the improved isolation forest," *Appl. Sci.*, vol. 13, no. 2, 702, Jan. 2023. doi: 10.3390/app13020702

[44] R. R. Guimaraes *et al.*, "Intelligent network security monitoring based on optimum-path forest clustering," *IEEE Netw.*, vol. 33, no. 2, pp. 126–131, Mar. 2019. doi: 10.1109/MNET.2018.1800151

[45] H. T. Nguyen and N. H. Thai, "Temporal and spatial outlier detection in wireless sensor networks," *ETRI J.*, vol. 41, no. 4, pp. 437–451, Aug. 2019. doi: 10.4218/etrij.2018-0261

[46] Z.-M. Wang, G.-H. Song, and C. Gao, "An isolation-based distributed outlier detection framework using nearest neighbor ensembles for wireless sensor networks," *IEEE Access*, vol. 7, pp. 96319–96333, 2019. doi: 10.1109/ACCESS.2019.2929581

[47] Y. Qiao, X. Cui, P. Jin, and W. Zhang, "Fast outlier detection for high-dimensional data of wireless sensor networks," *Int. J. Distrib. Sens. Netw.*, vol. 16, no. 10, 155014772096383, Oct. 2020. doi: 10.1177/1550147720963835

[48] L. Chen, G. Li, and G. Huang, "A hypergrid based adaptive learning method for detecting data faults in wireless sensor networks," *Inf. Sci.*, vol. 553, pp. 49–65, Apr. 2021. doi: 10.1016/j.ins.2020.12.011

[49] C. Gao, G. Song, Z. Wang, and Y. Chen, "A mobile edge-cloud collaboration outlier detection framework in wireless sensor networks," *IET Commun.*, vol. 15, no. 15, pp. 2007–2020, Sep. 2021. doi: 10.1049/cmu2.12231

[50] S. Saraswathi, G. R. Suresh, and J. Katiravan, "False alarm detection using dynamic threshold in medical wireless sensor networks," *Wirel. Netw.*, vol. 27, no. 2, pp. 925–937, Feb. 2021. doi: 10.1007/s11276-019-02197-y

[51] N. M. Zamry, A. Zainal, M. A. Rassam, E. H. Alkhammash, F. A. Ghaleb, and F. Saeed, "Lightweight anomaly detection scheme using incremental principal component analysis and support vector machine," *Sensors*, vol. 21, no. 23, 8017, Nov. 2021. doi: 10.3390/s21238017

[52] B. Sarangi and B. Tripathy, "Outlier detection technique for wireless sensor network using GAN with autoencoder to increase the network lifetime," *I. J. Computer Network and Information Security*, vol. 15, issue 1, 2023.

[53] A. E. Ezugwu, J. O. Agushaka, L. Abualigah, S. Mirjalili, and A. H. Gandomi, "Prairie dog optimization algorithm," *Neural Comput. Appl.*, vol. 34, no. 22, pp. 20017–20065, Nov. 2022. doi: 10.1007/s00521-022-07530-9

[54] L. Abualigah, S. Ekinci, D. Izci, and R. A. Zitar, "Modified elite opposition-based artificial hummingbird algorithm for designing FOPID controlled cruise control system," *Intell. Autom. Soft Comput.*, 2023. doi: 10.32604/iasc.2023.040291

[55] J. O. Agushaka, A. E. Ezugwu, and L. Abualigah, "Dwarf mongoose optimization algorithm," *Comput. Methods Appl. Mech. Eng.*, vol. 391, 114570, Mar. 2022. doi: 10.1016/j.cma.2022.114570