Computer
& Robotics

A Novel Ensemble Approach for Anomaly Detection in Wireless Sensor Networks Using Time-overlapped Sliding Windows

Zahra Malmir, Mohammad Hossein Rezvani *

Faculty of Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran

Received 03 November 2018; Revised 25 April 2019; Accepted 12 May 2019; Available online 27 June 2019

Abstract

One of the most important issues concerning the sensor data in the Wireless Sensor Networks (WSNs) is the unexpected data which are acquired from the sensors. Today, there are numerous approaches for detecting anomalies in the WSNs, most of which are based on machine learning methods. In this research, we present a heuristic method based on the concept of "ensemble of classifiers" of data mining. Our proposed algorithm, at first applies a fuzzy clustering approach using the well-known C-means clustering method to create the clusters. In the classification step, we created some base classifiers, each of which utilizes the data of overlapping windows to utilize the correlation among data over time by creating time-overlapped batches of data. By aggregating these batches, the classifier proceeds to find an appropriate label for future incoming instance. The concept of "Ensemble of Classifiers" with majority voting scheme has been used in order to combine the judgment of all classifiers. The results of our implementation with MATLAB toolboxes shows that the proposed majority-based ensemble learning method attains more efficiency compared to the case of the single classifier method. Our proposed method enhances the performance of the system in terms of major criteria such as False Positive Rate, True Positive Rate, False Negative Rate, True Negative Rate, Sensitivity, Specificity and also the ROC curve.

Keywords: *Wireless Sensor Networks, Anomaly detection, Data Mining, Ensemble of Learners, Performance Evaluation.*

1. Introduction

Wireless sensor networks (WSNs) are small, low-cost, low-energy and multi-role sensor systems that are used for monitoring, tracking or controlling the processes. The limitations of WSN nodes like memory, processing, consumed power and bandwidth have been a driving force to change the traditional computation methods in the area of WSN applications. Traditionally, in WSNs, the data are sent from different sources, called sensor nodes, to a central processing entity, called sink. In a dense WSN, many volumes of data are produced, each of which has very high repetition and frequency. Furthermore, local processing methods ignore the correlations and the dependencies among the streaming data [1, 2]. These could result in a waste of the energy and the bandwidth so that cause to have a shorter lifetime of the network.

Data mining is a program-oriented process with powerful mathematical tools for analyzing large volumes of data streams. Recently, data mining approaches have been used broadly in order to detect the anomalies in WSNs [1, 3].

* Corresponding author. Email: rezvani@qiau.ac.ir

Simply, the anomaly in WSN is defined as "an important data analysis task that detects anomalous or anomalous data from a given dataset." [4]. In the following sections, we will explain the anomaly in details. Since the sensor data can be destroyed for many reasons such as reading errors, malfunctioning of the sensors or destructive attacks, one of the most important motivations for anomaly detection in WSNs is to provide trustable and high-quality data. Other motivations for anomaly detection include frequent usage in applications such as discrete event monitoring, climate change monitoring and fire detection [5, 6]. Using anomaly detection in WSNs also helps in predicting upcoming events in the area of disaster management and smart cities equipped with sensors of the Internet of Things (IoT).

The main challenges concerning anomaly detection methods in WSNs are preventing resource limitation [2], Distributed data streaming, reducing false alarm rates, and so on.

According to the above explanations, an efficient anomaly detection technique for WSNs should be able to detect anomalies in an online and distributed form with high detection precision and low false alarm rates. In this way, the WSN limitations could be alleviated in terms of communication overhead, computation complexity, and memory consumption [8]. In general, the significant limitations of current anomaly detection models are as following:

Although recent anomaly detection models are designed for online streams, the computational cost of these approaches is still a challenging issue. Also, most of current anomaly detection models ignore space and time correlations among the features of WSNs' data. The features' correlations are essential for the efficiency of detection. Moreover, considering correlations is a critical factor in order to decrease the consumed power of the sensor.

Many approaches have been presented for anomaly detection, among them, the online ensemble methods are the most efficient ones [9, 10, 11, 12]. Ensemble process aims at combining all assumptions concerning all anomaly classes for creating a new combinatorial anomaly classification. Usually, ensemble learners are built in two steps. At first, several base learners are created and then they are combined. We will discuss the specifications of ensemble methods in depth later in Section 2.

Based on the advantages mentioned in the literature, we have proposed a window-based approach for anomaly detection. Our proposed algorithm, at first applies a fuzzy clustering approach using the well-known C-means clustering method to create the clusters. In this way, after determining the center of the clusters, the read-out instances are fed into the system and then the label of each data record is determined as "normal" or "anomalous". We have used decision tree classification method for the classification step to create the base learners. Then, for the ensemble of multiple classifiers, we have used a majority voting approach. We have used the concept of overlapping windows for ensemble step in which for each data interval, we create some overlapping windows. For each window, we have used a base decision tree learner.

The rest of this paper is as follows: In Section 2 the existing anomaly detection approaches and their application in WSNs will be addressed. In Section 3, we will present our proposed approach in order to detect anomalies in WSNs. In this section, we present the details of C-means fuzzy clustering, the process of creating base classifiers with decision trees, and also an ensemble of learners with a majority voting approach. In Section 4, we will demonstrate the evaluation results using MATLAB toolboxes. Finally, we will conclude the paper in Section 5 and will address future research trends.

2. Related Works

In this section, we will proceed to explain the general properties of anomaly detection methods in the literature. Fig. 1 shows the important types of anomaly detection methods in the area of WSNs. Interested readers can refer to [13] for further readings.

Clustering means to divide the read-out instances to some clusters in such a way that the instances located in each cluster have major similarities to each other [14]. Cluster-based methods are explained in [4, 15, 16]. These techniques also called semi-supervision techniques. The core idea of clustering approaches in anomaly detection is that if a sample doesn't belong to any of the defined clusters, it will be identified as anomalous [17].

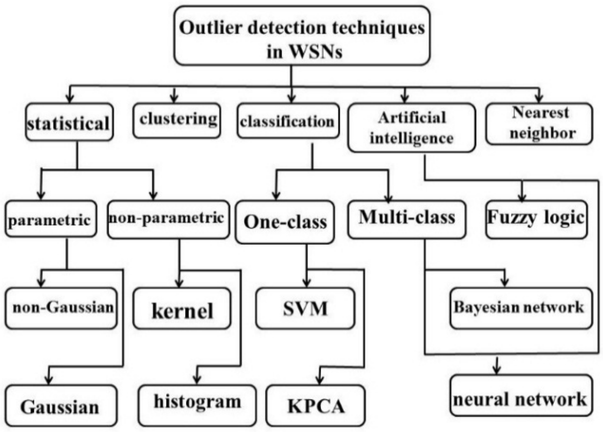

Fig. 1. Major anomaly detection methods in the area of WSNs [13].

Another major category of data mining schemes is classification [18]. One of the important researches carried out in the area of anomaly detection in WSNs has been presented in [19]. This research has used Support Vector Machine (SVM) scheme to classify the read-out data of sensors. Although SVM has rather low computation cost and memory usage, it's processing time increases dramatically with the number and volume of the read-out data. The authors of [20] presented a framework for collective conceptual anomaly detection (CCAD). This framework uses sliding window and sensor data history with conceptual features for detecting anomalous conceptual patterns in sensor data. The authors of [21] created a theoretical framework for sequential learning on cortex based on cortical and hierarchy temporary memory (HTM). Interested readers can refer to [10, 22] to see previous research in this area.

The nearest neighbor methods use machine learning and data mining for analyzing data instances according to the proximity of the neighbors. This procedure is used for different aims such as classification, clustering and anomaly detection. In these types of schemes, any instance data which is located outside the vicinity of its neighbor is called anomalous point [23]. In paper [24], unsupervised data mining is used with considering neighborhood and correlation information. The authors of [24] have sued anomaly detection method to decrease the consumed energy in WSNs.

Recently metaheuristic approaches have been used extensively to detect anomalies in the area of WSNs. The major advantage of such algorithms lies in its ability to reduce the time complexity of the data stream mining system. The authors of [25] have proposed a framework inspired by the genetic algorithm that greatly reduces the time complexity, even for millions of records. In paper [26] the ensemble clustering is modeled as multi-objective optimization problem and then a multi-objective genetic evolutionary algorithm has been presented to address the problem. Also, interested readers can find a valuable survey on nature-inspired metaheuristic algorithms in [27].

In general, according to the above explanations, it can be inferred that statistic methods need exact distribution model and their parametric forms are not suitable for WSNs while their non-parametric forms are not suitable for real-time programs due to high computation costs. On the other hand, the nearest neighbor methods suffer from scalability, proper threshold values, and high computation cost. Clustering schemes suffer from issues such as reference model updating, communication overhead, and high computation complexity. Specially, multi-variable clustering methods have shortcomings concerning computation complexity as well as inefficiency due to frequent changes in read-out streams. Artificial intelligence methods need high memory for processing basic rules. Finally, classification methods need self-learning while facing new read-out instances as well as shortcoming concerning high computation complexity recently, anomaly detection with ensemble approaches has raised much attention of the researchers in the area of WSNs. Ensemble methods aim at combining all hypotheses of all anomaly classes in order to create a recombinant anomaly classification. Clearly, there exist numerous approaches in order to combine the base classifiers. The majority voting is the most popular ensemble technique, which is used in the literature ([10] [11] [12]).

Interested readers can refer to [28, 29] to see comparative studies concerning the anomaly detection techniques in the area of smart city WSNs.

3. The Proposed Method

In this section, we describe our proposed method based on the sliding window in conjunction with the concept of ensemble of classifiers. The incentive behind using the concept of sliding windows is that a small change in read-out data (such as voltage fluctuations) may cause dramatic changes in system behavior. Moreover, we have made some minor modifications in the way that sliding window works to make our scheme agile. As stated before, According to previous researches in the area of multi-variable sensors, it is proven that using classification methods combined with ensemble methods such as majority voting could provide more efficiency in anomaly detection process in terms of criteria such as precision rate, speed, and so on.

The contribution of our research lies in the approach by which we have used the sliding window concept. Using the novel idea of overlapping sliding windows, we have designed ensemble learners which can robustly react against sudden changes in sensor read-out data. Such an agile scheme lends itself to better management of the alarming system. Roughly speaking, by this heuristic scheme, small changes in read-out data (e.g. voltage fluctuations) could not result in remarkable changes in the behavior of the WSN system at all. Thus, our approach takes into account the correlations between consecutive intervals of read-out data. Also, with regard to the existence of multivariable read-out data (moisture, heat, light and voltage parameters) we have used the well-known decision tree classifier due to its high speed in classification. In order to reach the aim of speed and precision simultaneously, we have used majority voting ensemble approach. To the best of our knowledge, this research is the first attempt to explore the overlapping sliding windows to detect anomalies in WSNs in conjunction with the concept of ensemble of classifiers.

We have used a well-known dataset which is used in the previous researches in [29, 30, 31]. Our proposed method includes two steps: our algorithm in the first step attempts to tag the data as “normal” or “anomalous” classes. To this end, we have used fuzzy clustering scheme using the C-means method. In this step, the samples according to their feature values (voltage, light, moisture, and temperature) are placed in one of the two primary clusters and are tagged as "normal" or " anomalous"). The goal of clustering scheme is to create two clusters in such a way that for each cluster the inter-cluster similarity is more than intra-cluster similarities. Then, it uses the concept of sliding window to create overlapping batches of data. To this end, it uses a sliding window of length and slides it by size to form a new independent base classifier. In fact, each window represents an independent classifier. So, by each movement of the window, we have a new classifier. By aggregating these batches, the classifier proceeds to find an appropriate label for future incoming instances. We have used the concept of “ensemble of classifiers” with majority voting scheme in order to combine the judgment of all classifiers.

As it is well-known in machine learning literature, the fuzzy clustering approach convergence possibility is very high. Furthermore, using fuzzy clustering approach let each record of read-out data belong to both clusters. This allows the designer to flexibly decide about the membership degree of each instance to each of the two clusters. Therefore, in this algorithm, error correction of instances is much easier than K-means algorithm. In [32], a thorough comparison between C-means fuzzy clustering and older algorithms such as K-means is done. At the end of the clustering step, every instance will have a class label as “normal” or “anomalous”. Now, the classification step starts. Our main contribution lies in this step, where we have used “decision tree” as base classifiers and “majority voting” as the ensemble of these classifiers. Due to space limitation, we omit the details of decision tree classification. Interested readers can refer to data mining textbooks (such as [18]) to see the details of the decision tree approaches. Since the sensors readout samples belong to different moments and also since the time intervals between each consecutive point may not equal with other points we decided to create each classifier based on the data gathered in T recent seconds. In other words, we have classified the points in time basis rather than number basis! As shown in Fig. 2 we denote the last registered time for read-out data as. Then the interval should be divided into some time windows. The length of each time window is denoted by. In order to capture the correlation between data points and we use the concept of sliding window. In each step, the window is sliding points. So, as it is shown in Fig.2, every window will have overlap with size m with its previous and next windows.

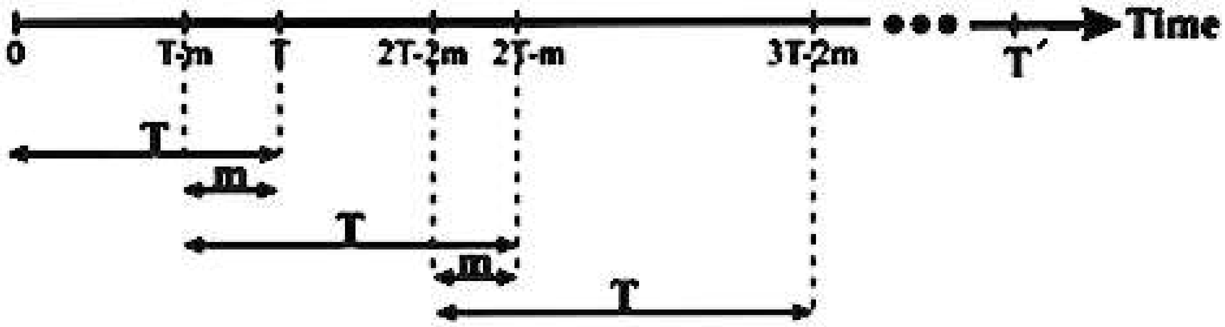

Fig. 2. Sliding window scheme in the proposed method: Each interval T contains tagged read-out samples of WSN sensors. Each window has m-size overlaps with its previous and next windows.

As it is shown in Fig.2, the first window T is located in interval $[0, T]$, the second window T is located in interval $[T-m, 2T-m]$, the third window T is located in interval $[2T-2m, 3T-2m]$ and so on. Let's K denotes the total number of windows. With similar reasoning, the last window lies in interval $[(K-1)T-(K-1)m, KT-(K-1)m]$. Notice that the point $KT-(K-1)m$ is located on the point T'. As it can be seen, Each window has m-size overlaps with its previous and next windows. Let's calculate the parameter m based on parameters K, T, and T'. According to the illustration in Fig. 2 and based on the above explanations we can write:

$$\mathrm{T}+(\mathrm{T}-\mathrm{m})+(\mathrm{T}-\mathrm{m})+\ \dots+\mathrm{T}=\mathrm{T}' \tag{1}$$

By simplifying the above equation we have:

$$\mathrm{KT}-(\mathrm{K}-1)\ \mathrm{m}=\mathrm{T}' \tag{2}$$

Finally, the size of time overlap between windows can be calculated as:

$$\mathrm{m}=\frac{\mathrm{KT}-\mathrm{T}'}{\mathrm{K}-1} \tag{3}$$

As it will be mentioned in the next section, we have ten base classifiers ($K=10$). Also, we have used data points from 24 hours of WSN sensors ($\mathrm{T}'=24_{\mathrm{hours}}$). For choosing a proper value for T, the last equation should be considered. It is clear that the necessary condition for the m to be non-negative is that $\mathrm{KT}-\mathrm{T}'>0$. In other words, we must have $\mathrm{T}>\frac{\mathrm{T}'}{\mathrm{K}}$. Since we have chosen $K=10$, it must be that $\mathrm{T}>\frac{24_{\mathrm{hours}}}{10}=2.4_{\mathrm{hours}}$. For simulations, we have used $\mathrm{T}=3_{\mathrm{hours}}$. It means that we have considered the time interval of each window equal to 3 hours. Now, using Eq. (3), the size of time overlap between windows is calculated as following:

$$\mathrm{m}=\frac{10\times 3-24}{10-1}{}_{\mathrm{hour}}=40_{\mathrm{min}} \tag{4}$$

As stated above, each classifier uses the data points located in its associated window of size T. Since the total number of windows is equal to K, there are K base classifiers, each of which uses the decision tree to classify the associated read-out data. Fig. 3 shows the structure of the base classifiers in detail. As it is shown in Fig.3, the first classifier is created based on data points located in the interval $[0, T]$. The second classifier is created based on data points located in the interval $[T-m, 2T-m]$, and so on. For every read-out data sample, the output of each classifier is a judgment which shows that this point is identified as "normal" or "anomalous".

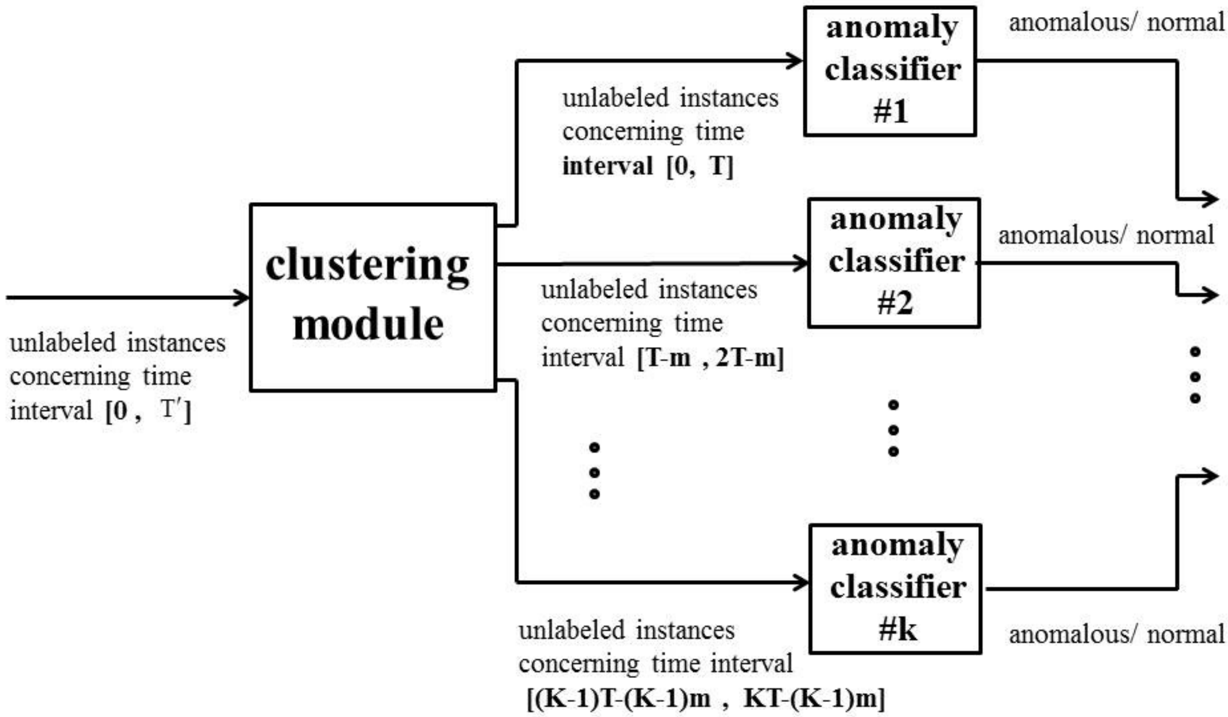

Fig. 3. Creation of K classifiers in the proposed method: The interval $[0, T']$ is divided into K sliding windows, each of which with length *T*. In each step, the window *T* is sliding *m* points. So, every window will have overlap with size *m* with its previous and next windows.

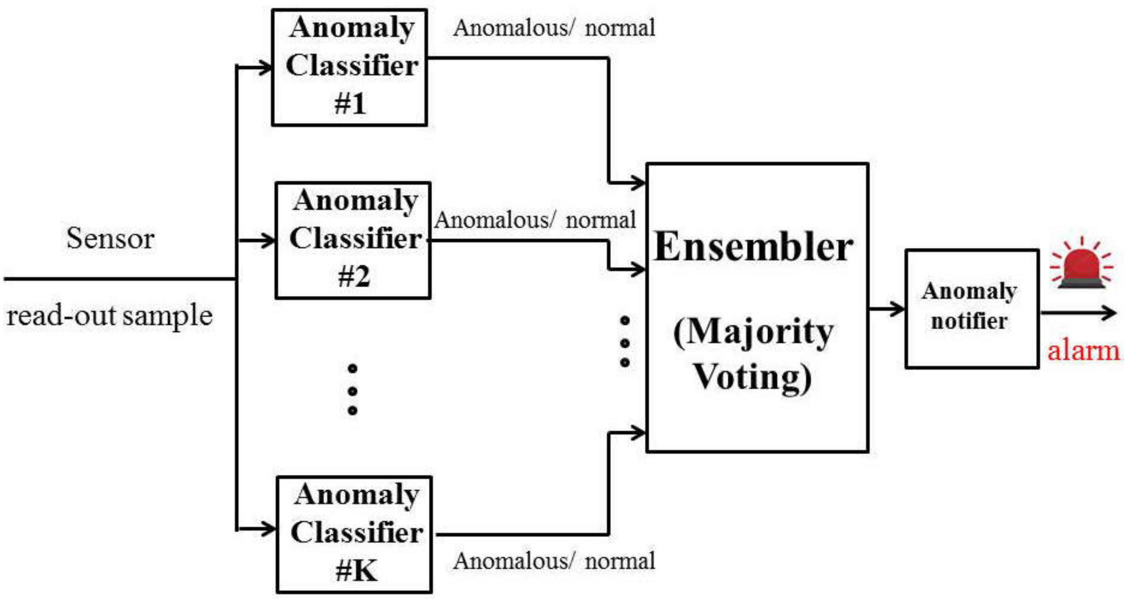

Fig. 4. The process of diagnosis and notification of anomalies in the proposed system based on majority voting.

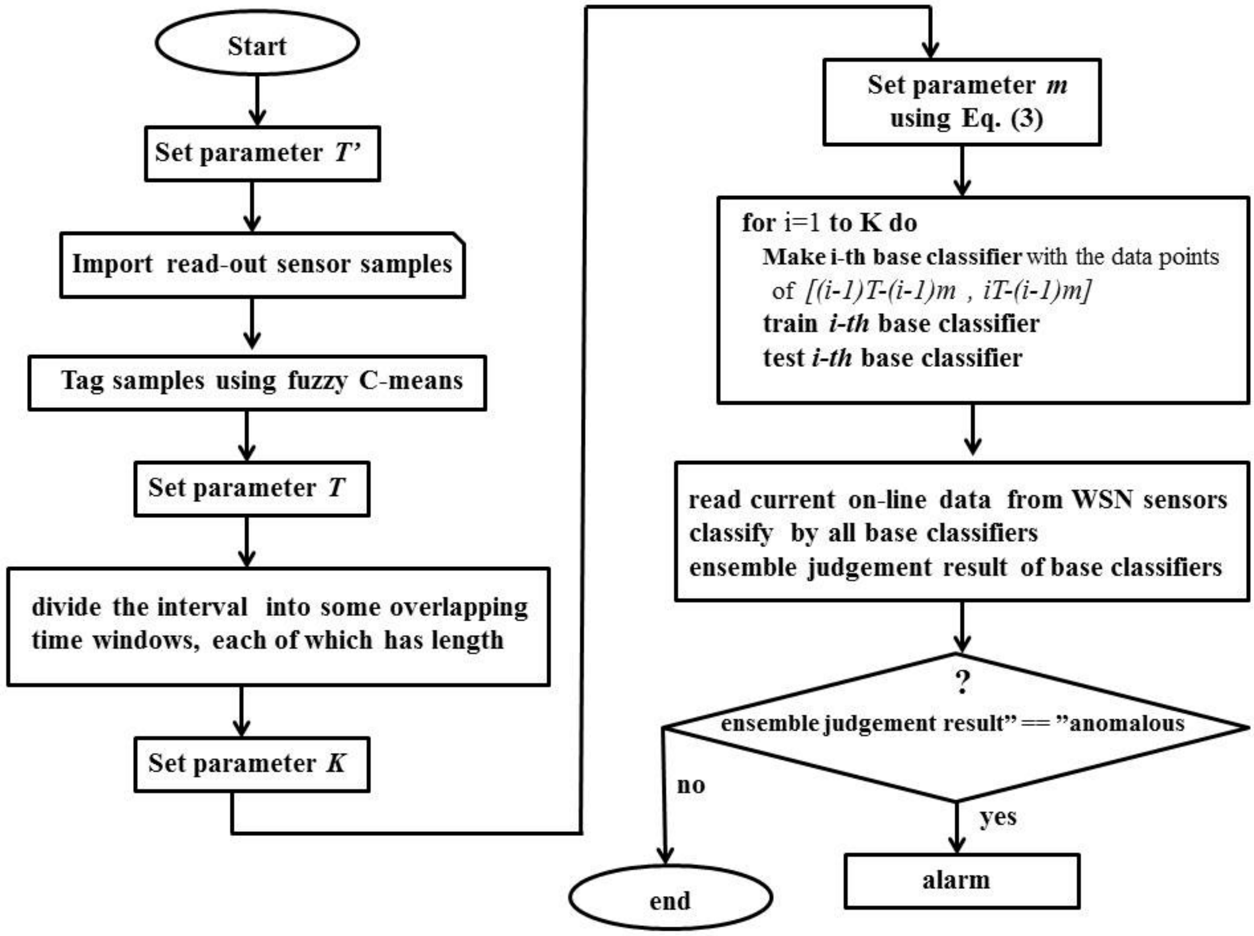

Fig. 5. Flowchart of the proposed approach.

After developing K classifier and training and testing them, finally by utilizing majority voting, the final judgment of the anomaly detection system for new read-out data is issued. Based on this output, if the read-out data is diagnosed as anomalous, the alarm system is activated. Fig. 4 shows the process of diagnosis and notification of anomalies in the proposed system. In Fig. 4, first of all, new read-out data from WSN sensors comes in. Then, the judgments of all classifiers are submitted to ensemble node and after majority voting, the final judgment of the system is delivered to the alarm system.

The flowchart of the proposed anomaly detection system is shown in Fig. 5.

Now we proceed to explain the criteria which we will use in order to evaluate the proposed anomaly detection system. As is the case in data mining at first we ought to define four base criteria namely, *true positive (TP), false negative (FN), true negative (TN),* and *false positive (FP).* Interested readers can refer to data mining textbooks such as [18] to see a detailed explanation of these metrics.

- *True Positive (TP)*: The number of "normal" read-out samples which are correctly identified as "normal" data by the proposed system.
- *False Negative (FN)*: The number of "normal" read-out samples which are misidentified as "anomalous" data by the proposed system.
- *True Negative (TN)*: The number of "anomalous" read-out samples which are correctly identified as "anomalous" data by the proposed system.
- *False Positive (FP)*: The number of "anomalous" read-out samples which are misidentified as "normal" data by the proposed system.

In order to evaluate the proposed method, we must use other operational metric which is defined based on the above definitions [18]. The metric "accuracy" in Eq. (5) is defined as the ratio of correct detections to the total number of data. Here, the term *"TP+TN"* represents the total number of correct detections, whether "normal" or "anomalous" [18]:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (5)$$

The metric *"false positive ratio"* or in abbreviate *"FPR"* in Eq. (6), is defined as the ratio of anomalies which are misidentified as normal to the total number of anomalous data. Here, the term *"TN+FP"* represents the total number of anomalous data, whether "correctly identified anomalies", i.e. *TN*, or "the anomalies which are misidentified as normal", i.e., *FP* [18]:

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad (6)$$

The metric *"false negative ratio"* or in abbreviate *"FNR"* in Eq. (7), is defined as the ratio of normal data which are misidentified as anomalies to the total number of normal data. Here, the term *"TP+FN"* represents the total number of normal data, whether "correctly identified normal data", i.e. *TP*, or "the normal data which are misidentified as anomalies", i.e., *FN* [18]:

$$\text{FNR} = \frac{\text{FN}}{\text{TP} + \text{FN}} \quad (7)$$

The metric *"true positive ratio"* or in abbreviate *"TPR"* in Eq. (8), is defined as the ratio of correctly identified normal data to the total number of normal data. Here, the term *"TP+FN"* represents the total number of normal data, whether "correctly identified normal data", i.e. *TP*, or "the normal data which are misidentified as anomalies", i.e., *FN*. In data mining terminology, the terms *"TPR","recall",* and *"sensitivity"* are often used interchangeably [18]:

$$\text{Sensitivity} = \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (8)$$

The metric *"true negative ratio"* or in abbreviate *"TNR"* in Eq. (9), is defined as the ratio of correctly identified anomalies to the total number of anomalies. Here, the term *"TN+FP"* represents the total number of anomalous data, whether "correctly identified anomalies", i.e. *TN*, or "the anomalies which are misidentified as normal data", i.e., *FP*. In data mining terminology, the terms *"TNR"* and *"specificity"* are often used interchangeably [18]:

$$\text{Specificity} = \text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (9)$$

Clearly, a good anomaly detection system must attain high accuracy, low FPR, low FNR, high sensitivity, and high specificity values.

In the researches carried out in the scope of machine learning and data mining, a receiver operating characteristic curve, i.e. ROC curve, is a graphical plot that illustrates the diagnostic ability of a binary classifier system. As it is shown in Fig. 6, the ROC curve plots TPR against FPR. This curve is a good graphical tool in order to analyze the performance of a typical binary anomaly detection system. The best possible anomaly detection method would yield a point in the upper left corner or coordinate (0,1) of the ROC space, representing 100% sensitivity (no FNs) and 100% specificity (no FPs). The shortest distance d from a coordinate on the ROC to coordinate (0,1) as shown in Fig. 6 can be computed as following [10]:

$$d^2 = (1 - \text{sensitivity})^2 + (1 - \text{specificity})^2 \quad (10)$$

Parameter d specifies the optimum threshold value concerning both the sensitivity and specificity values of the system. Interested readers can refer to [10, 18] to see an in-depth discussion of this topic.

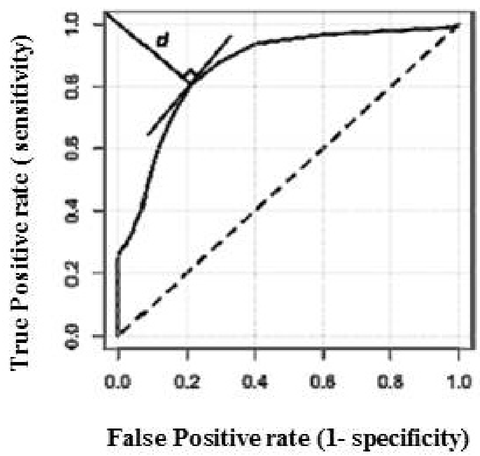

Fig. 6. The "ROC" curve and the optimum threshold value concerning the sensitivity and specificity [10].

4. Evaluation

In order to evaluate the proposed anomaly detection system, we have used MATLAB R2014 software tool installed on a personal computer with a CPU Core i5, 4GByte RAM and Windows 7 Operating system. We have used the evaluation dataset containing 20000 read-out data samples, which is collected by environmental sensors previously by Intel Lab in [25]. This dataset was used in previous researches such as [29, 30, 31]. As stated before in Section 4 (see Fig. 5), we have used the logged data points relevant to 24 hours of working WSN sensors ($T' = 24_{hours}$). These data contain 500 points in the dataset. After performing the clustering with fuzzy C-means in the first step, the data samples got labels as *"normal"* or *"anomalous."*

In the classification step, we chose ten base classifiers ($K = 10$) each of which created using the data points located in a window with time interval length equal to three hours ($\mathrm{T} = 3_{\mathrm{hours}}$). Regard the computations which discussed in detail in Eqs. (3)-(4), the size of time overlap between every two consecutive windows indicated as 40 minutes ($\mathrm{m} = 40_{\mathrm{min}}$). Then, using step 8 of the flowchart in Fig.5, we created ten base classifiers, each of which utilizing a decision tree classification scheme. After training and testing the classifiers, we entered the step (9) of the algorithm where we fed the base classifiers 10% of data points of WSN sensors as current on-line data. After combining the judgment outputs of the base classifier using a majority voting approach, if the result equals "anomalous" the alarm system will be activated.

Now, we proceed to analyze and discuss the evaluation results of the proposed system using the metrics which explained before in Eqs. (5)-(9). In order to evaluate the proposed system, we compared it with the case in which we have employed a single classifier using decision tree classification approach. Fig. 7 shows the *"accuracy"* of the proposed system against the single-classifier case. As it is shown in the figure, our proposed method attains 80% accuracy, which is almost 32% better than the case in which a single classifier is used. Recall from Eq. (5) that *"accuracy"* reflects the rate of correct detections. Thus, Fig. 7 confirms the success of the proposed system in detecting the anomalies efficiently.

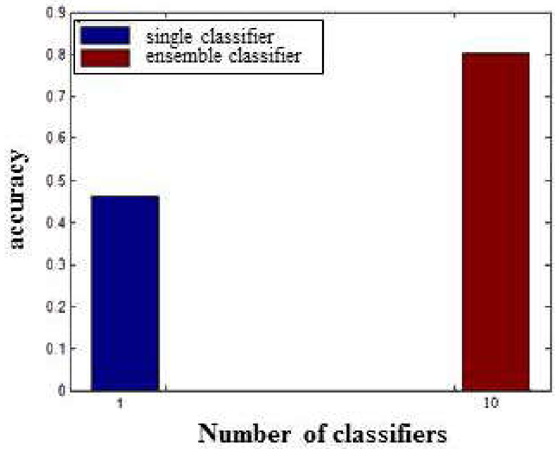

Fig. 7. The "accuracy" of the proposed system against the single-classifier case.

Fig. 8 shows the *FPR* of the proposed system against the single-classifier case. As it is shown in the figure, our proposed method attains less *FPR* compared to the case in which a single classifier is used. Recall from Eq. (6) that *FPR* reflects the rate of anomalies which are misidentified as normal. Thus, Fig. 8 confirms that the proposed system has low misdetection.

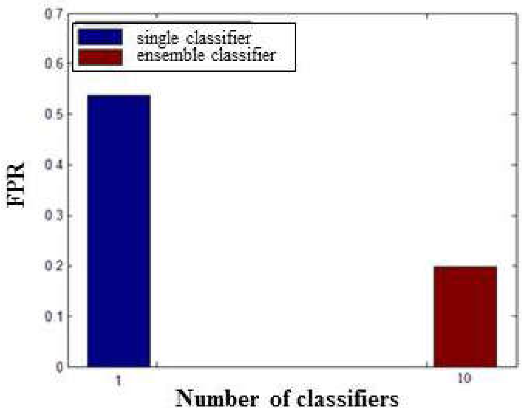

Fig. 8. The "false positive rate" of the proposed system against the single-classifier case.

Fig. 9 and Fig. 10 show the *"sensitivity"* and the *"specificity"* of the proposed system against the single-classifier case, respectively. As the figures show, our proposed method attains better specificity compared to the case in which a single classifier is used, while its sensitivity does not change considerably. Recall from Eq. (8) and Eq. (9) that "sensitivity" reflects the rate of correctly identified normal data while the "specificity" reflects the rate of correctly identified anomalies. It is often claimed that a highly specific test is effective at an anomaly detector system when the result is suspicious to be "anomalous", while a highly sensitive test is deemed effective at ruling out an anomaly detector system when the result is "normal". As we will discuss later, the tradeoff between "specificity" and "sensitivity" is explored in *ROC* analysis as a trade-off between *TPR* and *FPR*. Fig. 9 and Fig. 10 confirm that although the proposed system can detect the anomalies better than the single classifier case, it fails to improve the rate of correctly identified normal data against its rival! Note that, since the major aim of all anomaly detection systems is discovering anomalies rather than normal cases, it can be inferred that fail to have a high sensitivity rate does not considerably affect the performance of these types of systems! Note that the sensitivity rate of or proposed system is at least equal to that of the single classifier case.

Now we proceed to better understand the way by which the proposed system distinguishes the anomalies from the normal points. To this end, in Fig. 11, we have shown the original real labels of read-out data of WSN sensors in conjunction with the labels which are produced by the anomaly detection system. This comparison has been carried out for both proposed method and single classifier case. The red squares in Fig. 11 show original real data labels, including normal and anomalous, while the squares depicted with the other color represent the classification judgment. The red squares are in fact 500 collected data points which we denoted it before by T' . Let's assume numbers "3" and "4" represent normal data points and anomalous data points respectively. The classification judgment result of the single classifier case and the proposed method are depicted with blue and green color squares respectively. In both figures, at first, we have drawn the red squares. In this way, if the classification judgment result is the same as the original real data labels, then the blue or green squares will be overwritten on the red squares which are located at that position. Clearly, the remaining red squares indicate that the judgment of the classifier differs with the original label of those data samples! In other words, the number of remaining red points represents the summation of FN and FP values. Needless to say, it is better for an anomaly detection system to have less red points in Fig. 11. As it can be seen from figures, the proposed method has less red points (FP+FN value) in comparison with the single classifier case. Inversely, it can be concluded that the proposed method has higher TP+TN value compared to the single classifier case.

Fig. 12 shows the *ROC* curve of the proposed system (blue) against the single-classifier case (red). Recall from Eq. (10) that Parameter d reflects the shortest threshold value from any coordinate on the ROC to coordinate (0,1). Thus, as it is shown in Fig. 6, our proposed method attains a shorter d value concerning both the sensitivity and specificity values of the system compared to the case in which a single classifier is used.

In order to better understanding the way in which the proposed system manages to optimize the value of d parameter in each case, we have shown in Fig. 13 the concerning values of d parameter for the data points during the time period $[0, T']$ for each of 500 collected data. In this figure, the values of d parameter concerning the proposed method and the single-classifier are depicted with blue and red colors respectively. As it can be seen from the figure, the proposed method can gradually decrease the amount of d parameter with fewer oscillations over time, compared to those of single classifier case. This confirms that the proposed method is a self-managing scheme over time in order to detect the anomalies in sensor read-out data.

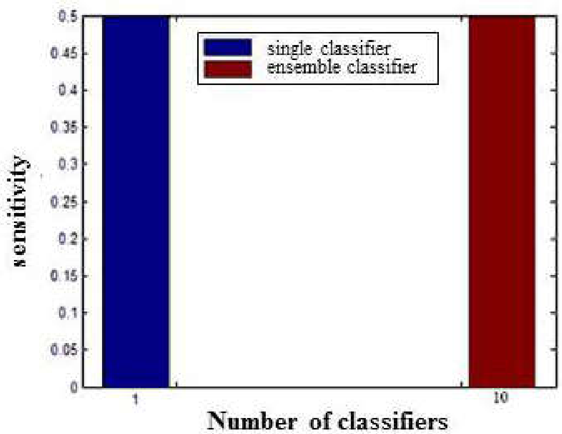

Fig. 9. The "sensitivity" of the proposed system against the single-classifier case.

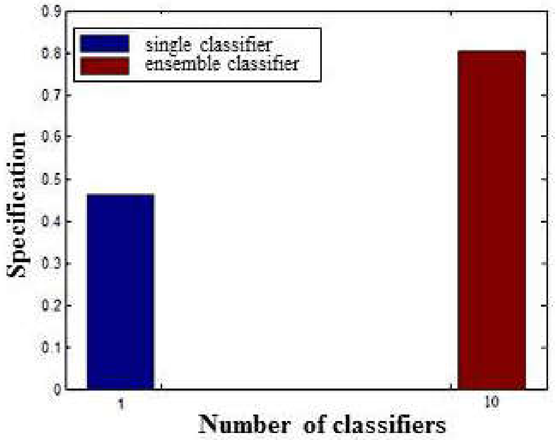

Fig. 10. The "specificity" of the proposed system against the single-classifier case.

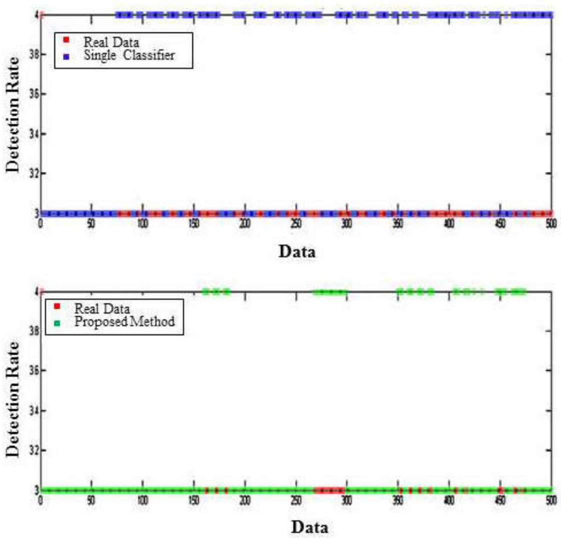

Fig. 11. The comparison of false detection (FN+FP) of the proposed method against the single-classifier case. The red points show the number of FP+FN.

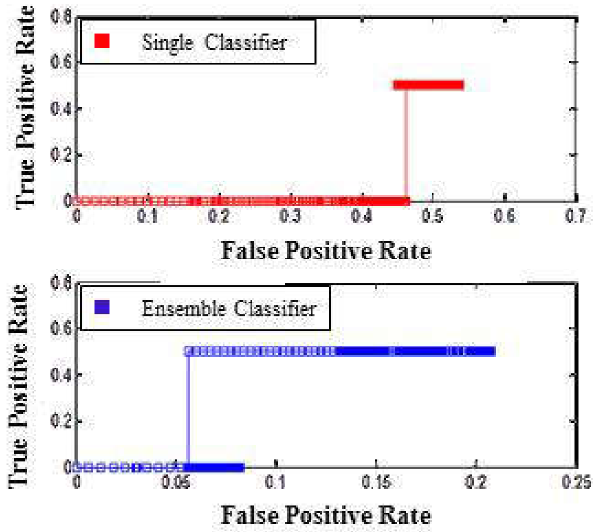

Fig. 12. *ROC* curve of the proposed system (blue) against the single-classifier case (red). The proposed method attains a shorter threshold value to coordinate (0,1).

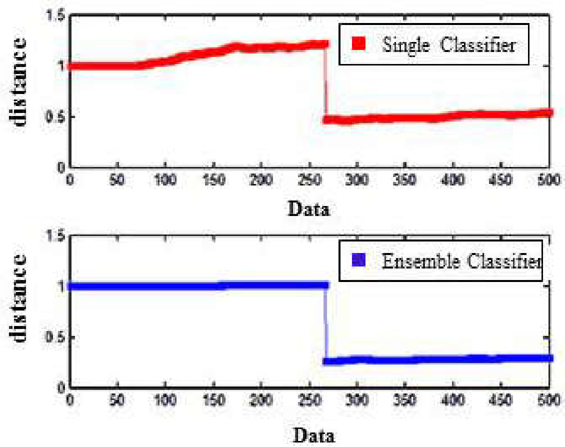

Fig. 13. The values of d parameter concerning the data points during the time period $[0, T']$.

5. Concluding Remarks and Future Trends

In this paper, we focused on anomaly detection in read-out data of sensors of WSN. We investigated and addressed outstanding research activities in this area as well as studying the most significant schemes of anomaly detection using data mining and artificial intelligence. As it was discussed the major problem in the area focuses on low cost, high speed, and high detection rates. Literature survey revealed that the ensemble of classifiers has been an ever-increasing rise in the research activities.

We have proposed a window-based ensemble approach based on majority voting among classifiers. Our proposed algorithm, at first applies a fuzzy clustering approach using the well-known C-means clustering method to create the clusters. In the classification step, we created some base classifiers each of which utilizes the data of overlapping windows to utilize the correlation among data over time. Evaluation results confirmed that the proposed method enhances the performance of the system in terms of convenient metrics in the area of anomaly detection systems.

Our future trend in research in this area will be concentrated on using weighted methods for majority voting among classifiers. We intend to carry out a comprehensive and comparative research in order to improve precision and decrease the computation complexity and memory consumption. Another future research trend is to decrease the dimension of read-out data of sensors using the concept of self-similarity in such a way that the correlation among data is considered. Roughly speaking, we ought to reduce the dimensions of the data regard similar patterns observed in the time range. The existence of self-similarity in network workloads has been explored by the researchers in past decades [33] and the subject is activated again in some other areas in order to detect anomalies [34]. We hope that it is probable to find some types of similarity among WSN sensor data in such a way that could accelerate the process of detecting anomalies by machine learning methods.

References

[1] Malik, N.; Kumar, P., "Distributed Data Mining in Wireless Sensor Network Using Fuzzy Naïve Byes." International Journal of Engineering and Computer Science, vol. 6, no. 8, pp. 22327-22332 (2017).

[2] Guo, X.; Wang, D.; Chen, F., "An Anomaly Detection Based on Data Fusion Algorithm in Wireless Sensor Networks." International Journal of Distributed Sensor Networks, pp. 1-10 (2015).

[3] Tripathi, R.; Dwivedi, S. K., "A Quick Review of Data Stream Mining Algorithms." Imperial Journal of Interdisciplinary Research, vol. 2. No. 7, pp. 870-873 (2016).

[4] Ahmed, M.; Mahmood, A. N.; Hu, J., "A Survey of Network Anomaly Detection Techniques." Journal of Network and Computer Applications, vol. 60, pp. 19–31 (2016).

[5] Thuc, K.-X.; Insoo, K., "A Collaborative Event Detection Scheme Using Fuzzy Logic in Clustered Wireless Sensor Networks." AEU-International Journal of Electronics and Communications, vol. 65, no. 5, pp. 485–488 (2011).

[6] Islam, R.; Shahadat Hossain, M.; Andersson, K., "A Novel Anomaly Detection Algorithm for Sensor Data under Uncertainty." Soft Computing, vol. 22, no.5, pp. 1623-1639 (2018).

[7] Gil, P.; Martins, H.; Januário, F., "Outliers Detection Methods in Wireless Sensor Networks." Artificial Intelligence Review, Springer, pp. 1-26 (2018).

[8] Zhang, Y.; Meratnia, N.; Havinga, P. J.M., "Distributed Online Outlier Detection in Wireless Sensor Networks Using Ellipsoidal Support Vector Machine." Ad Hoc Networks, vol. 11, no.3, pp. 1062–1074 (2013).

[9] Knorr, E.; Ng, R.T., "Algorithms for Mining Distance-based Outliers in Large Data Sets." VLDB 1998. In: Proceedings of the 24th International Conference on Very Large Databases pp. 392-403. New York City, USA (1998).

[10] Araya, D. B.; Grolinger, K.; ElYamany, H. F.; Capretz, M. A.; Bitsuamlak, G., "An Ensemble Learning Framework for Anomaly Detection in Building Energy Consumption." Energy and Buildings, vol. 144, pp. 191-206 (2017).

[11] Zhang, J.; Gardner, R.; Vukotic, I., "Anomaly Detection in Wide Area Network Meshes Using Two Machine Learning Algorithms." Future Generation Computer Systems, in press, accepted manuscript, (2018).

[12] Zhou, Z.-H., Ensemble learning. Encyclopedia of Biometrics, Springer, Berlin, Germany, pp. 270–273 (2009).

[13] Ayadi, A.; Ghorbel, O.; Obeid, A. M.; Abid, M., "Outlier Detection Approaches for Wireless Sensor Networks: A Survey." Computer Networks, vol. 129, no. 1, pp. 319-333 (2017).

[14] Agrawal, S.; Agrawal, J., "Survey on Anomaly Detection Using Data Mining Techniques." Procedia Computer Science, vol. 60, pp. 708-713 (2015).

[15] Ghorbel, O.; Abid, M.; Snoussi, H., "Improved KPCA for Outlier Detection in Wireless Sensor Networks." ATSIP 2014, 1st International Conference on Advanced Technologies for Signal and Image Processing. Sousse, Tunisia, IEEE, pp. 507–511 (2014).

[16] Asmuss, J.; Lauks, G., "Network Traffic Classification for Anomaly Detection Fuzzy Clustering-based Approach." FSKD 2015. 12th International Conference on Fuzzy Systems and Knowledge Discovery, Zhangjiajie, China. IEEE, pp. 313-318 (2017).

[17] Dromard, J.; Roudière, G.; Owezarski, P., "Online and Scalable Unsupervised Network Anomaly Detection Method." IEEE Transactions on Network and Service Management, vol. 14, no. 1, pp. 34-47 (2017).

[18] Tan, P.N.; Steinbach, M.; Kumar, V., "Introduction to Data Mining." Addison-Wesley (2005).

[19] Bhargava, A.; Raghuvanshi, A., "Anomaly Detection in Wireless Sensor Network Using S-Transform in Combination with SVM." IEEE International Conference on Computational Intelligence and Communication Networks. Mathura, India, IEEE, pp. 111-116 (2013).

[20] Araya, D. B.; Grolinger, K.; ElYamany, H. F.; Capretz, M. A.; Bitsuamlak, G., "Collective Contextual Anomaly Detection Framework for Smart Buildings." IJCNN 2016. IEEE International Joint Conference on Neural Networks. Vancouver, BC, Canada, IEEE, pp. 511–518 (2016).

[21] Ahmad, S.; Lavin, A.; Purdy, S.; Agha, Z., "Unsupervised Real-Time Anomaly Detection for Streaming Data." Neurocomputing, vol. 262, pp. 134-147 (2017).

[22] Padilla, D.E.; Brinkworth, R.; McDonnell, M.D., "Performance of a Hierarchical Temporal Memory Network in Noisy Sequence Learning." CYBERNETICSCOM 2013. Proceedings of the IEEE International Conference on Computational Intelligence and Cybernetics. Yogyakarta, Indonesia, IEEE, pp. 45-51 (2013).

[23] Dominguesa, R.; Filipponea, M.; Michiardia, P.; Zouaoui, J., "A Comparative Evaluation of Outlier Detection Algorithms:

Experiments and Analyses." Pattern Recognition, Elsevier, vol. 74, pp. 406-421 (2018).

[24] Bosman, H. H.; Iacca, G.; Tejada, A.; Wörtche, H. J.; Liotta, A., "Spatial Anomaly Detection in Sensor Networks Using Neighborhood Information." Information Fusion, vol. 33, pp. 41-56 (2017).

[25] Intel Lab Data (http://db.csail.mit.edu/labdata/labdata.html).

[26] Chatterjee, S.; Mukhopadhyay, A., "Clustering Ensemble: A Multiobjective Genetic Algorithm Based Approach." Procedia Technology, vol. 10, pp. 443-449 (2013).

[27] Jagannath Nanda, S.; Panda, G., "A Survey on Nature Inspired Metaheuristic Algorithms for Partitional Clustering." Swarm and Evolutionary Computation, vol. 16, pp. 1-18 (2014).

[28] Garcia-Font, V.; Garrigues, C.; RifÃ–Pous, H., "A Comparative Study of Anomaly Detection Techniques for Smart City Wireless Sensor Networks." Sensors, vol. 16, no. 6, 868 (2016).

[29] O'Reilly, C.; Gluhak, A.; Imran, M. A.; Rajasegarar, S., "Anomaly Detection in Wireless Sensor Networks in A Non-stationary Environment." IEEE Communications Surveys & Tutorials, vol. 16, no. 3, pp. 1413-1432 (2014).

[30] Rajasegarar, S.; Leckie, C.; Palaniswami, M., "Hyperspherical Cluster-based Distributed Anomaly Detection in Wireless Sensor Networks." Journal of Parallel and Distributed Computing, vol. 74, no. 1, pp. 1833-1847 (2014).

[31] Kumarage, H.; Khalil, I.; Tari, Z.; Zomaya, A., "Distributed Anomaly Detection for Industrial Wireless Sensor Networks Based on Fuzzy Data Modelling." Journal of Parallel and Distributed Computing, vol. 73, no. 6, pp. 790-806 (2013).

[32] Kapoor, A.; Singhal, A., "A comparative study of K-Means, K-Means++ and Fuzzy C-Means clustering algorithms." CICT 2017. 3rd International Conference on Computational Intelligence & Communication Technology. Ghaziabad, India, IEEE, pp. 1-6 (2017).

[33] Erramilli, A.; Roughan, M.; Veitch, D.; Willinger, W., "Self-Similar Traffic and Network Dynamics." Proceedings of the IEEE, vol. 90, no. 5, pp. 800-819 (2002).

[34] Napoletano, P.; Piccoli, F.; Schettini, R., "Anomaly Detection in Nanofibrous Materials by CNN-Based Self-Similarity." Sensors, vol. 18, no. 1, pp. 1-15 (2018).