*Article*

# A Systematic Literature Review on Outlier Detection in Wireless Sensor Networks

**Mahmood Safaei [1], Shahla Asadi [2], Maha Driss [3,4], Wadii Boulila [3,4,*], Abdullah Alsaeedi [3], Hassan Chizari [5], Rusli Abdullah [2] and Mitra Safaei [6]**

[1]  School of Computing Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia; ma.safaei@gmail.com
[2]  Department of Software Engineering & Information System,
    Faculty of Computer Science & Information Technology, Universiti Putra Malaysia, Seri Kembangan 43400, Malaysia ; asadi.shahla2003@gmail.com (S.A.); rusli@upm.edu.my (R.A.)
[3]  College of Computer Science and Engineering, Taibah University, Medina 42353, Saudi Arabia; maha.idriss@riadi.rnu.tn (M.D.); aasaeedi@taibahu.edu.sa (A.A.)
[4]  RIADI Laboratory, University of Manouba, Manouba 2010, Tunisia
[5]  Department of Computing, University of Glouctershire, Cheltenham GL50 2RH, UK; hchizari@glos.ac.uk
[6]  Fakultät Electronic und Informatik, Gottfried Wilhelm Leibniz Universität Hannover, Hannover 30167, Germany; mitra.safaei@stud.uni-hannover.de
*   Correspondence: wadii.boulila@riadi.rnu.tn (W.B.)

**Abstract:** A wireless sensor network (WSN) is defined as a set of spatially distributed and interconnected sensor nodes. WSNs allow one to monitor and recognize environmental phenomena such as soil moisture, air pollution, and health data. Because of the very limited resources available in sensors, the collected data from WSNs are often characterized as unreliable or uncertain. However, applications using WSNs demand precise readings, and uncertainty in data reading can cause serious damage (e.g., health monitoring data). Therefore, an efficient local/distributed data processing algorithm is needed to ensure: (1) the extraction of precise and reliable values from noisy readings; (2) the detection of anomalies from data reported by sensors; and (3) the identification of outlier sensors in a WSN. Several works have been conducted to achieve these objectives using several techniques such as machine learning algorithms, mathematical modeling, and clustering. The purpose of this paper is to conduct a systematic literature review to report the available works on outlier and anomaly detection in WSNs. The paper highlights works conducted from January 2004 to October 2018. A total of 3520 papers are reviewed in the initial search process. Later, these papers are filtered by title, abstract, and contents, and a total of 117 papers are selected. These papers are examined to answer the defined research questions. The current paper presents an improved taxonomy of outlier detection techniques. This will help researchers and practitioners to find the most relevant and recent studies related to outlier detection in WSNs. Finally, the paper identifies existing gaps that future studies can fill.

**Keywords:** systematic literature review; outlier detection; wireless sensor networks

## 1. Introduction

The wireless sensor network (WSN) consists of a set of distributed and interconnected sensors located in a target area. It aims to monitor and recognize environmental phenomena such as soil moisture, air pollution, and health data [1]. Low-cost devices and easy-to-deploy sensor nodes have found a variety of applications in positioning and tracking [2], health care [3], environmental monitoring [4], etc.

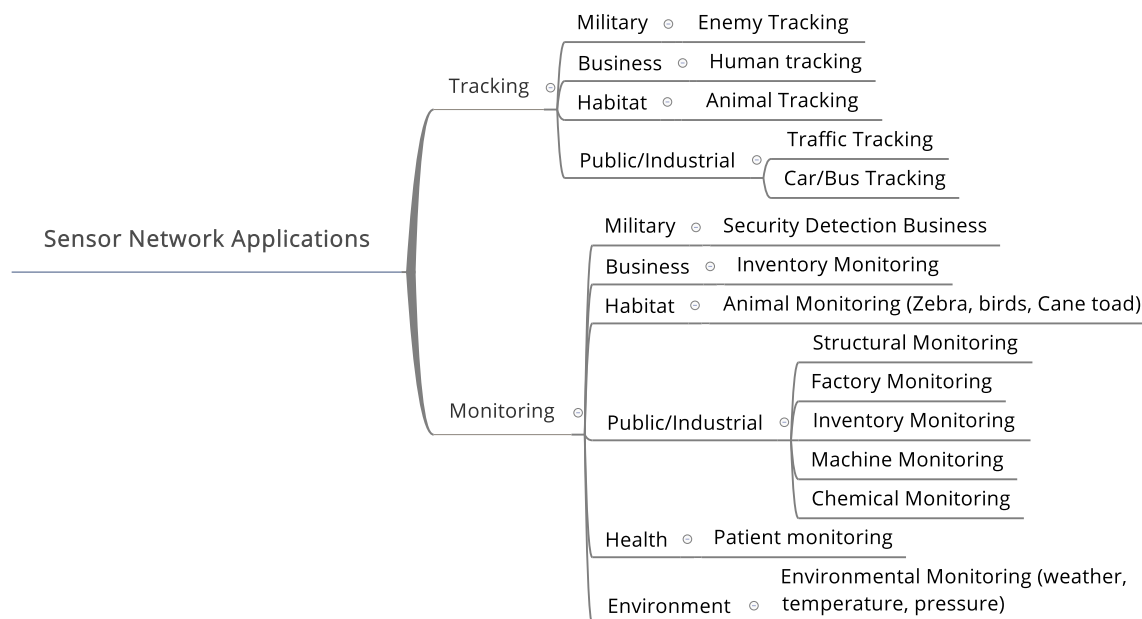Figure 1 shows some WSN applications in different fields.



**Figure 1.** Wireless sensor network applications categories.

However, there are still many critical challenges that need to be tackled via reliable technology. Usually, sensors are deployed in harsh environments with an unattended operation, which may lead to the sensor or network failures. Therefore, it is important for sensors to have not only a fault tolerance system but also the ability to do self-calibrating, self-recovering, self-repairing, and self-testing. In some scenarios such as health applications, it is important to have accurate data collection in the network. Data reliability in sensor networks is the area of focus for many applications.

Usually, data retrieved from WSNs have low reliability due to missing values, inconsistent or duplicate data, errors, noise, and malicious attacks. Low-quality sensors may compromise memory, battery functionality, communication efficacy, and computation ability, thus leading to inaccurate WSN sensory data [5]. Sensor nodes are vulnerable to the effects of the environment as well. A WSN with high density employs hundreds or thousands of sensor nodes within a setting, which may eventually result in malfunction nodes, leading to inaccurate and insufficient data. These nodes are susceptible to malevolent attacks such as eavesdropping, black holes, and denial of service (DoS) [6].

In the field of WSNs, measurements that significantly differ from the normal pattern of sensed data are declared as outliers [7]. The potential causes of outliers are noise and errors, events, and malicious attacks. Outlier detection in WSNs is the process of identifying data instances that deviate from the rest of the data patterns based on certain measurements [8].

Outliers can occur for different reasons, and understanding their source helps to decide what actions to take after detecting them [9]. Many studies have investigated abnormal data detection under various terms such as anomaly detection, fraud detection, and outlier detection [10]. In the WSN context, the outlier also is defined as an anomaly or divergence which is unusual behavior in comparison with the majority of sensory data as indicated in Figure 2. The outlier data can be classified into two main classes, including single and batch outlier data. An outlier is single when data are far from a group of sensory data, whereas batch outliers are bulk data points that continuously occurred over a period. According to the related literature, there are no general definitions of outliers or anomalies. Therefore, in Table 1, this study shows a set of common definitions of anomalies and outliers proposed by several researchers.
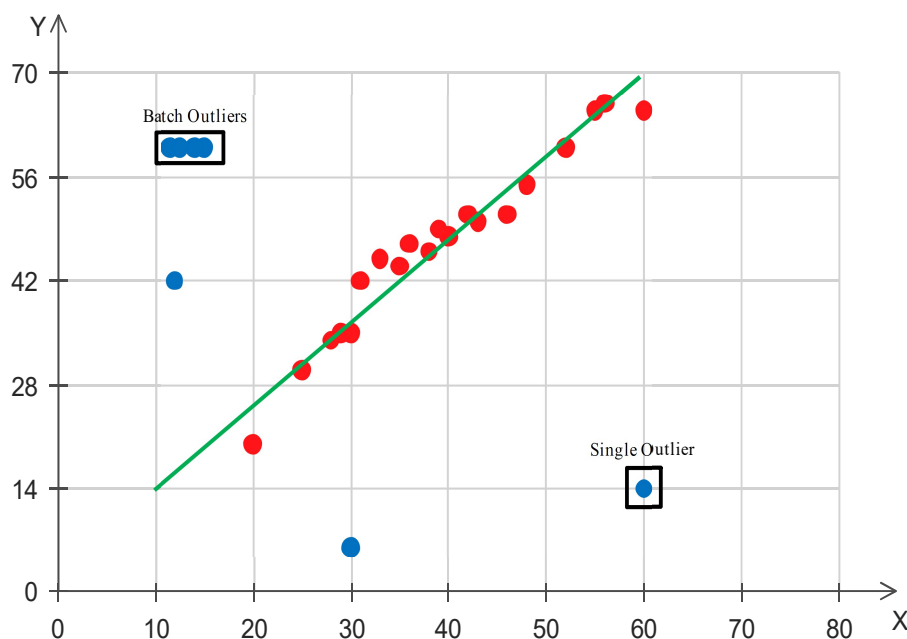
**Figure 2.** Example of single and batch outliers in sensory data.

**Table 1.** Outlier detection definitions from previous studies.

| Reference | Definition |
|---|---|
| [11] | "A process to identify data points that are very different from the rest of the data based on a certain measure." |
| [12] | "An observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism." |
| [13] | "An observation that deviates a lot from other observations and can be generated by a different mechanism." |
| [14] | "An outlier is an observation or subset of observations that appears to be inconsistent with the rest of the set of data." |
| [15] | "An outlier is a data point which is significantly different from other data points, or does not conform to the expected normal behavior, or conforms well to a defined abnormal behavior." |
| [16] | "A spatial-temporal point, which non-spatial attribute values are significantly different from those of other spatially and temporally referenced points in its spatial or/and temporal neighborhoods, is considered as a spatial-temporal outlier." |
| [17] | "A point is considered to be an outlier if, in some lower-dimensional projection, it is present in a local region of abnormal low density." |
| [18] | "If the removal of a point from the time sequence results in a sequence that can be represented more briefly than the original one, then the point is an outlier." |
| [19] | "Outliers are points that do not belong to clusters of a dataset or clusters that are significantly smaller than other clusters." |
| [15] | "Outliers are points that lie in the lower local density with respect to the density of their local neighborhoods." |

As shown in Figure 3, several sources for outliers have been categorised as follows: noise or error, events, and malicious attacks [20]. An event-based sensor network sends information to the base station after an event occurs in the network. Query and data-driven methods are different from event detection. In query and data-driven methods, sensor nodes reply to queries issued by sink nodes.

- An event-based network is different from a monitoring sensor network. Some typical event examples are earthquake monitoring, flood, volcanic eruption alarm, rainfall and flood detection, weather changes, chemical hazardous alert, air pollution, air quality monitoring, and fireplace detection. In mutuality with inaccurate data, outliers generated by events tend to have an

especially smaller probability of incidence [21]. Deleting the outlier event from the dataset can cause loss of necessary and important data from relevant events [22]. Several techniques are proposed for event detection such as [23–25].

- Noise or error that is based on noise in measurement may occur because of several sources, like a sensor fault or sensor misbehavior [20]. Faulty data are ordinarily described as a modification in the dataset that is disparate from the rest of the data. Error or noise can result in several changes associated with the environment, including harshness and the difficulties of the preparation areas. If possible, faulty data, as well as noisy data, must be corrected or deleted [20].
- Malicious attacks are associated with the security of the network. Outliers based on malicious attacks begin with a sensor node that is compromised by the attacker and the injection of unreliable or corrupt data into the network topology. Malicious attacks are classified into passive and active attacks. A passive attack changes sensory data with the aim of interrupting the decision-making system of the network [20], whereas an active attack has an effect on network functionality and performance. This attack can slow or even shut down the network [26].

Usually, identifying outliers amongst a vast data are a difficult task [27]. The two primary challenges in detecting outliers within WSNs are ensuring less resource consumption and achieving high accuracy. These challenges should be overcame to ensure the accuracy and the reliability of data retrieved from sensors for further processes [27].
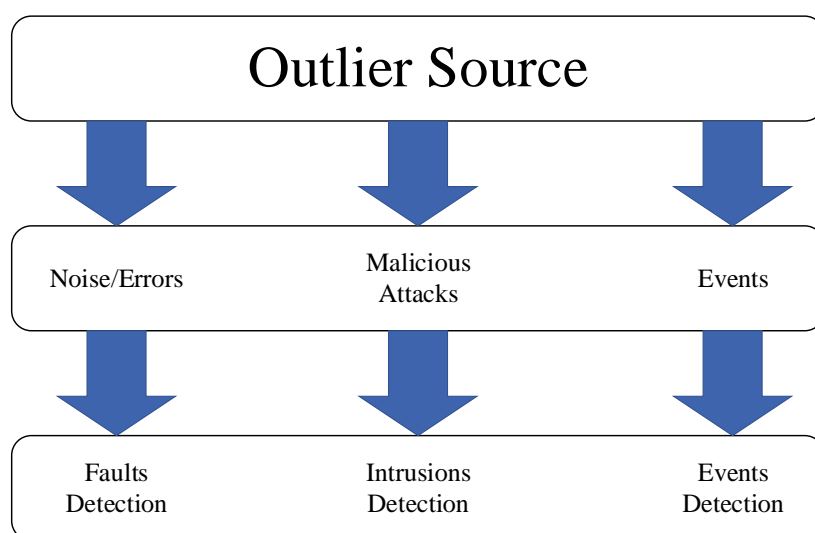


**Figure 3.** Different types of outlier sources in WSNs.

This paper presents a detailed overview of techniques that are dedicated to detecting outliers in WSNs, compares existing methods, and discusses future research prospects. Although some works have used prior studies' outcomes to assess the present state of the work in this area, no work has been conducted to systematically synthesize and review outlier detection in WSNs. Therefore, this study systematically collects, analyzes, and synthesizes all papers linked with outlier detection in WSNs in order to highlight emerging methods, themes, taxonomies, and datasets. This paper presents a systematic literature review (SLR) conducted on a large pool of papers proposing anomaly detection techniques across several research parts and domain applications. The remainder of this study is organized as follows: Section 2 describes applications of outlier detection in WSN. Section 3 illustrates the methodology that is employed in this study, whereas Section 4 discusses the planning review, and Section 5 explains how the review was conducted. Next, Section 6 provides answers to research questions (RQ), and Section 7 compares methods for detecting outliers. Finally, the study is concluded in Section 9.

## 2. Application of Outlier Detection in WSNs

Anomaly or outlier detection is a main function of the data mining procedure, as illustrated in Figure 4. Outlier detection can help in preventing malicious attacks and identifying sensors with outlier data to provide reliable data for decision-makers. Many lifetime and real-time applications use outlier detection:

- Environmental monitoring: Many sensors such as temperature, humidity, air pollution, and wind speed sensors are deployed in harsh environments to monitor and analyze environmental factors.
- Industrial monitoring: Sensors such as vibration, pressure, or temperature sensors are installed on sensitive equipment to monitor the state of this equipment.
- Healthcare monitoring: Small sensors are used to monitor patients' vital state. These sensors are implemented in the patient's body in different positions to monitor blood pressure, heart rate, or enzymes and minerals.
- Smart cities: Different kinds of sensors such as parking sensors, dustbin sensors, and pedestrian sensors are used to make cities more comfortable for citizens.
- Forest fire detection: Forests are monitored to prevent fires using a variety of sensor nodes. Thousands of them are deployed in the target area to predict and prevent forest fires.
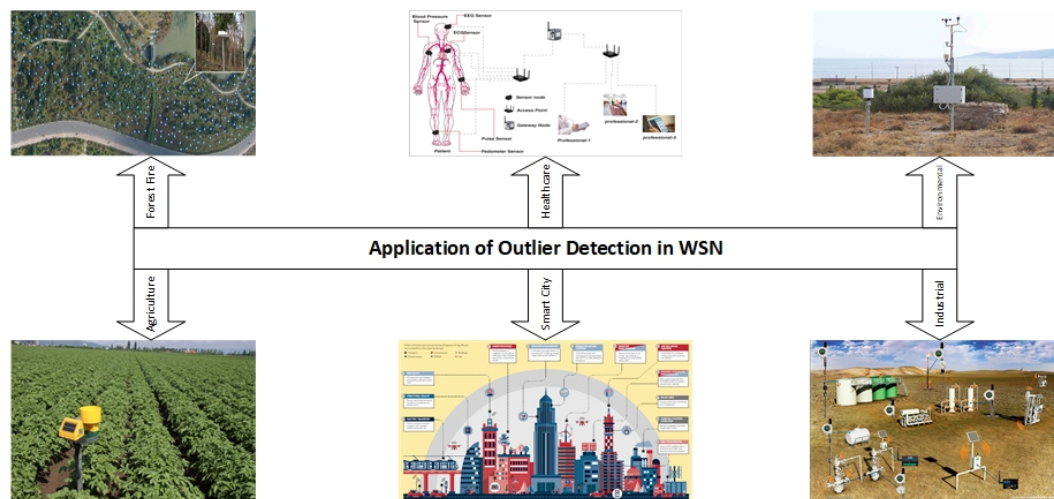


**Figure 4.** Application of outlier detection in WSNs.

## 3. Review Method

We used SLR as the methodology to study current research work regarding outlier detection. The 'systematic literature review provides a means for the evaluation and interpretation of the available research which is pertinent to a specific topic area, RQ, or a phenomenon of interest' [28–30]. This study employed the SLR guidelines and standards proposed by Kitchenham [31], which consist of a set of well-defined stages conducted in line with a predefined protocol. The aim of performing SLR is to systematically collect, evaluate, and interpret all the published studies relevant to the predefined RQs in order to deliver comprehensive information for the research community. The SLR was selected to gather data regarding cutting-edge notions, to list the benefits of certain approaches, and to find a research gap that may be bridged via investigation [32]. According to [31], the SLR approach has three phases: 'planning, conducting, and reporting the review'. These phases consist of the following processes: (1) identifying RQs; (2) developing a review protocol; (3) determining both exclusion and inclusion criteria; (4) selecting search strategy and study process; (5) quality assessment (QA); and (6) extracting and synthesizing data. As illustrated in Figure 5, for performing SLR, we summarized the methodological steps. In the following section, the details of these steps are explained.
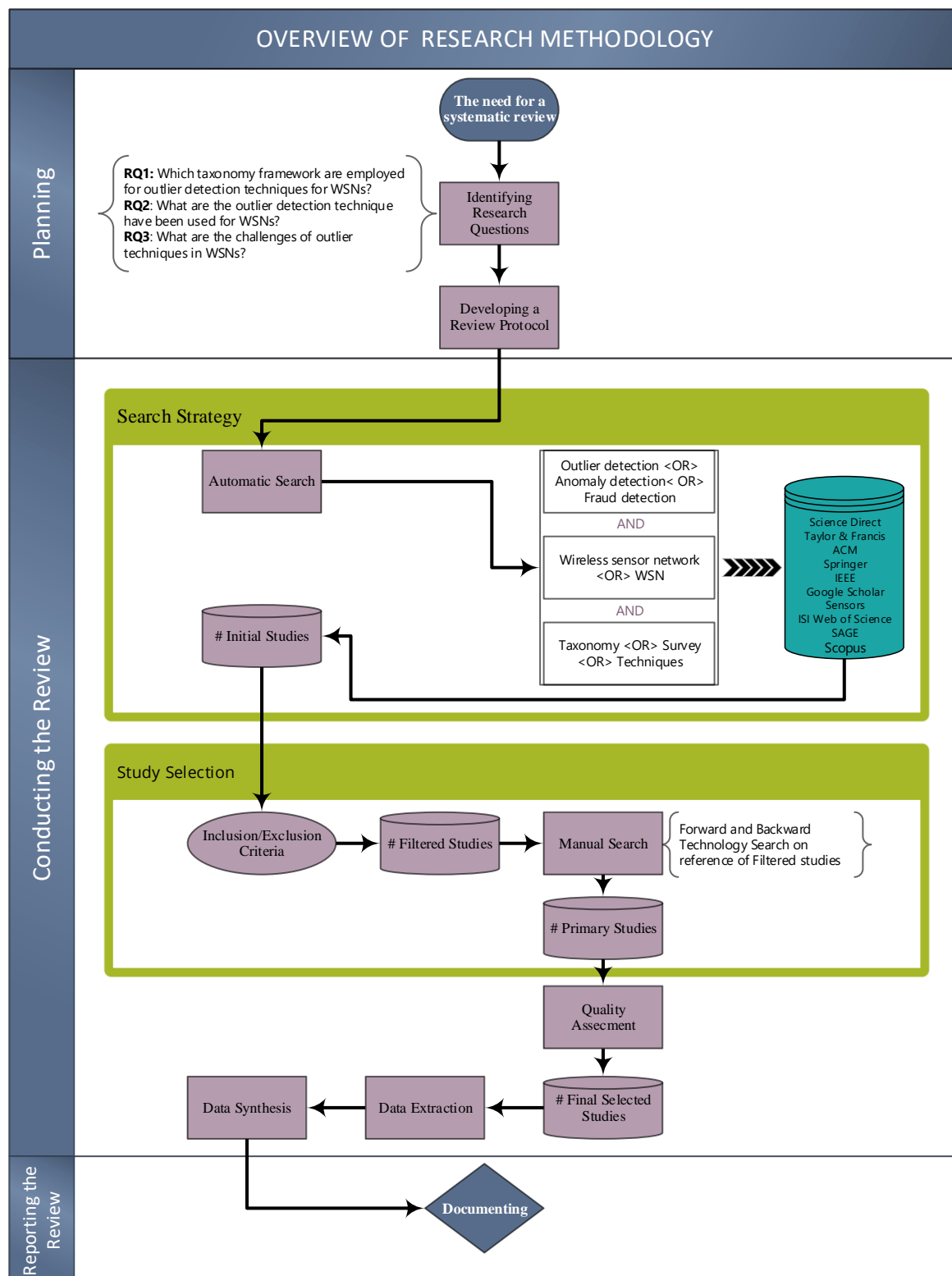
**OVERVIEW OF RESEARCH METHODOLOGY**

Planning

The need for a systematic review

**RQ1:** Which taxonomy framework are employed for outlier detection techniques for WSNs?
**RQ2:** What are the outlier detection technique have been used for WSNs?
**RQ3:** What are the challenges of outlier techniques in WSNs?

Identifying Research Questions

Developing a Review Protocol

Conducting the Review

Search Strategy

Automatic Search

Outlier detection <OR> Anomaly detection< OR> Fraud detection

AND

Wireless sensor network <OR> WSN

AND

Taxonomy <OR> Survey <OR> Techniques

Science Direct
Taylor & Francis
ACM
Springer
IEEE
Google Scholar
Sensors
ISI Web of Science
SAGE
Scopus

# Initial Studies

Study Selection

Inclusion/Exclusion Criteria

# Filtered Studies

Manual Search

Forward and Backward Technology Search on reference of Filtered studies

# Primary Studies

Quality Assecment

Data Synthesis

Data Extraction

# Final Selected Studies

Reporting the Review

Documenting

**Figure 5.** Overview of research methodology.

## 4. Planning the Review

The planning phase begins by determining the need for SLR, identifying RQs, and developing a review protocol. The review protocol is as follows:

*4.1. The Need for a Systematic Review*

Although many strategies have been suggested for detecting specific subsets of WSN outliers, there is still a need for more comprehensive outlier detection strategies. This study looked into the various methods that have been developed for outlier detection in the literature review, besides those that have tried to provide an overview of the vast literature on techniques, classifications, taxonomies, and comparisons. Numerous techniques for detecting outliers have been developed for a specific application or a single study area. This survey significantly expands the discussion in several directions according to the following research questions.

*4.2. Identifying Research Questions*

To achieve the main objectives of this study, we propose three key research questions:

- RQ1: What is the designed taxonomy and framework for outlier detection techniques in WSNs?
- RQ2: What are the outlier detection techniques that have been used in WSNs?
- RQ3: What are the challenges in current outlier detection techniques in WSNs?

*4.3. Developing a Review Protocol*

The review protocol is considered an important step in conducting the SLR. It helps to determine the methods that will be applied in the systematic review. The main aim of the review protocol is to decrease study bias and differentiate SLR from traditional methods of reviewing the literature [31]. This review protocol categorizes the 'review background, search strategy, development of RQs, extraction of data, criteria for study selection, and data synthesis'. The relevant RQs and review background are explained above. The following section provides details about other elements.

**5. Conducting the Review**

The review begins with a study selection and extraction and synthesis of data.

*5.1. Search Strategy*

The search strategy has a significant impact on data extraction from selected papers. A search strategy can assist scholars in obtaining as many relevant studies as possible [33]. Figure 5 illustrates the two steps of search strategies: manual and automatic. Both manual and automatic search approaches are employed for investigating the content of a review. This allows more studies to be incorporated and a wide range of academic publications to be covered. An automatic search can be employed to find primary studies on anomaly detection in WSNs. Web searches can be conducted based on search keywords in online library databases. Based on [34]'s suggestions, the search strategy was not limited to only a certain type of article; rather, it included a wide range of relevant and high-impact-factor publications in online libraries. The following online databases (with their assigned link) were included in the search strategy:

- Science Direct (http://www.sciencedirect.com/),
- SpringerLink (http://www.springer.com/in/),
- IEEE Explorer (http://www.ieee.org/index.html),
- Taylor and Francis Online (http://www.tandfonline.com/),
- ACM Digital Library (https://dl.acm.org/),
- MDPI (https://www.mdpi.com/).

The proposed study aimed to identify articles that were relevant to the domain. The main research keywords included are: 'anomaly detection in WSN', 'outlier detection in WSN', and 'anomaly detection techniques in WSN'. A string of words was used to make sure that no relevant publication was missed. The search was limited to the year range of 2004 to October 2018 (more than 10 years). The search exposed a big volume of literature, including journal publications, conference proceedings,

and many other published materials. All included digital repositories were manually searched using the predefined keywords.

The details of the overall search process based on the defined keywords in the given libraries are shown in Figure 6.
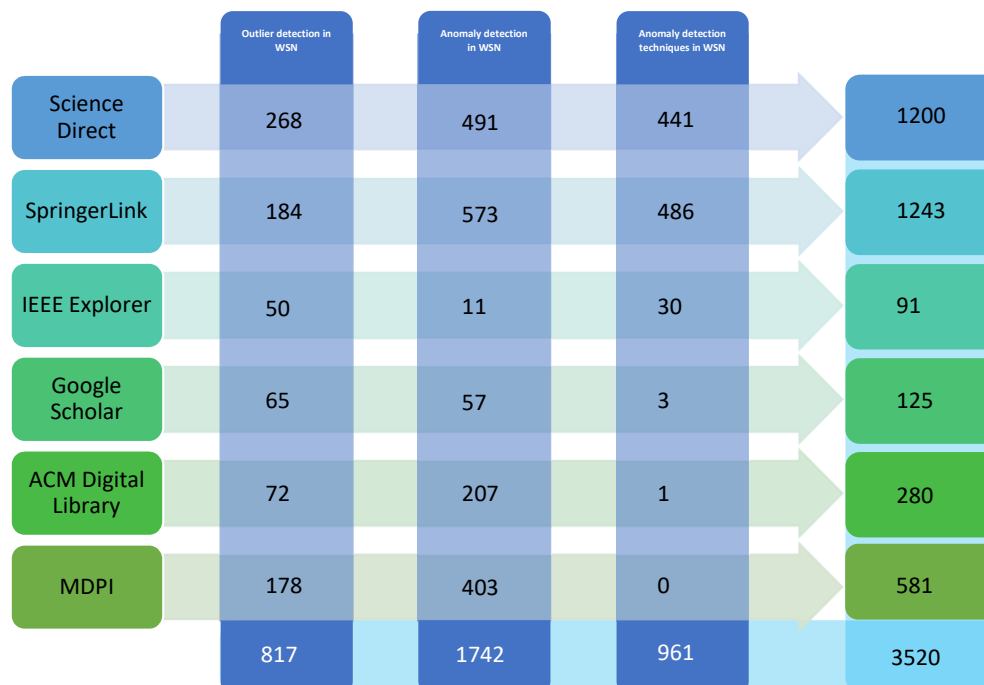
| | Outlier detection in WSN | Anomaly detection in WSN | Anomaly detection techniques in WSN | |
|---|---|---|---|---|
| Science Direct | 268 | 491 | 441 | 1200 |
| SpringerLink | 184 | 573 | 486 | 1243 |
| IEEE Explorer | 50 | 11 | 30 | 91 |
| Google Scholar | 65 | 57 | 3 | 125 |
| ACM Digital Library | 72 | 207 | 1 | 280 |
| MDPI | 178 | 403 | 0 | 581 |
| | 817 | 1742 | 961 | 3520 |

**Figure 6.** Search process based on the defined keywords for articles.

## 5.2. Criteria for Inclusion and Exclusion Articles

Exclusion and inclusion criteria ensure that only relevant studies are incorporated in data analysis. Because this review focused on understanding outlier detection in WSNs, only papers published in the English language from 2004 to 2018 were included in this study. The reason for selecting this particular time-frame was that the term 'outlier' has been gradually utilized in many studies since 2004, and several articles have covered the topic of outlier detection as of 2014. Thus, this study aimed to systematically collect, analyze, and synthesize articles until 2018. Studies unrelated to outlier detection in WSNs were discarded. Table 2 shows the criteria applied.

**Table 2.** Criteria for inclusion and exclusion of the articles.

| Inclusion Criteria | Exclusion Criteria |
|---|---|
| Studies are written in English | Studies whose full text is not available |
| Studies are published between 2004−2018 | Duplicated studies |
| Studies are published in the above selected database | Studies that are not related to outlier detection in wireless network domain |
| Studies that provide answers to the research questions | Articles that did not match the inclusion criteria |

## 5.3. Manual Search

Based on [34], a forward and backward search was employed to trace the citations of primary studies. We used the Google Scholar search engine to find studies that were cited in the selected primary studies. The manual search also ensured that the systematic review of the research was relatively complete and comprehensive and that we did not miss anything. Mendeley (https://www.mendeley.com) was employed for sorting and managing all the studies and to remove duplicate studies.

### 5.4. Process for Selection of Studies

The primary aim of the selection process (primary studies) was to identify relevant studies to SLR. This search was performed by adhering to the steps outlined in the previous section. As a result, 3520 research articles were retrieved via the automatic search. Using Mendeley, the duplicated articles were removed. Initially, each folder of the library was checked manually, and all the articles were properly named by their titles. The duplications in these publications were removed by checking the titles in each folder. The initial selection filtering process was performed manually for all the libraries by title, and a total of 247 articles were obtained. Based on Kitchenham's [35] recommendations, these articles were then filtered manually by abstract, and a total of 208 articles were included. In the last step, these articles were again filtered manually by content, and finally, a total of 117 articles were selected. The details of the selected papers by title, abstract, and contents are given in Figure 7. The list of year-wise publications is shown in Table 3. The list of final selected papers along with the titles and citations is given in Table 4.
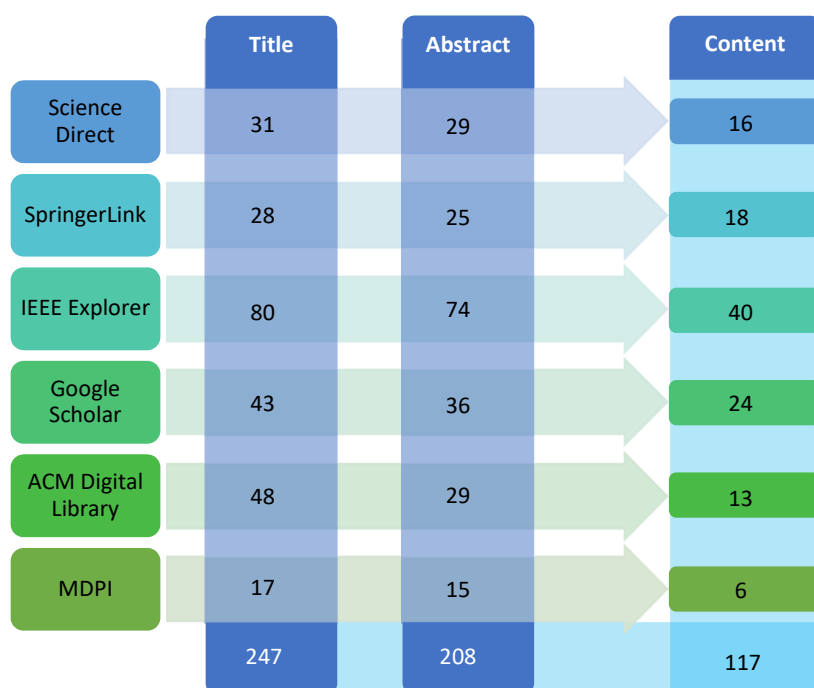


**Figure 7.** Filtering papers by title, abstract, and contents.

**Table 3.** Year-wise breakup of selected publications.

| 2018 | 2017 | 2016 | 2015 | 2014 | 2013 | 2012 | 2011 | 2010 | 2009 | 2008 | 2007 | 2006 | 2005 | 2004 |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| | | | | | | | | | | | | [82] | | |
| | | | | | | | | | | | | [158] | | |
| | | | | | | | | | | | | [39] | | |
| | | | | | | | | | [74] | | | [140] | | |
| | | | | | | | | [104] | [68] | | | [4] | | |
| | | | | | | | | [59] | [41] | | | [22] | | |
| | [186] | [184] | | | [134] | | | [176] | [50] | | | [43] | | |
| | [100] | [157] | | | [75] | | | [98] | [156] | [177] | | [2] | | |
| [183] | [189] | [162] | | [168] | [58] | [193] | | [80] | [78] | [61] | [11] | [178] | | |
| [190] | [188] | [96] | [38] | [77] | [161] | [101] | | [6] | [106] | [45] | [97] | [46] | | |
| [181] | [170] | [155] | [94] | [146] | [51] | [26] | [148] | [173] | [74] | [102] | [66] | [65] | | |
| [187] | [99] | [95] | [13] | [114] | [70] | [119] | [1] | [141] | [81] | [108] | [53] | [105] | [159] | |
| [182] | [57] | [37] | [163] | [153] | [180] | [40] | [165] | [49] | [164] | [172] | [103] | [174] | [42] | |
| [151] | [93] | [152] | [185] | [160] | [76] | [143] | [135] | [154] | [48] | [169] | [67] | [5] | [24] | |
| [191] | [20] | [56] | [139] | [84] | [142] | [192] | [69] | [175] | [8] | [137] | [118] | [64] | [17] | [149] |

**Table 4.** Primary Studies.

| S-ID | Reference | Year | Type | Methodology | Taxonomy | Dataset |
|------|-----------|------|------|-------------|----------|---------|
| S1 | [151] | 2018 | Journal | Support Vector Machines | Classification | Smart city datset |
| S2 | [94] | 2015 | Journal | Quarter-sphere support vector machine (QSSVM) | Classification | – |
| S3 | [95] | 2015 | Journal | Bayesian network | Classification | Mica2Dot sensor nodes dataset at Berkeley Lab |
| S4 | [134] | 2013 | Journal | Survey | – | – |
| S5 | [96] | 2016 | Conference | Support vector machine technique within a sliding window-based learning algorithm | Classification and Spectral Decomposition | univariate datasets: an artificial dataset, in addition to the Well-Log and Dow Jones dataset |
| S6 | [152] | 2016 | Journal | Support vector machine and a sliding window learning | Classification and Spectral Decomposition | Benchmark three-tank system |
| S7 | [153] | 2014 | Journal | Review Paper | – | – |
| S8 | [56] | 2016 | Conference | Nearest neighbor | Classification | Intel Berkeley base |
| S9 | [13] | 2015 | Journal | Naïve bayesian | Classification | Intel Berkeley Research Lab |
| S10 | [37] | 2016 | Conference | Kernel principal component analysis (KPCA) | Statistical | Intel Berkeley (IBRL), Grand-St- Bernard (GStB), and Sensor-scope(LUCE) |
| S11 | [154] | 2010 | Journal | Rule, LLSE, time series forecasting, and HMMs | – | Sensor Scope, INTEL Lab, GDI, NAMOS |
| S12 | [6] | 2010 | Journal | Survey | – | – |
| S13 | [26] | 2012 | Journal | Survey | – | – |
| S14 | [38] | 2015 | Journal | KPCA based Mahalanobis kernel | Statistical and Classification | Intel Berkeley Research Lab (IBRL), Grand St. Bernard (GStB), Sensorscope Lausanne Urban Canopy Experiment (LUCE) |
| S15 | [155] | 2016 | Journal | STODM algorithm and the fuzzy logic | – | St.Bernard wireless sensor network |
| S16 | [156] | 2009 | Conference | Review | – | – |
| S17 | [82] | 2006 | Journal | K-Means | Clustering | Dataset generated from multivariate Gaussian distribution |
| S18 | [157] | 2016 | Journal | Cross calibration | – | Simulated dataset |
| S19 | [158] | 2006 | Journal | Localized fault detection | – | – |
| S20 | [159] | 2005 | Journal | Bayesian algorithm | Classification based | – |
| S21 | [160] | 2014 | Journal | Multi criteria | statistical | Dataset acquired from a real world |
| S22 | [24] | 2005 | Conference | Boundary detection | – | – |
| S23 | [161] | 2013 | Journal | Fault-tolerant | Clustering | – |
| S24 | [162] | 2016 | Journal | Data compression | – | Events collected data from different locations |
| S25 | [163] | 2015 | Journal | Automatic data quality control | – | Real dataset |
| S26 | [149] | 2004 | Journal | review Survey | – | – |
| S27 | [164] | 2009 | Journal | Pareto algebra | Statistical | Simulated dataset |
| S28 | [57] | 2017 | Journal | Dynamically aggregated neighboring information | Nearest Neighbor based | Dataset from Sensor Scope Grand St. Bernard scenario |
| S29 | [165] | 2011 | Journal | – | – | – |
| S30 | [8] | 2009 | Journal | Survey | – | – |
| S31 | [66] | 2007 | Journal | Aggregation tree | Classification | Dataset provided by Berkeley research lab |
| S32 | [39] | 2006 | Journal | Bayesian and Neyman Pearson | Statistical | Simulated dataset |
| S33 | [166] | 2013 | Journal | Survey | – | – |
| S34 | [97] | 2007 | Conference | Bayesian | Classification | Simulated dataset and actual environmental dataset collected in the forest |
| S35 | [119] | 2012 | Journal | Hierarchical Bayesian spatio temporal (HBST) modeling | Classification | Three simulated datasets |
| S36 | [167] | 2014 | Journal | Systematic Literature Review | – | – |
| S37 | [22] | 2006 | Conference | Clustering | Clustering | Simulated dataset from the Great Duck Island project |
| S38 | [168] | 2014 | Journal | Review of anomaly detection methods | – | – |

**Table 4.** *Cont.*

| S-ID | Reference | Year | Type | Methodology | Taxonomy | Dataset |
|------|-----------|------|------|-------------|----------|---------|
| S39 | [192] | 2012 | Journal | Support vector machine | Classification | synthetic and real |
| S40 | [11] | 2007 | Journal | Histogram | Statistical | Dataset of temperature records |
| S41 | [98] | 2010 | Journal | Bayesian network | Classification | Simulated dataset |
| S42 | [169] | 2008 | Journal | Bayesian network | Classification | Dataset gathered from deployed sensor networks in existing Australian coal mines |
| S43 | [53] | 2007 | Journal | Two localized algorithms | – | Simulated dataset |
| S44 | [58] | 2013 | Journal | k-nearest neighbor | Nearest neighbor | Real WSN dataset |
| S45 | [81] | 2009 | Journal | Clustering | Clustering | Dataset acquired from the UCI Machine Learning Repository |
| S46 | [40] | 2012 | Journal | Time series analysis and geostatistics | Statistics | Real dataset from the Swiss Alps |
| S47 | [93] | 2017 | Journal | Bayesian network | Classification | Dataset of Intel Lab |
| S48 | [99] | 2017 | Journal | Support vector machine | Classification | UCI dataset and IBRL dataset of WSNs |
| S49 | [170] | 2017 | Journal | Segment-based anomaly detection | – | Dataset of real-word received signal strength (RSS) |
| S50 | [100] | 2017 | Conference | Five different classifiers: bayesian network, neural network, nearest neighbors, support vector machine, and decision tree | Classification | Dataset from WSN in static and dynamic environments |
| S51 | [171] | 2014 | Journal | Voronoi diagram based network | – | Dataset of IRBL |
| S52 | [1] | 2011 | Journal | Survey | – | – |
| S53 | [41] | 2009 | Journal | Non-parametric and unsupervised methods | Statistical | Simulated data |
| S54 | [172] | 2008 | Journal | Survey | – | – |
| S55 | [101] | 2012 | Journal | Support vector machine | Classification | Two synthetic datasets and a real dataset gathered at the Grand St. Bernard, Switzerland |
| S56 | [5] | 2006 | Conference | Non Parametric | Statistical | Simulated dataset and real dataset from Pacific Northwest region |
| S57 | [25] | 2009 | Conference | Bayesian network and support vector machine | Classification | Simulated dataset and real dataset from Grand-St-Bernard, Switzerland |
| S58 | [102] | 2008 | Conference | Quarter sphere Support vector machine (SVM) | Classification | Dataset of Intel Berkeley Research Laboratory |
| S59 | [42] | 2005 | Conference | Gaussian distribution | Statistical | Simulated data |
| S60 | [173] | 2010 | Conference | Survey | – | – |
| S61 | [174] | 2006 | Journal | Lightweight methods | – | Simulated data |
| S62 | [175] | 2010 | Conference | Discrete Wavelet Transform (DWT) and the self-organizing map (SOM) | Classification | Synthetic dataset and actual dataset collected from a wireless sensor network |
| S63 | [59] | 2010 | Journal | Neighborhood | Nearest neighbor | Real life datasets (Annealing and Cancer) |
| S64 | [176] | 2010 | Conference | Optimization | – | Real and synthetic datasets |
| S65 | [103] | 2007 | Conference | Quarter sphere support vector machines | Classification | Real dataset gathered from a deployment of wireless sensors in the Great Duck Island project |
| S66 | [177] | 2008 | Conference | Hyperellipsoidal support vector machine | Classification | Real dataset from the Great Duck Island Project |
| S67 | [104] | 2010 | Journal | Support vector machine (CESVM) and one class quarter sphere support vector machine | Classification | Synthetic and real datasets: GDI, Ionosphere, Banana, and Synth |
| S68 | [105] | 2006 | Conference | Bayesian Networks | Classification | Dataset of habitat monitoring on Great duck island |
| S69 | [106] | 2009 | Conference | One class support vector machine | Classification | Synthetic and real datasets of the Sensor Scope System |
| S70 | [65] | 2006 | Conference | Wavelet based outlier correction and DTW distance | – | Simulated dataset |
| S71 | [43] | 2006 | Journal | Outlier detection algorithm | Statistical | – |
| S72 | [44] | 2007 | Journal | kernel density estimation and mico cluster | Statistical and classification | – |
| S73 | [70] | 2013 | Journal | k-nearest neighbor | Clustering based and nearest neighbor based | Intel Berkeley research lab and synthetic dataset |

**Table 4.** *Cont.*

| S-ID | Reference | Year | Type | Methodology | Taxonomy | Dataset |
|------|-----------|------|------|-------------|----------|---------|
| S74 | [178] | 2006 | Conference | Unsupervised learning | Classification and nearest neighbor | Dataset of the KDD-Cup 1999 network |
| S75 | [107] | 2007 | Journal | Hierarchical Bayesian model within a decision theoretic framework | Classification | Simulated dataset |
| S76 | [45] | 2008 | Conference | Density estimation | Statistical | Datasets from the UCI Machine Learning Repository and a number of synthetic datasets |
| S77 | [108] | 2008 | Conference | One class Support Vector machine | Classification | Simulated datasets |
| S78 | [46] | 2006 | Conference | linear regression and control chart | statistical | Dataset of the observation of the air pollution taken in Kuala Lumpur |
| S79 | [47] | 2007 | Journal | One sided and two sided median | Statistical | Dataset of a flight data recorder (FDR) |
| S80 | [109] | 2006 | Journal | Bayesian network | Classification | Simulated datasets |
| S81 | [8] | 2009 | Journal | Survey | – | – |
| S82 | [179] | 2005 | Journal | Local search heuristic | – | Real life datasets (lymphography and cancer) and synthetic datasets |
| S83 | [48] | 2009 | Journal | Squence-based method | Statistical | Real life datasets (lymphography and cancer) |
| S84 | [193] | 2012 | Journal | Neural Network and Rough set | Classification | Simulated Data |
| S85 | [154] | 2010 | Journal | Rules, Time series analysis, learning, and estimation methods | – | Real World datasets |
| S86 | [110] | 2006 | Conference | Dynamic bayesian networks | Classification | SERF windspeed sensor dataset streams from Corpus Christi Bay |
| S87 | [111] | 2011 | Journal | Bayesian networks | Classification | – |
| S88 | [60] | 2006 | Conference | Neighboring network | Nearest neighbor | Dataset of meteorological from various neighboring ground stations in the island of Crete in Greece |
| S89 | [61] | 2008 | Journal | Distance | Nearest neighbor | Real and synthetic datasets |
| S90 | [49] | 2010 | Journal | Kernel Density Estimation | Statistical | Real dataset from Intel Berkeley Research lab |
| S91 | [80] | 2010 | Journal | Fuzzy clustering | Clustering | Three datasets |
| S92 | [112] | 2011 | Journal | Support vector machine | Classification | Real dataset collected from a closed neighborhood from a WSN deployed in Grand-St-Bernard |
| S93 | [79] | 2011 | Journal | Clustering | Clustering | Real dataset obtained from Intel Lab's web site and synthetic dataset |
| S94 | [78] | 2009 | Conference | Hyper-ellipsoidal | Clustering | Real life dataset called the IBRL and a synthetic dataset |
| S95 | [50] | 2009 | Journal | Statistical analysis | Statistical | Dataset from a real sensor network obtained from the Intel Berkeley Research Laboratory (IBRL) |
| S96 | [51] | 2013 | Journal | Linear regression | Statistical | Real medical dataset with many (both real and synthetic) anomalies |
| S97 | [77] | 2014 | Journal | Hyperspherical clusters | Clustering | Two real sensor network deployment datasets and two synthetic datasets for evaluation purposes, namely the IBRL, GDI, Banana and Gaussmix datasets |
| S98 | [76] | 2013 | Journal | Fuzzy clustering | Clustering and Statistical | Real dataset from 54 sensors deployed at the Intel Berkeley Research Lab and artificial datasets from Intel Lab |
| S99 | [75] | 2013 | Journal | Principal component analysis (PCA) | Clustering | Real sensed dataset collected by 54 Mica2Dot sensors deployed in Intel Berkeley Research Lab |
| S100 | [113] | 2006 | Journal | Support vector machine | Classification | Dataset of the wide area airborne mine detection (WAAMD) and hyperspectral digital imagery collection experiment (HYDICE) |

**Table 4.** *Cont.*

| S-ID | Reference | Year | Type | Methodology | Taxonomy | Dataset |
|------|-----------|------|------|-------------|----------|---------|
| 101 | [52] | 2011 | Conference | Tukey and relative entropy statistics | Statistical | Dataset from RUBiS |
| 102 | [74] | 2009 | Journal | Sequence Miner | Clustering | Synthetic dataset and real dataset |
| 103 | [180] | 2013 | Journal | Survey | – | – |
| 104 | [114] | 2014 | Journal | Decision tree | Classification | Intel Berkley lab dataset |
| 105 | [27] | 2011 | Conference | Temporal technique | Statistical technique combined with nearest neighbor technique | Real datasets in different fields |
| 106 | [20] | 2017 | Journal | Survey | – | – |
| 107 | [181] | 2018 | Journal | DUCF protocol of based on fuzzy logic interface system | Clustering | Real dataset |
| 108 | [182] | 2018 | Journal | Case Study | Machine learning&Classification | – |
| 109 | [183] | 2018 | Journal | Support vector machine | Classification | Real dataset |
| 110 | [184] | 2016 | Journal | Kriging | Clustering | Real dataset |
| 111 | [185] | 2015 | Journal | Support vector machine | Classification | Dataset of multiple intelligent monitoring in intensive care (MIMIC) |
| 112 | [186] | 2017 | Conference | support vector machine | Classification | IBRL dataset |
| 113 | [187] | 2018 | Conference | support vector machine | Classification | IBRL real dataset |
| 114 | [188] | 2017 | Conference | support vector machine | Classification | Intel Berkeley Research Laboratory (IBRL) |
| 115 | [189] | 2017 | Conference | Neural network | Classification | Real dataset |
| 116 | [190] | 2018 | Conference | Bayesian network | Classification | – |
| 117 | [191] | 2018 | Journal | K-medoids | Clustering | Synthetic datasets provided by NS2 and R studio |

## 5.5. Applying Quality Assessment (QA)

The next stage was involved in assessing the quality of the selected studies by using QA, as it is recommended by Kitchenham [35]. The QA was performed for all the articles and with respect to each research question. For assessing the quality of each article, this study used four RQs as QA criteria:

- QA1: Is the topic addressed in the paper related to anomaly detection in WSN?
- QA2: Is the research methodology defined in the article?
- QA3: Is there a sufficient explanation of the background in which the study was performed?
- QA4: Is there a clear declaration concerning the research objectives?

These four QA criteria were tested among the 117 research papers to determine their reliability. The QA was comprised of three stages of quality schema, high, medium, and low [36], in which the quality of the paper relied on its loading score. For example, papers that satisfied the criteria were awarded a score of 2, papers that partially satisfied the criteria were awarded a score of 1, and papers that did not provide any information regarding the question and did not satisfy the criteria were awarded a score of 0. Consequently, based on the four defined criteria, studies with a score of 5 or above were considered with high quality, studies with a score of 4 were considered with medium quality, and studies with a score below 4 were considered with low quality. Table 5 presents the QA list of every study.

## 5.6. Data Extraction and Synthesis

A form for data extraction was developed at this phase to accurately key in all data. This was done by cautiously analyzing each study and extracting appropriate information through Mendeley and Microsoft Excel spreadsheets. The columns considered in the review were as follows: Study ID, authors, publication date, type (e.g., journal, conference proceeding), methodology, technique-based taxonomy, and datasets. Retrieval of this information was related to both research objectives and RQs. Table 6 presents the items embedded in the form, whereas Table 4 shows the data extracted from the selected 117 research papers based on the form. The extracted data were synthesized for discursive analysis to address several issues related to WSN, including advantages and disadvantages, classification, and methods.

**Table 5.** Quality assessment criterion.

| S_ID | QA1 | QA2 | QA3 | QA4 | Score |
|------|-----|-----|-----|-----|-------|
| S1 | 2 | 2 | 2 | 2 | 8 |
| S2 | 2 | 2 | 2 | 2 | 8 |
| S3 | 2 | 1 | 1 | 2 | 6 |
| S4 | 2 | 2 | 1 | 1 | 6 |
| S5 | 2 | 2 | 1 | 2 | 7 |
| S6 | 2 | 1 | 2 | 2 | 7 |
| S7 | 1 | 2 | 1 | 1 | 5 |
| S8 | 2 | 2 | 2 | 2 | 8 |
| S9 | 2 | 2 | 2 | 2 | 8 |
| S10 | 2 | 2 | 2 | 2 | 8 |
| S11 | 1 | 1 | 2 | 2 | 6 |
| S12 | 2 | 1 | 2 | 2 | 7 |
| S13 | 2 | 2 | 2 | 2 | 8 |
| S14 | 2 | 2 | 2 | 2 | 8 |
| S15 | 2 | 1 | 2 | 2 | 7 |
| S16 | 2 | 2 | 2 | 2 | 8 |
| S17 | 2 | 2 | 2 | 2 | 8 |
| S18 | 2 | 1 | 2 | 2 | 7 |
| S19 | 2 | 2 | 2 | 2 | 8 |
| S20 | 2 | 2 | 2 | 2 | 8 |
| S21 | 2 | 1 | 2 | 2 | 7 |
| S22 | 2 | 2 | 2 | 2 | 8 |
| S23 | 2 | 2 | 2 | 2 | 8 |
| S24 | 2 | 1 | 2 | 2 | 7 |
| S25 | 2 | 2 | 2 | 2 | 8 |
| S26 | 2 | 2 | 2 | 2 | 8 |
| S27 | 2 | 1 | 2 | 2 | 7 |
| S28 | 2 | 2 | 2 | 2 | 8 |
| S29 | 2 | 2 | 2 | 2 | 8 |
| S30 | 2 | 1 | 2 | 2 | 7 |
| S31 | 2 | 2 | 2 | 2 | 8 |
| S32 | 2 | 2 | 2 | 2 | 8 |
| S33 | 2 | 1 | 2 | 2 | 7 |
| S34 | 2 | 2 | 2 | 2 | 8 |
| S35 | 2 | 2 | 2 | 2 | 8 |
| S36 | 2 | 1 | 2 | 2 | 7 |
| S37 | 2 | 2 | 2 | 2 | 8 |
| S38 | 2 | 2 | 2 | 2 | 8 |
| S39 | 2 | 1 | 2 | 2 | 7 |
| S40 | 2 | 2 | 2 | 2 | 8 |
| S41 | 2 | 2 | 2 | 2 | 8 |
| S42 | 2 | 1 | 2 | 2 | 7 |
| S43 | 1 | 2 | 1 | 1 | 5 |
| S44 | 2 | 1 | 1 | 1 | 4 |
| S45 | 1 | 2 | 1 | 1 | 5 |
| S46 | 1 | 1 | 2 | 1 | 5 |
| S47 | 1 | 2 | 1 | 2 | 6 |
| S48 | 2 | 2 | 1 | 2 | 7 |
| S49 | 2 | 1 | 1 | 1 | 5 |
| S50 | 2 | 1 | 2 | 1 | 6 |
| S51 | 1 | 1 | 1 | 1 | 4 |
| S52 | 1 | 1 | 1 | 1 | 4 |
| S53 | 1 | 2 | 2 | 1 | 6 |
| S54 | 1 | 2 | 1 | 1 | 5 |
| S55 | 2 | 1 | 2 | 1 | 6 |
| S56 | 2 | 1 | 1 | 1 | 5 |
| S57 | 2 | 1 | 1 | 2 | 6 |
| S58 | 2 | 1 | 1 | 1 | 5 |

**Table 5.** *Cont.*

| S_ID | QA1 | QA2 | QA3 | QA4 | Score |
|------|-----|-----|-----|-----|-------|
| S59 | 2 | 1 | 1 | 2 | 6 |
| S60 | 2 | 1 | 2 | 2 | 7 |
| S61 | 2 | 2 | 2 | 1 | 7 |
| S62 | 2 | 1 | 1 | 1 | 5 |
| S63 | 2 | 1 | 1 | 2 | 6 |
| S64 | 2 | 1 | 2 | 2 | 7 |
| S65 | 2 | 1 | 1 | 2 | 6 |
| S66 | 2 | 2 | 2 | 1 | 7 |
| S67 | 2 | 1 | 1 | 1 | 5 |
| S68 | 1 | 2 | 2 | 2 | 7 |
| S69 | 2 | 2 | 1 | 2 | 7 |
| S70 | 2 | 1 | 1 | 1 | 5 |
| S71 | 2 | 1 | 1 | 1 | 5 |
| S72 | 2 | 2 | 1 | 1 | 6 |
| S73 | 2 | 2 | 2 | 2 | 8 |
| S74 | 2 | 1 | 1 | 2 | 6 |
| S75 | 2 | 1 | 1 | 1 | 5 |
| S76 | 1 | 2 | 2 | 1 | 6 |
| S77 | 1 | 2 | 1 | 1 | 5 |
| S78 | 2 | 1 | 1 | 1 | 5 |
| S79 | 2 | 1 | 2 | 2 | 7 |
| S80 | 1 | 2 | 2 | 1 | 6 |
| S81 | 1 | 2 | 2 | 2 | 7 |
| S82 | 2 | 1 | 2 | 1 | 6 |
| S83 | 1 | 1 | 2 | 1 | 5 |
| S84 | 1 | 1 | 0 | 1 | 3 |
| S85 | 1 | 2 | 0 | 1 | 4 |
| S86 | 1 | 1 | 1 | 1 | 4 |
| S87 | 1 | 1 | 1 | 1 | 4 |
| S88 | 1 | 2 | 0 | 1 | 4 |
| S89 | 2 | 2 | 1 | 1 | 6 |
| S90 | 1 | 1 | 1 | 1 | 4 |
| S91 | 2 | 1 | 1 | 1 | 5 |
| S92 | 2 | 1 | 1 | 1 | 5 |
| S93 | 1 | 1 | 1 | 1 | 4 |
| S94 | 2 | 0 | 1 | 1 | 4 |
| S95 | 2 | 1 | 1 | 1 | 5 |
| S96 | 2 | 2 | 2 | 2 | 8 |
| S97 | 1 | 1 | 2 | 1 | 5 |
| S98 | 2 | 1 | 1 | 1 | 5 |
| S99 | 2 | 1 | 1 | 2 | 6 |
| S100 | 2 | 1 | 1 | 1 | 5 |
| S101 | 2 | 1 | 1 | 1 | 5 |
| S102 | 2 | 2 | 1 | 1 | 6 |
| S103 | 2 | 1 | 1 | 2 | 6 |
| S104 | 2 | 2 | 1 | 1 | 6 |
| S105 | 2 | 2 | 2 | 1 | 7 |
| S106 | 1 | 1 | 2 | 2 | 6 |
| S107 | 2 | 1 | 2 | 1 | 6 |
| S108 | 2 | 2 | 1 | 2 | 7 |
| S109 | 1 | 2 | 2 | 1 | 6 |
| S110 | 1 | 2 | 1 | 1 | 5 |
| S111 | 2 | 1 | 1 | 1 | 5 |
| S112 | 2 | 1 | 2 | 2 | 7 |
| S113 | 1 | 1 | 2 | 1 | 5 |
| S114 | 2 | 1 | 1 | 1 | 5 |
| S115 | 2 | 1 | 1 | 2 | 6 |
| S116 | 1 | 1 | 2 | 2 | 6 |
| S117 | 2 | 1 | 2 | 1 | 6 |

**Table 6.** Data extraction of primary studies.

| Extracted data | Description |
| --- | --- |
| Study ID | Unique identity for each article |
| Authors | authors' names |
| Year | Publication Date |
| Type | Journal or conference |
| Methodology | E.g. bayesian network, k-nearest neighbor (kNN), support vector machine, etc |
| Taxonomy | Comparative techniques that are addressed in each paper |
| datasets | e.g. simulated data, real data, etc. |

## 5.7. Publication Sources Overview

The list of selected papers published from the year 2004 until the year 2018 is presented in Figure 8. A gradual increase was noted in the number of papers published in the field of detecting outlier detection in WSNs. This shows the increasing interest in this domain, particularly after 2005. The year 2012 saw an increase of 12 studies, compared to only seven studies for 2014 and 2015. However, 2006 had the highest publication rate with 15 studies based on their reliability, an increase in the number of impact factor journals, and an increase in the number of computer science conferences. Figure 8 shows that 83 research papers were retrieved from journals (71%), and 34 papers were from conference proceedings (29%).
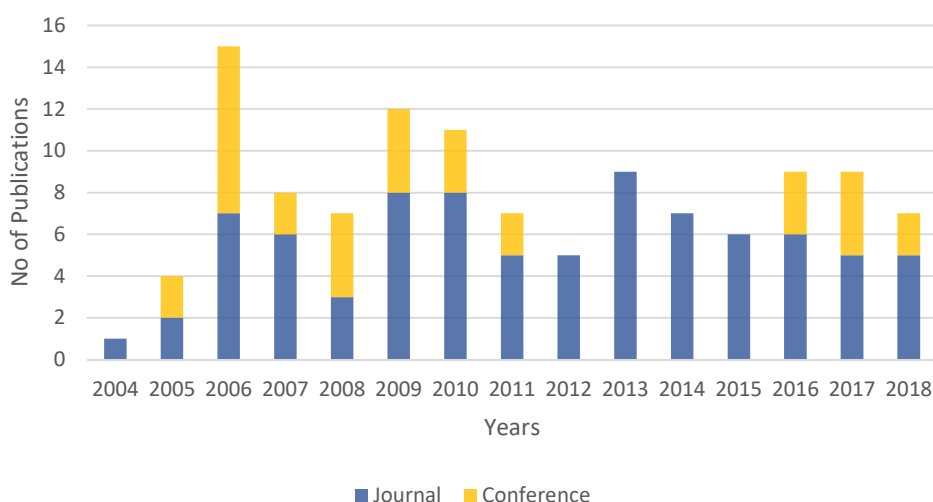


**Figure 8.** Studies distribution per year.

## 5.8. Classification of Outlier Detection Techniques Used in Previous Studies

The techniques for outlier detection in WSNs are presented in Figure 9: classification, nearest neighbor, statistical analysis, clustering, and spectral technique. A total of 38 studies use the classification approach, whereas 17 use statistical analysis, 13 use clustering, 10 use hybrid techniques, and 5 use nearest neighbor techniques.

| | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ■ Classification based | | 1 | 4 | 4 | 4 | 2 | 3 | 2 | 2 | | 1 | 4 | 1 | 6 | 4 |
| ■ Hybrid | | | 2 | 1 | | | | 1 | | 2 | | 1 | 2 | | 1 |
| ■ Statistical based | | 1 | 3 | 2 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | | 1 | | |
| ■ Clustering based | | | 2 | | | 3 | 1 | 1 | | 2 | 1 | | 1 | | 2 |
| ■ Nearest Neighbor based | | | 1 | | 1 | | 1 | | | 1 | | | | 1 | |
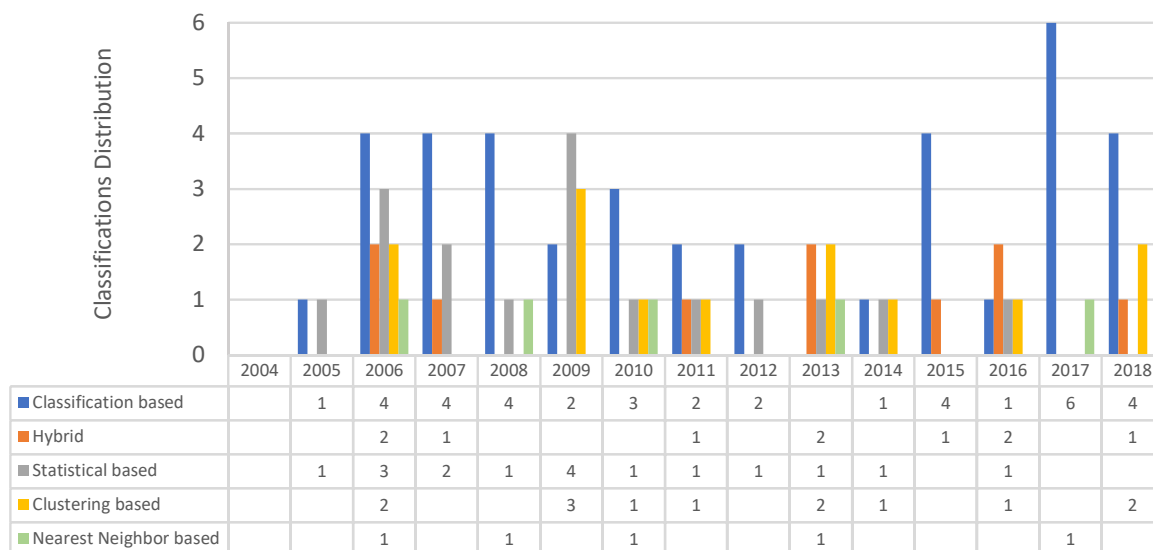
**Figure 9.** Distribution of outlier detection techniques per year.

## 6. RQ Results

The RQs of this study were addressed after extracting essential data from 117 selected research papers. Every study was mapped to the most relevant question and grouped based on similarity. The upcoming sections answer the RQs outlined in Section 4.2.

### 6.1. What is the Complete Taxonomy Framework for Outlier Detection Techniques for WSNs? (RQ1)

Lately, several studies have been employed for detecting outlier detection in WSNs. This highlights the need for a taxonomy to address all the techniques and requirements of WSNs. Figure 10 presents a taxonomy for detecting outlier detection techniques in WSN. For WSNs, outlier detection techniques can be classified into nearest neighbor-based, information-theoretic-based, statistical-based, clustering-based, spectral, classification-based, and spectral decomposition-based approaches. The statistics approach was divided into parametric, non-parametric, and hybrid methods based on the probability of the distribution model. Gaussian-based, regression-based, mixture-of-parametric-distribution-based, and non-Gaussian-based approaches are parametric approaches, whereas kernel-based and histogram-based approaches are non-parametric approaches. Furthermore, classification-based approaches are Bayesian network-based, support vector machine (SVM)-based, neural network-based, and rule-based approaches. The Bayesian network can be divided into naïve Bayesian network and dynamic Bayesian network (DBN) based on the degree of probability in dependencies among the variables. Spectral decomposition-based techniques apply principal component analysis (PCA) for outlier detection. The nearest neighbor-based methods employ distance to $K^{th}$ nearest neighbor and relative density for outlier detection. Therefore, in this study, we provide a comprehensive taxonomy framework and highlight the advantages and disadvantages of each class of outlier detection techniques under this taxonomy framework [6].
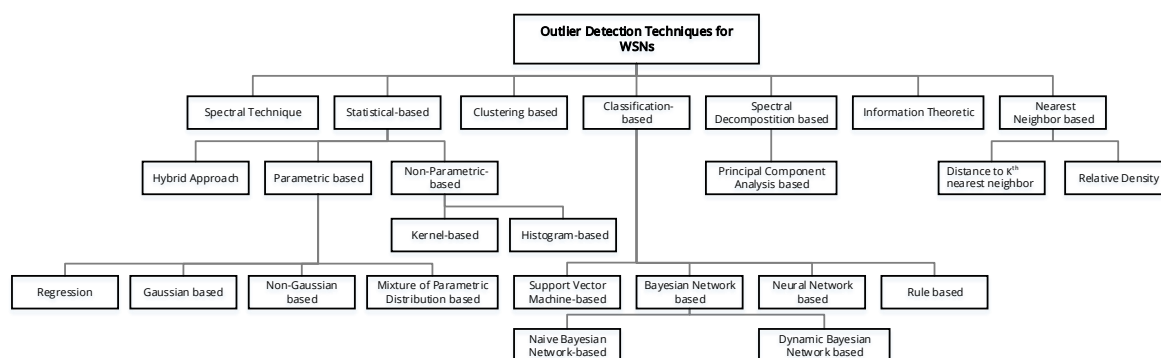
**Figure 10.** A comprehensive taxonomy for outlier detection techniques.

*6.2. What Are the Outlier Detection Techniques That Have Been Used for WSNs? (RQ2)*

Outlier detection methods for WSNs are classified in this section based on their respective disciplines. Figure 10 provides a description for each discipline.

6.2.1. Statistical-Based Approaches

Statistical-based approaches require a model for data distribution to detect outliers. A statistical model looks into data distribution and assesses the fit of data instances to the model. Data instances become outliers when the data probability produced by the model appears in distance measures. The methods are grouped into non-parametric and parametric. Parametric methods produce data from an acknowledged distribution that is presumed from data that are available based on either the Gaussian or non-Gaussian model. Meanwhile, non-parametric models dismiss data dispersion availability as a distance measure that is calculated with a statistical model and new data instances of other parameters to determine the outlier. Some of the statistical-based techniques that are considered in this paper are [11,37–52].

1. Parametric-Based Approaches: These strategies consider the accessibility of information from the fundamental data distribution. It is followed by approximation of distribution limitations using the available data. Data distribution is classified as Gaussian-based models or non-Gaussian-based models. Gaussian models are characterized by a normal distribution of data.

   - Gaussian-Based Models: Outlying sensors and sensor networks' event boundaries are identified by using two specific strategies described by [53]. These strategies depend on the spatial correspondence of the evaluation of adjacent sensor nodes to compare outlining sensors with the event boundary. The difference between readings of a node and the mean of the readings of its adjacent nodes is calculated by each node in the strategy employed for recognition of outlining sensors. This is followed by the regulation of every difference from the adjacent nodes. If the extent of variation of a reading of a node's absolute value is considerably higher than the predetermined criteria, then the node is said to be an outlying node. The event boundary recognition strategy depends on the preceding outcomes of distant sensor recognition. In this case, the node is said to be an event node if there is a significant variation in the absolute value of the extent of divergence of the node in different geological areas. These strategies do not consider the temporal association of sensor readings, so their precision is not very high.

   - Non-Gaussian-Based Models: A mathematically supported strategy is proposed by [54], where the outliers in the shape of spontaneous noise are modeled using a symmetric $\alpha$-stable ($S_\alpha S$) distribution. In this strategy, the time-space associations of sensor data are employed to recognize outliers. The anticipated data and sensing data are contrasted by every group node for identifying and correcting the temporal outliers. This corrected data from nodes

are gathered by the cluster-head to identify spatial outliers that show significant divergence from regular data. There is a reduction in communication costs that can be attributed to local transfer. Moreover, costs incurred on calculation are minimized because a major part of computations is conducted by cluster-heads. However, it may not be appropriate to apply $S_{\alpha}S$ distribution to real sensor data. Powerful alterations of network topology may be experienced by the cluster-based model.

2. Non-Parametric-Based Approaches: Accessibility of data distribution is not considered by non-parametric strategies. The space between new test cases and mathematical models is usually identified by these strategies. To identify whether the observation is an outlier or not, some criteria are applied to the measured space. Histograms and kernel density estimators are famous strategies in this regard. In histogram models, the rate of incidence of various data instances is determined by calculating the probable incidence of a data instance. Afterward, the test is contrasted with every type of histogram to determine the type to which it is associated. The probability distribution function (pdf) for regular instances is evaluated by kernel-density estimators and by employing the kernel functions. An outlier is found to be any new instance in a pdf that is found in a region characterized by a low probability.

   - Histogramming: Worldwide outliers in applications of sensor networks that are responsible for the collection of data are recognized by a strategy developed on the basis of a histogram proposed by [11]. This histogram is characterized by a minimization of the cost incurred for communication because it focuses on gathering histogram data instead of unprocessed data for further processing. Histogram information help to extract data distribution from the network and sort out non-outliers. Additional histogram data can be gathered from the network for recognizing outliers. Outliers are determined by a predetermined standard distance or by their position amongst the outliers. One shortcoming of this strategy is that communication expenses are increased because of the need to gather additional histogram data from the entire network. Moreover, merely single-dimensional data are considered by this strategy.

   - Kernel Functions: It is a strategy used for the detection of outliers online in transferring sensor data, it was recommended by [55]. It is based on kernels and is independent of the predetermined data distribution. The strategy uses the kernel density evaluator in order to estimate the fundamental distribution of sensor data. Thus, outliers are recognized by nodes in case of major divergence of value from the pre-set model of data distribution. An outlier is the value of a node whose adjacent node values do not meet the criteria set by the user. This strategy is also applicable to complex nodes for recognition of outliers overall. This strategy is highly dependent on pre-set criteria. This makes it problematic because it is very complicated to select suitable criteria. Moreover, identification of outliers in data with multiple variables may not be possible using a single criterion.

3. Evaluation of Statistical-Based Techniques: These strategies have been proved mathematically to effectively recognize outliers when an accurate model of the probability distribution is given. Additionally, the basic information on which the model is constructed is not needed afterward. However, in reality, previous information on sensor stream distribution is usually unavailable. Hence, in the absence of a predetermined distribution to be followed by sensor data, parametric strategies are deemed to be ineffective. Non-parametric strategies are more efficient because they do not depend on distribution features. Histogram models are suitable for single variable data, but in the case of multiple variables, this model fails to consider the correlation between various aspects of data. For data with multiple variables, a kernel function is a better option, specifically in terms of computation cost.

6.2.2. Nearest Neighbor Based Techniques

These techniques are widely applied to analyze data instances based on the nearest neighbor via previous machine learning and data mining. Some acknowledged distances are employed to calculate the distance of data instances. If the data instance is positioned further from the neighbor, it is called an outlier. Univariate data prefer Euclidean distance, whereas multivariate data prefer Mahalanobis distance. Some examples of these methods are outlined in [56–61]. However, these methods are not popular and have several shortcomings, as depicted in the upcoming sections.

In cases where the distant positioned data instance is deemed to be an outlier, [62,63], many processes including categorization, clustering, and outlier identification are performed using this strategy. Data distribution is not considered by these strategies, but many mathematical strategies are simplified. An outlier identification strategy based on the closest node has a clear idea of closeness. Various clear distant notions are considered as a couple of data instances, a group of instances, or a series of instances. Euclidean distance is the optimum choice for both the univariate and multivariate constant features. The strategy for resolving the issue of uncontrolled worldwide outlier identification in a system of wireless sensors was recommended by [64]. Data similarity was the basis of this strategy. Distance correspondence is used by every node for the recognition of local outliers. These outliers are subsequently transmitted to adjacent nodes for rectification.

The process continues until every sensor node in the system finally corresponds to worldwide outliers. However, the cost of communication is increased because every node employs broadcast for facilitating communication between nodes in the system. Consequently, this algorithm is suitable for systems that evaluate outlier rating confidence by tuning the sliding window to the part where the precision of the algorithm is observed, a significant communication load is exerted, and significant power consumption is required. Moreover, [65] proposed an in-network strategy for outlier clean-up to be applied to sensor system data-gathering applications. Outlier correction based on wavelets and adjacent dynamic time warping distance based on the exclusion of outliers with respect to space-time related data that are used in this strategy. This ensures efficient clean-up of the sensor data by minimizing the transfer of outliers. Thus, many outliers are corrected or eliminated from broadcast in a maximum of two steps. However, this strategy is dependent on appropriate criteria that are difficult to determine. In 2007, a new uncontrolled strategy based on distance was given by [66] for identifying worldwide outliers in a snapshot and implementing a sensor system to handle queries. An arrangement similar to that of an aggregation tree is observed here when nodes gather data from their children and then forward the constructive data to their parents.

The sink is responsible for sorting the world's leading outliers and forwarding these outliers to nodes in the system so that they can be checked. When a node does not correspond to worldwide outcomes determined by the sink, the process is performed again. Because only one dimension is considered, the cost of communication is minimized. The model of the sliding window is employed to conduct outlier queries. This identifies irregularities in the present window. To renew the addition or the removal of a present window, a single scan is conducted by the algorithm. Consequently, system efficacy is enhanced. The contribution of Angiulli et al. [67,68] was supported and broadened by Kontaki et al. [69]. They are known for their contribution to detecting universal outliers based on distance inflow of data, consequently resolve the issues of complication and use of memory. A new algorithm allowing the identification of outliers independently from the existing limitations was suggested by Yang et al. [70] proposed to calculate the ordered distance with a difference outlier factor. This strategy is based on the computation of a new outlier score for every point of data. This is done by considering the divergence between structured paces employed for the calculation of outlier scores.

The success of Local Outlier Factor (LOF) strategy and its recognition in high detection activity in dissimilar densities have proved that it is a significant strategy that can be modified in many ways. The precision of identification of LOF strategy is enhanced by some other strategies. Time complications are resolved to make the strategy precise by altering k-NNs or by conducting estimations [71]. Another strategy is to compare the efficiency of techniques based on mathematics and those

based on the closest neighborhood to recognize outliers in the process of extraction of useful data. The comparison has revealed that the mathematical strategy of the histogram-based outlier score has more points of outliers compared to neighbor-based strategy, including LOF, class outlier factors, LOOP, and improving influenced outlierness. All these works showed only some outliers with severe divergence. An uncontrolled outlier identifier based on DNOD was recommended by [56], and it allowed to examine data collected by sensors for considering dimensions of outliers.

### 6.2.3. Clustering-Based Techniques

Clustering involves grouping data instances with similar attributes into clusters [72,73]. The algorithms of clustering can be distributed or centralized. The nodes transmit all data to the central node for clustering in the centralized algorithms, which is ineffective in communication. As for distributed algorithms, the nodes can cluster the data and send certain parameters to the gateway node to minimize overhead in communication. The distance measure is employed from the nearest cluster to determine the outlier [22,70,74–82]. Euclidean distance serves as a measurement of correspondence between two data instances, but the calculation of this correspondence in data with multiple variables is very costly. The strategy is based on clustering, and outliers are recognized on this basis. The data instances are deemed to be outliers if they have no relation to clusters or if their dimensions are smaller relatively to other clusters [6,19,83]. These strategies do not have former data regarding data distribution and can be applied to the incremental model. It is plagued by issues with determining the dimensions of the cluster.

Refs. [8,80,84] detail the benefits of clustering-based techniques. These partially controlled strategies are appropriate for the innovation's identification [85], wherein regular data are used to create clusters signifying the normal form of data conduct [86,87]. Moreover, threats to the system are identified by K-means clustering, Self-Organising Maps (SOM), and expectation maximization. These methods employ clusters for categorizing test data. Similarly, a strategy has been proposed by Vinueza and Grudic [88] to detect local and universal outliers on the basis of the cluster. A data point is pronounced to be an outlier if it is located away from the clusters or if its class is located away from other points. Correspondingly, the clustering algorithm is used for the categorization of clustering-based strategies as an uncontrolled strategy. Afterward, data instances are evaluated on the basis of clusters.

In the clustering learning anomaly detectors algorithm employed by [89], an arbitrary sample was taken for calculating the mean distance between the nearest points to obtain data dimensions. The cluster was pronounced to be a local outlier if it had a density lower than that specified in the criteria, and the cluster was pronounced to be a universal outlier if it was located away from other clusters. A strategy was proposed by [90] that employed the recurrent point set mining for obtaining clusters by differentiating regular data from outliers and the COOLCAT strategy [91]. The strategy is called COOLCAT because it decreases the entropy of clusters and ultimately cools the clusters. Furthermore, a universal strategy was proposed by [22] for recognizing the offline dimensions of outliers in sensor nodes. A fixed-width algorithm for clustering is employed by every measured value of the sensor cluster. This is followed by the transmission of cluster conclusions to parent nodes. Outliers are recognized by the sink once the later receives the collected cluster statistics of the children clusters from the head cluster. An anomalous cluster is fixed in case the mean inter-distance of the clusters is more than one standard value of the group of inter-cluster distances.

The cost of communication is reduced, and energy-saving is ensured in such a way that the identification of irregularity is implemented only at the base station. However, one of the drawbacks of this strategy is that it does not apply to local and real-time decision-making. Moreover, a spatiotemporal strategy for the identification of outliers was proposed by [140]. This strategy is based on the concept of clustering known as the spatiotemporal density-based clustering in spatial databases (ST-DBSCAN), which is an extensive adaptation of the clustering strategy DBSCAN [92].

### 6.2.4. Classification-Based Techniques

Classification-based techniques can be supervised or unsupervised. The unsupervised methods learn the boundary (called sphere or quarter-sphere) during training and declare data instances outside the boundary as outliers. Nevertheless, classifiers need training for new datasets.

Classification methods are divided into SVM-based and Bayesian approaches [13,25,93–114].

Multi-class is the first group of categorization and includes neural networks and Bayesian networks. These strategies are based on the supposition that marked instances relating to multiple regular groups create the training data [115,116]. The discrimination between regular classes and other classes can only be pointed out if one has knowledge regarding classifiers. Classifiers get a confidence score from multi-category techniques. The instance is considered to be an outlier that is not belonging to any of the classifiers and this with taking into consideration that the test data are regular (i.e., none of the classifiers get a good score).

A probabilistic graphical model is employed by the strategies based on the Bayesian network for modification of a group of variables and their probable independence. Data are collected from various instances, and the probability of an instance is computed to be a part of the learned group. In 2004, a strategy was proposed by [117] to ensure structuring and learning mathematical data in WSNs. This was helpful for identifying local outliers and sorting defective sensors by applying the strategy to Bayesian model-based technique. The issues involved in understanding space-time correlations and limitations of the Bayesian classifier can be resolved by this strategy, which makes use of the classifier for probabilistic supposition. In the given model, the observed value for every sensor is controlled by the former reading of that particular sensor, and the whole values interval divides the subsequent readings in every class.

The next step is the prediction of the maximum probability class of the next reading. Here, a reading is pronounced to be an outlier if it has a lower probability in its own class as compared to other classes. A specific criterion is not needed for the recognition of outliers. This strategy can identify the lost readings in the system, but no consideration is given to multidimensional data. Bayesian networks are capable of telling if an observed value is related to class or not but do not consider provisional relation between the observed values of the sensory attributes. Similarly, a strategy based on BN was proposed for the recognition of local sensors in the flow of sensor data. BN is employed for understanding the spatiotemporal relations between various aspects and for evaluating the values that are lost from the flow of data emitting from the sensors. A year later, Ref. [118] came up with another strategy based on using DBNs along with a network topology. It developed over time to detect the local outliers in a sensor data flow. Inconsistent data can be recognized by two strategies, namely the Bayesian credible interval and the maximum posteriori measurement status. These strategies have the capacity to function in various data flows simultaneously. A Bayesian credible interval is structured for the latest dimensions and observations by hidden distributions, which are minimizing stepwise by a method known as Kalman filtering. In this method, the sensors provide the latest observed values. Outliers are the measured values that exceed the value of the anticipated interval. The second method involves more intricate DBN. This DBN identifies the outliers with the help of a couple of measured state variables. Moreover, another strategy has been proposed by experts: Hierarchical Bayesian Space-Time (HBST) [119]. In this strategy, the relations between time and space are only presumed and not computed. A tagging system is used for spotting data that do not meet the given criteria.

HBST is complicated, but it is accurate; its rate of fake identification is very low. It is much more appropriate for divergence models and unmodeled dynamics compared to linear auto-regression models. A Bayesian strategy for recognition of outliers within the data gathered using WSNs was recommended by [95]. This algorithm has many benefits: it enhances precision by resolving issues of categorization, time, and communication complications. It also makes relative improvements in the measure of latency period and uses energy in contrast to non-adaptive approaches. Various masses connected to the system are examined with the help of neural networks to create classifiers.

The neural network is a network of integrated nodes functioning similarly to the human brain. Every node is linked with adjacent nodes in closely located levels. The Replicator Neural Network (RNN) is a triple-layered network with three output and three input neurons. This neural network was used by [120] for data modeling. The input and output variables are the same in this network in order to form a clear and compact data model. The aim of this study was to measure distant data records in order to detect errors that are reforming from separate points of data. A graded score evaluator was employed to analyze the activity of the RNN. The efficiency of RNN in identifying outliers is demonstrated in two records that are accessible to the general public. This is similar to Smart Sifter [121], which creates models for recognizing outliers.

The difference lies in the technique of ranking the individuals, which is dependent on their extent of offense with the model. Sykacek [122] proposed another strategy to identify outliers using a multiple layer perception to serve as a regression model. Subsequently, outliers are perceived as data with their remaining parts located outside the error bars. WSNs models are also proposed based on RNNs for identifying outliers. Ref. [175] also proposed a general method for the identification of outliers. The purpose of this study is to recommend an algorithm allowing to identify irregularities. This method examines the identification of irregularity in sensor readings. For this purpose, SOM employing wavelet coefficients must be trained.

### 6.2.5. Information Theoretic

Various tools such as Kolmogorov complexity, entropy, and relative entropy are employed by data theoretic strategies for examining dataset components. Both physically organized data instances that are spatial and sequential data are considered. Data are simplified into simple components wherein component $I$ is identified by the outlier recognition strategy. Component $I$ has the utmost value of $C(D) - C(D - I)$. It is applicable to spatial, graphic, and sequential data. However, the determination of the most favorable dimension for components is the main concern regarding this strategy.

### 6.2.6. Spectral Decomposition-Based Approaches

PCA employs the strategies of spectral simplification [95] to reduce the volume of the data and develop patterns of regular data by proposing a model. An outlier is a data that is not capable of corresponding to the proposed model. However, PCA requires complex calculation activities to reduce the volume of data before recognizing outliers. Specifically, some main parts learn the data model, and, in the case of non-correspondence, that instance of data is regarded as an outlier. These spectral simplification strategies estimate data with characteristics ensuring the learning of inconsistencies in the data [8]. The key strategy for recognizing outliers is the determination of sub-spaces (for instance, embeddings and projections) that are appropriate for both controlled and uncontrolled circumstances.

Ref. [105] proposed a PCA-based technique to solve the data integrity and the accuracy problem caused by compromising or malfunctioning sensor nodes. This technique uses PCA to efficiently model spatiotemporal data correlations in a distributed manner and identify local outliers spanning through neighboring nodes. Each primary node that is offline builds a model of the normal condition by selecting appropriate principal components (PCs) and then obtaining sensor readings from other nodes in its group to conduct local real-time analysis. The readings that significantly vary from the modeled variation value under normal conditions are declared as outliers. The primary nodes eventually forward the information about the outlier data to the sink. The offline procedure for selecting appropriate PCs is computationally very expensive. PCA-based approaches tend to capture the normal pattern of the data using the subset of dimensions, and they can be applied to high-dimensional data. However, selecting suitable principal components, which is necessary to accurately estimate the correlation matrix of normal patterns, is computationally very expensive.

*6.3. What Are the Challenges of Outlier Techniques in WSNs? (RQ3)*

Extracting essential data from raw sensor data is vital [6]. Extracting sensor data embedded in networks designed to detect outliers is a difficult task. Common techniques are inappropriate to detect outliers in WSNs because of the following reasons:

- **Resource limitations:** Low-quality and cheap sensor nodes present several barriers, such as limited memory and energy, narrow communication bandwidth, and poor computational ability. Many common outlier detection techniques hesitate to probe into higher computational capabilities because of the high cost involved as well as the extensive storage and analysis that are needed. Thus, common sensors are inadequate to identify outliers in WSNs [6].
- **High communication cost:** A lot of energy in WSNs is channeled to radio communication, and the non-computation costs for communication in sensor nodes are higher than those for computation costs. Most common outlier detecting techniques employ centralized steps to analyze data, which causes higher energy use and communication overhead, decreasing network lifetime and blocking network traffic.
- **Distributed streaming data:** Sensor data that originate from varied channels may shift in a dynamic manner. Moreover, no model seems to spell out the distribution of these data. Additionally, calculating probabilities is a challenging task. Most techniques that identify outliers fail to satisfy the fixed criteria to process dispersion of stream data. Theoretical conceptions are unsuitable for sensor data and thus are inappropriate for WSNs.
- **Heterogeneity and mobility of nodes, frequent communication failures, dynamic network topology:** Sensor nodes placed in frenzy settings are deemed to fail because of dynamic network topology and frequent communication. Sensor nodes with varied capacities can move into different positions because each node may contain various kinds of sensors. Thus, the intricacy of generating a viable outlier detecting method for WSNs is heightened because of such dynamic and complex features.
- **Large-scale deployment:** The scale of WSNs may be massive and may thus require the higher task of detecting outliers, which cannot be performed by common sensors.
- **Identifying outlier sources:** A sensor network monitors activities and provides raw data. Nevertheless, it is difficult to determine outliers in complex and intricate WSNs. Common methods may not even be able to identify events from outliers. Hence, it is more challenging to identify outliers in WSNs from other normal events.

## 7. Advantages and Disadvantages of Existing Outlier Detection Techniques

This section compares outlier detection techniques used by previous studies and highlights the advantages and disadvantages of each algorithm.

*7.1. Statistical-Based Techniques*

Detection of outliers via the statistical method incorporates the production of observed profiles. The generated profile embeds several measures, such as activity intensity, audit record distribution, and ordinal measures (CPU usage). Two types of profiles are generated for the subjects: stored and current profiles. For the processing of network events (e.g., audit log records, incoming packets), the outlier detection system constantly updates the current system and outlier (degree of irregular activities). This is done after comparing the stored profile with the current one, that of current by employing the abnormality function of all related profile measures. When outliers exceed a particular aspect, the detection system signals an alert. Some benefits of outlier detection via statistical methods are listed in the following points:

1. The systems, similar to many outlier detection systems, do not require prior knowledge of security flaws and attacks. Hence, the systems can detect '0 day' or the latest attacks.

2. The statistical techniques offer accurate alert regarding attacks for extended periods. Thus, they are excellent signals for forthcoming DoS attacks (e.g., port scan).

Some shortcomings of the statistical methods in WSNs are as follows:

1. Skilled attackers can train a statistical outlier detection to accept abnormal behavior as normal.
2. It is challenging to determine thresholds that balance the likelihood of false positives with that of false negatives.
3. Statistical techniques demand accurate statistical distributions. However, not all behaviors can be modeled statistically. Most of the suggested outlier detection methods demand the assumption of a quasi-stationary process that cannot be estimated for most data [123].

### 7.2. Nearest-Neighbor-Based Techniques

The nearest neighbor-based outlier detection method demands distance/similarity measures based on dual data instances that can be calculated for various methods. Euclidean distance is the preferred choice for continuous features [124]. For multivariate data instances, distance/similarity is calculated for every feature and is later amalgamated [124]. In fact, numerous methods, including the clustering-based method, dismiss distance measure as a compulsory aspect. Although the measure has to be symmetric and positive, there is a need to meet the triangle disparity.

The two categories of the nearest neighbor-based outlier detection methods are: (1) methods that apply distance of data instance to its $k^{th}$ nearest neighbor as the outlier score; and (2) methods that calculate the relative density of every data instance to determine outlier score.

The benefits of nearest neighbor-based techniques are: (1) it is unsupervised and does not make any assumption about the underlying data distribution and (2) it is a straightforward method for varied types of data and requires appropriate distance measure for data [70].

### 7.3. Clustering-Based Techniques

The clustering technique is a popular choice in data mining to cluster data with similar traits [125,126]. In fact, clustering is a significant instrument for the analysis of outliers [127]. The primary presumption in many methods based on the clustering approach is that normal data are also linked to dense and huge clusters, whereas outliers are isolated or clustered in minute groups [125,127]. The benefits of clustering-based methods [6,8,70,80] are:

1. Easy to adapt with incremental mode (after learning the clusters, new points can be inserted into the system and tested for outliers).
2. Do not require supervision.
3. Appropriate to detect outliers from temporal data.
4. Have a rapid testing stage because the number of clusters that require comparisons is normally small.

Meanwhile, the drawbacks of these clustering-based techniques are:

1. Rely highly on the efficiency of clustering algorithms to capture cluster structure in normal instances.
2. Most methods that detect outliers are by-products of clustering and are thus non-optimized to detect outliers.
3. Several clustering algorithms force every instance to be assigned to some clusters. This might result in anomalies getting assigned to a large cluster and being considered as normal instances by techniques that operate under the assumption that anomalies do not belong to any cluster.
4. Some clustering algorithms insist on assigning each instance to a cluster. Thus, outliers may be linked to a large cluster and seen as a normal instance by methods that assume that outliers are always in isolation.
5. Some clustering-based methods are effective only when outliers are not a part of essential clusters.
6. There is bottleneck computation intricacy, particularly when O(N2d) clustering algorithm is applied.

### 7.4. Classification-Based Techniques

These methods can be supervised or unsupervised. The unsupervised methods learn the boundary (called sphere or quarter-sphere) at training and declare data instances outside the boundary as outliers. Nevertheless, classifiers need training for new datasets. The classification methods are divided into SVM-based and Bayesian approaches [13,25,93–114,128].

The benefits of the classification-based methods are as follows:

1. Classification-based methods, particularly multi-class approaches, apply powerful algorithms that can differentiate instances from varied classes.
2. The testing stage is rapid because the data instances are only compared with a pre-computed model.

The drawbacks of these classification-based methods are as follows:

1. They rely on the availability of accurate labels for varied normal classes, which is difficult to obtain.
2. Classification-based methods have a label for every test instance that turns into a drawback if an outlier score is desired for test instances. Several classification methods that gain probabilistic estimation scores from classifier outputs can be employed to overcome this issue [8].

### 7.5. Information Theoretic

These methods analyze information content from a dataset via information-theoretic measures such as Kolmogorov complexity, entropy, and relative entropy. Outliers in data generate irregularities in the information content of the dataset. Let $C(D)$ denote the intricacy of a given dataset, $D$. The fundamental information-theoretic method is elaborated as follows: given a dataset $D$, find the minimal subset of instances, $I$, such that $C(D) - C(D - I)$ is maximum. All the instances found in the subset are assumed to be outliers. The issue is overcome through this fundamental method of determining a Pareto-optimal solution that is not optimum, as other varied objectives require optimization. This method promotes dual optimization to reduce the size of the subset and to decrease dataset intricacy. The local search algorithm was employed by [129] to identify a subset in a linear manner by applying the entropy for intricate cases. Meanwhile, Ando proposed a method that applied the measure of information bottleneck [130]. Although the approximate methods have linear time intricacy, fundamental information-theoretic outlier detecting methods have exponential time intricacy [8]. The benefits of information-theoretic methods are as follows:

1. Do not require supervision.
2. Discard assumptions regarding underlying statistical data distribution.

The drawbacks of information-theoretic methods are as follows:

1. High reliance on the selection of information-theoretic measures. These measures often identify outliers when they are present in large numbers.
2. The information-theoretic methods used in spatial and sequence datasets depend on sub-structure size, which is challenging to determine.
3. It is challenging to link test instances with outlier scores via the information-theoretic method.

### 7.6. Spectral Decomposition-Based Approaches

These methods seek the normal behavior of data via PCA [131]. PCA minimizes dimensionality prior to the detection of outliers. A technique that incorporates data derived from varied nodes in WSNs was developed by [60]. This technique amalgamates sensor data in a distributed manner to detect outliers from several neighboring nodes. A method that is based on PCA can address issues related to the integrity of data and accuracy due to malfunctioning nodes. This method has two phases: online and offline phases. The sub-space approach is used for the online phase [132] to segregate the

data into two spaces: (1) contains normal data and reflects the modeled data trends, and (2) contains residual data. In the presence of an outlier, the residual domain has varied parameters, whereas the system can identify paths with outliers after choosing the parameters. The squared prediction error (SPE) [133] has been employed to detect abnormal settings. In the presence of an outlier, the SPE is greater than normal thresholds, whereas the system can detect nodes that have outliers. The selection of variables can vastly contribute to huge modifications in SPE. Moreover, multivariate data are weighed in for this technique, and spatiotemporal correlations are applied to identify outliers [134]. The benefits of spectral anomaly detection methods are as follows:

1.  Spectral methods can automatically minimize dimensionality and are thus adequate to handle datasets with high dimensions. They can be also applied as a pre-processing step, and they are followed by the use of existing outlier detection methods in the transformed space.
2.  Spectral methods do not require supervision.

    The drawbacks of the spectral anomaly detection methods are as follows:

1.  Spectral methods are useful if both normal data and outliers are segregated for data at lower dimensions.
2.  The methods demand computation that is highly intricate.

Table 7 reveals the common features of the current strategy for recognition of outliers. These strategies are specifically formulated for WSNs. Table 7 shows a comparative analysis of various strategies with respect to their dimension outlier (i.e., whether there are single or multiple variables involved), the status of recognition (i.e., online or offline), structural design, and space-time association. There are three main classifications of the current works according to Table 1: (1) Relation between sensor data of adjacent nodes, with respect to space, is employed by many strategies, but the problem lies in the selection of suitable adjacent ranges; (2) Relation between the sensor data, with respect to time, is considered by some strategies, but the appropriate selection of the sliding window dimension is an issue; (3) Some strategies consider space-time relation in the sensor data, completely ignoring the dependencies of various features of the sensor nodes on each other. These results have low precision in recognizing the outliers while they enhance the difficulty in calculations.

The formulation of an outlier recognition strategy that can be applied to diverse domains on the basis of various significant features is the main aim. These features include the flow of data and data involving multiple variables, the characteristics of the sensor node and its dependence on adjacent nodes, the determination of satisfactory and adaptable criteria for decision-making, and the power of renewal of sensor data and network topology. High-dimensional data and online approach for transfer of data with multiple variables ensure lower communication costs, and simplified computations can be managed by the outlier strategy of recognition under the specified criteria.

Additionally, for a better understanding of the WSN techniques, this study provides comparisons based on the algorithm, characteristics, and usability of these techniques, presented in Table 8. This table reveals how each defined technique can be applied for outlier detection in WSN based on their characteristics, usability, and drawbacks.

**Table 7.** Comparison of anomaly detection in WSN based on the previous studies.

| References | Detection Technique | Outlier Dimensional | | Detection Mode | | Detection Model | | |
|---|---|---|---|---|---|---|---|---|
| | | Univariate | Multivariate | Online | Offline | Local | Global | Centralized |
| [135] | Clustering | x | ✓ | x | ✓ | x | ✓ | x |
| [136] | Hybrid | ✓ | x | x | ✓ | x | ✓ | x |
| [137] | Statistical | - | - | x | ✓ | ✓ | ✓ | x |
| [112] | Classification | x | ✓ | ✓ | x | ✓ | x | x |
| [49] | Statistical | ✓ | x | x | ✓ | - | - | - |
| [138] | Nearest neighbor | - | - | x | ✓ | x | ✓ | x |
| [139] | Nearest neighbor | - | - | x | ✓ | x | x | ✓ |
| [40] | Statistical | - | - | ✓ | x | x | ✓ | x |
| [56] | Nearest neighbor | - | - | x | ✓ | x | ✓ | x |
| [140] | Clustering | - | - | x | ✓ | ✓ | ✓ | x |
| [141] | Classification | x | ✓ | - | - | ✓ | x | x |
| [142] | Hybrid | x | ✓ | ✓ | x | - | - | - |
| [84] | Classification | - | - | ✓ | x | - | - | - |
| [95] | Classification | - | - | ✓ | x | - | - | - |
| [143] | Hybrid | ✓ | x | - | - | ✓ | ✓ | x |
| [78] | Clustering | x | - | x | ✓ | x | ✓ | x |

**Table 8.** Comparison of techniques for anomaly detection in WSNs.

| References | Detection Technique | Characteristics | Usability/Limitations |
|---|---|---|---|
| [144] | kNN | The complexity of this technique is depending on the number of dimensions | Valuable, scalable, efficient, and human independent solution |
| [145,146] | Spectral | The detection performance is highly depending on the choices of features and distance measure | Robust to parameter perturbations and good performances with different anomaly scoring metrics |
| [147,148] | Gaussian | Use of the spatial correlation to determine outlying sensors and event boundaries | The accuracy is not relatively high due to the ignorance of the temporal correlation of sensor readings |
| [134,147] | Non-Gaussian | Use of the spatio-temporal correlations of data to locally detect outliers | Reduction of the communication cost (due to local transmission) and of the computational cost (due to the execution of tasks by the cluster-heads) |
| [147,149] | kernel | Use of kernel density estimator to approximate the underlying distribution of sensor data | High dependency on threshold definition (the choice of an appropriate threshold is quite difficult and a single threshold may also not be suitable for outlier detection in multi-dimensional data) |
| [20,147] | Histogram | Reduction of the communication cost by collecting histogram information rather than collecting raw data for centralized processing | The collection of more histogram information from the whole network will cause a communication overhead. In addition, this technique only considers one-dimensional data |
| [134,147] | Naïve Bayesian Network | Computation of the probabilities of each node locally | The spatial neighborhood under the dynamic change of network topology is not specified. In addition, this technique deals only with one-dimensional data |
| [134,147] | Bayesian Network (BN) | Use of BN to capture the spatio-temporal correlations that exist between the observations of sensor nodes and the conditional dependence between the observations of sensor attributes | Improvement of the accuracy in detecting outliers as it considers conditional dependencies between the attributes |
| [134,147] | Dynamic Bayesian Network | Identification of outliers by computing the posterior probability of the most recent data values in a sliding window | Possibility of operation on several data streams at once |
| [26,134,147] | Support vector machine | Mapping of the data into a higher dimensional feature space where it can be easily separated by a hyperplane | Identification of outliers from the data measurements collected after a long-time window and is not performed in real-time. In addition, this technique ignores the spatial correlation between neighboring nodes, which leads to inaccurate results of local outliers |

## 8. Evaluation of Outlier Detection Techniques

In this section, we provide an overview of used techniques for outlier detection techniques for WSNs and the requirements that an optimal outlier detection technique should meet.

**Statistical-based approaches**: They are more adapted when a small number of outliers exist in the WSN data. Statistical-based approaches work in an unsupervised way by building statistically-based models and applying descriptive statistics to detect outliers.

**Parametric-based approaches**: They are suitable for underlying WSN data that can be modeled by a probability distribution. Generally, parametric-based approaches can be used in Gaussian and non-Gaussian models. Gaussian models are used when the WSN data are compared with the neighbor in spatial correlation mode. In this case, Gaussian models need a pre-selected threshold to detect anomaly data. However, non-Gaussian models are used for local outlier detection. In this case, they use temporal correlation for outlier detection.

**Non-parametric-based approaches**: These approaches are interesting since no assumption about the distribution of WSNs data are required. Non-parametric-based approaches include histogram-based and kerned-based models. The first models involve determining the frequency of occurrence of different data instances. They can achieve excellent results for univariate WSNs data but less for multivariate data with interactions between the attributes. The second type, kerned-based models, uses kernel density to estimate the probability distribution function of sensor data. They can achieve excellent results with multivariate WSNs data with a good computational time.

**Nearest Neighbor-based approaches**: They are very convenient when the distance between two neighbor sensors is the key matter for the analysis of the WSN data. The nearest neighbor technique is one of the well-known techniques not only in WSN but also in data mining and machine learning. This technique requires the use of several distances between two sensor nodes. The goal of using nearest neighbor-based approaches is to assume that normal WSN data occur in dense neighborhoods, while outliers are far away from their closest neighbors.

**Clustering-based approaches**: They are used when similar WSN data instances are very important for data mining. These techniques provide WSN data in clusters with similar behavior. After that, points that are not within clusters can be considered as anomalies.

**Classification-based approaches**: They are divided into two types: supervised and unsupervised. Supervised techniques require labeling the WSN data and dividing it into training and testing parts. Unsupervised techniques do not require labeling the data; they determine the boundary of the normal instances and identify new instances existing outside this boundary as an outlier.

The SLR conducted in this work indicates an important need to design techniques related to outlier detection for WSN. The summary of the studied works can result in the following requirements that an optimal outlier detection technique should meet:

- High outlier detection rate.
- High scalability.
- High distinction between erroneous measurements and events.
- Low computational complexity and easy implementation.
- Consideration of correlation between attributes, spatial/spatiotemporal, and multivariate sensory data.
- Unsupervised techniques are preferred since the learning phase for WSN sensory data are a difficult task for supervised methods.
- Non-parametric methods are preferred for WSN sensory data due to the absence of knowledge about the data distribution.
- Energy-efficient and robust to communication failures.

## 9. Conclusions

The proposed study discussed outlier detection in WSNs. The study also provided information regarding WSN applications and definitions of outliers in previous studies. Moreover, different types

of outlier sources in WSNs were discussed in detail. The study endeavored to provide a comprehensive report on outlier detection in the field of WSNs. The study used the systematic literature protocol and guidelines presented by Kitchenham. Data were collected from primary studies published from 2004 to October 2018 in the form of conference proceedings and journal articles. The study summarized and organized the existing literature related to outlier and anomaly detection in WSNs based on the defined keywords and RQs. A total of 117 primary studies were included based on the defined exclusion, inclusion, and quality criteria. The results of the proposed study presented the complete taxonomy framework for outlier detection techniques for WSNs. This study also introduced the key characteristics and brief explanations of existing outlier detection techniques, which were applied in the anticipated taxonomy framework. The study presented a list of techniques, and compared outlier detection techniques and their advantages and disadvantages used in each application domain. In addition, the challenges of outlier techniques in WSNs were explained.

Finally, the study provided a comparison of the defined techniques in terms of their characteristics, usability, and drawbacks for outlier detection in WSNs. The limitations of the existing techniques for WSNs call for new anomaly detection techniques that take into account multivariate data and the dependencies of attributes of the sensor node to offer reliable, real-time adaptive detection while considering the unique characteristics of WSNs. An interesting perspective of the proposed work would be to conduct a review of deep-learning-based methods [150] for outlier detection in WSNs.

**Author Contributions:** Methodology, M.S., S.A., M.D., W.B. and A.S.; Project administration, W.B., R.A., and H.C.; Writing-original draft, M.S., S.A., M.D., and W.B.; Writing-review & editing, M.S., S.A., M.D., W.B., A.S and MI.S. All authors have read and agreed to the published version of the manuscript.

## References

1. Xie, M.; Han, S.; Tian, B.; Parvin, S. Anomaly detection in wireless sensor networks: A survey. *J. Netw. Comput. Appl.* **2011**, *34*, 1302–1325, doi:10.1016/j.jnca.2011.03.004.
2. Myllyla, R.; The Mendeley Support Team. Vital Sign Monitoring System with Life Emergency Event Detection using Wireless Sensor Network. In Proceedings of the 2006 5th IEEE Conference on Sensors, Daegu, Korea, 22–25 October 2006; pp. 518–521.
3. You, Z.; Mills-Beale, J.; Pereles, B.D.; Ong, K.G. A Wireless, Passive Embedded Sensor for Real-Time Monitoring of Water Content in Civil Engineering Materials. *IEEE Sensors J.* **2008**, *8*, 2053–2058.
4. Hao, Q.; Brady, D.J.; Guenther, B.D.; Burchett, J.B.; Shankar, M.; Feller, S. Human tracking with wireless distributed pyroelectric sensors. *IEEE Sensors J.* **2006**, *6*, 1683–1695.
5. Subramaniam, S.; Palpanas, T.; Papadopoulos, D.; Kalogeraki, V.; Gunopulos, D. Online outlier detection in sensor data using non-parametric models. In Proceedings of the VLDB '06: Proceedings of the 32nd International Conference on Very Large Data Bases, Seoul, Korea, 12–15 September 2006; pp. 187–198.
6. Meratnia, N.; Havinga, P. Outlier Detection Techniques for Wireless Sensor Networks: A Survey. *IEEE Commun. Surv. & Tutorials* **2010**, *12*, 159–170, doi:10.1109/SURV.2010.021510.00088.
7. Ganguly, A.R.; Gama, J.; Omitaomu, O.A.; Gaber, M.; Vatsavai, R.R. *Knowledge Discovery From Sensor Data*; CRC Press: Boca Raton, FL, USA, 2008.
8. Chandola, V.; Banerjee, A.; Vipin, K. Anomaly Detection: A Survey. *ACM Comput. Surv.* **2009**, *41*, 1–6, doi:10.1145/1541880.1541882.
9. Esnaola-Gonzalez, I.; Bermúdez, J.; Fernández, I.; Fernández, S.; Arnaiz, A. Towards a Semantic Outlier Detection Framework in Wireless Sensor Networks. In Proceedings of the 13th International Conference on Semantic Systems–Semantics2017, Amsterdam, Netherlands, 12-13 September 2017; pp. 152–159, doi:10.1145/3132218.3132226.
10. Fontugne, R.; Ortiz, J.; Tremblay, N.; Borgnat, P.; Flandrin, P.; Fukuda, K.; Culler, D.; Esaki, H. Strip, bind, and search: A method for identifying abnormal energy consumption in buildings. In Proceedings of the 2013

ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Philadelphia, PA, USA, 2013; pp. 129–140.

11. Sheng, B.; Li, Q.; Mao, W.; Jin, W. Outlier detection in sensor networks. In Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing—MobiHoc '07, Montreal, Quebec, Canada, 2007; pp. 219–228, doi:10.1145/1288107.1288137.

12. Hawkins, D.M. *Identification of Outliers*; Springer Netherlands: Dordrecht, The Netherlands, 1980, doi:10.1007/978-94-015-3994-4.

13. Titouna, C.; Aliouat, M.; Gueroui, M. Outlier Detection Approach Using Bayes Classifiers in Wireless Sensor Networks. *Wirel. Pers. Commun.* **2015**, *85*, 1009–1023, doi:10.1007/s11277-015-2822-3.

14. Barnett, V.; Lewis, T. *Outliers in Statistical Data*; Wiley: Hoboken, NJ, USA, 1974.

15. Breunig, M.M.; Kriegel, H.P.; Ng, R.T.; Sander, J. LOF: Identifying density-based local outliers. *ACM Sigmod Rec.* **2000**, *29*, 93–104.

16. Cheng, T.; Li, Z. A multiscale approach for spatio-temporal outlier detection. *Trans. GIS* **2006**, *10*, 253–263.

17. Aggarwal, C.C.; Philip, S.Y. An effective and efficient algorithm for high-dimensional outlier detection. *VLDB J.* **2005**, *14*, 211–221.

18. Muthukrishnan, S.; Shah, R.; Vitter, J.S. Mining deviants in time series data streams. In Proceedings of the 16th International Conference on Scientific and Statistical Database Management, **2004**; pp. 41–50.

19. Jiang, M.F.; Tseng, S.S.; Su, C.M. Two-phase clustering process for outliers detection. *Pattern Recognit. Lett.* **2001**, *22*, 691–700.

20. Ayadi, A.; Ghorbel, O.; Obeid, A.M.; Abid, M. Outlier detection approaches for wireless sensor networks: A survey. *Comput. Netw.* **2017**, *129*, 319–333, doi:10.1016/j.comnet.2017.10.007.

21. Buratti, C.; Conti, A.; Dardari, D.; Verdone, R. An overview on wireless sensor networks technology and evolution. *Sensors* **2009**, *9*, 6869–6896.

22. Rajasegarar, S.; Leckie, C.; Palaniswami, M.; Bezdek, J. Distributed Anomaly Detection in Wireless Sensor Networks. In Proceedings of the 2006 10th IEEE Singapore International Conference on Communication Systems, 2006; pp. 1–5, doi:10.1109/ICCS.2006.301508.

23. Krishnamachari, B.; Iyengar, S. Distributed Bayesian algorithms for fault-tolerant event region detection in wireless sensor networks. *IEEE Trans. Comput.* **2004**, *53*, 241–250.

24. Ding, M.; Chen, D.; Xing, K.; Cheng, X. Localized Fault-Tolerant Event Boundary Detection in Sensor Networks. In Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies, 2005; pp. 902–913, doi:10.1109/INFCOM.2005.1498320.

25. Bahrepour, M.; Zhang, Y.; Meratnia, N.; Havinga, P.J. Use of event detection approaches for outlier detection in wireless sensor networks. In Proceedings of the 2009 International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), Melbourne, VIC, Australia, 2009, pp. 439–444.

26. Shahid, N.; Naqvi, I.H.; Qaisar, S.B. Characteristics and classification of outlier detection techniques for wireless sensor networks in harsh environments: A survey. *Artif. Intell. Rev.* **2012**, *43*, 193–228, doi:10.1007/s10462-012-9370-y.

27. Ghaddar, A.; Razafindralambo, T.; Simplot-Ryl, I.; Tawbi, S.; Hijazi, A. Algorithm for temporal anomaly detection in WSNs. In Proceedings of the 2011 IEEE Wireless Communications and Networking Conference, WCNC 2011, Cancun, Quintana Roo, Mexico, 28-31 March 2011; pp. 743–748, doi:10.1109/WCNC.2011.5779255.

28. Hanafizadeh, P.; Keating, B.W.; Khedmatgozar, H.R. A systematic review of Internet banking adoption. *Telemat. Inform.* 2014, 31, 492–510, doi:10.1016/j.tele.2013.04.003.

29. Asadi, S.; Hussin, A.R.C.; Dahlan, H.M. Organizational research in the field of Green IT: A systematic literature review from 2007 to 2016. *Telemat. Informatics* **2017**, *34*, 1191–1249, doi:10.1016/j.tele.2017.05.009.

30. Asadi, S.; Abdullah, R.; Yah, Y.; Nazir, S. Understanding Institutional Repository in Higher Learning Institutions: A systematic literature review and directions for future research. *IEEE Access* **2019**, *7*, 35242–35263.

31. Kitchenham, B. Procedures for performing systematic reviews. *Keele, UK, Keele Univ.* **2004**, *33*, 2004.

32. Keele, S. Guidelines for performing systematic literature reviews in software engineering. In *Technical Report, Ver. 2.3 EBSE Technical Report. EBSE, Keele University and Durham University, UK*; 2007.

33. Bandara, W.; Miskon, S.; Fielt, E. A Systematic, Tool-Supported Method for Conducting Literature Reviews in IS. *Inf. Syst. J.* **2011**, pp. 1 – 14.

34. Webster, J.; Watson, R.T. Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Q.* **2002**, *26*, 13–23, doi:10.1.1.104.6570.

35. Kitchenham, B.A.; Charters, S. Guidelines for performing Systematic Literature Reviews in Software Engineering. *Keele Univ. Univ. Durh.* **2007**, *2*, 1–65.

36. Nidhra, S.; Yanamadala, M. Knowledge Transfer Challenges and Mitigation Strategies in Global Software Development. *Int. J. Inf. Manag.* **2012**, *33*, 333–355.

37. Ghorbel, O.; Obeid, A.M.; Abid, M.; Snoussi, H. One class outlier detection method in wireless sensor networks: Comparative study. In Proceedings of the 2016 24th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, 22-24 Sept. 2016; pp. 1–8, doi:10.1109/SOFTCOM.2016.7772168.

38. Ghorbel, O.; Ayedi, W.; Snoussi, H.; Abid, M. Fast and efficient outlier detection method in wireless sensor networks. *IEEE Sensors J.* **2015**, *15*, 3403–3411, doi:10.1109/JSEN.2015.2388498.

39. Luo, X.; Dong, M.; Huang, Y. On distributed fault-tolerant detection in wireless sensor networks. *IEEE Trans. Comput.* **2006**, *55*, 58–70, doi:10.1109/TC.2006.13.

40. Zhang, Y.; Hamm, N.; Meratnia, N.; Stein, a.; van de Voort, M.; Havinga, P.J.M. Statistics-based outlier detection for wireless sensor networks. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1373–1392, doi:10.1080/13658816.2012.654493.

41. Branch, J.W.; Giannella, C.; Szymanski, B.; Wolff, R.; Kargupta, H. In-Network Outlier Detection in Wireless Sensor Networks *Knowl. Inf. Syst.* **2009**, *34*, 23–54, doi:10.1007/s10115-011-0474-5.

42. Du, W.; Fang, L. LAD : Localization Anomaly Detection for Wireless Sensor Networks. In Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium, Denver, CO, USA, 4-8 April 2005, 2005, doi:10.1109/IPDPS.2005.267.

43. Otey, M.E.; Ghoting, A.; Parthasarathy, S. Fast distributed outlier detection in mixed-attribute data sets. *Data Min. Knowl. Discov.* **2006**, *12*, 203–228, doi:10.1007/s10618-005-0014-6.

44. Su, L.; Han, W.; Yang, S.; Zou, P.; Jia, Y. Continuous Adaptive Outlier Detection on Distributed Data Streams. *Lect. Notes Comput. Sci.* **2007**, pp. 74–85.

45. Aggarwal, C.; Philip, S. Outlier Detection with Uncertain Data. In Proceedings of the 2008 SIAM International Conference on Data Mining, Atlanta, Georgia, USA, April 24-26 2008.

46. Bakar, Z.A.; Mohemad, R.; Ahmad, A.; Deris, M.M. A Comparative Study for Outlier Detection Techniques in Data Mining. In Proceedings of the IEEE Conference on Cybernetics and Intelligent Systems, Bangkok, Thailand, 7-9 June 2006; pp. 1–6, doi:10.1109/ICCIS.2006.252287.

47. Basu, S.; Meckesheimer, M. Automatic outlier detection for time series: An application to sensor data. *Knowl. Inf. Syst.* **2007**, *11*, 137–154, doi:10.1007/s10115-006-0026-6.

48. Jiang, F.; Sui, Y.; Cao, C. Some issues about outlier detection in rough set theory. *Expert Syst. Appl.* **2009**, *36*, 4680–4687, doi:10.1016/j.eswa.2008.06.019.

49. Samparthi, V.S.K.; Verma, H.K. Outlier Detection of Data in Wireless Sensor Networks Using Kernel Density Estimation. *Int. J. Comput. Appl.* **2010**, *5*, 975–8887, doi:10.5120/924-1302.

50. Rajasegarar, S.; Bezdek, J.C.; Leckie, C.; Palaniswami, M. Elliptical anomalies in wireless sensor networks. *ACM Trans. Sens. Netw.* **2009**, *6*, 1–28, doi:10.1145/1653760.1653767.

51. Salem, O.; Mehaoua, A. Anomaly Detection in Medical Wireless Sensor Networks. *J. Comput. Sci. Eng.* **2013**, *7*, 272–284, doi:10.5626/JCSE.2013.7.4.272.

52. Wang, C.; Viswanathan, K.; Choudur, L.; Talwar, V.; Satterfield, W.; Schwan, K. Statistical techniques for online anomaly detection in data centers. In Proceedings of the 12th IFIP/IEEE International Symposium on Integrated Network Management, IM 2011, Dublin, Ireland, 23-27 May 2011; pp. 385–392, doi:10.1109/INM.2011.5990537.

53. Wu, W.; Cheng, X.; Ding, M.; Xing, K.; Liu, F.; Deng, P. Localized outlying and boundary data detection in sensor networks. *IEEE Trans. Knowl. Data Eng.* **2007**, *19*, 1145–1156, doi:10.1109/TKDE.2007.1062.

54. Hida, Y.; Huang, P.; Nishtala, R. *Aggregation Query Under Uncertainty in Sensor Networks*; Department of Electrical Engineering and Computer Science. University of California: Berkeley, CA, USA, 2007; pp. 1–17.

55. Palpanas, T.; Papadopoulos, D.; Kalogeraki, V.; Gunopulos, D. Distributed deviation detection in sensor networks. *ACM Sigmod Rec.* **2003**, *32*, 77–82, doi:10.1145/959060.959074.

56. Abid, A.; Kachouri, A.; Mahfoudhi, A. Anomaly detection through outlier and neighborhood data in Wireless Sensor Networks. In Proceedings of the 2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Monastir, Tunisia, 2016; pp. 26–30, doi:10.1109/ATSIP.2016.7523045.

57. Hwj, H.; Iacca, G.; Tejada, A.; Wörtche, H.J.; Liotta, A. Spatial anomaly detection in sensor networks using neighborhood information. *Inf. Fusion* **2017**, *33*, 41–56, doi:10.1016/j.inffus.2016.04.007.

58. Xie, M.; Hu, J.; Han, S.; Chen, H.H. Scalable hypergrid k-NN-based online anomaly detection in wireless sensor networks. *IEEE Trans. Parallel Distrib. Syst.* **2013**, *24*, 1661–1670, doi:10.1109/TPDS.2012.261.

59. Chen, Y.; Miao, D.; Zhang, H. Neighborhood outlier detection. *Expert Syst. Appl.* **2010**, *37*, 8745–8749, doi:10.1016/j.eswa.2010.06.040.

60. Chatzigiannakis, V.; Papavassiliou, S.; Grammatikou, M.; Maglaris, B. Hierarchical anomaly detection in distributed large-scale sensor networks. In Proceedings of the 11th IEEE Symposium on Computers and Communications, ISCC'06, Pula-Cagliari, Sardinia, Italy, 2006; pp. 761–767.

61. Ghoting, A.; Parthasarathy, S.; Otey, M.E. Fast mining of distance-based outliers in high-dimensional datasets. *Data Min. Knowl. Discov.* **2008**, *16*, 349–364, doi:10.1007/s10618-008-0093-2.

62. Knorr, E.M.; Ng, R.T. Algorithms for Mining Distance-Based Outliers in Large Datasets. *Proc. 24th VLDB Conf.* **1998**, *98*, 392–403.

63. Ramaswamy, S.; Rastogi, R.; Shim, K. Efficient algorithms for mining outliers from large data sets. *ACM Sigmod Rec.* **2000**, pp. 427–438, doi:10.1145/342009.335437.

64. Branch, J.; Szymanski, B.; Giannella, C.; Wolff, R.W.R.; Kargupta, H. In-Network Outlier Detection in Wireless Sensor Networks. In Proceedings of the 26th IEEE International Conference on Distributed Computing Systems (ICDCS'06), 2006; pp. 51–51, doi:10.1109/ICDCS.2006.49.

65. Zhuang, Y.; Chen, L. In-network Outlier Cleaning for Data Collection in Sensor Networks. *Workshop in VLDB*, Seoul, Korea, 12-15 September **2006**.

66. Kejia Zhang Hong Gao, Jianzhong Li, S.S. Unsupervised Outlier Detection in Sensor Networks Using Aggregation Tree. *Adv. Data Min. Appl.* **2007**, *4632*, 158–169.

67. Angiulli, F.; Fassetti, F. Detecting distance-based outliers in streams of data. In Proceedings of the Sixteenth ACM Conference on Conference On Information and Knowledge Management, CIKM '07, Lisbon, Portugal, November 6-10, 2007, doi:10.1145/1321440.1321552.

68. Yang, D.; Rundensteiner, E.a.; Ward, M.O. Neighbor-based pattern detection for windows over streaming data. In Proceedings of the 12th International Conference on Extending Database Technology Advances in Database Technology, EDBT'09, March 2009, Pages 529–540, doi:10.1145/1516360.1516422.

69. Kontaki, M.; Gounaris, A.; Papadopoulos, A.N.; Tsichlas, K.; Manolopoulos, Y. Continuous monitoring of distance-based outliers over data streams. In Proceedings of the IEEE 27th International Conference on Data Engineering, Hannover, Germany, 11-16 April 2011, doi:10.1109/ICDE.2011.5767923.

70. Fawzy, A.; Mokhtar, H.M.O.; Hegazy, O. Outliers detection and classification in wireless sensor networks. *Egypt. Inform. J.* **2013**, *14*, 157–164, doi:10.1016/j.eij.2013.06.001.

71. Papadimitriou, S.; Kitagawa, H.; Gibbons, P.B.; Faloutsos, C. LOCI: Fast outlier detection using the local correlation integral. In Proceedings of the 19th International Conference on Data Engineering, Bangalore, India, 5-8 March 2003, pp. 315–326, doi:10.1109/ICDE.2003.1260802.

72. Boulila, W.; Farah, I.R.; Ettabaa, K.S.; Solaiman, B.; Ghézala, H.B. Spatio-Temporal Modeling for Knowledge Discovery in Satellite Image Databases. CORIA 2010, 7th French Information Retrieval Conference, Sousse, Tunisia, March 18-20, 2010, pp. 35–49, doi:10.24348/coria.2010.35

73. Boulila, W. A top-down approach for semantic segmentation of big remote sensing images. *Earth Sci. Informatics* **2019**, *12*, 295–306.

74. Budalakoti, S.; Srivastava, A.N.; Otey, M.E. Anomaly detection and diagnosis algorithms for discrete symbol sequences with applications to airline safety. *IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev.* **2009**, *39*, 101–113, doi:10.1109/TSMCC.2008.2007248.

75. Livani, A.A.; Abadi, M.; Alikhani, M. Outlier detection in wireless sensor networks using distributed principal component analysis *J. Data Min.* **2013**. *1*, 1–11, doi:10.22044/jadm.2013.98.

76. Kumarage, H.; Khalil, I.; Tari, Z.; Zomaya, A. Distributed anomaly detection for industrial wireless sensor networks based on fuzzy data modelling. *J. Parallel Distrib. Comput.* **2013**, *73*, 790–806, doi:10.1016/j.jpdc.2013.02.004.

77. Rajasegarar, S.; Leckie, C.; Palaniswami, M. Hyperspherical cluster based distributed anomaly detection in wireless sensor networks. *J. Parallel Distrib. Comput.* **2014**, *74*, 1833–1847.

78. Moshtaghi, M.; Rajasegarar, S.; Leckie, C.; Karunasekera, S. Anomaly Detection by Clustering Ellipsoids in Wireless Sensor Networks Masud. *Intell. Sensors, Sens. Networks Inf. Process.* **2009**, *15*, 331–336.

79. Hassan, A.F.; Mokhtar, H.M.O.; Hegazy, O. A Heuristic Approach for Sensor Network Outlier Detection. *Science* **2011**, *1*, 66–72.

80. Al-Zoubi, M.; Al-Dahoud, A.; Yahya, A. New outlier detection method based on fuzzy clustering. *WSEAS Trans. Inf.* **2010**, *7*, 681–690.

81. Yang, P.; Zhu, Q.; Zhong, X. Subtractive clustering based RBF neural network model for outlier detection. *J. Comput.* **2009**, *4*, 755–762, doi:10.4304/jcp.4.8.755-762.

82. Bandyopadhyay, S.; Giannella, C.; Maulik, U.; Kargupta, H.; Liu, K.; Datta, S. Clustering distributed data streams in peer-to-peer environments. *Inf. Sci.* **2006**, *176*, 1952–1985, doi:10.1016/j.ins.2005.11.007.

83. Yu, D.; Sheikholeslami, G.; Zhang, A. FindOut: Finding Outliers in Very Large Datasets. *Knowl. Inf. Syst.* **2002**, doi:10.1007/s101150200013.

84. Ghorbel, O.; Abid, M.; Snoussi, H. Improved KPCA for outlier detection in Wireless Sensor Networks. In Proceedings of the 2014 1st International Conference on Advanced Technologies for Signal and Image Processing, ATSIP 2014, Sousse, Tunisia, 17-19 March 2014, doi:10.1109/ATSIP.2014.6834666.

85. Allan, J.; Carbonell, J.; Doddington, G.; Yamron, J.; Yang, Y. Topic detection and tracking pilot study: Final report. *DARPA Broadcast News Transcr. Underst. Workshop* **1998**, doi:10.1007/BF02985802.

86. Marchette, D.J. A Statistical Method for Profiling Network Traffic. *1st Workshop on Intrusion Detection and Network Monitoring* , **1999**, 119–128.

87. Wu, N.; Zhang, J. Factor analysis based anomaly detection. *IEEE Syst. Man Cybern. Soc. Inf. Assur. Workshop*, 2003, doi:10.1109/SMCSIA.2003.1232408.

88. Vinueza, A.; Grudic, G. *Unsupervised Outlier Detection and Semi-Supervised Learning*; Technical Report CU-CS-976-04; 2004.

89. Chan, P.K.; Mahoney, M.V.; Arshad, M.H. *A Machine Learning Approach to Anomaly Detection*; Florida Institute of Technology: Melbourne, FL, USA, 2003.

90. Barbará, D.; Li, Y.; Couto, J.; Lin, J.L.; Jajodia, S. Bootstrapping a data mining intrusion detection system. In Proceedings of the 2003 ACM Symposium on Applied Computing, SAC '03, Melbourne, Florida, March, 2003, pp. 421-425, doi:10.1145/952532.952616.

91. Barbará, D.; Li, Y.; Couto, J. COOLCAT: An Entropy-Based Algorithm for Categorical Clustering. *Entropy* **2002**, doi:10.1145/584792.584888.

92. Sander, J.; Ester, M.; Kriegel, H.P.; Xu, X. Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications. *Data Min. Knowl. Discov.* **1998**, [10.1.1.71.1980], doi:10.1023/A:1009745219419.

93. Wang, M.; Xue, A.; Xia, H. Abnormal Event Detection in Wireless Sensor Networks Based on Multiattribute Correlation. *J. Electr. Comput. Eng.* **2017**, *2017*.

94. Yao, H.; Cao, H.; Li, J. Comprehensive Outlier Detection in Wireless Sensor Network with Fast Optimization Algorithm of Classification Model. *Int. J. Distrib. Sens. Netw.* **2015**. *2015*, doi:10.1155/2015/398761.

95. De Paola, A.; Gaglio, S.; Re, G.L.; Milazzo, F.; Ortolani, M. Adaptive distributed outlier detection for WSNs. *IEEE Trans. Cybern.* **2015**, *45*, 888–899, doi:10.1109/TCYB.2014.2338611.

96. Gil, P.; Martins, H.; Cardoso, A.; Palma, L. Outliers detection in non-stationary time-series: Support vector machine versus principal component analysis. In Proceedings of the 2016 12th IEEE International Conference on Control and Automation (ICCA), Kathmandu, Nepal, 1-3 June 2016, Volume 1, pp. 701–706, doi:10.1109/ICCA.2016.7505361.

97. Ni, K.; Pottie, G. Bayesian selection of non-faulty sensors. *IEEE Int. Symp. Inf. Theory Proc.* **2007**, pp. 616–620, doi:10.1109/ISIT.2007.4557293.

98. Takruri, M.; Challa, S.; Chakravorty, R. Recursive bayesian approaches for auto calibration in drift aware wireless sensor networks. *J. Netw.* **2010**, *5*, 823–832, doi:10.4304/jnw.5.7.823-832.

99. Feng, Z.; Fu, J.; Du, D.; Li, F.; Sun, S. A new approach of anomaly detection in wireless sensor networks using support vector data description. *Int. J. Distrib. Sens. Networks* **2017**, *13*, 155014771668616, doi:10.1177/1550147716686161.

100. Ayadi, A.; Ghorbel, O. Performance of outlier detection techniques based classification in Wireless Sensor Networks. In Proceedings of the 13th International Wireless Communications and Mobile Computing Conference (IWCMC), Valencia, Spain, 26-30 June 2017; pp. 687–692, doi: 10.1109/IWCMC.2017.7986368.

101. Zhang, Y.; Meratnia, N.; Havinga, P.J.M. Ad Hoc Networks Distributed online outlier detection in wireless sensor networks using ellipsoidal support vector machine. *Ad Hoc Netw.* **2013**, *11*, 1062–1074, doi:10.1016/j.adhoc.2012.11.001.

102. Zhang, Y.; Meratnia, N.; Havinga, P. An online outlier detection technique for wireless sensor networks using unsupervised quarter-sphere support vector machine. In Proceedings of the 2008 International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), Sydney, NSW, Australia, 15-18 Dec. 2008; pp. 151–156, doi:10.1109/ISSNIP.2008.4761978.

103. Rajasegarar, S.; Leckie, C.; Palaniswami, M.; Bezdek, J.C. Quarter Sphere Based Distributed Anomaly Detection in Wireless Sensor Networks. In Proceedings of the 2007 IEEE International Conference on Communications, Glasgow, UK, 24-28 June 2007; pp. 3864-3869, doi:10.1109/ICC.2007.637.

104. Rajasegarar, S.; Leckie, C.; Bezdek, J.C.; Palaniswami, M. Centered hyperspherical and hyperellipsoidal one-class support vector machines for anomaly detection in sensor networks. *IEEE Trans. Inf. Forensics Secur.* **2010**, *5*, 518–533, doi:10.1109/TIFS.2010.2051543.

105. Janakiram, D.; Reddy, V.; Kumar, a.V.U.P.; V, A.M.R. Outlier Detection in Wireless Sensor Networks using Bayesian Belief Networks. In Proceedings of the 2006 1st International Conference on Communication Systems Software & Middleware, New Delhi, India, 8-12 Jan. 2006; pp. 1–6, doi:10.1109/COMSWA.2006.1665221.

106. Zhang, Y.; Meratnia, N.; Havinga, P. Adaptive and Online One-Class Support Vector Machine-Based Outlier Detection Techniques for Wireless Sensor Networks. In Proceedings of the 2009 International Conference on Advanced Information Networking and Applications Workshops, Bradford, UK, 26-29 May 2009; pp. 990–995, doi:10.1109/WAINA.2009.200.

107. Agarwal, D. Detecting anomalies in cross-classified streams: A Bayesian approach. *Knowl. Inf. Syst.* **2007**, *11*, 29–44, doi:10.1007/s10115-006-0036-4.

108. Agovic, A.; Banerjee, A.; Ganguly, A.; Protopopescu, V. Anomaly Detection in Transportation Corridors Using Manifold Embedding. *ACM Workshop on Knowledge Discovery from Sensor Data: The 13th International Conference on Knowledge Discovery and Data Mining*; San Jose, CA, USA, 2007.

109. Budalakoti, S.; Srivastava, A.; Akella, R.; Turkov, E. *Anomaly Detection in Large Sets of High-Dimensional Symbol Sequences*. Tech. Rep. NASA TM-2006-214553; NASA Ames Research Center: Mountain View, CA, USA, 2006.

110. Hill, D.J.; Minsker, B.S.; Amir, E. Real-time Bayesian anomaly detection for environmental sensor data. *Proc. Congr.-Int. Assoc. Hydraul. Res.* **2006**, *32*, 503.

111. Farruggia, A.; Lo Re, G.; Ortolani, M. Probabilistic Anomaly Detection for Wireless Sensor Networks. In *In Congress of the Italian Association for Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 438–444, doi:10.1007/978-3-642-23954-0_44.

112. Mohamed, M.S.; Kavitha, T. Outlier Detection Using Support Vector Machine in Wireless Sensor Network Real Time Data. *Int. J. Soft Comput. Eng. (IJSCE)* **2011**, *1*, 68–72.

113. Banerjee, a.; Burlina, P.; Diehl, C. A support vector method for anomaly detection in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2282–2291, doi:10.1109/TGRS.2006.873019.

114. Govindarajan, M.; Abinaya, V. An Outlier detection approach with data mining in wireless sensor network. *Int. J. Curr. Eng. Technol.* **2014**, *4*, 929–932.

115. De Stefano, C.; Sansone, C.; Vento, M. To reject or not to reject: That is the question - an answer in case of neural classifiers. *IEEE Trans. Syst. Man Cybern. Part Appl. Rev.* **2000**, doi:10.1109/5326.827457.

116. Barbará, D.; Wu, N.; Jajodia, S. Detecting Novel Network Intrusions Using Bayes Estimators. In Proceedings of the 2001 SIAM International Conference on Data Mining, 2001, doi:10.1137/1.9781611972719.28.

117. Elnahrawy, E.; Nath, B. Context-aware sensors. *Wirel. Sens. Netw. Proc.* **2004**, doi:10.1007/978-3-540-24606-0_6.

118. Hill, D.J.; Minsker, B.S.; Amir, E. Real-Time Bayesian Anomaly Detection for Environmental Sensor Data. Water Resources Research, Volume 45, Issue 4, April 2009, doi: 10.1029/2008WR006956.

119. Ni, K.; Pottie, G. Sensor network data fault detection with maximum a posteriori selection and bayesian modeling. *ACM Trans. Sens. Netw.* **2012**, *8*, 1–21, doi:10.1145/2240092.2240097.

120. Hawkins, S.; He, H.; Williams, G.; Baxter, R. Outlier Detection Using Replicator Neural Networks. *Data Warehous.* **2002**, 170–180, doi:10.1007/978-3-540-74553-2.

121. Yamanishi, K.; Takeuchi, J.I.; Williams, G.; Milne, P. On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms. *Data Min. Knowl. Discov.* **2004**, doi:10.1023/B:DAMI.0000023676.72185.7c.

122. Sykacek, P. Equivalent Error Bars For Neural Network Classifiers Trained By Bayesian Inference. *European Symposium on Artificial Neural Networks*, Bruges, Belgium, 16-18 April 1997; pp. 121-126, doi:10.1109/CVPR.2016.438.

123. Patcha, A.; Park, J.M. An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Comput. Networks* **2007**, *51*, 3448–3470, doi:10.1016/j.comnet.2007.02.001.

124. Tan, P.N. *Introduction to data mining*; Pearson Education India, University of Minnesota, USA, 2006.

125. Jain, A.K.; Dubes, R.C. *Algorithms for clustering data*; Prentice-Hall, Inc.: Michigan State University, USA, 1988.

126. Boulila, W.; Farah, I.R.; Ettabaa, K.S.; Solaiman, B.; Ghézala, H.B. A data mining based approach to predict spatiotemporal changes in satellite images. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 386–395.

127. Han, J.; Pei, J.; Kamber, M. *Data mining: Concepts and techniques*; Elsevier: Amsterdam, The Netherlands, 2011.

128. Safaei, M.; Ismail, A.S.; Chizari, H.; Driss, M.; Boulila, W.; Asadi, S.; Safaei, M. Standalone noise and anomaly detection in wireless sensor networks: A novel time-series and adaptive Bayesian-network-based approach. *J. Softw. Pract. Exp.* 2020, doi:10.1002/spe.2785.

129. He, Z.; Deng, S.; Xu, X.; Huang, J.Z. A fast greedy algorithm for outlier mining. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 567–576.

130. Ando, S. Clustering needles in a haystack: An information theoretic analysis of minority and outlier detection. In Proceedings of the Seventh IEEE International Conference on Data Mining, Omaha, NE, 2007; pp. 13–22.

131. Jolliffe, I.T. Principal component analysis and factor analysis. In *Principal Component Analysis*; Springer: New York, NY, USA, 2002; pp. 150–166.

132. Dunia, R.; Joe Qin, S. Subspace approach to multidimensional fault identification and reconstruction. *AIChE J.* **1998**, *44*, 1813–1831.

133. Jackson, J.E.; Mudholkar, G.S. Control procedures for residuals associated with principal component analysis. *Technometrics* **1979**, *21*, 341–349.

134. Ritika.; Kumar, T.; Kaur, A. Outlier Detection in WSN- A Survey. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **2013**, *3*, 609–617.

135. Bezdek, J.C.; Rajasegarar, S.; Moshtaghi, M.; Leckie, C.; Palaniswami, M.; Havens, T.C. Anomaly detection in environmental monitoring networks [application notes]. *IEEE Comput. Intell. Mag.* **2011**, *6*, 52–58.

136. Fawzy, A.; Mokhtar, H.M.; Hegazy, O. Outliers detection and classification in wireless sensor networks. *Egypt. Inform. J.* **2013**, doi:10.1016/j.eij.2013.06.001.

137. Shuai, M.; Xie, K.; Chen, G.; Ma, X.; Song, G. A kalman filter based approach for outlier detection in sensor networks. In Proceedings of the 2008 International Conference on Computer Science and Software Engineering, Hubei, 2008; Volume 4, pp. 154–157.

138. Lijun, C.; Xiyin, L.; Tiejun, Z.; Zhongping, Z.; Aiyong, L. A data stream outlier delection algorithm based on reverse k nearest neighbors. In Proceedings of the 2010 International Symposium on Computational Intelligence and Design, Hangzhou, 2010; Volume 2, pp. 236–239.

139. Kannan, K.; Manoj, K.; Sakthivel, E. A comparative study on nearest-neighbor based outlier detection in data mining. *A J. Manag. NISMA Noorul Islam Strateg. Manag. Ambience01/2015* **2015**, *1*, 203–204. cited By 1.

140. Birant, D; Kut, A. Spatio-temporal outlier detection in large databases. 28th International Conference on Information Technology Interfaces. Cavtat/Dubrovnik, Croatia, 19-22 June 2006, Doi: 10.1109/ITI.2006.1708474.

141. Peng, X.; Chen, J.; Shen, H. Outlier detection method based on SVM and its application in copper-matte converting. In Proceedings of the 2010 Chinese Control and Decision Conference, Xuzhou, 2010; pp. 628–631.

142. Chandore, P.; Chatur, D. Hybrid approach for outlier detection over wireless sensor network real time data. *Int. J. Comput. Sci. Addit. Appl.* **2013**, *6*.

143. Warriach, E.U.; Nguyen, T.A.; Aiello, M.; Tei, K. Notice of Violation of IEEE Publication Principles A hybrid fault detection approach for context-aware wireless sensor networks. In Proceedings of the 2012 IEEE 9th International Conference on Mobile Ad-Hoc and Sensor Systems (MASS 2012), Las Vegas, NV, 2012; pp. 281–289.

144. Rizwan, R.; Khan, F.A.; Abbas, H.; Chauhdary, S.H. Anomaly detection in wireless sensor networks using immune-based bioinspired mechanism. *Int. J. Distrib. Sens. Networks* **2015**, *11*, 684952.

145. Egilmez, H.E.; Ortega, A. Spectral anomaly detection using graph-based filtering for wireless sensor networks. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, 2014; pp. 1085–1089.

146. Shukla, D.S.; Pandey, A.C.; Kulhari, A. Outlier detection: A survey on techniques of WSNs involving event and error based outliers. In Proceedings of the International Conference on Innovative Applications of Computational Intelligence on Power, Energy and Controls with Their Impact on Humanity, CIPECH 2014, Ghaziabad, 2014; pp. 113–116, doi:10.1109/CIPECH.2014.7019101.

147. Abukhalaf, H.; Wang, J.; Zhang, S. Outlier detection techniques for localization in wireless sensor networks: A survey. *Int. J. Future Gener. Commun. Netw.* **2015**, *8*, 99–114.

148. Farruggia, A. A Probabilistic Approach to Anomaly Detection for Wireless Sensor Networks Abstract. PhD Thesis, Università degli Studi di Palermo PA, Italy, 2011.

149. Hodge, V.J.; Austin, J. A Survey of Outlier Detection Methodoligies. *Artif. Intell. Rev.* **2004**, *22*, 85–126, doi:10.1007/s10462-004-4304-y.

150. Al-Sarem, M.; Boulila, W.; Al-Harby, M.; Qadir, J.; Alsaeedi, A. Deep Learning-Based Rumor Detection on Microblogging Platforms: A Systematic Review. *IEEE Access* **2019**, *7*, 152788–152812.

151. Garcia-Font, V.; Garrigues, C.; Rifà-Pous, H. Difficulties and challenges of anomaly detection in smart cities: A laboratory analysis. *Sensors (Switzerland)* **2018**, *18*, 1–21, doi:10.3390/s18103198.

152. Gil, P.; Martins, H.; Januário, F. Detection and accommodation of outliers in Wireless Sensor Networks within a multi-agent framework. *Appl. Soft Comput. J.* **2016**, *42*, 204–214, doi:10.1016/j.asoc.2015.12.042.

153. Pimentel, M.A.F.; Clifton, D.A.; Clifton, L.; Tarassenko, L. Review: A Review of Novelty Detection. *Signal Process.* **2014**, *99*, 215–249, doi:10.1016/j.sigpro.2013.12.026.

154. Sharma, A.B.; Golubchik, L.; Govindan, R. Sensor faults. *ACM Trans. Sens. Networks* **2010**, *6*, 1–39, doi:10.1145/1754414.1754419.

155. Kamal, S.; Ramadan, R.; El-Refai, F. Smart outlier detection of wireless sensor network. *Facta Univ. - Ser. Electron. Energetics* **2016**, *29*, 383–393, doi:10.2298/FUEE1603383K.

156. Bal, M.; Shen, W.; Ghenniwa, H. Collaborative signal and information processing in wireless sensor networks: A review. In Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics, San Antonio, TX, 2009; pp. 3151–3156, doi:10.1109/ICSMC.2009.5346152.

157. Can, A.; Guillaume, G.; Picaut, J. Cross-calibration of participatory sensor networks for environmental noise mapping. *Appl. Acoust.* **2016**, *110*, 99–109, doi:10.1016/j.apacoust.2016.03.013.

158. Chen, J.; Kher, S.; Somani, A. Distributed Fault Detection of Wireless Sensor Networks. In Proceedings of the 2006 Workshop on Dependability Issues in Wireless Ad Hoc Networks and Sensor Networks, 2006; pp. 65–72, doi:10.1145/1160972.1160985.

159. Chen, Q.; Lam, K.Y.; Fan, P. Comments on "Distributed Bayesian algorithms for fault-tolerant event region detection in wireless sensor networks". *IEEE Trans. Comput.* **2005**, *54*, 1182–1183, doi:10.1109/TC.2005.140.

160. Dauwe, S.; Oldoni, D.; De Baets, B.; Van Renterghem, T.; Botteldooren, D.; Dhoedt, B. Multi-criteria anomaly detection in urban noise sensor networks. *Environ. Science. Process. & Impacts* **2014**, *16*, 1–10, doi:10.1039/c4em00273c.

161. Duh, D.r.; Li, S.p.; Cheng, V.W. Distributed Fault-Tolerant Event Region Detection of Wireless Sensor Networks *J. Distrib. Sens. Netw.* **2013**. *2013*, 160523.

162. Dziengel, N.; Seiffert, M.; Ziegert, M.; Adler, S.; Pfeiffer, S.; Schiller, J. Deployment and evaluation of a fully applicable distributed event detection system in Wireless Sensor Networks. *Ad Hoc Netw.* **2016**, *37*, 160–182, doi:10.1016/j.adhoc.2015.08.017.

163. Guo, J.; Liu, F. Automatic data quality control of observations in wireless sensor network. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 716–720, doi:10.1109/LGRS.2014.2359685.

164. Hoes, R.; Basten, T.; Tham, C.K.; Geilen, M.; Corporaal, H. Quality-of-service trade-off analysis for wireless sensor networks. *Perform. Eval.* **2009**, *66*, 191–208, doi:10.1016/j.peva.2008.10.007.

165. Jurdak, R.; Wang, X.R.; Obst, O.; Valencia, P. Wireless Sensor Network Anomalies: Diagnosis and Detection Strategies. *Intell. Syst. Ref. Libr.* **2011**, *10*, 309–325, doi:10.1007/978-3-642-17931-0_12.

166. McDonald, D.; Sanchez, S.; Madria, S.; Ercal, F. A Survey of Methods for Finding Outliers in Wireless Sensor Networks. *J. Netw. Syst. Manag.* **2013**, *23*, 163–182.

167. Portocarrero, J.M.T.; Delicato, F.C.; Pires, P.F.; Gámez, N.; Fuentes, L.; Ludovino, D.; Ferreira, P. Autonomic Wireless Sensor Networks: A Systematic Literature Review. *J. Sensors* **2014**, *2014*, doi:10.1155/2014/782789.

168. O'Reilly, C., Gluhak, A., I.; Rajasegarar, S. Anomaly Detection in Wireless Sensor Networks in a Non-Stationary Environment Colin. *IEEE Commun. Surv. Tutorials* **2014**, *16*, 1413–1432, doi:10.1109/RSTSCC.2010.5712878.

169. Wang, X.; Lizier, J.; Obst, O. Spatiotemporal anomaly detection in gas monitoring sensor networks. In Proceedings of the 5th European conference on Wireless sensor networks, Bologna, Italy, 30 January - 1 February 2008; pp. 90–105, doi:10.1007/978-3-540-77690-1.

170. Xie, M.; Hu, J.; Guo, S.; Zomaya, A.Y. Distributed Segment-Based Anomaly Detection with Kullback–Leibler Divergence in Wireless Sensor Networks. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 101–110, doi:10.1109/TIFS.2016.2603961.

171. Mamun, Q.; Islam, R.; Kaosar, M. Anomaly Detection in Wireless Sensor Network. *J. Networks* **2014**, *9*, 2914–2924, doi:10.4304/jnw.9.11.2914-2924.

172. Rajasegarar, S.; Leckie, C.; Palaniswami, M. CESVM: Centered hyperellipsoidal support vector machine based anomaly detection. *IEEE Int. Conf. Commun.* **2008**, 1610–1614, doi:10.1109/ICC.2008.311.

173. a.S. Sangari. Anomaly detection in wireless sensor networks. *Recent Adv. Space Technol. Serv. Clim. Chang.* **2010**, *16*, 1413–1432, doi:10.1109/RSTSCC.2010.5712878.

174. Bhuse, V.; Gupta, A. Anomaly Intrusion Detection in Wireless Sensor Networks. *J. High Speed Netw.* **2006**, *Volume 15*, 33–51, doi:10.1007/978-3-540-77871-4_14.

175. Siripanadorn, S.; Siripanadorn, S.; Hattagam, W.; Teaumroong, N. Anomaly detection using self-organizing map and wavelets in wireless sensor networks. In Proceedings of the 10th WSEAS international conference on Applied computer science, Stevens Point, Wisconsin, United States, 2010; pp. 291–297.

176. Orair, G.H.; Teixeira, C.H.; Meira, W.J.; Wang, Y.; Parthasarathy, S. Distance-based outlier detection: Consolidation and renewed bearing. *Vldb* **2010**, *3*, 1469–1480, doi:10.14778/1920841.1921021.

177. Rajasegarar, S.; Leckie, C.; Palaniswami, M. Anomaly detection in wireless sensor networks. *Wirel. Commun. IEEE* **2008**, *15*, 34–40.

178. Abe, N.; Zadrozny, B.; Langford, J. Outlier detection by active learning. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2006; pp. 504–509, doi:10.1145/1150402.1150459.

179. He, Z.; Deng, S.; Xu, X. An Optimization Model for Outlier Detection in Categorical Data. In *International Conference on Intelligent Computing*; Springer: Berlin/ Heidelberg, Germany, 2005; pp. 400–409, doi:10.1007/11538059_42.

180. Rassam, M.A.; Zainal, A.; Maarof, M.A. Advancements of data anomaly detection research in Wireless Sensor Networks: A survey and open issues. *Sensors (Switzerland)* **2013**, *13*, 10087–10122, doi:10.3390/s130810087.

181. Li, Q.; Sun, R.; Wu, H.; Zhang, Q. Parallel distributed computing based wireless sensor network anomaly data detection in IoT framework. *Cogn. Syst. Res.* **2018**, *52*, 342–350, doi:10.1016/j.cogsys.2018.07.007.

182. Ramotsoela, D.; Abu-Mahfouz, A.; Hancke, G. A survey of anomaly detection in industrial wireless sensor networks with critical water system infrastructure as a case study. *Sensors (Switzerland)* **2018**, *18*, 1–24, doi:10.3390/s18082491.

183. Granjal, J.; Silva, J.M.; Lourenço, N. Intrusion detection and prevention in CoAP wireless sensor networks using anomaly detection. *Sensors (Switzerland)* **2018**, *18*, doi:10.3390/s18082445.

184. Li, G.; He, B.; Huang, H.; Tang, L. Temporal data-driven sleep scheduling and spatial data-driven anomaly detection for clustered wireless sensor networks. *Sensors (Switzerland)* **2016**, *16*, doi:10.3390/s16101601.

185. Haque, S.A.; Rahman, M.; Aziz, S.M. Sensor anomaly detection in wireless sensor networks for healthcare. *Sensors (Switzerland)* **2015**, *15*, 8764–8786, doi:10.3390/s150408764.

186. Trinh, V.V.; Tran, K.P.; Huong, T.T. Data driven hyperparameter optimization of one-class support vector machines for anomaly detection in wireless sensor networks. In Proceedings of the International Conference on Advanced Technologies for Communications, Quy Nhon, Vietnam, 18-20 Oct. 2017; pp. 6–10, doi:10.1109/ATC.2017.8167642.

187. Yahyaoui, A.; Abdellatif, T.; Attia, R. READ: Reliable Event and Anomaly Detection System in Wireless Sensor Networks. In Proceedings of the 2018 IEEE 27th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2018; pp. 193–198, doi:10.1109/WETICE.2018.00044.

188. Trinh, V.V.; Tran, K.P.; Mai, A.T. Anomaly Detection in Wireless Sensor Networks via Support Vector Data Description with Mahalanobis Kernels and Discriminative Adjustment. In Proceedings of the 4th NAFOSTED Conference on Information and Computer Science Anomaly, Hanoi, Vietnam, 24-25 Nov. 2017; pp. 567–586, doi:10.1007/978-1-4615-3792-2_30.

189. Kanev, A.; Nasteka, A.; Bessonova, C.; Nevmerzhitsky, D.; Silaev, A.; Efremov, A.; Nikiforova, K. Anomaly Detection in Wireless Sensor Network of the "Smart Home" System. In Proceedings of the 2017 20th Conference of Open Innovations Association (FRUCT), St. Petersburg, Russia, 3–7 April 2017; pp. 118–124.

190. Kumar Dwivedi, R.; Pandey, S.; Kumar, R. A Study on Machine Learning Approaches for Outlier Detection in Wireless Sensor Network. In Proceedings of the 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 11-12 Jan. 2018; pp. 189–192, doi:10.1109/CONFLUENCE.2018.8442992.

191. Ahmad, B.; Jian, W.; Ali, Z.A.; Tanvir, S.; Khan, M.S.A. Hybrid Anomaly Detection by Using Clustering for Wireless Sensor Network. *Wirel. Pers. Commun.* **2018**, pp. 1–13, doi:10.1007/s11277-018-5721-6.

192. Nauman Shahid, Ijaz Haider Naqvi, and Saad B Qaisar. Real time energy efficient approach to outlier & event detection in wireless sensor networks. In *2012 IEEE International Conference on Communication Systems (ICCS)*, pages 162–166. IEEE, 2012.

193. Weizheng Ren and Yansong Cui. A parallel rough set tracking algorithm for wireless sensor networks. *Journal of Networks*, 7(6):972–979, 2012.